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

A Panel Data Spatial Regression Approach for Modeling Poverty Data In Southern Sumatra

Nurul Hidayati [1*], Elisabeth Evelin Karuna [2], Muhammad Arib Alwansyah [2]
[1] S1 Statistics Study Program, Bengkulu University, Bengkulu
[2] S2 Statistics Study Program, Bengkulu University, Bengkulu

* Corresponding Author: nurulhidayati@unib.ac.id

Article Info

Article History:
Received: 23 04 2025
Revised: 24 04 2025
Accepted: 24 04 2025
Available Online: 25 04 2025

Key Words:
Poverty
Spatial Analysis
Spatial Panel Regression
Panel Data

Abstract

This research examines the use of spatial panel data regression approach to model poverty data in the Southern Sumatra region. The main objective of the study is to model poverty in the Southern Sumatra region using spatial panel data regression. Panel data from districts/cities in South Sumatra, Jambi, Lampung, Bengkulu, and Bangka Belitung during the 2015-2021 period were used in the analysis. The spatial panel models used in this study are panel SAR regression and panel SEM. The results show that the spatial panel data approach is better at explaining variations in poverty levels compared to non-spatial models. A significant spatial spillover effect was found, where the poverty level of an area is influenced by the conditions of its neighboring areas. The results of the analysis show that the best model to use in modeling the Poverty Percentage data in the Southern Sumatra region is the Spatial Autoregressive Fixed Effect (SAR-FE) model based on the smallest AIC and BIC values. Factors such as average years of schooling and life expectancy are proven to have a significant influence on the percentage of poverty in the SAR Fixed Effect model.

1. INTRODUCTION

Poverty remains one of the main challenges in socio-economic development in Indonesia, especially in regions outside Java such as Southern Sumatra. This region, which includes the provinces of South Sumatra, Lampung, Bengkulu, Jambi and Bangka Belitung Islands, shows significant variations in poverty levels between regions and over time [1]. Although poverty alleviation efforts have been made, the different geographical characteristics, natural resources, and economic development patterns in this region create complex poverty dynamics [2].

Traditional poverty analysis often ignores two important aspects: spatial and temporal dimensions. In fact, poverty in a region is not only influenced by the characteristics of the region itself, but also by the conditions of surrounding regions, a phenomenon known as spatial dependence [3]. In addition, poverty patterns also change over time, reflecting the dynamics of the economy and the policies implemented [4].

The panel data spatial regression approach offers a promising solution to overcome this limitation. This method integrates spatial analysis with panel data, allowing for more comprehensive modeling of the poverty phenomenon [5]. Panel data spatial regression can capture spillover effects between regions and control for unobserved heterogeneity, both spatial and temporal [6].

Several previous studies have shown the advantages of this approach. For example, Jajang et al. (2013) used a panel spatial autoregressive (SAR) model to analyze poverty in West Java and found that this model is better at explaining poverty variation than non-spatial models [7]. At the international level, Ahlburg (2017) applied a spatial Durbin model for panel data in the analysis of poverty in Pacific countries, demonstrating the importance of spillover effects in regional poverty dynamics [8].

However, the application of this method in the Southern Sumatra region is still limited. In fact, the characteristics of this region - with high geographical variation and diverse development patterns - make it an ideal candidate for spatial-temporal analysis [9]. For example, the significant differences between coastal and inland areas in terms of access to infrastructure and markets can create distorted spatial patterns of poverty [10].

This study aims to model poverty in the Southern Sumatra region using spatial panel data regression. The results of this study are expected to provide new insights into the dynamics of poverty in Southern Sumatra and provide a more accurate analytical tool for the formulation of focused and effective poverty alleviation policies.

1.1 Research Design and Research Variables

This study uses two types of variables, namely response variables and predictor variables as presented in Table 1 below:

Table 1. Research Variables

Variable	Symbol	Indicator	Unit
Response (Y)	Y	Poverty Percentage	Persen
Predictor (X)	X_1	Average Years of Schooling	Persen
	X_2	GRDP	Persen
	X_3	Unemployment Rate	Persen
	X_4	Life Expectancy	Persen

1.2 Stages of Analysis

1. Estimating panel data regression model parameters with the Common Effect Model, Fixed Effect Model, Random effect Model estimation approach.

   i. Common Effect Model (CEM)

   The Common Effect Regression Model aims to estimate panel data, by combining time series and cross section data without seeing differences between time and individuals, using the ordinary Least Square (OLS) method, which assumes the intercept and slope coefficient values for all cross section and time series units are the same. The CEM model is as follows:

   $$y_{it} = \alpha + \boldsymbol{\beta}'\boldsymbol{x}_{it}^{'}{}' + \varepsilon_{it} \quad ; i = 1, \dots, N \; and \; t = 1, \dots, T \tag{1}$$

   ii. Fixed Effect Model (FEM)

   The fixed effect model structure is a model that considers the diversity of independent variables by individual. The objects used (N) are mostly aggregate objects or only focus on N objects only. The assumptions that must be met are, μ is assumed to be fixed, so that it can be estimated; ε_{it} spreads freely stochastic identical normal $(0.\sigma_\varepsilon^2)$; $E(x_{it}, \varepsilon_{it}) = 0$ or x_{it} is mutually independent with ε_{it} for each i and t [13]. The FEM model for cross section can be expressed as follows:

   $$y_{it} = D_i\alpha_i + \boldsymbol{\beta}'\boldsymbol{x}_{it}^{'} + \varepsilon_{it} \; ; i = 1, \dots, N \; and \; t = 1, \dots, T \tag{2}$$

   iii. Random Effect Model (REM)

   Random effect model will estimate panel data where disturbance variables may be interconnected over time or between individuals. The random effect model equation is expressed as follows:

   $$y_{it} = \alpha + \boldsymbol{\beta}'\boldsymbol{x}_{it}^{'} + \eta_{it} \; ; i = 1, \dots, N \; and \; t = 1, \dots, T \tag{3}$$

2. Determine the best panel data model selection using the Chow test, Haustman test and Breusch-Pangan test. Model selection is statistically carried out so that the estimates obtained can be as efficient as possible. Tests in determining the model to be used in panel data processing [8], namely:

   i. *Chow Test*

   The Chow test is used to choose between the Fixed Effect Model (FEM) or the Common Effect Model (CEM). The test procedure is as follows [7]. The hypothesis is as follows:
   $H_0: \beta_{01} = \beta_{02} = \dots = \beta_{0n} = 0$ (the best model is CEM)

H_1: $there\ is\ at\ least\ one\ \beta_{0i} \neq 0$ (the best model is FEM) with $i = 1,2, \ldots, n$
The test statistic used is the F test, which is:

$$F_{hitung} = \frac{\frac{\left(R^2_{LSDV} - R^2_{pooled}\right)}{n-1}}{\left(1 - R^2_{LSDV}\right)(nT - n - k)} \sim \quad F_{\alpha;n-1;n(T-1)-k} \tag{4}$$

with n is the number of individual units; T is the observation time period; k is the number of independent variables in the fixed effects model. The test criteria used are reject H_0 if the value of $F_{count} > F_{table}$, with $F_{tabel} = F_{\alpha;n-1;n(T-1)-k}$ or reject H_0 [11]. If the test is significant, the appropriate model is the FEM model. Conversely, if the test is not significant, then the appropriate model is the CEM model.

ii. Hausman Test

The Hausman Test aims to see if there is a random effect in the panel data [12]. The Hausman test is used to select the REM and FEM models with the test statistic:

$$\chi 2 = \boldsymbol{\hat{q}}[\boldsymbol{Var}(\boldsymbol{\hat{q}})]^{-\boldsymbol{1}}\boldsymbol{\hat{q}} \tag{5}$$

with

$$\hat{q} = \hat{\beta}_{acak} - \hat{\beta}_{tetap}$$

The decision rejects H_0 if $\chi^2 > \chi^2_{k,\alpha}$ with k being the number of explanatory variables, or rejects H_0 if $p < \alpha$ [9]. If the test is significant then the appropriate model is the FEM model. Conversely, if the test is not significant, then the appropriate model is the REM model.

3. Classical assumption testing, in the form of normality test using Kolmogorov Smirnov, multicollinearity test using Variance Factor (VIF), heteroscedasticity test using Breusch-Pagan-Godfrey method and autocorrelation test using Durbin Watson test.

i. Normality Test

In this research data, the normality test used is the Kolmogorov-Smirnov test. The hypothesis used is as follows:
H_0: error is normally distributed
H_1: error are not normally distributed
Test Statistics:

$$D_{count} = \max_{1<i<N}\left(F(Y_i) - \frac{i-1}{N}, \frac{i-1}{N} - F(Y_i)\right) \tag{6}$$

Rejection Criteria
Reject H_0 if the probability value < α, meaning the errors are not normally distributed

ii. Multicollinearity Test

This test is to determine whether the independent variables in the regression equation are not correlated with each other. One indicator to detect multicollinearity is by calculating the Variance Factor (VIF) value with the formula:

$$VIF_j = \frac{1}{1 - R_j^2} \tag{7}$$

Hypothesis

H_0: there is no multicollinearity

H_1: there is multicollinearity

Test Statistics

$VIF_j = \frac{1}{1-R_j^2}$; with $j = 1,2,..,k$

Rejection Criteria

Reject H_0, if the VIF value > 10, meaning there is multicollinearity

iii. Autocorrelation Test

Autocorrelation is the correlation between the error of one observation and the error in another observation. The autocorrelation test aims to determine whether or not there is a correlation between errors in period t and errors in the previous period (t-1) [14]. The autocorrelation test is a statistical analysis conducted to determine whether there is a variable correlation in the prediction model with changes in time, using the Durbin Watson test (DW-test).

Hypothesis:

$H_0 : d = 0$ (there is autocorrelation)

$H_1 : d > 0$ atau $d < 0$ (there is no autocorrelation)

Test Statistics:

$$DW = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \tag{8}$$

Rejection Criteria:

Table 2. Autocorrelation Assumption Decision Making

Null Hypothesis	Decision	Criteria
No positive autocorrelation	Reject	0<dw<dl
No positive autocorrelation	No decision	$dl \leq dw \leq du$
No negative autocorrelation	Reject	$4 - dl < dw < 4$
No negative autocorrelation	No decision	$4 - du \leq dw \leq 4 - dl$
Failure to Reject	No autocorrelation, positive or negative	$du < dw < 4 - du$

iv. Heteroscedasticity Test

The heteroscedasticity test is used to determine whether there is an unequal variance from the error of an observation to another observation. In this study using the Breusch-Pagan-Godfrey test, with the hypothesis:

$H_0 : \sigma_i^2 = 0$ (there is no heteroscedasticity)

$H_1 : \sigma_i^2 \neq 0$ (there is heteroscedasticity)

Test Statistics:

$$LM = \frac{NT}{2(t-1)} \sum_{i=1}^{N} \left[\frac{\left[\sum_{t=1}^{T} \varepsilon_{it}\right]^2}{\sum_{t=1}^{T} \varepsilon_{it}^2} - 1 \right]^2 \tag{9}$$

with, N being the number of data, and T being the number of time periods and ε being the residuals. Test criteria if $LM > \chi^2_{t\alpha, N-1}$ or p-value < significance level then reject H_0 so that the variance-covariance structure of residuals is heteroscedasticity.

4. Use of parameter significance using simultaneous test (F-test and partial test (t-test))

   i. Test the significance of regression coefficients simultaneously.

Simultaneous test or F-test is a test conducted to determine the effect of independent variables with dependent variables together. The hypothesis used:
$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (there is no effect between all independent variables on the percentage of poor people simultaneously)
$H_1: minimal\ ada\ satu\ \beta_j \neq 0\ \ ; j = 1,2,3,4$; (there is an influence between all independent variables on the percentage of poor people simultaneously)
Test Statistics:

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \tag{10}$$

with; N = number of observations, $R^2 =$ coefficient of determination, and k = number of independent variables.
The test criteria used are if $F_{count} > F_{tabel\,(k;NT-N-K,\alpha)}$ or $p - value\ < \alpha$, which means that there is at least one independent variable that has a significant effect on the dependent variable [10].

ii. Partial regression coefficient significance test

Partial test or t test is conducted to determine the significance of independent variables individually on the dependent variable. The hypothesis used is as follows:
$H_0: \beta_j = 0$
$H_1:\ \beta_j \neq 0$; with $j = 1,2,3,4$
Test Statistics:

$$t_{count} = \frac{\widehat{\beta}_p}{se(\widehat{\beta}_p)} \tag{11}$$

The test criterion is to reject H_0 if $|t_{hitung}| > t_{(\frac{\alpha}{2}, N-k-1)}$ or $p - value\ < \alpha$, which means that the p-th independent variable has a significant effect on the dependent variable [11].

5. Sets the spatial weight matrix

The spatial weighting matrix to be used is queen contiquity (side-corner intersection) which defines $\boldsymbol{w_{ij} = 1}$ for regions whose common side or common vertex meets the corner of the region of interest, while $\boldsymbol{w_{ij} = 0}$ for other regions (Lesage, 1999), with the formula:

$$\boldsymbol{w_{ij}} = \frac{\boldsymbol{c_{ij}}}{\sum \boldsymbol{c_{ij}}} \tag{12}$$

6. Moran'I Test

Moran'I test, which is a statistical test to see the value of spatial autocorrelation used to identify a location of spatial clustering.
Hypothesis:
$\boldsymbol{H_0}$: $\boldsymbol{I}$ =0 (No autocorrelation between regions)
$\boldsymbol{H_1}$: $\boldsymbol{I}$ ≠0 (there is autocorrelation between regions)
Test Statistics:

$$\boldsymbol{Z_{count}} = \frac{\boldsymbol{I - I_0}}{\sqrt{\boldsymbol{var(I)}}} \sim \boldsymbol{N(0,1)} \tag{13}$$

where $\boldsymbol{I}$: Moran's I value, $\boldsymbol{I_0}$: Expected value of Moran's I, $var(I)$: Variance of Moran's, $\boldsymbol{I_n}$: number of regions of occurrence. Reject $\boldsymbol{H_0}$ if $|Zhitung| > \boldsymbol{Z_{\frac{\alpha}{2}}}$ or $p\text{-}value$ < α. The value for the index I is between -1 and 1. If $I > \boldsymbol{I_0}$, the data has positive autocorrelation, if $I > \boldsymbol{I_0}$, the data has negative autocorrelation.

7. Langrange Multiplier Test

Performing the Langrange Multiplier test, which is the test used as a basis for selecting the appropriate spatial regression model. there are two types of spatial interaction, namely spatial lag and error. One of the statistical tests to determine the existence of spatial dependency is by using the Lagrange Multiplier (LM) test and the Robust Lagrange Multiplier (RLM) test. The spatial lag LM test aims to determine whether a model is said to be a spatial lag model while the test to determine the spatial error model is the spatial error LM test.
Hypothesis testing on the spatial lag or Autoregressive (SAR) model is:
$\boldsymbol{H_0}$: $\boldsymbol{\delta}$=0 (No spatial lag dependence)
$\boldsymbol{H_1}$: $\boldsymbol{\delta}$ ≠0 (There is spatial lag dependence)

$$\boldsymbol{LM_{lag}} = \frac{\boldsymbol{e'(I_T \otimes W)e/\hat{\sigma}_e^2}}{\boldsymbol{J}} \tag{14}$$

and for hypothesis testing on the Spatial Error model (SEM)
$\boldsymbol{H_0}$: $\boldsymbol{\tau = 0}$ (No spatial error dependency)
$\boldsymbol{H_1}$: $\boldsymbol{\tau \neq 0}$ (There is spatial error dependence)

$$\boldsymbol{LM_{error}} = \frac{\boldsymbol{e'(I_T \otimes W)e/\hat{\sigma}_e^2}}{\boldsymbol{JxT_w}} \tag{15}$$

The LM test statistic is $\boldsymbol{\chi^2_{(p)}}$ distributed, so reject $\boldsymbol{H_0}$ if the LM value > $\boldsymbol{\chi^2_{(q)}}$ or the $\boldsymbol{p - value} < \propto$, with q defined as the number of spatial parameters which is 1.

8. Estimation of panel spatial model

i. Spatial Autoregressive panel fixed effect (SAR-FE)

Spatial lag model or spatial autoregressive (SAR) model shows that the dependent variable depends on the observed independent variable and the dependent variable in the nearest unit. The SAR-FE model is as follows:

$$\boldsymbol{y} = \boldsymbol{\delta W_{nt} y} + \boldsymbol{X\beta} + (\boldsymbol{l_t} \otimes \boldsymbol{I_n})\boldsymbol{\mu} + \boldsymbol{\varepsilon} \quad (16)$$

ii. Spatial Error panel fixed effect (SEM-FE)

Spatial Error Model (SEM) shows that the dependent variable depends on the observed independent variables and errors that are correlated between neighboring places. The SEM-FE model is as follows:

$$\boldsymbol{y} = \boldsymbol{X\beta} + (\boldsymbol{l_t} \otimes \boldsymbol{I_n})\boldsymbol{\mu} + \emptyset \quad (17)$$

$$\emptyset = \boldsymbol{\tau W_{nt}} \emptyset + \boldsymbol{\varepsilon} \quad (18)$$

9. Significance test of SAR-FE and SEM-FE

The significance test is one of the most important stages in a study. The significance test is used to determine whether the hypothesis made at the beginning of the study will be accepted or rejected. The test statistic used is the Z test.

10. Best Model Selection

Model selection criteria are carried out using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values. If the AIC and BIC values have a smaller value, then the model is said to be good. The AIC equation can be shown as follows:

$$AIC = -2l + 2p \quad (19)$$

where l is the log likelihood and p is the number of parameters. while for the BIC equation as follows:

$$BIC = -2l + kLn(n) \quad (20)$$

Where l is the log likelihood, k is the number of parameters and n is the sample size.

2. RESULTS AND DISCUSSION

3.1 Descriptive Statistics

Descriptive statistics are statistical techniques used to describe the data that has been collected. Descriptive statistics aim to provide useful information about the data to be analyzed. Therefore, before conducting further analysis, it is necessary to know the descriptive statistics of the existing data as follows:

Table 3. Descriptive Statistics

Variable	Minimum	Mean	Maximum
Poverty Percentage (Y)	2.67	11.66	23.20
Average Years of Schooling (X1)	5.88	8.04	11.80
GRDP (X2)	2.28	17.96	162.65
Unemployment Rate (X3)	0.65	4.17	12.31
Life Expectancy (X4)	62.04	68.53	73.41

After knowing the descriptive statistics of the existing data, further analysis will be carried out as follows.

3.1 Spatial Dependence Test

Moran's index is a statistical test used to see the value of spatial autocorrelation in the observed area with its neighboring areas. The Moran index value ranges from $-1 < I < 1$. If the Moran index value is in the range $-1 \leq I < 0$ then the spatial autocorrelation that occurs is negative spatial autocorrelation and if the Moran index value is in the range $0 < I \leq 1$ then the spatial autocorrelation that occurs is positive spatial autocorrelation, while if the Moran index value shows zero, then there is no spatial autocorrelation. The following shows the results of the calculation of the Moran Index of the Percentage of Poor Population in the Regency/City in the Southern Sumatra region using the queen contiguity spatial weighting matrix:

Table 4. Moran Index Test

Year	I	E(I)	Var (I)	Z(I)	p-value
2015	0.25583	-0.01695	0.00646	3.39330	0.00069
2016	0.26493	-0.01695	0.00646	3.50730	0.00045
2017	0.27260	-0.01695	0.00647	3.60110	0.00032
2018	0.26132	-0.01695	0.00647	3.45890	0.00054
2019	0.28647	-0.01695	0.00648	3.76960	0.00016
2020	0.30416	-0.01695	0.00648	3.98970	0.00007
2021	0.30553	-0.01695	0.00647	4.01030	0.00006

Table 4 shows the results of the Moran Index test of the Percentage of Poor Population in Districts/Cities in the Southern Sumatra region using the queen contiguity spatial weighting matrix. Based on these results, it can be seen that there is significant spatial autocorrelation between districts/cities in the Southern Sumatra region for each year based on the $p_{value} < \alpha = 0.05$. This spatial correlation can also be seen from the resulting Moran Index value, where all Moran Index (I) values are in the range $0 \leq I < 1$, indicating that there is positive spatial autocorrelation between districts/cities in the Southern Sumatra region from 2015 to 2021. Furthermore, the correlation between variable Y and each variable X will be identified using the Pearson correlation test. The test results can be seen in Table 5:

Table 5. Pearson Correlation Test

Variable	Correlation	Test Statistics	df	p-value
Y and X1	-0.0353	-0.7223	418	0.4705
Y and X2	0.0431	0.8830	418	0.3778
Y and X3	-0.0439	-0.8988	418	0.3693
Y and X4	-0.4851	-11.3420	418	0.0000

The correlation test results in Table 5 show that overall, there is a correlation between variable Y and each variable X although the resulting correlation can be said to be quite low. In the variables Y and X1, Y and X3, and Y and X4 there is a negative correlation, while for the variables Y and X2 there is a positive correlation. However, if statistical testing is carried out on the resulting correlation value, it can be concluded that for the variables Y and X1, Y and X2, and Y and X3, the resulting correlation value is not statistically significant. Meanwhile, for the variables Y and X4, the resulting correlation value is statistically significant. This indicates that the correlation in the analyzed data is not sufficient to be explained in general without considering spatial effects, but requires the influence or spatial effects of each region in explaining the correlation of each variable used.

3.2 Multicollinearity Test

Multicollinearity test is a test conducted to determine whether there is a relationship or correlation between independent variables. One way to detect multicollinearity is to calculate the variance inflation factor (VIF) value. The multicollinearity test results can be seen in the following table:

Table 6. Multicollinearity Test

Independent Variable	VIF	Indication of Multicollinearity
X1	1.4567	No Multicollinearity
X2	1.1708	No Multicollinearity
X3	1.4070	No Multicollinearity
X4	1.2853	No Multicollinearity

The multicollinearity test results in Table 6 show the conclusion that there is no multicollinearity in all independent variables used with VIF values <10. This leads to the conclusion that all independent variables can be used for panel data spatial regression modeling.

3.3 Parameter Estimation of CEM, FEM, and REM

In this study, the parameter estimation of panel data regression models will be carried out using 3 models, namely the Common Effect Model (CEM), Fixed Effect Model (FEM) and Random Effect Model (REM). The parameter estimation results of each model applied aim to determine which panel data regression model is most suitable for modeling the Poverty Percentage data in the South Sumatra region. Therefore, the parameter estimation results can be seen in the following table.

Table 7. Parameter Estimation of CEM, FEM, and REM

	CEM				FEM				REM			
	Estimate	Std. Error	t-value	p-value	Estimate	Std. Error	t-value	p-value	Estimate	Std. Error	z-value	p-value
(Intercept)	43.7139	2.8923	15.1137	0.0000					23.2852	3.5735	6.5161	0.0000
X1	0.1559	0.0253	6.1663	0.0000	-0.2104	0.0353	-5.9544	0.0000	-0.1333	0.0314	-4.2400	0.0000
X2	0.8112	0.1606	5.0503	0.0000	0.1377	0.1482	0.9294	0.3533	0.1276	0.1395	0.9146	0.3604
X3	-0.0203	0.0540	-0.3751	0.7077	0.0094	0.0081	1.1658	0.2445	0.0127	0.0082	1.5463	0.1220
X4	-10.2668	0.7170	-14.3198	0.0000	-3.1699	0.9729	-3.2582	0.0012	-4.9384	0.8875	-5.5645	0.0000

Based on the results in Table 7, it can be seen that variable X1 provides significant and different results in the CEM and FEM models, with the opposite direction of influence or relationship. X2 is not significant in the FEM and REM models, but significant in the CEM model. X3 is not significant in all models. X4 is significant in all models with a consistent direction of influence or relationship (negative) on the dependent variable. The different results between the CEM, FEM, and REM models indicate that the effects of independent variables may vary depending on the type of model used. Based on Table 7, the following modeling can be made:

CEM Model:

$$\widehat{Y}_{it} = 43.7139 + 0.1559X_{1it} + 0.8112X_{2it} - 0.0203X_{3it} - 10.2668X_{4it}$$

FEM Model:

$$\widehat{Y}_{it} =- 0.2104X_{1it} + 0.1377X_{2it} + 0.0094X_{3it} - 3.1699X_{4it}$$

REM Model**:**

$$\widehat{Y}_{it} = 23.2852 - 0.1333X_{1it} + 0.1276X_{2it} + 0.0127X_{3it} - 4.9384X_{4it}$$

3.4 Panel Data Regression Model Selection

The Chow test is conducted to select the Common Effect model or Fixed Effect model to be used. The Hausman test is conducted to select the Random Effect model or the Fixed Effect model to be used. The test results can be seen in the following table:

Table 8. Chow Test and Hausman Test Results

Chow Test				Hausman Test		
F-count	df1	df2	p-value	Chi-Square	df	p-value
532.1300	59	356	0.0000	14.7940	4	0.0051

Based on the test results in Table 8, it is known that for the Chow test results, the $F_{count} = 532.1300$, with a $p_{value} = 0.0000 < \alpha = 0.05$, so H_0 is rejected. This indicates that the panel data regression model that can be used is FEM. Meanwhile, for the Hausman test results, the value of $\chi^2_{count} = 14.7940$, with a $p_{value} = 0.0051 < \alpha = 0.05$, so H_0 is rejected. This indicates that the panel data regression model that can be used is FEM. Furthermore, testing for spatial effects or spatial dependence will be carried out which aims to see which model is suitable for use between the Spatial Autoregressive Model Fixed Effect (SAR-FE) or Spatial Error Model Fixed Effect (SEM-FE) using the Lagrange Multiplier test.

3.5 Lagrange Multiplier Test

Spatial effects are spatial dependencies that occur due to correlations between regions consisting of lag dependence and spatial errors. These two effects can be tested using the Lagrange Multiplier (LM) test which aims to serve as the basis for the formation of spatial regression models. The Lagrange Multiplier test results are as follows:

Table 9. *Lagrange Multiplier Test Results*

Year	LM	LM Test Statistics	df	p-value
2015	LM Lag	7.1285	1	0.0076
	LM Error	0.1008	1	0.7508
2016	LM Lag	5.6260	1	0.0177
	LM Error	0.0271	1	0.8692
2017	LM Lag	9.0298	1	0.0027
	LM Error	0.7540	1	0.3852
2018	LM Lag	7.6456	1	0.0057
	LM Error	0.9111	1	0.3398
2019	LM Lag	8.2865	1	0.0040
	LM Error	0.9076	1	0.3408
2020	LM Lag	8.8510	1	0.0029
	LM Error	0.7484	1	0.3870
2021	LM Lag	10.1470	1	0.0014
	LM Error	1.1071	1	0.2927

Table 9 shows the results of the Lagrange Multiplier test. Based on the test results, it can be seen that the $p_{value} < \alpha$, then H_0 is rejected, meaning that there is a spatial lag dependency, so the spatial regression model that can be made is the Spatial Autoregressive Model (SAR). Based on the test results for panel data regression model selection conducted with 2 tests, namely the Chow test and the Hausman test, it was found that the selected model was FEM. This means that the model that can be used to model the Poverty Percentage data in the South Sumatra region is the Spatial Autoregressive Fixed Effect (SAR-FE) model.

3.6 Modeling the Percentage of Poor Population with Spatial Panel Data Model

In this study, the spatial panel data model used to model the percentage of poor people in districts/municipalities in the Southern Sumatra region is the SAR and SEM models using queen contiguity spatial weights. In this study, the SAR model is used to see whether the percentage of poor people in districts/cities in the Southern Sumatra region is related to the percentage of poor people in other districts/cities in the Southern Sumatra region. Meanwhile, the SEM model is used to see whether there is a spatial correlation between the error model of a district/city and other districts/cities in the Southern Sumatra region. To estimate the SAR and SEM models, this study will use three approaches, namely pooled, fixed effect, and random effect. The parameter estimation results are presented in the following table:

Table 10. SAR model parameter estimation

Parameter	Pooling				Fixed Effects Model				Random Effects Model			
	Estimate	Std. Error	t-value	p-value	Estimate	Std. Error	t-value	p-value	Estimate	Std. Error	t-value	p-value
Intercept	38.3019	2.9166	13.1322	0.0000					23.2301	3.6977	6.2824	0.0000
X1	0.1304	0.0250	5.2086	0.0000	-0.1517	0.0319	-4.6158	0.0000	-0.1095	0.0329	-3.3311	0.0009
X2	0.8405	0.1711	4.9130	0.0000	0.1436	0.1312	1.0942	0.2739	0.1329	0.1336	0.9950	0.3198
X3	-0.0461	0.0540	-0.8541	0.3931	0.0073	0.0071	1.0269	0.3045	0.0084	0.0079	1.0731	0.2832
X4	-8.9797	0.7266	-12.3588	0.0000	-2.6972	0.8749	-3.0828	0.0021	-4.9767	0.9140	-5.4447	0.0000
ρ	0.0580	0.0159	3.6473	0.0003	-	-	-	-	-	-	-	-
λ	0.0010	0.0044	0.2203	0.8256	0.0558	0.0107	5.2239	0.0000	0.0140	0.0097	1.4351	0.1513

Table 10 shows the parameter estimation results of the SAR model for each effect, namely SAR-Pooling, SAR-Fixed Effects Model, and SAR-Random Effects Model. Based on these results, the regression equation for each model can be made as follows:

SAR-*Pooling*

$$y_{it} = 38.3019 + 0.1304x_{1it} + 0.8405x_{2it} - 0.0461x_{3it} - 8.9797x_{4it} + \mu_i + \varepsilon_{it}$$

SAR-*Fixed Effects Model*

$$y_{it} = 0.0558\sum_{j=1}^{N} W_{ij}y_{jt} - 0.1517x_{1it} + 0.1436x_{2it} + 0.0073x_{3it} - 2.6972x_{4it} + \mu_i + \varepsilon_{it}$$

SAR-*Random Effects Model*

$$y_{it} = 0.0140\sum_{j=1}^{N} W_{ij}y_{jt} + 23.2301 - 0.1095x_{1it} + 0.1329x_{2it} + 0.0084x_{3it} - 4.9767x_{4it} + \mu_i + \varepsilon_{it}$$

After making the SAR model, then model the SEM based on the estimated values in Table 11.

Table 11. Estimation of SEM model parameters

Parameter	Pooling				Fixed effects model				Random effects model			
	Estimate	Std. Error	t-value	p-value	Estimate	Std. Error	t-value	p-value	Estimate	Std. Error	t-value	p-value
Intercept	38.2424	2.9178	13.1067	0.0000					24.3223	3.7512	6.4838	0.0000
X1	0.1303	0.0250	5.2061	0.0000	-0.1834	0.0350	-5.2452	0.0000	-0.1127	0.0335	-3.3678	0.0008
X2	0.8461	0.1714	4.9367	0.0000	0.1395	0.1314	1.0615	0.2884	0.1420	0.1334	1.0645	0.2871
X3	-0.0463	0.0540	-0.8576	0.3911	0.0073	0.0073	0.9984	0.3181	0.0084	0.0079	1.0616	0.2884
X4	-8.9662	0.7270	-12.3337	0.0000	-3.6525	0.9498	-3.8457	0.0001	-5.2016	0.9263	-5.6152	0.0000
ρ	0.0595	0.0141	4.2066	0.0000	0.0689	0.0128	5.3684	0.0000	-	-	-	-

Table 11 shows the parameter estimation results of the SEM model for each effect, namely SEM-Pooling, SEM-Fixed Effects Model, and SEM-Random Effects Model. Based on these results, the regression equation for each model can be made as follows:

SEM-*Pooling*

$$y_{it} = 38.2424 + 0.1303x_{1it} + 0.8461x_{2it} - 0.0463x_{3it} - 8.9662x_{4it} + \mu_i + \varepsilon_{it}$$

SEM-*Fixed Effects Model*

$$y_{it} = 0.0689 \sum_{j=1}^{N} W_{ij}\phi_{jt} - 0.1834x_{1it} + 0.1395x_{2it} + 0.0073x_{3it} - 3.6525x_{4it} + \mu_i + \varepsilon_{it}$$

SEM-*Random Effects Model*

$$y_{it} = 0.0689 \sum_{j=1}^{N} W_{ij}\phi_{jt} + 24.3223 - 0.1127x_{1it} + 0.1420x_{2it} + 0.0084x_{3it} - 5.2016x_{4it} + \mu_i + \varepsilon_{it}$$

3.7 Significance Testing of Spatial Panel Data Model

Parameter significance test is one of the important things in Spatial Panel Data modeling. This of course aims to determine whether the estimated parameters obtained have a significant effect on the model or not. To see the significance of the parameters can be seen in Table 10 and Table 11.

Based on Table 10 of the SAR model parameter estimation, it can be seen that for the SAR-Pooling model, the variables that have a significant effect on the Poverty Percentage (Y) are Average Years of Schooling (X1), GRDP (X2), and Life Expectancy (X4), this can be seen from the $p_{value} < \alpha = 0.05$. Meanwhile, the Open Unemployment Rate variable (X3) has no significant effect on the Poverty Percentage variable (Y), this can be seen from the $p_{value} > \alpha = 0.05$. For the SAR-Fixed Effects Model and SAR-Random Effects Model, it can be seen that the Average Years of Schooling (X1) and Life Expectancy (X4) variables have a significant effect on the Poverty Percentage (Y), while the GRDP (X2) and Open Unemployment Rate (X3) variables do not have a significant effect on the Poverty Percentage (Y) variable.

Based on Table 11 of the SEM model parameter estimation, it can be seen that for the SEM-Pooling model, the variables that have a significant effect on the Poverty Percentage (Y) are Average Years of Schooling (X1), GRDP (X2), and Life Expectancy (X4), this can be seen from the $p_{value} < \alpha = 0.05$. Meanwhile, the Open Unemployment Rate variable (X3) has no significant effect on the Poverty Percentage variable (Y), this can be seen from the $p_{value} > \alpha = 0.05$. For the SEM-Fixed Effects Model and SEM-Random Effects Model, it can be seen that the Average Years of Schooling (X1) and Life Expectancy (X4) variables have a significant effect on the Poverty Percentage (Y), while the GRDP (X2) and Open Unemployment Rate (X3) variables do not have a significant effect on the Poverty Percentage (Y) variable.

In general, to determine the best model, it can be seen based on the AIC and BIC values in the following table:

Table 12. Goodness of Fit

Goodness of Fit	SAR			SEM		
	Pooling	Fixed Effects Model	Random Effects Model	Pooling	Fixed Effects Model	Random Effects Model
loglikelihood	-182.865	765.8135	535.6126	-182.895	94.54415	534.504
AIC	375.7309	-1523.627	-1061.23	375.7899	-181.088	-1059.01
BIC	395.9322	-1507.466	-1041.02	395.9911	-164.927	-1038.81

Based on the results of Table 10 of the Goodness of Fit criteria, the best model to model the poverty percentage data in the South Sumatra region is the Spatial Autoregressive Fixed Effect (SAR-FE) model, because it has the highest log-likelihood value as well as the lowest AIC and BIC values, which indicates that this model is most suitable in capturing the characteristics of the data.

3. CONCLUSION

Based on the results of existing research, it can be concluded that the best model that can be used to model data on the percentage of poverty in the South Sumatra region is the Spatial Autoregressive Fixed Effect (SAR-FE) model, with variables that are significant to the Poverty Percentage (Y) are the Average Years of Schooling (X1) and Life Expectancy (X4) variables. This best model was chosen based on the highest log-likelihood value and the smallest AIC and BIC values, which indicates that this model is most suitable for capturing the characteristics of the data in modeling the percentage of poverty in the South Sumatra region.

REFERENCES

[1] Badan Pusat Statistik. (2023). Statistik Kemiskinan Regional Indonesia: Fokus Sumatera. Jakarta: BPS.
[2] Kementerian PPN/Bappenas. (2022). Peta Jalan Pengentasan Kemiskinan di Wilayah Sumatera. Jakarta: Bappenas.
[3] Anselin, L. (2013). Spatial Econometrics: Methods and Models. Dordrecht: Springer Science & Business Media.
[4] World Bank. (2021). Indonesia Poverty Assessment: Towards a Comprehensive Approach to Poverty Reduction. Washington, DC: World Bank Group.
[5] LeSage, J., & Pace, R. K. (2019). Introduction to Spatial Econometrics. Boca Raton: CRC Press.
[6] Elhorst, J. P. (2014). Spatial Panel Data Models. In Spatial Econometrics (pp. 37-93). Berlin: Springer.
[7] Jajang, Saefuddin, A., Mangku, I. W., & Siregar, H. (2013). Analisis Kemiskinan menggunakan Model Spatial Autoregressive Panel di Jawa Barat. Forum Statistika dan Komputasi, 18(1), 21-27.
[8] Ahlburg, D. A. (2017). Regional Poverty Dynamics in the Pacific: A Spatial Panel Analysis. Pacific Economic Bulletin, 32(2), 80-94.
[9] Kementerian PUPR. (2022). Pemetaan Infrastruktur dan Pengembangan Wilayah Sumatera Bagian Selatan. Jakarta: PUPR.
[10] Bank Indonesia. (2023). Kajian Ekonomi dan Keuangan Regional: Sumatera. Jakarta: Bank Indonesia.
[11] Greene, W. H. (2018). Econometric Analysis (8th ed.). New York: Pearson.
[12] Baltagi, B. H. (2021). Econometric Analysis of Panel Data (6th ed.). Cham: Springer.
[13] Wooldridge, J. M. (2020). Introductory Econometrics: A Modern Approach (7th ed.). Boston: Cengage Learning.
[14] Gujarati, D. N. (2009). Basic Econometrics (5th ed.). New York: McGraw-Hill.