BAREKENG: Journal of Mathematics and Its Applications
September 2023 Volume 17 Issue 3 Page 1203–1214
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss3pp1203-1214

POVERTY PANEL DATA MODELING IN SOUTH SUMATRA

Nurul Hidayati[1*], **Herlin Fransiska**[2], **Winalia Agwil**[3]

[1,2,3] *Statistics Study Program, Faculty of Mathematics and Natural Sciences, Bengkulu University*
Wr. Supratman Street, Kandang Limun, Bengkulu City, 38371, Indonesia

*Corresponding author's e-mail: * **nurulhidayati@unib.ac.id***

Article History:

Received: 17th November 2022
Revised: 17th June 2023
Accepted: 28th June 2023

Keywords:

Fixed Effect Model (FEM);
Panel data;
Poverty;
Southern Sumatra.

ABSTRACT

Poverty is still a major problem on Sumatra's island, despite abundant natural resources potential, such as mining and plantation products. Sumatra Island consists of 10 provinces divided into three regions: Northern Sumatra, Central Sumatra, and Southern Sumatra. The island of Sumatra has the highest number of poor people in the Southern Sumatra region, which reaches 2.88 million people. Poverty is an integrated concept with five dimensions: poverty, powerlessness, vulnerability to emergencies, dependency, and alienation, both geographically and sociologically. One method that can be used to analyze poverty data problems is panel data regression analysis, which combines two data, namely cross-sectional data and time series data. It is expected to produce more in-depth and comprehensive information, both the interrelationships between the variables and their development within a certain period. The panel data was related to poverty and included 60 districts in the Southern Sumatra region from 2018 to 2020. This study aimed to model poverty panel data in Southern Sumatra. Three estimation methods were used in the panel data regression, including the Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM). The results of the model specification test show that the best model for estimating the percentage of poor people in the Southern Sumatra region is the Fixed Effect Model (FEM), with a value of R2 = 75.57%. The results of the significance test show that the variables that significantly influence the percentage of poverty in the Southern Sumatra region using the FEM model are the open unemployment rate (X_1), life expectancy (X_3), and the average length of schooling (X_4).

How to cite this article:
N. Hidayati, H. Fransiska and W. Agwil., " POVERTY PANEL DATA MODELING IN SOUTH SUMATRA," *BAREKENG: J. Math. & App.*, vol. 17, iss. 3, pp. 1203-1214, September, 2023.

1. INTRODUCTION

One of the islands located in Indonesia is Sumatra, with an area of approximately 473,481 km^2, making it the sixth largest island in the world and the second most populous island after Java in Indonesia. Sumatra also has Indonesia's second-largest population; one-third of Indonesia's provinces are located within it. The Central Bureau of Statistics (BPS) noted in the 2020 population census that 21.68 percent of Indonesia's population *de facto* resides on the island of Sumatra. When viewed from the number of poor people, there are 5.83 million on Sumatra Island, or 22.06 percent of the total number of poor people in Indonesia. Sumatra Island also has the potential for abundant natural resources, such as mining and plantation products. However, it is not proportional to the welfare of its people. Poverty is still a major problem on the island of Sumatra. The high poverty rate in Sumatra Island is "homework" for the government to continue to launch various appropriate and targeted reduction policies for the people of Sumatra Island [1].

Sumatra Island consists of ten provinces divided into three parts of the region: Northern Sumatra, Central Sumatra, and Southern Sumatra. Northern Sumatra consists of the provinces of Aceh and North Sumatra, and Central Sumatra consists of the provinces of West Sumatra, Riau, and Riau Islands. Southern Sumatra consists of the provinces of Jambi, Bengkulu, Lampung, Bangka Belitung Islands, and South Sumatra. Based on the 2021 publication of the Central Bureau of Statistics, the highest number of poor people on the island of Sumatra is in the Southern Sumatra region, reaching 2.88 million.

To measure poverty, BPS (Central Bureau of Statistics) uses the ability to fulfill basic needs (basic needs approach). Poverty is seen as an economic inability to fulfill basic food and non-food needs measured in expenditure. Therefore, the definition of poor is a population with an average per capita expenditure below the poverty line. In addition, other factors that become BPS measurement tools to calculate the poverty percentage (%) are the number of poor people (thousand people), the Poverty Severity Index (%), and the Depth Index (%)[1].

An *integrated concept* that has five dimensions, namely poverty, powerlessness, vulnerability to emergencies, dependency, and isolation both geographically and sociologically, is called poverty[2]. Panel data regression analysis is one method that is very suitable for analyzing poverty data problems because it combines two data, namely data across individuals (cross-section) and time series data (time series). Therefore, combining the two data can control individual diversity and is expected to produce more in-depth and comprehensive information, both the relationship between variables and their development in a certain period [3].

Previous research conducted by "Wahyudi" [4] modeled and found the factors affecting poverty in Aceh Province in 2016-2020. The results showed that the panel data regression model with fixed effects (fixed effect model) is a model that fits the poverty data of Aceh Province in 2016-2020. The factors that affect poverty are the human development index, morbidity rate, and population growth rate with a high R squared value of 97.14%. Further research was conducted to model and determine the significant factors affecting the percentage of poor people in Bali Province by applying panel data analysis. The results of the study "Budinirmala et al." [5] showed that the best model to describe the causal relationship between variables is the panel data regression model with fixed effects, and the factors that significantly affect the percentage of poor people in Bali Province are the economic growth rate and the human development index. Previous research only used one approach in estimating parameters, namely fixed effect. Therefore, this study will use three estimation approaches: the Common Effect Model, Fixed Effect Model, and Random Effects Model, which will later select the best model with the Chow and Hausman tests in making Poverty Panel Data Modeling in Southern Sumatra.

2. RESEARCH METHODS

2.1 Research Objects and Research Variable

The variables used in this study are presented in Table 1, below:

Table 1. Research Variables

Variables	Symbol	Indicator	Unit
Response (Y)	Y	Percentage of Poor Population	Percent
Predictor (X)	X_1	Open Unemployment Rate	Percent
	X_2	Gross Regional Domestic Product Growth Rate	Percent
	X_3	Life Expectancy Rate (AHH)	Percent
	X_4	Average years of schooling	Percent

2.2 Stages of analysis

1. Performing parameter estimation of panel data regression models using the Common Effect Model, Fixed Effect Model, and Random Effects Model estimation approaches.

In estimating the parameters of the panel data regression model, the method highly depends on the assumptions about the intercept, slope coefficient, and error [6]. There are three methods of estimating panel data regression models that can be viewed from various assumptions and factors of its formation [7], namely:

i. Common Effect Model (CEM)

The Common Effect Regression Model aims to estimate panel data by combining time series and cross-section data without looking at differences between time and individuals. The CEM model is as follows:

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

ii. Fixed Effect Model (FEM)

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

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

iii. Random Effect Model (REM)

This method estimates panel data with residuals that may be correlated across time and individuals, assuming each subject has a different intercept. This model is particularly useful if the randomly selected individuals represent the population[8]. The objects used are usually randomly selected objects from a large population. The assumptions that must be met are μ_i stochastic free spread, normal $(0.\sigma_\mu^2), \varepsilon_{it}$ identical stochastic free spread, normal $(0.\sigma_\varepsilon^2)$; $E(x_{it}, \mu_i) = 0$ and $(x_{it}, \varepsilon_{it}) = 0$ or x_{it} are mutually independent with, μ_i and ε_{it} for every *i* and *t* [9]. The estimation method for the REM model is Generalized Least Square (GLS). The REM model equation is as follows:

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

2. Conduct the best Panel Data Model selection test using the Chow, Hausman, and Breusch-Pagan tests.

Model selection is statistically carried out so that the estimates obtained can be as efficient as possible. Testing in determining the model to be used in panel data processing [10], namely:

i. Chow *Test*

The Chow test chooses between the Fixed Effect Model (FEM) or the Common Effect Model (CEM). The test statistic used is the F test, which is:

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

With n is the number of individual units, T is the observation period, and k is the number of independent variables in the fixed effects model. The test criteria used are rejected H_0 if the value $F_{count} > F_{table}$, with $F_{table} = F_{\alpha;n-1;n(T-1)-k}$ or reject H_0 [9]. If the test is significant, then the appropriate model is the FEM model. Conversely, if the test is insignificant, then the CEM model is the appropriate model.

ii. Hausman Test

The Hausman Test aims to see if there is a random effect in the panel data[10]. The Hausman test is used to select the REM and FEM models with the test statistic:

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

with

$$\hat{q} = \hat{\beta}_{random} - \hat{\beta}_{constant}$$

The decision of reject H_0 if $\chi^2 > \chi^2_{k,\alpha}$ with k being the number of explanatory variables, or reject H_0 if $p < \alpha$ [11]. If the test result is significant, then the selected model is the FEM model. Vice versa, if the test results are not significant, then the selected model is the REM model.

3. Conduct significance tests of model parameters, namely simultaneous test (F test) and partial test (t-test).

i. Test the significance of regression coefficients simultaneously.

A simultaneous test or F-test is conducted to determine the effect of independent variables with the dependent variable together. The hypothesis used:

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (There is no influence between all independent variables on percentage of the poor population simultaneously)

H_1: there is at least 1 $\beta_j \neq 0$; $j = 1,2,3,4$; (there is an influence between all independent variables on the percentage of poor people simultaneously.

Statistics Test:

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \quad (6)$$

with:

N : total number of observations,
R^2 : coefficient of determination, and
k : number of independent variables.

The test criteria used are if $F_{count} > F_{table\,(k;NT-N-K,\alpha)}$ or $p - value < \alpha$, which means that at least one independent variable has a significant effect on the dependent variable [11].

ii. Partial regression coefficient significance test

A 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 used:

$$t_{count} = \frac{\hat{\beta}_p}{se(\hat{\beta}_p)} \quad (7)$$

The test criterion is to reject H_0 if $|t_{count}| > t_{(\frac{\alpha}{2},N-k-1)}$ or $p - value < \alpha$, which means that the p-th independent variable significantly affects the dependent variable[12].

4. Testing panel data regression assumptions, namely Normality Test and Multicollinearity Test.

The panel data regression model is good if the model meets the Best, Linear, Unbiased, and Estimator (BLUE) criteria. BLUE can be achieved if the classical assumptions are met [13]. The classical assumption test includes normality, multicollinearity, linearity, heteroscedasticity, and autocorrelation tests. If CEM with OLS estimation and FEM with LSDV estimation are the best models, then normality and autocorrelation tests do not need to be done. Meanwhile, if the best model is REM with GLS estimation, heteroscedasticity, and autocorrelation tests do not need to be done because REM is estimated using the GLS method, where the GLS method is assumed to be able to solve heteroscedasticity problems. Autocorrelation is not required in panel data (CEM, FEM, and REM) because autocorrelation is only in time series data [14].

a. Normality Test

The normality test used in this study is the Kolmogorov-Smirnov test. The hypothesis used is as follows:

H_0: normally distributed errors
H_1: abnormally distributed errors
Test Statistics:

$$D_{count} = \max_{1<i<N}\left(F(Y_i) - \frac{i-1}{N}, \frac{i-1}{N} - F(Y_i)\right) \quad (8)$$

Rejection Criteria
Reject H_0 if probability value < α, meaning the error is not normally distributed.
Accept H_0 if probability value > α, meaning the error is normally distributed.

b. 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 Inflation Factor (VIF) value with the formula:

$$VIF_j = \frac{1}{1-R_j^2} \quad (9)$$

Hypothesis:
H_0: there is no multicollinearity
H_1: there is multicollinearity
Test Statistics
$VIF_j = \frac{1}{1-R_j^2}$; where j = 1,2,...k
Rejection Criteria
Reject H_0if the VIF value > 10, meaning that there is multicollinearity
Accept H_0 if the VIF value < 10, meaning there is no multicollinearity.

c. Autocorrelation Test

Autocorrelation is the correlation between one observation's error and another's error. 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)[15]. The autocorrelation test is a statistical analysis conducted to determine whether variables are correlated 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:

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \quad (10)$$

Rejection Criteria

Table 2. Decision-making on autocorrelation assumptions

Null Hypothesis	Decision	If
There is negative autocorrelation	Reject	0<dw<dl
There is negative autocorrelation	No decision	$dl \leq dw \leq du$
There is positive autocorrelation	Reject	$4 - dl < dw < 4$
There is positive autocorrelation	No decision	$4 - du \leq dw \leq 4 - dl$
Don't reject	No autocorrelation, positive or negative	$du < dw < 4 - du$

Source: [16]

d. Heteroscedasticity Test

To determine whether there is an unequal variance from the error of one observation to another, a heteroscedasticity test is performed, 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:

$$LM = \frac{NT}{2(t-1)}\sum_{i=1}^{N}\left[\frac{[\sum_{t=1}^{T}\varepsilon_{it}]^2}{\sum_{t=1}^{T}\varepsilon_{it}^2} - 1\right]^2 \tag{11}$$

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

3. RESULTS AND DISCUSSION

3.1 Data Description

The descriptive results of the research variables are presented in Table 3 below:

Table 3. Descriptive Statistics of Research Data

Variables	Average	Median	Minimum Value	Maximum Value
Y	11.215	11.47	2.67	20.85
X1	4.114	3.84	1.41	10.49
X2	2.937	4.745	-5.43	8.65
X3	68.7	68.73	62.73	73.3
X4	8.121	7.85	6.34	11.79

Source: Analysis Result (2022)

Based on Table 3, the number of observations is 180 from 60 districts/cities in the Southern Sumatra region during the 2018-2020 research period. Based on Table 2, it can be seen that the average value of the percentage of poor people in the Southern Sumatra Region is 11.215 percent. The highest percentage of poor people was in North Lampung at 20.85 percent in 2018, while the lowest was in West Bangka Regency, which amounted to 2.67 percent in 2019. The open unemployment rate variable (X_1) has an average percentage of 4.114 percent in the 2018-2020 period in the Southern Sumatra region. The lowest open unemployment rate was in West Tanjung Jabung Regency, which amounted to 1.41 percent in 2020, and the highest was in Jambi City in 2020, which amounted to 10.49 percent. Furthermore, the average GRDP value (X_2) in the Southern Sumatra Region in 2018-2020 amounted to 2,937 percent. The lowest GRDP growth rate was in the West Bangka district/city in 2020, which was -5.43, and the highest was in Muara Enim in 2018, at 8.65. The life expectancy variable (X_3) has an average of 68 percent in the 2018-2020 period in the Southern Sumatra Region. The lowest life expectancy was in Lebong Regency in 2018 at 62.73 percent, while the highest was in Bengkulu City in 2020 at 11.79 percent. Furthermore, the average length of schooling variable (X_4) in the 2018-2020 period in the Southern Sumatra Region has an average of 8.121 percent. The lowest average length of schooling was in East Tanjung Jabung Regency in 2018, which amounted to 6.34 percent, and the highest was in Bengkulu City in 2020, at 11.79 percent.

3.2 Panel Data Regression Estimation Method

1. Common Effect Models (CEM)

The CEM method is a simple technique for estimating the parameters of panel data regression models by combining time series and cross-section data without looking at differences between time and individuals[17]. The CEM method assumes that the behavior of data between spaces is the same in various periods. The parameter estimation results of the CEM model with the assistance of R software are presented in Table 4 as follows:

Table 4. Parameter Estimation in CEM

Variables	Estimated value	Standard Deviation	Value of t	*p-value*
Intercept	84.342785	9.573058	8.8104	$1.202\text{x}10^{-15}$
X1	-0.154553	0.216050	-0.7154	0.475341
X2	0.066067	0.092618	0.7133	0.476594
X3	-1.158224	0.148202	-7.8152	$4.901\text{x}10^{-13}$
X4	0.847181	0.297286	2.8497	0.004901

Source: Analysis Result (2022)

Based on **Table 4**, the equation of the panel data regression model for CEM is as follows:

$$Y_{it} = 84.342785 - 0.154553X_{1_{it}} + 0.066067X_{2_{it}} - 1.158224X_{3_{it}} + 0.847181X_{4_{it}}$$

The results of the analysis show that the value of R^2 is very small, namely 0.287, which means that the percentage of poor people is influenced by the open unemployment rate, GRDP growth rate, life expectancy, and average years of schooling amounted to 28.7% and other variables outside this study explained the rest.

2. Fixed Effect Models (FEM)

The estimation technique using dummy variables to capture differences in intercepts between variables but simultaneously intercepts is called the FEM estimation model. This model assumes that the regression coefficient (slope) is fixed between variables and between times. The estimation results of FEM obtained with the assistance of R software are presented in **Table 5** below:

Table 5. Parameter Estimation in FEM

Variables	Estimated value	Standard Deviation	Value of t	*p-value*
X1	0.145075	0.051117	2.8381	0.005359
X2	-0.011545	0.013919	-0.8294	0.408589
X3	-2.090121	0.177974	-11.7439	2.2×10^{-16}
X4	1.131925	0.394825	2.8669	0.004924

Source: Analysis Result (2022)

The panel data regression model estimation for FEM based on the results from **Table 5** can be written in the form of the following equation:

$$Y_{it} = 0.145075X_{1_{it}} - 0.011545X_{2_{it}} - 2.090121X_{3_{it}} + 1.131925X_{4_{it}}$$

The R^2 value from the FEM model estimation of 0.757, shown in **Table 5**, explains that the percentage of poor people is influenced by the open unemployment rate, GRDP growth rate, life expectancy, and average years of schooling by 75.6% and other variables outside the model can explain the rest.

3. Random Effect Model (REM)

The REM method estimates panel data with residuals that may be interconnected between time and individuals, assuming each subject has a different intercept. REM estimation was done with the GLS technique and viewed the slope α as a random variable, in addition to the residual component μ_{it}. The estimation results of the REM model are as follows:

Table 6. Parameter Estimation in REM

Variables	Estimated value	Standard Deviation	Value of t	*p-value*
Intercept	130.497872	7.968846	16. 3760	2.2×10^{-16}
X1	0.158743	0.052200	3.0410	0.002358
X2	-0.009148	0.014032	-0.6520	0.514424
X3	-1.834672	0.140590	-13.0498	2.2×10^{-16}

Source: Analysis Result (2022)

The equation form of the panel data regression model for REM from **Table 6** is as follows:

$$Y_{it} = 130.497872 + 0.158743X_{1_{it}} - 0.009148X_{2_{it}} - 1.834672X_{3_{it}} + 0.754427X_{4_{it}}$$

Based on the regression results with the REM model in **Table 6**, the R^2 value is 0.664, which explains that the percentage of poor people is influenced by the open unemployment rate, GRDP growth rate, life expectancy, and average years of schooling by 66.4% and other variables outside this study explain the rest.

3.3 Selection Best Model of Panel Data Regression

The model selection to be used in research is necessary based on statistical considerations. There are two methods in model selection, namely the Chow Test and the Hausman Test.

1. Chow Test

The Chow test aims to select the model to be used between the CEM estimation model and the FEM estimation model, with the hypothesis test:

$H_0: \beta_{01} = \beta_{02} = \cdots = \beta_{0n} = 0$ (The selected model is CEM)
H_1: there is at least 1 $\beta_{0i} \neq 0$ (The selected model is FEM) with $i = 1.2.\dots.n$

If the p-value < 0.05 (significant), then the model used is FEM. Conversely, if the p-value > 0.05 (not significant), then the model used is CEM. The results of the Chow test analysis are presented in **Table 7** below:

Table 7. Chow Test Results

Effect Test	Statistic	Df	p-value
Cross-section F	209.77	(59.116)	2.2 x10^{-16}

Based on **Table 7**, the value of the -F test statistic is 209.77 with a p-value of 2.2 x10^{-16} (less than 0.05), so the statistic is H_0 is rejected and accepts H_1. These estimation results prove that the selected model is the FEM estimation model. Furthermore, the Hausman test process can be carried out.

2. Hausman Test

The Hausman test is used to select a model, namely between the FEM estimation model or the REM estimation model, with the hypothesis test:

H_0: corr $(X_{it}.\epsilon_{it}) = 0$ (REM is better).
H_1: corr $(X_{it}.\epsilon_{it}) \neq 0$ (FEM is better).

This Hausman test can be done by looking at the p-value. If the p-value < 0.05 (significant), then the model used is the FEM estimation model. Conversely, if the p-value > 0.05 (insignificant), then the REM estimation model is used. The Hausman Test results are presented in **Table 8** as follows:

Table 8. Hausman Test Results

Test summary	Chi-square Statistic	Chi-square d.f	p-value
Random Period	14.786	4	0.005167

The value of the *chi-square* statistical distribution obtained from the calculation and presented in **Table 8** is 14.786, and the p-value is 0.005167 < 0.05 (significant), so statistically, H_0 rejected and accepted H_1. Then, the model used is the FEM estimation model.

From the tests conducted by comparing the CEM, FEM, and REM models through the cow test and Hausman test, the FEM panel data regression model is very good for modeling and explaining the poverty percentage data in the Southern Sumatra region. Therefore, the model used in this study used the FEM estimation model.

3.4 Significance Test of Panel Data Regression Parameters

The panel regression model parameter significance test is used to determine the causal relationship between the independent variable and the dependent variable. There are two stages in testing the panel regression model parameters, namely simultaneous test and partial test.

1. Simultaneous regression coefficient significance test (F-test)

The F-test determines the relationship between the independent and dependent variables by comparing the probability of F with an alpha value of 0.05. The hypothesis formulation for the F-test is as follows:

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (There is no effect between all independent variables on the percentage of poor people simultaneously)
$H_1: there\ is\ at\ least\ 1\ \beta_j \neq 0; j = 1,2,3,4$ (There is an influence between all independent variables on the percentage of poor people simultaneously)

The results of the F-test can be shown in **Table 9** below:

Table 9. F-test Results

F test	P-value
90.320	< 2.22e-16

In **Table 9**, a p-value obtained is $2.22 \times 10^{-16} < 0.05$, so it can be concluded that all independent variables simultaneously significantly affect the percentage of poor people.

2. Partial regression coefficient significance test (t-test)

The t-statistical test shows how far the influence of one independent variable individually is in explaining the variations in the dependent variable. The partial test aims to determine whether the dependent variable influences the independent variable by comparing the probability value of t with α=0.05. The hypothesis of the t-test is as follows:

$H_0: \beta_j = 0$
H_1: there is at least 1 $\beta_j \neq 0; j = 1,2,3,4$

The results of testing the significance of the model partially (t-test) are presented in **Table 10** below:

Table 10. T-Test Results

Parameters	Estimation	Std. Error	t-value	p-value
$\hat{\beta}_1$	0.145	0.051	2.838	0.0054**
$\hat{\beta}_2$	-0.012	0.014	-0.829	0.409
$\hat{\beta}_3$	-2.090	0.178	-11.744	<2.2e-16***
$\hat{\beta}_4$	1.132	0.395	2.867	0.0049**

(*) significance at real level $\alpha = 0.05$

Based on **Table 10**, panel data regression model obtained, namely:

$$Y_{it} = 0.145075X_{1_{it}} - 2.090121X_{2_{it}} + 1.131925X_{3_{it}}$$

The results of the T-test in **Table 10** show that the variables X_1, X_3, and X_4 have p-value < 0.05. It means that the open unemployment rate (X_1), life expectancy (X_3), and average years of schooling (X_4) have a significant effect on the percentage of poor people.

3.5 Panel Data Regression Classic Assumption Test

Based on the results of the T-test, variables of the open unemployment rate (X_1), life expectancy (X_3), and average years of schooling (X_4) have a significant effect on the percentage of poor people. The selected model is FEM as the best model, so the normality and autocorrelation tests do not need to be carried out. The classical assumption tests carried out in this study are as follows:

1. Multicollinearity Test

The classic assumption test used to see whether there is a correlation between the independent variables is called the multicollinearity test by calculating the Variance Inflation Factor (VIF) value. If the VIF value is > 10, it can be concluded that multicollinearity occurs in the independent variables. The VIF value for the variables of the open unemployment rate (X_1), life expectancy (X_3), and average years of schooling (X_4) are presented in **Table 11** below:

Table 11. Multicollinearity Test

Label	Variables	VIF
X_1	Open unemployment	1.553
X_3	Life expectancy	1.287
X_4	Average years of schooling	1.534

Based on the output of **Table 11**, it can be seen that the VIF value of the open unemployment rate variable (X_1), life expectancy (X_3), and average years of schooling (X_4) < 10, so it can be concluded that the resulting model does not have multicollinearity problems.

2. Heteroscedasticity Test

The heteroscedasticity test determines whether there is an inequality of variance from one observation's error to another. The results of the heteroscedasticity test can be seen in **Table 12** as follows:

Table 12. Breusch-Pagan test

Brush Pagan	p-value
4.566	0.0007235

The output results in **Table 12** show that the value of Breusch Pagan is 4.566, and the p-value is 0.0007235 > 0.05 (significant), rejecting the Breusch Pagan value. H_0. It means that there are no symptoms of heteroscedasticity.

3.6 Model Interpretation

From the results of the analysis, it can be concluded that the selected model in this study is the Fixed Effect Model (FEM), and the variables that have a significant effect are open unemployment, life expectancy ((X_1), life expectancy (X_3), and average years of schooling (X_4). The regression model can be expressed as follows:

$$\hat{Y}_{it} = 0.145075X_{1_{it}} - 2.090121X_{3_{it}} + 1.131925X_{4_{it}}$$

Based on the model obtained, if there is an increase in the level of open unemployment (X_1), the poverty percentage will increase by 0.145075 with the assumption that the other variables are considered constant. If life expectancy (X_3) increases by one unit, the percentage of poor people will decrease by 2.090121 because other independent variables are considered constant. If there is an increase in the average years of schooling (X_4), the percentage of poor people will increase by 1.131925, provided that the other independent variables are considered constant. Based on the FEM model, the R^2 value is 0.757 or 75.7%.

4. CONCLUSIONS

The results of the analysis show that the best and selected model for the percentage of poverty in South Sumatra in 2018-2020 is the FEM model with the following equation:

$$\hat{Y}_{it} = 0.145075X_{1_{it}} - 2.090121X_{3_{it}} + 1.131925X_{4_{it}}$$

Retrieved from $\hat{Y}_{it}$: the estimated percentage of poor people in the i-th district and year t; $X_{1_{it}}$: open unemployment rate of the i-th district/city and year t;$X_{3_{it}}$:life expectancy of the i-th district/city and year t; $X_{4_{it}}$ average years of schooling in the i-th district and year t.

Based on the model obtained, if there is an increase in the open unemployment rate of $(\boldsymbol{X_1})$ by one percent, then the poverty percentage will increase by 0.145075 percent with the assumption that other variables are considered constant. If life expectancy $(\boldsymbol{X_3})$ increases by one percent, the percentage of poor people will decrease by 2.090121 percent, provided that the other independent variables are held constant. Suppose there is an increase in the average years of schooling $(\boldsymbol{X_4})$ by one percent. In that case, the percentage of poor people will increase by 1.131925 percent, provided that the other independent variables are considered constant. Based on the FEM model, the R^2 value is 0.757 or 75.7%.

ACKNOWLEDGMENT

The authors would like to give their gratitude to the Faculty of Mathematics and Natural Sciences, Bengkulu University for funding this research with contract number: 1970/UN30.12/HK/2022

REFERENCES

[1] S. Abe, Support Vector Machines for Pattern Classification, Second ed., London: Springer-Verlag, 2010.

[2] Arifin, Sinta, Skripsi Pengambilan Keputusan untuk Penilaian Kinerja Menggunakan Analytic Hierarchy Process (AHP) [Thesis on Decision Making for Performance Appraisal Using Analytic Hierarchy Process (AHP)], Yogyakarta: Yogyakata State University, 2013.

[3] Brunelli, M, Introduction to the Analytic Hierarchy Process, New York: Springer, 2015.

[4] Wahyudi, "Analisis Regresi Data Panel Pada Kemiskinan Provinsi Aceh Tahun 2016-2020," ["Panel Data Regression Analysis on Poverty in Aceh Province 2016-2020"] *Journal of Mathematics and Applied Sciences*, vol. 18, no. 2, pp. 264-271, 2021.

[5] K. Budinirmala, P. Suciptawati, K. Jayanegara and E. N. Kencana, "Memodelkan Kemiskinan Penduduk Provinsi Bali dengan Regresi Data Panel," ["Modeling Poverty in Bali Province with Panel Data Regression"] *E-Journal of Mathematics,* vol. 7, no. 3, pp. 219-225, August 2018.

[6] Hastuti, D. Dwi Puji, "Proses Hierarki Analisis (PHA) dalam Pengambilan Keputusan Multikriteria. Jurusan Pendidikan Matematika: Universitas Negeri Yogyakata.," ["Hierarchical Process Analysis (PHA) in Multicriteria Decision Making. Department of Mathematics Education: State University of Yogyakata,"] 2010.

[7] Kelelufna, Solomon, "Skripsi Aplikasi Proses Hirarki Analitik (PHA) Sistem Pemilihan Keputusan Pembelian Sepeda Motor. Jurusan Matematika: Universitas Pattimura.," ["Thesis Application of Analytic Hierarchy Process (PHA) Motorcycle Purchase Decision Selection System. Department of Mathematics: Pattimura University,"] 2016.

[8] Manggiora, G. M, "Computer-Aided Decision Making in Pharmaceutical Research. Proceedings of the Beilstein-Institut Workshop, Bozen, Italy," 2002.

[9] E. Mu, "Practical Decision Making. Springer Briefs in Operations Research," 2017.

[10] Permadi, Bambang, "AHP Pusat Antar Universitas – Studi Ekonomi. Jakarta: Universitas Indonesia.," ["AHP Inter-University Center - Economic Studies. Jakarta: University of Indonesia,"] 1992.

[11] Sukarto, Haryono, "Pemilihan Model Transportasi di DKI Jakarta dengan AnalisisKebijakan “Proses Hierarki Analitik” ["Selection of Transportation Model in DKI Jakarta with "Analytic Hierarchy Process" Policy Analysis,"] *Journal of Civil Engineering,* p. Vol. 3 No.1, 2006.

[12] S.Abe, Support Vector Machines for Pattern Classification, second ed., London: Spinger-Verlag, 2010.

[13] L. Fausett, Fundamentals of Neural Networks: Architectures, Fundamentals, and Applications, New Jersey: Prentice-Hall, Inc. 1994.

[14] N. I. Pradasari, F. T. P. W. and D. Triyanto, "Aplikasi Jaringan Saraf Tiruan untuk memprediksi Penyakit Saluran Pernapasan dengan Metode Backpropagation," ["Application of Artificial Neural Network to predict Respiratory Disease with Backpropagation Method,"] *Untan Computer System Coding Journal*, vol. 1, no. 1, 2013.

[15] H. Sasrawan, "Kehidupan Awal Masyarakat Indonesia," ["The Early Life of Indonesian Society,"] January 2, 2013. [Online]. Available: http://hedisasrawan.blogspot.co.id/2013/01/kehidupan-awal-masyarakat-indonesia.html?m=1. [Accessed November 5, 2015].

[16] I. Ghozali, Application of Multivariate Analysis with the IBM SPSS 21 Update PLS Regression Program, Semarang: Publishing Agency: Diponegoro University, 2013.

[17] A. Widarjono, Ekonometrika Pengantar daan Aplikasinya [Econometrics Introduction and Application], Yogyakarta: EKONISIA, 2009.