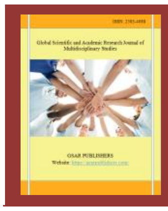

Global Scientific and Academic Research Journal of Multidisciplinary Studies
ISSN: 2583-4088 (Online)
Frequency: Monthly
Published By GSAR Publishers
Journal Homepage Link- https://gsarpublishers.com/journals-gsarjebm-home/

ANALYSIS OF FACTORS AFFECTING POVERTY IN PAPUA ISLAND USING SPATIAL PANEL DATA FOR 2017 - 2020

BY

Delia Oktaviani[1], Dian Lestari[2*], Yekti Widyaningsih[3]

[1,2,3]Department of Mathematics, Universitas Indonesia

Article History

Received: 01/12/2023
Accepted 05/12/2023
Published: 07/12/2023

Vol – 2 Issue –12

PP: - 07-12

Abstract

Poverty is a condition characterized by the inability of people to meet their basic needs. Based on the publication from Badan Pusat Statistik (BPS) Indonesia in 2017 – 2020, Papua and West Papua are the two provinces with the highest poverty rate in Indonesia. This study aims to analyze the variables that significantly affect poverty in Papua Island from 2017 - 2020. Poverty is a problem that does not only occur at one time and poverty can be influenced by spatial aspects, so it is necessary to conduct research involving spatial effects over several periods. The spatial heterogeneity test was performed using the Breusch-Pagan test. Based on the test, it was found that there is spatial heterogeneity in poverty data in Papua Island from 2017 - 2020, so further analysis was carried out using the Geographically Weighted Panel Regression (GWPR) model. The results showed that the GWPR model with the fixed gaussian kernel weighting function was the best model with a coefficient of determination of 84.93% and an RMSE of 0.0135. The variables of school year expectancy, life expectancy, gini ratio, consumption per capita, and labor force participation rate have a significant effect on poverty in at least one location in a district/city in Papua Island.

Keywords*: Geographically Weighted Panel Regression, Heterogeneity Spatial, Weighted Least Square method*

Introduction

Poverty is a social issue that affects every country in the world. Indonesia, as a developing country, cannot be separated from the problem of poverty. Poverty reduction is a big challenge faced by the government. Based on poverty data from Badan Pusat Statistik (BPS), the poverty rate in Indonesia from 2017 to 2019 tends to decrease, from 10.64% to 9.41% of the population (BPS, 2020). However, by 2020, this figure had risen to 9.78% of the population (BPS, 2021). Papua Island, which consists of the provinces of Papua and West Papua, is the region with the highest poverty rate in Indonesia. From 2017 to 2020, these two provinces always occupied the top two positions as the provinces with the highest poverty rate. The high percentage of poverty in Papua Island indicates the need for poverty reduction in this area. Therefore, it is necessary to analyze the factors that influence poverty in Papua Island.

Poverty can be influenced by spatial aspects, so that poverty in one area is related to poverty in another area. Research related to poverty that considered spatial effects was conducted by Nugroho and Slamet (2018) in the Central Java region using the Geographically Weighted Regression (GWR) method. Poverty is not a problem that occurs once, it has been a problem experienced by Indonesia for years, especially in Papua Island. Therefore, it is necessary to conduct research that is not only focused on one point in time but is carried out over several periods. The type of analysis that can be used for this type of data is panel data analysis. Panel data has several advantages, such as: (1) Panel data provides a greater number of observations, more informative data, more variability, and less collinearity among independent variables, (2) Panel data is able to study the dynamics of change in a phenomenon, (3) Panel data can identify effects that cannot be captured in pure cross-section or pure time-series data (Baltagi, 2021). Research related to poverty using panel data has been conducted by Widodo et al. (2019) in Indonesia. Based on the description above, it is known that the problem of poverty is a complex problem that does not only occur at one time and can be influenced by spatial aspects. It indicates the need to conduct a study involving spatial effects over several periods to investigate what factors influence poverty in Papua Island using Geographically Weighted Panel Regression (GWPR).

GWPR is a method that combines panel data regression and Geographically Weighted Regression (GWR). The GWPR model does not only consider spatial aspects but also the dynamics of change due to the use of data in a panel structure. GWPR was first applied by Yu (2010) in China. This research shows that GWPR produces a better model than cross-sectional GWR and panel data regression models. In this study, the model that has been formed with different kernel functions is compared to the values of the coefficient of determination and RMSE to choose the best model for modeling data. The best model is analyzed to determine the factors that influence poverty in Papua Island.

MATERIAL AND METHOD

Data

This study uses data consisting of 38 districts/cities on Papua Island from 2017 to 2020. The districts/cities used in this study are located on the same island, namely Papua Island. The data used in this study is secondary data obtained from the official website of Badan Pusat Statistik (BPS) for Papua Province and West Papua Province accessed on February 21, 2022. The variables in this study consist of the dependent variable, which is the percentage of poverty (y), and several independent variables that are thought to influence poverty, such as variables of school year expectancy (x_1), life expectancy (x_2), Gini ratio (x_3), consumption per capita (x_4), and labor force participation rate (x_5).

Method

Panel data is data that combines cross-sectional data and time-series data. In panel data, the same individual unit is observed repeatedly over several periods. Regression that uses panel data in the analysis process is called panel data regression. The general equation for the panel data regression model can be written as follows (Baltagi, 2021):

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

$$u_{it} = \mu_i + \lambda_t + \varepsilon_{it} \quad (2)$$

y_{it} : value of the dependent variable from i-th individual at t-th time
α : intercept
$\boldsymbol{x}'_{it}$: $(x_{it1}, x_{it2}, \dots, x_{itp})$ represents vector of the i-th individual independent variable on t-th time with size $1 \times p$
$\boldsymbol{\beta}'$: $(\beta_1, \beta_2, \dots, \beta_p)$ represents vector of the regression coefficient of size $1 \times p$
p : number of independent variables
N : number of individual units
T : number of time units
u_{it} : error component of i-th individual at t-th time
μ_i : unobserved specific effect of i-th individual
λ_t : unobserved specific effect of t-th time
ϵ_{it} : error of the i-th individual at t-th time

The error component in the panel data regression model can be divided into one-way and two-way error components. Models that only involve either individual-specific effects or time-specific effects are called one-way error component models. Meanwhile, a model that involves both individual-specific effects and time-specific effects is called a two-way error component model (Baltagi, 2021).

There are several approaches that can be used to estimate panel data regression models, namely the Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM). Judge et al., in Gujarati (2004) explain the considerations in choosing between FEM and REM. If the panel data has a large N and a small T, and individual units are the results of random selection from the larger population, then REM is appropriate. Meanwhile, if the panel data has a large N and a small T, and individual units are not the results of random sampling, then FEM is appropriate to use. This is the basis for selecting FEM in this study, where the panel data in this study has a large number of individual units (districts/cities on Papua Island) and small time units (years), as well as individual units in the form of districts/cities on Papua Island, are not selected randomly.

Fixed Effect Model

The fixed effect model implies that differences between individual units or times can be captured by different intercept values (Greene, 2012). This model assumes that the effect of individual units or units of time is a fixed parameter and part of the intercept (Baltagi, 2021). Here are some possible forms of FEM.

1. Fixed Effect Model with individual effect

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

$$u_{it} = \mu_i + \varepsilon_{it} \quad (4)$$

2. Fixed Effect Model with time effect

$$y_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta} + u_{it}, i = 1, \dots, N; t = 1, \dots, T \quad (5)$$

$$u_{it} = \lambda_t + \varepsilon_{it} \quad (6)$$

3. Fixed Effect Model with individual and time effect

$$y_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta} + u_{it}, i = 1, \dots, N; t = 1, \dots, T \quad (7)$$

$$u_{it} = \mu_i + \lambda_t + \varepsilon_{it} \quad (8)$$

$\boldsymbol{\beta}$ in the fixed effect model can be estimated using the within transformation, by eliminating unobserved specific effects in the model and then applying Ordinary Least Square (OLS) to the transformed model (Wooldridge, 2012).

Specific Effect Test

This test is carried out by comparing the FEM with the pooled model (CEM), which ignores individual and time-specific effects. The pooled model (CEM) can be stated as follows (Greene, 2012).

$$y_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}, i = 1, \dots, N; t = 1, \dots, T \quad (9)$$

1. Individual Effects Test

Hypothesis:
H_0: $\mu_i = 0$ (There were no significant individual effects)
H_1: At least there is one $\mu_i \neq 0$ (There are significant individual effects)
Test Statistic:

*Corresponding Author: **Delia Oktaviani1, Dian Lestari**

$$F = \frac{(RSS_0 - RSS_{FEMindividu})(N-1)}{(RSS_{FEMindividu})(NT - N - p)} \qquad (10)$$

With RSS_0 denotes residual sum of squares from estimation using the pooled model and $RSS_{FEMindividu}$ is the residual sum of squares from estimation using the FEM method with individual effects.

Decision rule: H_0 is rejected if the value $F > F_{\alpha,N-1,NT-N-p}$ or p-value $< \alpha$

1. Time Effects Test

Hypothesis:

$H_0: \lambda_t = 0$ (There were no significant individual effects)

H_1: At least there is one $\lambda_t \neq 0$ (There are significant individual effects)

Test Statistic:

$$F = \frac{(RSS_0 - RSS_{FEMtime})(T-1)}{(RSS_{FEMtime})(NT - T - p)} \qquad (11)$$

With $RSS_{FEMtime}$ is the residual sum of squares from estimation using FEM method with time effects.

Decision rule: H_0 is rejected if the value $F > F_{\alpha,T-1,NT-T-p}$ or p-value $< \alpha$

Panel Data Regression Model Assumption Test

a. Multicollinearity test

Multicollinearity occurs when there is a correlation among independent variables in the model. Multicollinearity can be identified by Variance Inflation Factor (VIF). In general, VIF with a value above 10 indicated the presence of high multicollinearity among independent variables (Mendenhall & Sincich, 2012).

b. Residual normality test

Residual normality test is used to determine whether residuals are normally distributed. Normality test can be performed with a visualization (a normal Q-Q plot) or hypothesis testing (Jarque-Bera test). Jarque-Bera test was carried out with the following hypothesis (Jauhari *et al*., 2020):

Hypothesis:

H_0 : residuals are normally distributed

H_1 : residuals are not normally distributed

Test statistic:

$$JB = \frac{NT}{6}\left(S^2 + \frac{(k-3)^2}{4}\right)$$

k : kurtosis

S : skewness

Decision rule : H_0 is rejected if the value $JB > \chi^2_{\alpha,2}$ or p-value $< \alpha$

c. Homoscedasticity test

Homoscedasticity can be tested by forming a plot between the estimated value of the independent variable and the residual value. If a plot is formed like in Figure 2.1. (a) where the points are spread irregularly, then it can be concluded that the assumption of homoscedasticity is met. If a certain pattern is formed on plot, for example, the funnel shape shown in Figure 2.1. (b) then there is an indication of variance inequality so that the assumption of homoscedasticity is not met.

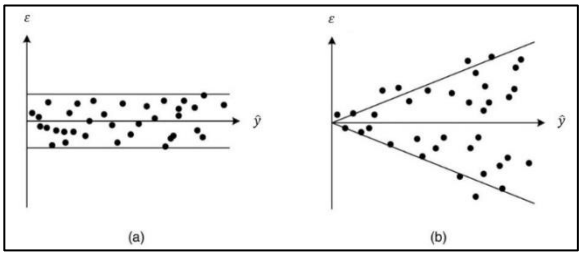

Figure 1: Homoscedasticity test plot

d. Autocorrelation test

Autocorrelation test is used to check whether the residuals from observation are independent. Autocorrelation test can be done with Durbin Watson test as follows (Jauhari *et al*., 2020):

Hypothesis:

H_0 : there is no autocorrelation

H_1 : there is autocorrelation

Test statistic:

$$D = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\hat{\varepsilon}_{it} - \hat{\varepsilon}_{i,t-1}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2}$$

Decision rule : H_0 is rejected if the value $0 < D < dL$ or $4 - dL < D < 4$. H_0 is not rejected if $dU < D < 4 - dU$. If $dL \leq D \leq dU$ or $4 - dU \leq D \leq 4 - dL$, then no decision has been made. dL is the lower limit for Durbin Watson and dU is the upper limit for Durbin Watson.

Geographically Weighted Panel Regression

Geographically Weighted Panel Regression (GWPR) is a combination of GWR and panel data regression. GWPR assumes that the time series of observations at a particular location is the realization of a smooth spatiotemporal process, in which adjacent observations are more related than distant observations, either in geographic space or time-space (Yu, 2010). In this study, the GWPR model is a combination of the fixed effect panel data regression model with GWR, which can be referred to as the GWPR fixed effect. The model is created by subtracting the fixed effects model from the model that has been averaged over time using the within-transformation method. The fixed effect model can be written as the following equation:

$$y_{it} = \beta_0(u_i, v_i) + \mu_i + \sum_{k=1}^{p}\beta_k(u_i, v_i)x_{itk} + \varepsilon_{it} \qquad (12)$$

The model equation for the average over time from equation (12) can be written as follows:

$$\bar{y}_{i.} = \beta_0(u_i, v_i) + \mu_i + \sum_{k=1}^{p}\beta_k(u_i, v_i)\bar{x}_{i.k} + \bar{\varepsilon}_{i.} \qquad (13)$$

Then, transformation is performed by subtracting equation (12) from (13) to obtain the results as follows:

$$\ddot{y}_{it} = \sum_{k=1}^{p}\beta_k(u_i, v_i)\ddot{x}_{itk} + \ddot{\varepsilon}_{it} \qquad (14)$$

*Corresponding Author: **Delia Oktaviani1, Dian Lestari**

$$\ddot{y}_{it} = y_{it} - \bar{y}_{i.}; \ddot{x}_{it} = x_{itk} - \bar{x}_{i.k}; \ddot{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_{i.} \quad (15)$$

GWPR model is estimated using the Weighted Least Square (WLS) method. This method is applied by adding different weights for each observation location. The parameter estimates for the GWPR model at location i are as follows:

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = [\ddot{\boldsymbol{X}}\boldsymbol{W}(u_i, v_i)\ddot{\boldsymbol{X}}]^{-1}\ddot{\boldsymbol{X}}'\boldsymbol{W}(u_i, v_i)\ddot{\boldsymbol{y}} \quad (16)$$

$\boldsymbol{W}(u_i, v_i)$ is spatial weight matrix for location i and it is a diagonal matrix of size $NT \times NT$.

$$\boldsymbol{W}(u_i, v_i) = \begin{bmatrix} w_{11}(u_i, v_i) & 0 & \dots & 0 \\ 0 & w_{12}(u_i, v_i) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{NT}(u_i, v_i) \end{bmatrix} \quad (17)$$

There are several kernel functions that can be used to determine the weight of spatial weighting matrix (Fotheringham et al., 2002).

1. *Fixed Gaussian*: $w_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right]$
2. *Fixed Bisquare*: $w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b}\right)^2\right]^2, d_{ij} < b \\ 0, other \end{cases}$
3. *Fixed Tricube*: $w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b}\right)^3\right]^3, d_{ij} < b \\ 0, other \end{cases}$

$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$ is the Euclidean distance between observation locations i and j, b is the bandwidth, u is the longitude, and v is the latitude.

The optimum bandwidth value is an important part because it can affect the accuracy of the model formed at each observation location. Fotheringham et al. (2002) describe several methods that can be used to determine optimum bandwidth, one of which is the cross-validation (CV) score. In GWPR, the CV score is calculated using the following formula (Yu, 2010):

$$CV = \sum_{i=1}^{N} \left(\bar{y}_i - \hat{\bar{y}}_{\neq i}(b)\right)^2 \quad (18)$$

where $\bar{y}_i$ is the average value of the time-dependent variable at the observation location i and $\hat{\bar{y}}_{\neq i}(b)$ is the estimated value of $\bar{y}_i$ using *bandwidth* b without involving observations at the i-th location in the estimation process. The optimum bandwidth is the bandwidth with the smallest CV value.

A parameter significance test was carried out to determine parameters that had an individual significant effect on the dependent variable at each location. The procedure for testing the GWPR model parameter coefficients is as follows.

Hypothesis:

$H_0 : \beta_k(u_i, v_i) = 0$

$H_1 : \beta_k(u_i, v_i) \neq 0$

Test statistic:

$$t_k(u_i, v_i) = \frac{\hat{\beta}_k(u_i, v_i)}{\hat{\sigma}\sqrt{c_{kk}}} \quad (19)$$

$\hat{\sigma} = \sqrt{\frac{RSS_{GWPR}}{\delta_1}}$, c_{kk} is the k-th diagonal element of the matrix $\boldsymbol{C}_i\boldsymbol{C}_i'$ with $\boldsymbol{C}_i = \left(\ddot{\boldsymbol{X}}'\boldsymbol{W}(u_i, v_i)\ddot{\boldsymbol{X}}\right)^{-1}\ddot{\boldsymbol{X}}\boldsymbol{W}(u_i, v_i)$, $\delta_i = tr\left((\boldsymbol{I} - \boldsymbol{L})^T(\boldsymbol{I} - \boldsymbol{L})\right)^i; i = 1, 2.$

L is a matrix that projects $\ddot{\boldsymbol{y}}$ to $\hat{\ddot{\boldsymbol{y}}}$

$$\boldsymbol{L}_{(NT,NT)} = \begin{bmatrix} x_{11}'[\ddot{\boldsymbol{X}}'\boldsymbol{W}(u_1, v_1)\ddot{\boldsymbol{X}}]^{-1}\ddot{\boldsymbol{X}}'\boldsymbol{W}(u_1, v_1) \\ x_{21}'[\ddot{\boldsymbol{X}}'\boldsymbol{W}(u_2, v_2)\ddot{\boldsymbol{X}}]^{-1}\ddot{\boldsymbol{X}}'\boldsymbol{W}(u_2, v_2) \\ \vdots \\ x_{N1}'[\ddot{\boldsymbol{X}}'\boldsymbol{W}(u_N, v_N)\ddot{\boldsymbol{X}}]^{-1}\ddot{\boldsymbol{X}}'\boldsymbol{W}(u_N, v_N) \\ \vdots \\ x_{1T}'[\ddot{\boldsymbol{X}}'\boldsymbol{W}(u_1, v_1)\ddot{\boldsymbol{X}}]^{-1}\ddot{\boldsymbol{X}}'\boldsymbol{W}(u_1, v_1) \\ x_{2T}'[\ddot{\boldsymbol{X}}'\boldsymbol{W}(u_2, v_2)\ddot{\boldsymbol{X}}]^{-1}\ddot{\boldsymbol{X}}'\boldsymbol{W}(u_2, v_2) \\ \vdots \\ x_{NT}'[\ddot{\boldsymbol{X}}'\boldsymbol{W}(u_N, v_N)\ddot{\boldsymbol{X}}]^{-1}\ddot{\boldsymbol{X}}'\boldsymbol{W}(u_N, v_N) \end{bmatrix} \quad (17)$$

Decision rule: H_0 is rejected if the value $|t_k(u_i, v_i)| > t_{\frac{\alpha}{2}, df}$ with $df = \frac{\delta_1^2}{\delta_2}$

Coefficient of Determination

The coefficient of determination provides information about how much variation of the dependent variable can be explained by the independent variables in the model. The formula of the coefficient of determination can be written as follows:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{t=1}^{T}\sum_{i=1}^{N}(y_{it} - \hat{y}_{it})^2}{\sum_{t=1}^{T}\sum_{i=1}^{N}(y_{it} - \bar{y})^2} \quad (20)$$

Root Mean Square Error

Root Mean Square Error (RMSE) is an error measurement technique used to measure model performance based on the difference between the estimated value, $\hat{y}$, and the actual value, y. The RMSE formula can be written as follows:

$$RMSE = \sqrt{\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}(y_{it} - \hat{y}_{it})^2} \quad (21)$$

The model with smaller RMSE value means that it has better performance than the other models (Georganos *et al*., 2017).

RESULTS AND DISCUSSION

First, a multicollinearity test is performed by calculating the VIF value for each independent variable. The results of the multicollinearity test show that the VIF value of all independent variables is less than 10, so it can be concluded that there is no multicollinearity. Furthermore, a specific effect test was carried out to see whether there were significant individual or time-specific effects on the data. The results show that there is a significant individual effect on the data with a value of F = 189.21 and p-value = 2.22×10^{-16}. Then, a fixed effect panel regression model with individual effects was formed. The estimation results are shown in Table 1.

*Corresponding Author: **Delia Oktaviani1, Dian Lestari**

Table 1: Estimation results of individual fixed effect model

Variable	Estimation	Std. Error	$t-value$	p-value
x_1	-0.5678	0.4448	-1.2765	0.2045
x_2	-1.8091	0.4525	-3.9983	0.0001
x_3	4.3950	1.9155	2.2944	0.0237
x_4	-0.0011	0.0005	-2.2401	0.0271
x_5	0.0025	0.0115	0.2213	0.8253
R-squared	0.4873			
F-statistic	20.72			

The residual normality assumption test was carried out using the Jarque-Bera test. Based on the test, the results show that the residuals are not normally distributed, so it is necessary to handle violations of this assumption. One way that can be used to deal with data that is not normally distributed is to transform the data using a natural logarithm transformation. As a result, the natural logarithmic transformation is applied to the y, x_4, and x_5 variables.

After that, a multicollinearity test was performed again by calculating the VIF value for each independent variable after the transformation. The results of the multicollinearity test show that the VIF value of all independent variables is less than 10, so it can be concluded that there is no multicollinearity. Furthermore, a specific effect test was carried out to see whether there were significant individual or time-specific effects on the data. The results show that there is a significant individual effect on the data with a value of F = 300.52 and p-value = 2.22×10^{-16}. Then, a fixed effect panel data regression model with individual effects was formed. The estimation results are shown in Table 2.

Table 2: Estimation results of individual fixed effect model after transformation

Variable	Estimation	Std. Error	$t-value$	p-value
x_1	-0.0127	0.0150	-0.8498	0.3973
x_2	-0.0772	0.0148	-5.2151	8.8×10^{-7}
x_3	0.1842	0.0647	2.8446	0.0053
x_4	0.2009	0.1105	-1.8187	0.0717
x_5	0.0076	0.0260	-0.2909	0.7717
R-squared	0.5225			
F-statistic	23.86			

Then, test the assumptions of normality residual, homogeneity, and non-autocorrelation. Based on the test, the results show that the residuals are normally distributed, heteroscedasticity, and non-autocorrelation. The spatial heterogeneity test was carried out using Breusch-Pagan test. The results of the Breusch-Pagan test showed that there was spatial heterogeneity in the data. It causes the global model to be unsuitable for modeling data, so the Geographically Weighted Panel Regression (GWPR) model is used.

Prior to analyzing data using GWR, it is necessary to transform the data by subtracting each value from the average value over time for the appropriated individual. The distance among observation locations was calculated using the Euclidean distance formula. Then, determine the optimum bandwidth for each weighting function as shown in Table 3.

Table 3: Bandwidth of each weighting function

Weighting Function	Estimation
Fixed Gaussian	1.4065
Fixed Bisquare	-3.7702
Fixed Tricube	3.9184

Table 4 shows that the GWPR model with various kernel weighting functions is better at modeling data when compared to the fixed effect model (global model). The GWPR model with the Fixed Gaussian kernel weighting function is the best model with largest R^2 value and smallest RMSE value.

Table 4: Coefficient of determination and RMSE for each model

Variable	R^2	RMSE
FEM (global model)	52.25%	0.0239
GWPR (Fixed Gaussian)	84.93%	0.0135
GWPR (Fixed Bisquare)	84.52%	0.0136
GWPR (Fixed Tricube)	84.00%	0.0139

Then, the parameter significance tests was calculated using the t-test at each location to obtain district/city groups with similarities significant independent variables as shown in Table 5.

Table 5: Group of significant independent variables

Group	Significant Variable	Districts/Cities
1	x_2, x_4	Fakfak, Kaimana, Teluk Wondama, Asmat, Boven Digoel, Jayapura, Jayawijaya, Keerom, Kota Jayapura, Lanny Jaya, Mamberamo Raya, Mamberamo Tengah, Mappi, Mimika, Nduga, Pegunungan Bintang, Puncak, Puncak Jaya, Sarmi, Tolikara, Yahukimo, Yalimo
x_2	x_1, x_2, x_3	Merauke

*Corresponding Author: **Delia Oktaviani1, Dian Lestari**

x_3	x_2, x_4, x_5	Manokwari, Sorong Selatan, Sorong, Tambrauw, Maybrat, Manokwari Selatan, Pegunungan Arfak, Kota Sorong
x_4	x_2, x_3, x_4	Deiyai, Intan Jaya, Paniai, Waropen
x_5	x_2, x_3, x_4, x_5	Dogiyai, Nabire

If the distribution of significant independent variables as listed in Table 5 is presented on the map, a map is formed as shown in Figure 1. It is informs that regions that have the same significant variables tend to have locations that are close to each other.

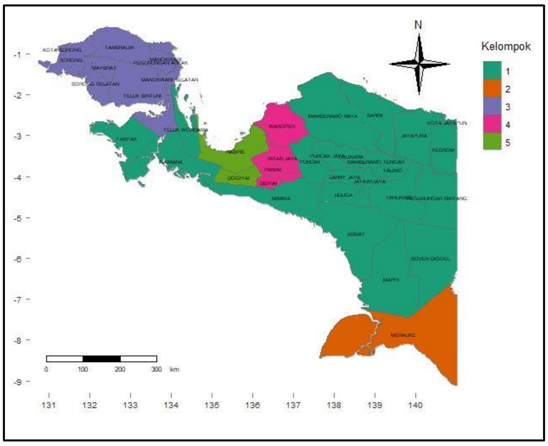

Figure 2: Map of the 5 groups of significant variables

CONCLUSION

Based on the results of the analysis and discussion conducted on poverty data in Papua Island, following conclusions can be obtained:

1. Based on the spatial heterogeneity test, the results show that there is heterogeneity spatial on poverty data in Papua Island from 2017 - 2020.
2. Analysis using Geographically Weighted Panel Regression (GWPR) provides parameter estimation results that vary for each observation location. These results lead to different variables that have a significant effect on each location. In general, variables of school year expectancy, life expectancy, gini ratio, consumption per capita, and labor force participation rate have a significant effect on poverty in at least one location in a district/city in Papua Island.

ACKNOWLEDGMENT

This work is fully funded by Hibah Riset Penugasan FMIPA UI 2022-2023 research grant from RPM FMIPA Universitas Indonesia.

(Project ID: 001/UN2.F3.D/PPM.00.02/2022).

REFERENCES

1. Baltagi, B. H. (2021). Econometric Analysis of Panel Data (Sixth Edition). New York: John Wiley & Sons.
2. BPS. (2020). Statistik Indonesia 2020. Jakarta: BPS RI
3. BPS. (2021). Statistik Indonesia 2021. Jakarta: BPS RI
4. Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). Geographically Weighted Regression: The Analysis of Spatial Varying Relationships. Chichester: John Wiley & Sons.
5. Georganos, S., Abdi, A. M., Tenenbaum, D. E., & Kalogirou, S. (2017). Examining the NDVI-rainfall relationship in the semi-arid Sahel using geographically weighted regression. Journal of Arid Environments, 146, 64–74.
6. Gujarati, D. (2004). Basic Econometrics (Fourth Edition). New York: McGraw-Hill/Irwin.
7. Greene, W. H. (2012). Econometric Analysis (Seventh Edition). New Jersey: Prentice Hall.
8. Jauhari, A. F., Suprapto, Y. K., & Mauludiyanto, A. (2020). Data Panel Modelling with Fixed Effect Model (FEM) Approach to Analyze the Influencing Factors of DHF in Pasuruan Regency. Proceedings of the 2nd International Conference on Applied Science, Engineering and Social Sciences, 224 – 232.
9. Mendenhall, W., & Sincich, T. (2012). A Second Course in Statistics: Regression Analysis (Seventh Edition). Boston: Pearson.
10. Nugroho, N. F. T. A., & Slamet, I. (2018). Geographically Weighted Regression Model with Kernel Bisquare and Tricube Weighted Function on Poverty Percentage Data in Central Java Province. Journal of Physics: Conference Series, 1025(1), 012099.
11. Widodo, E., Suriani, E., Ristyaningrum, I. P., & Gita Evi. (2019). Analisis Regresi Panel pada Kasus Kemiskinan di Indonesia. PRISMA, Prosiding Seminar Nasional Matematika, 2, 710–717.
12. Wooldridge, J. M. (2012). Introductory Econometrics: A Modern Approach (Fifth Edition). Mason: South-Western Cengage Learning.
13. Yu, D. (2010). Exploring spatiotemporally varying regressed relationships: the geographically weighted panel regression analysis. The international archives of the photogrammetry, remote sensing, and spatial information sciences, 38(Part II), 134-139.

*Corresponding Author: **Delia Oktaviani1, Dian Lestari**