International Journal of Scientific Research in Science, Engineering and Technology

Print ISSN - 2395-1990
Online ISSN : 2394-4099

Available Online at : www.ijsrset.com
doi : https://doi.org/10.32628/IJSRSET2411126

Modeling Poverty Data in Indonesia with Spatial Hierarchy Structure Using HLM, GWR, and HGWR Methods

Cintia Septemberini[1], Muhammad Nur Aidi[2], Anang Kurnia[3]
[1,2,3]Department of Statistics, IPB University, Bogor, West Java, Indonesia
Email: cintiaseptemberini@apps.ipb.ac.id[1]

ARTICLEINFO

Article History :

Accepted: 02 Feb 2024
Published: 19 Feb 2024

Publication Issue :
Volume 11, Issue 1
January-February-2024
Page Number :
260-271

ABSTRACT

Poverty causes the majority of the Indonesian population to face challenges in fulfilling basic needs such as clothing, food, and shelter. The factors that play a role in determining the poverty rate in Indonesia tend to vary in each province; this is due to the diverse conditions resulting from spatial heterogeneity. However, poverty in Indonesia is not only influenced by factors from various regions but also by the conditions of the districts/cities within them. Districts/cities within a province form a spatial hierarchy structure. Therefore, in this study, the Hierarchical Linear Model (HLM), Geographically Weighted Regression (GWR), and Hierarchical and Geographically Weighted Regression (HGWR) methods were applied to determine the best model among the three methods in analyzing the factors affecting the poverty rate in Indonesia with a spatial hierarchy structure. The results of the analysis show that the HGWR method is the best model compared to HLM and GWR, as evidenced by the higher R^2 value of 0.8004 compared to HLM and GWR. Based on the HGWR model, most of the local estimators for population dependency ratio (G_1), adjusted per capita expenditure (G_2), and economic growth rate (G_3) showed significance in provinces located in eastern Indonesia. In addition, the fixed effects and random effects estimators, namely the percentage of households without access to electricity (X_1), the ratio of per capita normative consumption to net product (X_2), and the percentage of households without access to clean water (X_3), also have a significant influence on the poverty rate in Indonesia.

Keywords: Poverty, Spatial Heterogeneity, Hierarchical Structure, GWR, HLM, HGWR

I. INTRODUCTION

Poverty is a problem in development in various parts of the world, including in Indonesia. The impact of poverty involves the difficulty of a large number of Indonesians in meeting basic needs such as clothing, food, and adequate shelter [1]. Therefore, poverty reduction has become a top priority in various regions around the world, especially in areas with high poverty rates. Data from [2] shows that in September 2022, the poverty rate in Indonesia was still relatively high. Although there was a decrease when compared to September 2021, the reduction was not so significant. The poverty rate in September 2022 reached 9.57 percent, or around 26.36 million people living below the poverty line.

Various factors cause the high poverty rate in Indonesia, and analysis of these factors cannot be done simultaneously in each region because each region may have different causes of poverty. Each province in Indonesia has further potential in terms of economy, education, health, and others, indicating a spatial effect called spatial heterogeneity. One method that is often used to analyze spatial heterogeneity is Geographically Weighted Regression (GWR). GWR is the development of a global regression model into a local regression [3].

The approach used by GWR is a point approach where each parameter value is estimated at each observation location point, resulting in different parameter estimates at each location. To produce other parameter estimates at each location, the GWR model requires a weight matrix formed from a weight function that is influenced by bandwidth [4]. Bandwidth is the radius of a circle in which points within it are still considered to affect the formation of model parameters at an observation location [5].

However, the poverty rate in Indonesia is not only influenced by factors from various provinces but also by the conditions of the districts/municipalities within them, such as the difficulty of access to the area, which can be the cause of the lack of infrastructure supporting economic activities. Therefore, the poverty rate in Indonesia can be influenced by factors from two levels, namely provinces and districts/municipalities. The kabupaten/kota level within the province level forms a spatial hierarchy structure, where several kabupaten/kota are grouped based on the spatial location of the province. This results in several districts/municipalities being allocated to the exact coordinates.

One method that can be used to analyze data with a hierarchical structure is the Hierarchical Linear Model (HLM), also known as a multilevel model [6], [7]. HLM is a development of regression analysis that aims to handle homogeneity within groups and heterogeneity between groups [8]. In HLM, two types of variables are commonly used, namely group-level variables and sample-level variables. Group-level variables describe the characteristics of specific groups (such as province, block, school, etc.). In contrast, sample-level variables include individual observations within the group (such as city, house, student, etc.). The effects of some sample-level variables are similar across groups, so they are modeled with fixed coefficients (effects), while others are modeled individually, i.e., with random effects. However, for group-level variables, HLM can only be modeled with fixed effects.

With the development of statistical analysis methods, [9] proposed the Hierarchical and Geographically Weighted Regression method, abbreviated as HGWR, which is a spatial modeling method designed for data with a spatial hierarchical structure. As the name implies, the HGWR method is a combination of HLM and GWR. In the HGWR model, spatial effects are divided into three types, namely global fixed, local fixed, and random. Therefore, the purpose of this study is to determine the best model between the HLM, GWR, and HGWR methods in analyzing the factors that influence the poverty rate in Indonesia with a spatial hierarchy structure.

II. METHODS AND MATERIAL

A. Material

The data used in this study is poverty data in Indonesia obtained from the publication of the Central Statistics Agency in 2022. The data structure studied consists of two levels, namely district/city as the sample level and province as the group level. The research was conducted on 514 districts/cities from 34 provinces in Indonesia. The response variable used is the percentage of poor people per district/city in Indonesia in 2022, while the explanatory variables used:

- District/city level:
  - X_1 : percentage of households without access to electricity (%)
  - X_2 : ratio of per capita normative consumption to net product
  - X_3 : percentage of households without access to clean water (%)
- Province level:
  - G_1 : population dependency ratio
  - G_2 : adjusted per capita expenditure (thousand rupiah)
  - G_3 : economic growth rate (%)

Based on the variables used, the data structure formed can be seen in Table 1.

B. Methods

Data analysis was carried out using R software version 4.2.2 with the steps of data analysis in this study are as follows:

1. Exploring the data visually to determine the characteristics of each variable, both response variables and explanatory variables, based on the district/city level and provincial level.
2. Conduct HLM analysis with the following steps:

a) Estimating the parameter values of the model to be built using the following equation:

$$
\begin{aligned}
y_{ij} &= \beta_{j0} + \beta_{jX_1}X_{1ij} + \beta_{jX_2}X_{2ij} + \beta_{jX_3}X_{3ij} + \epsilon_{ij} \qquad (1)\\
\beta_{j0} &= \gamma_{00} + \gamma_{0G_1}G_{1j} + \gamma_{0G_2}G_{2j} + \gamma_{0G_3}G_{3j} + \delta_{j0}\\
\beta_{jX_1} &= \gamma_{x_1 0}\\
\beta_{jX_2} &= \gamma_{X_2 0} + \delta_{jX_2}\\
\beta_{jX_3} &= \gamma_{x_3 0}
\end{aligned}
$$

where

y_{ij}	:	observation value of the response variable in the i-th sample in the j-th group
$X_{1ij}, X_{2ij}, X_{3ij}$	:	observation value of the explanatory variable in the i-th sample in the j-th group
$\beta_{j0}, \beta_{jX_1}, \dots, \beta_{jX_3}$	:	coefficient estimated using the sample in the j-th group
G_{1j}, G_{2j}, G_{3j}	:	observation value of the explanatory variable in the j-th group
$\gamma_{00}, \gamma_{0G_1}, \dots, \gamma_{X_3 0}$	:	coefficient for $\beta_{j0}, \beta_{jX_1}, \beta_{jX_2}$ and β_{jX_3}
ϵ_{ij}	:	error value on the i-th sample in the j-th group
$\delta_{j0}, \delta_{jX_2}$	:	error value β_{j0} and β_{jX_2}

Table 1: Two-level spatial hierarchy data structure

Province	**District/city**	**Response variable ($\boldsymbol{Y}$)**	**Level Explanatory Variables District/City**			**Province Level Explanatory Variables**			**Coordinates**	
			$\boldsymbol{X_1}$	$\cdots$	$\boldsymbol{X_3}$	$\boldsymbol{G_1}$	$\cdots$	$\boldsymbol{G_3}$	$\boldsymbol{u_j}$	$\boldsymbol{v_j}$
1	1	y_{11}	x_{111}		x_{113}	g_{11}		g_{13}	u_1	v_1
	$\vdots$	$\vdots$	$\vdots$		$\vdots$					
	n_1	$y_{n_1 1}$	$x_{n_1 11}$		$x_{n_1 13}$					
2	1	y_{12}	x_{121}		x_{123}	g_{21}		g_{23}	u_2	v_2
	$\vdots$	$\vdots$	$\vdots$		$\vdots$					
	n_2	$y_{n_2 2}$	$x_{n_2 21}$		$x_{n_2 23}$					
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$
34	1	$y_{1;34}$	$x_{1;34;1}$		$x_{1;34;3}$	$g_{34;1}$		$g_{34;3}$	u_{34}	v_{34}
	$\vdots$	$\vdots$	$\vdots$		$\vdots$					
	n_{34}	$y_{n_{34};34}$	$x_{n_{34};34;1}$		$x_{n_{34};34;3}$					

b) Conduct a significance test using the t_{count}value.

c) Calculating the R^2 value is categorized into 2 types, namely marginal R^2 and conditional R^2[10]. The marginal R^2 represents the variance explained by the factors affected by the fixed effects and is defined as follows:

$$R^2_{marginal} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\delta^2 + \sigma_\epsilon^2} \tag{2}$$

The conditional R^2 is interpreted as the variance explained by factors influenced by fixed effects and random effects and is calculated by the following equation:

$$R^2_{conditional} = \frac{\sigma_f^2 + \sigma_\delta^2}{\sigma_f^2 + \sigma_\delta^2 + \sigma_\epsilon^2} \tag{3}$$

where σ_f^2 is the variance of the fixed effects component.

d) Conclude from the HLM analysis results.

3. Perform GWR analysis with the following steps:

a) Finding the optimal bandwidth value based on the minimum cross-validation (CV) value. The CV value is calculated with the following formula:

$$CV(b) = \sum_{j=1}^{n} \left[y_j - \hat{y}_{\neq j}(b)\right]^2 \tag{4}$$

where $\hat{y}_{\neq j}(b)$ is the estimated value of y_j with the parameter at the j-th location omitted in the estimation process. The optimal bandwidth is obtained through an iteration process by changing the bandwidth value (b) until the minimum CV is accepted [3].

b) Calculate the weight matrix $\mathbf{W}_j$ using the Gaussian kernel and Bisquare kernel functions that produce the minimum CV value. The spatial weight matrix $\mathbf{W}_j$ can be described as follows:

$$\mathbf{W}_j = \begin{bmatrix} w_{j1} & 0 & \dots & 0 \\ 0 & w_{j2} & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & w_{jn} \end{bmatrix} \tag{5}$$

Gaussian kernel and Bisquare kernel weighting functions, according to [11], are as follows:

- Adaptive Gaussian Kernel

$$w_{ji} = \exp\left[-\frac{1}{2}\left(\frac{d_{ji}}{b_{j(q)}}\right)^2\right] \quad (6)$$

- Adaptive Bisquare Kernel

$$w_{ji} = \begin{cases} \left(1 - \frac{d_{ji}^2}{b_{j(q)}{}^2}\right)^2, d_{ji} \le b_j \\ 0, d_{ji} > b_j \end{cases} \quad (7)$$

where d_{ji} is the distance of the j-th location to the i-th location and $b_{j(q)}$ is an adaptive bandwidth that sets q as the nearest neighbor distance from the j-th observation location point.

c) Estimating the parameter values of the model to be built using the following equation:

$$y_j = \beta_{j0} + \beta_{jG_1}G_{1j} + \beta_{jG_2}G_{2j} + \beta_{jG_3}G_{3j} + \beta_{jX_1}X_{j1} + \beta_{jX_2}X_{2j} + \beta_{jX_3}X_{3j} + \in_j \quad (8)$$

where

j	:	location index expressed by a pair of coordinates (u_j, v_j)
y_j	:	observation value of the response variable at the j-th location
$G_{1j}, G_{2j}, \dots, X_{3j}$	:	observation value of the explanatory variable at the j-th location
β_{j0}	:	intercept at the j-th location
$\beta_{jG_1}, \beta_{jG_2}, \dots, \beta_{jX_3}$	:	the regression coefficient for the explanatory variable at the j-th location
$\in_{ij}$	:	error value at the j-th location, which is assumed to be free, identical, and normally distributed

[12] stated that the estimation of parameter coefficients at each location uses the Weighted Least Square (WLS) method, which can be represented as follows:

$$\widehat{\boldsymbol{\beta}}_j = \left(\boldsymbol{X}^T\mathbf{W}_j\,\mathbf{X}\right)^{-1}\left(\boldsymbol{X}^T\mathbf{W}_j\boldsymbol{y}\right) \quad (9)$$

where $\widehat{\boldsymbol{\beta}}_j = \left(\hat{\beta}_0, \hat{\beta}_{G_1}, \dots, \hat{\beta}_{X_2}\right)^T$ is a vector containing the estimated parameter coefficients for each location, $\boldsymbol{X}$ is a matrix of explanatory variables, $\boldsymbol{y} = (y_1, y_2, \dots, y_n)^T$ is a vector containing response variables, while $\mathbf{W}_j$ is a diagonal matrix that shows the spatial weights between the j-th location and other locations.

d) Test the significance of each observed location using the t_{count} value.

e) Conclude from the results of the GWR analysis.

4. Conduct HGWR analysis with the following steps:

a) Find the optimal bandwidth value and calculate the weighting matrix using the Gaussian kernel and Bisquare kernel weighting functions according to equation (6) and (7) based on the minimum CV value.

b) Estimating the parameter values of the model to be built using the following equation:

$$\begin{aligned} y_{ij} &= \beta_{j0} + \beta_{jX_1}X_{1ij} + \beta_{jX_2}X_{2ij} + \beta_{jX_3}X_{3ij} + \in_{ij} \\ \beta_{j0} &= \gamma_{j00} + \gamma_{j0G_1}G_{1j} + \gamma_{j0G_2}G_{2j} + \gamma_{j0G_3}G_{3j} + \delta_{j0} \\ \beta_{jX_1} &= \gamma_{x_10} \\ \beta_{jX_2} &= \gamma_{X_20} + \delta_{jX_2} \\ \beta_{jX_3} &= \gamma_{x_30} \end{aligned} \quad (10)$$

where

y_{ij}	:	observation value of the response variable in the i-th sample in the j-th group
$X_{1ij}, X_{2ij}, X_{3ij}$	:	observation value of the explanatory variable in the i-th sample in the j-th group
$\beta_{j0}, \beta_{jX_1}, \dots, \beta_{jX_3}$	:	coefficient estimated using the sample in the j-th group

G_{1j}, G_{2j}, G_{3j}	:	observation value of the explanatory variable in the j-th group
$\gamma_{j00}, \gamma_{j0G_1}, \cdots, \gamma_{j0G_3}$	:	the coefficient for β_{j0} each j -th group or location
$\gamma_{X_10}, \gamma_{X_20}, \gamma_{X_30}$	:	coefficient for $\beta_{jX_1}, \beta_{jX_2}$, and β_{jX_3}
ϵ_{ij}	:	error value on the i-th sample in the j-th group
$\delta_{j0}, \delta_{jX_2}$	:	error value β_{j0} and β_{jX_2}

Estimation of $\gamma_{j00}, \gamma_{j01}, \gamma_{j02}, \dots, \gamma_{j0q}$ for each j-th group or location using the GWR model based on equation (11) below:

$$\widehat{\boldsymbol{\gamma}}_{j0} = \left(\boldsymbol{G}^{\boldsymbol{T}}\boldsymbol{W}_j\,\mathbf{G}\right)^{-1}\left(\boldsymbol{G}^{\boldsymbol{T}}\boldsymbol{W}_j\boldsymbol{\beta}_0\right) \qquad (11)$$

where $\widehat{\boldsymbol{\gamma}}_{j0} = (\hat{\gamma}_{j00}, \hat{\gamma}_{j01}, \dots, \hat{\gamma}_{j0q})$ is the vector of estimated $\gamma_{j01}, \gamma_{j02}, \dots, \gamma_{j0q}$; $\boldsymbol{\beta}_0 = (\beta_{10}, \beta_{20}, \dots, \beta_{j0})^T$ is the coefficient β_{j0} estimated from the sample level regression model using HLM; $\boldsymbol{W}_j$ is a diagonal matrix showing the spatial weights between the j-th location and other locations or groups; j is the group or location level index expressed by a pair of coordinates (u_j, v_j); $\mathbf{G}$ is the group level variable matrix.

c) Conduct a significance test using the t_{count} value.
d) Conclude from the results of the HGWR analysis.

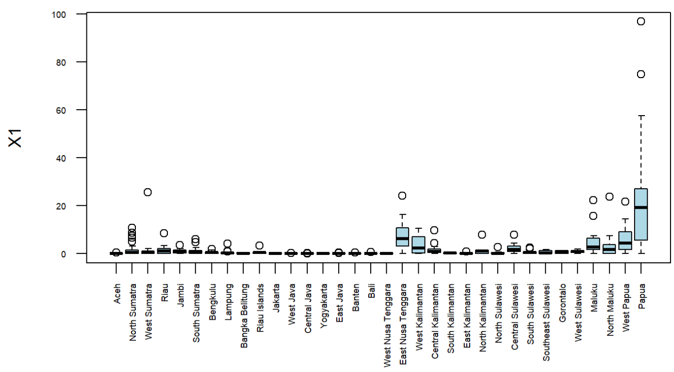
III.RESULTS AND DISCUSSION

Data Exploration

Figure 1 shows that high poverty rates are primarily found in eastern Indonesia, such as Papua, Maluku, Nusa Tenggara, and parts of Sulawesi. In contrast, Bali and North Maluku, which are close to these regions, show relatively low poverty rates in their districts/municipalities. This reflects the diversity of poverty between provinces in Indonesia in 2022, where districts/municipalities within one province tend to show more homogeneous rates compared to other provinces.

Figure 1: Distribution of the percentage of poor people per district/city in Indonesia

Figure 2 reflects that districts/municipalities in Papua have a higher percentage of households without access to electricity (X_1) and clean water (X_3) than other provinces. Limited access to electricity and clean water can indicate a decline in welfare, which has the potential to increase the percentage of poverty [13], which is in line with what is seen in Figure 1, where areas with high poverty rates are found in Papua. In addition, Figure 2 also reflects that there are still many provinces where districts/cities have abysmal normative consumptive ratios (X_2). A ratio that exceeds 1 indicates that production is lower than consumption, reflecting low food security and potentially increasing poverty rates [14].



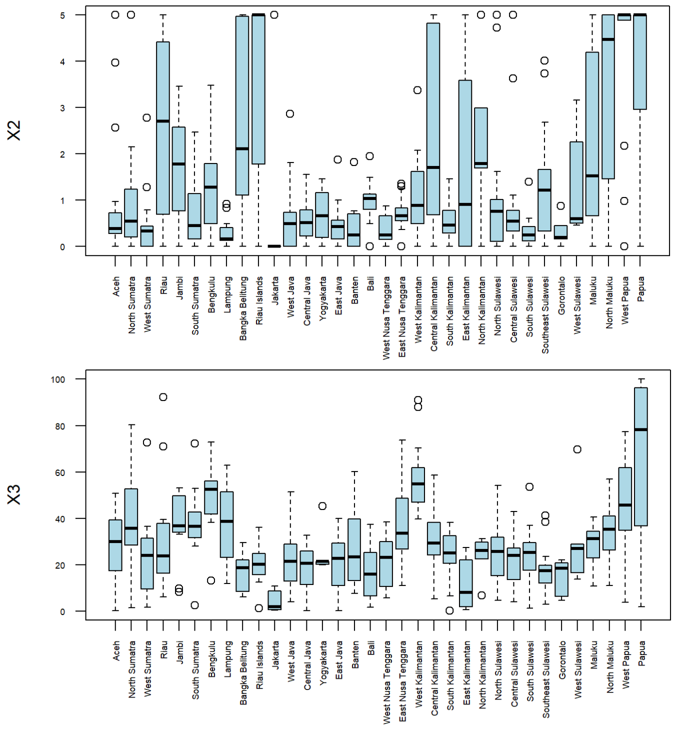


Figure 2: Boxplot of district level explanatory variables

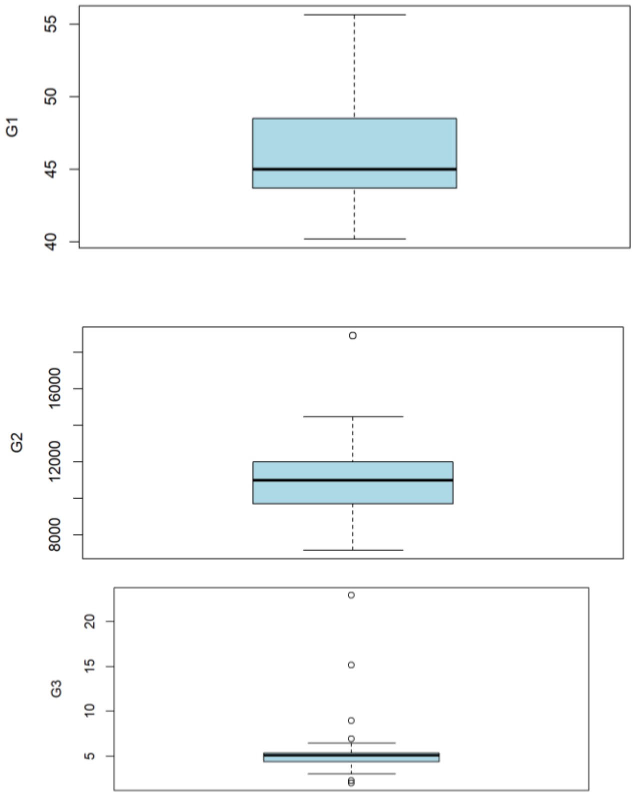


Figure 3: Boxplot of province-level explanatory variables

From Figure 3, it can be seen that the economic growth rate (G_3) in Indonesian provinces is not evenly distributed and tends to be centered in the range of 4% to 5%, and several provinces have values that are far different from these values. In contrast, the dependency ratio (G_1) and per capita expenditure (G_2) are evenly distributed and not concentrated. However, there is still one province that is an outlier in per capita expenditure, and the dependency ratio data is skewed to the right. In theory, high economic growth and per capita expenditure may reflect prosperity as a result of poverty reduction [15]. However, a high dependency ratio may reflect the burden that must be borne by the productiveage population, which could potentially increase poverty [12].

Results of HLM

The poverty data in Indonesia has a hierarchical structure, so it is modeled using HLM based on equation (1), which combines fixed effects and random effects. Table 2 shows the coefficient estimates with fixed effects, as well as diagnostic information such as the marginal R^2 and conditional R^2 of the HLM model. The coefficient estimates for the intercept and X_2 are presented in map form in Figure 4 as they vary between groups affected by random effects.

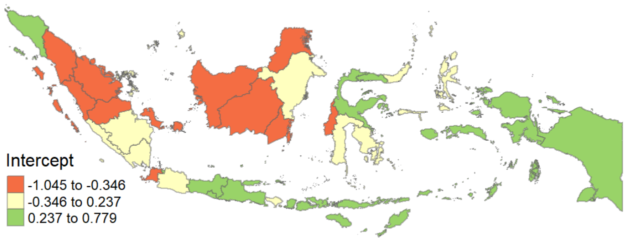


Figure 4: HLM coefficients estimator with random effects

Table 2 illustrates that the percentage of households without access to electricity (X_1) and clean water (X_3) as explanatory variables at the sample level (district/city) have a significant influence on the percentage of poor people. The coefficient estimates for X_1 and X_3 show positive values, indicating a unidirectional effect on the poverty rate. This illustrates that the higher these two explanatory variables, the higher the percentage of poor people. On the other hand, G_1, G_2, and G_3 as explanatory variables at the group level (province) do not show a significant effect on the poverty rate.

Table 2: Diagnostic information from results of HLM

Explanatory Variables	Coefficient	t_{count}	Diagnostic	Value
G_1	0.14505	1.271	Conditional R^2	0.709
G_2	-0.13959	-1.462	Marginal R^2	0.340
G_3	-0.09766	-1.194		
X_1	0.13321	3.715*		
X_3	0.28854	9.047*		
Note: * Significance at $\alpha = 5\%$ level				

Meanwhile, Table 2 also shows that the conditional R^2 is much higher than the marginal R^2, illustrating that the random effects that cause differences in the coefficients in each group (province) contribute significantly to the goodness of the model. This reinforces the importance of considering spatial heterogeneity issues in this study. However, in the HLM method, the resulting estimator is global, which does not consider the problem of spatial heterogeneity.

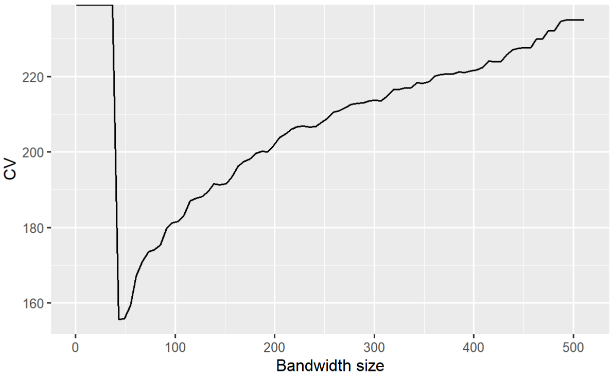
Results of GWR

GWR is one of the effective methods to overcome the problem of spatial heterogeneity in data. The selection of the Kernel function with optimum bandwidth is essential in the GWR model, and that is why, in this study, parameter estimation was performed using Adaptive Gaussian and Bisquare Kernels, the results of which are presented in Table 3.

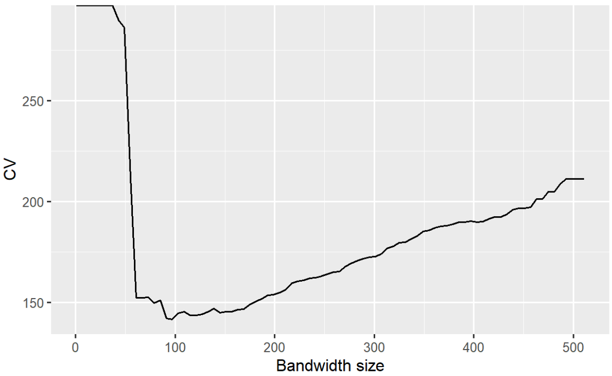
Table 3: Diagnostic information from results of GWR with different Kernel functions

Indicator	Kernel Function	
	Adaptive Gaussian	Adaptive Bisquare
Bandwidth	41	145
R^2	0.7674	0.7842

The results from Table 3 show that the Adaptive Bisquare Kernel performs better than the Adaptive Gaussian Kernel because it produces higher R^2 values. However, judging from the visualization in Figure 5, none of the curves appear normal. Most likely, this abnormality is due to the characteristics of the data where a large number of observations use the same spatial coordinates. This causes a slight change in bandwidth to result in the involvement of very different observations in each estimation process at any given point. Therefore, there can be significant fluctuations in the CV values in Figure 5 may occur. This is a serious problem of optimizing the bandwidth of the GWR model because it is still being determined whether a truly optimum bandwidth value can be found or there may not even be an optimum value due to non-convergence. Thus, for this type of data that has a hierarchical structure, the use of the GWR method should be done carefully by considering the potential problems in bandwidth optimization.



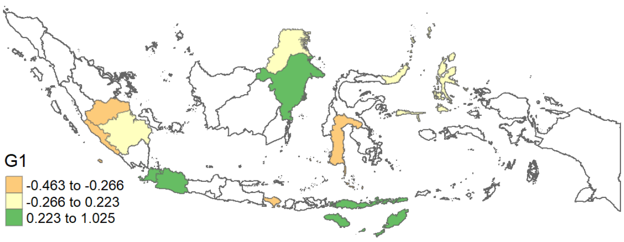
(a)



(b)

Figure 5: Bandwidth optimization of GWR model via CV using (a) Adaptive Gaussian Kernel and (b) Adaptive Bisquare Kernel

Based on the GWR performance listed in Table 3, Figure 6 only displays the coefficient estimates using the Adaptive Bisquare Kernel. In Figure 6, the local t_{count} values are also used to show the significance level of each coefficient estimator based on location. Insignificant locations are shown without color, while significant locations are colored according to their expected value.



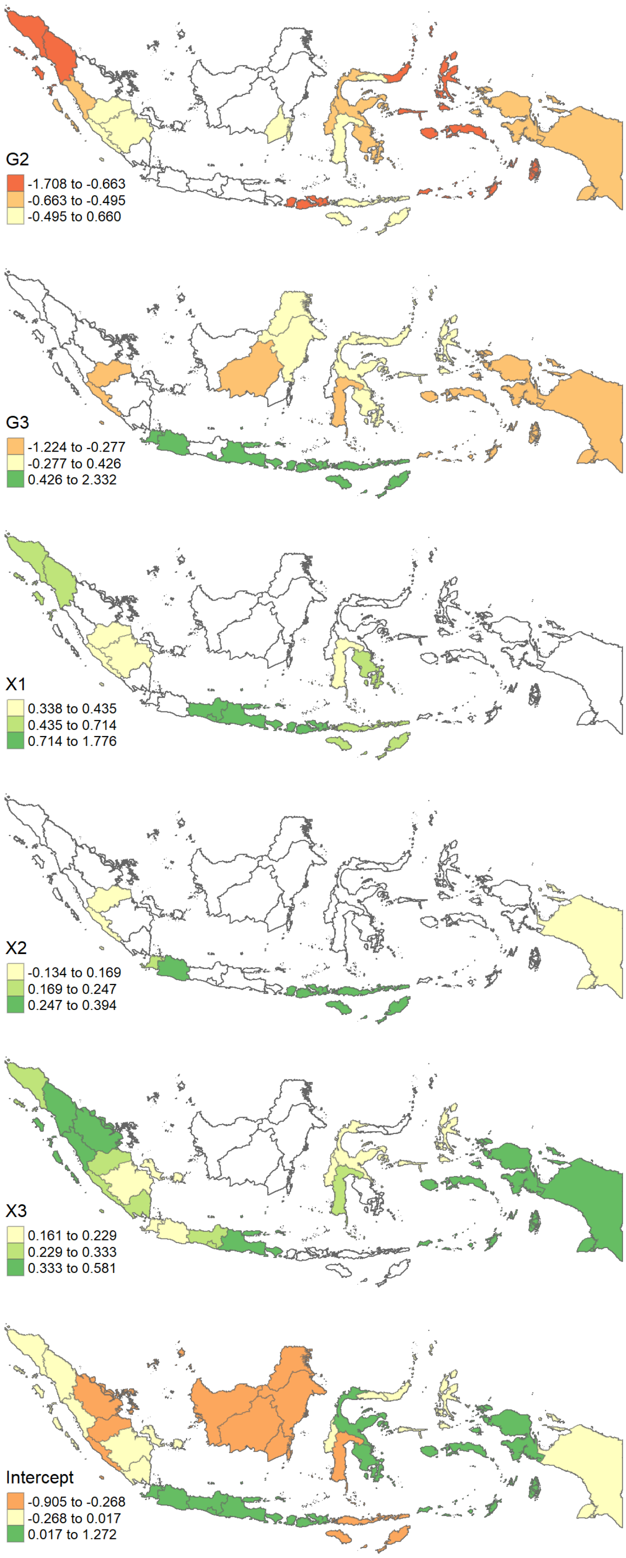


Figure 6: GWR coefficient estimator with Adaptive Bisquare Kernel

Based on Figure 6, there is information that districts/cities in West Kalimantan and Riau Islands provinces do not show the presence of variables that have a significant effect. In contrast, all explanatory variables appear important in districts/cities in Bengkulu and Jambi provinces. The economic growth rate (G_3) and the percentage of households without access to clean water (X_3) are variables that are influential in almost all districts/municipalities in Indonesia. However, it is essential to note that the accuracy of the standard error calculation can be affected by the repeated use of data in estimating the coefficients and determining the bandwidth value, so there is a possibility of inference errors from the significant results of the GWR model [16].

A. Results of HGWR

Implementing the method proposed by [9], namely HGWR, which is a combination of HLM and GWR. In this study, the HGWR model was built according to equation (10) using Adaptive Gaussian and Bisquare Kernels, as has been applied in the previous analysis stage using GWR, and the results are contained in Table 4.

Table 4: Diagnostic information from results of HGWR with different Kernel functions

Indicator	Kernel Function	
	Adaptive Gaussian	Adaptive Bisquare
Bandwidth	19	26
R^2	0.8000	0.8004

Table 4 illustrates that the R^2 values of both Kernel types are almost precisely the same. However, based on Figure 7, the curve of the Adaptive Bisquare Kernel appears more stable as it does not show significant fluctuations. Therefore, in this HGWR model, it was decided to use the Adaptive Bisquare Kernel function as it is considered to provide more reliable results. A comparison between Figures 5(b) and 7(b) also shows that bandwidth optimization in HGWR offers more reliable results, overcoming the problems that arise in bandwidth optimization in GWR. In addition, the HGWR model has a higher R^2 value compared to HLM and GWR. Thus, HGWR is the best method compared to HLM and GWR in modeling the poverty rate in Indonesia, which has a spatial hierarchy structure.

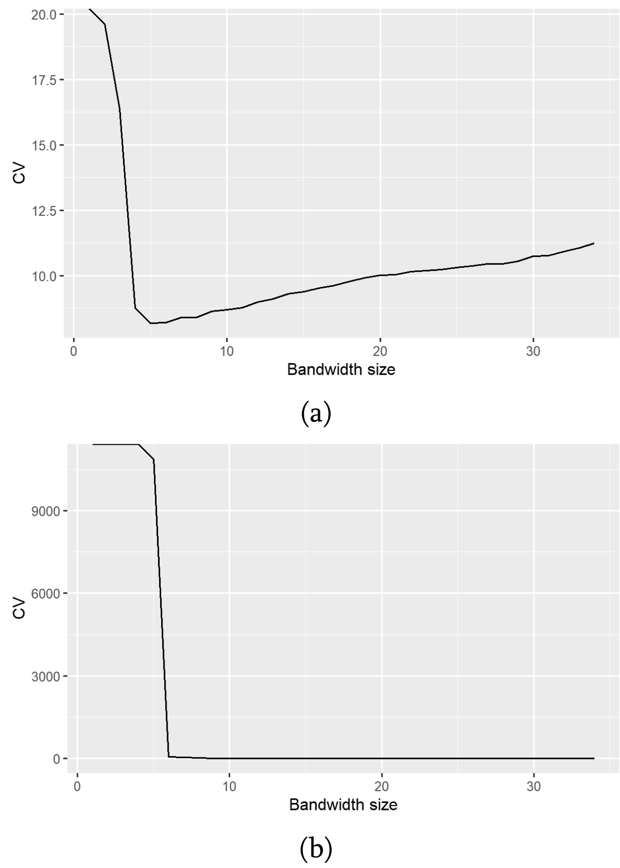
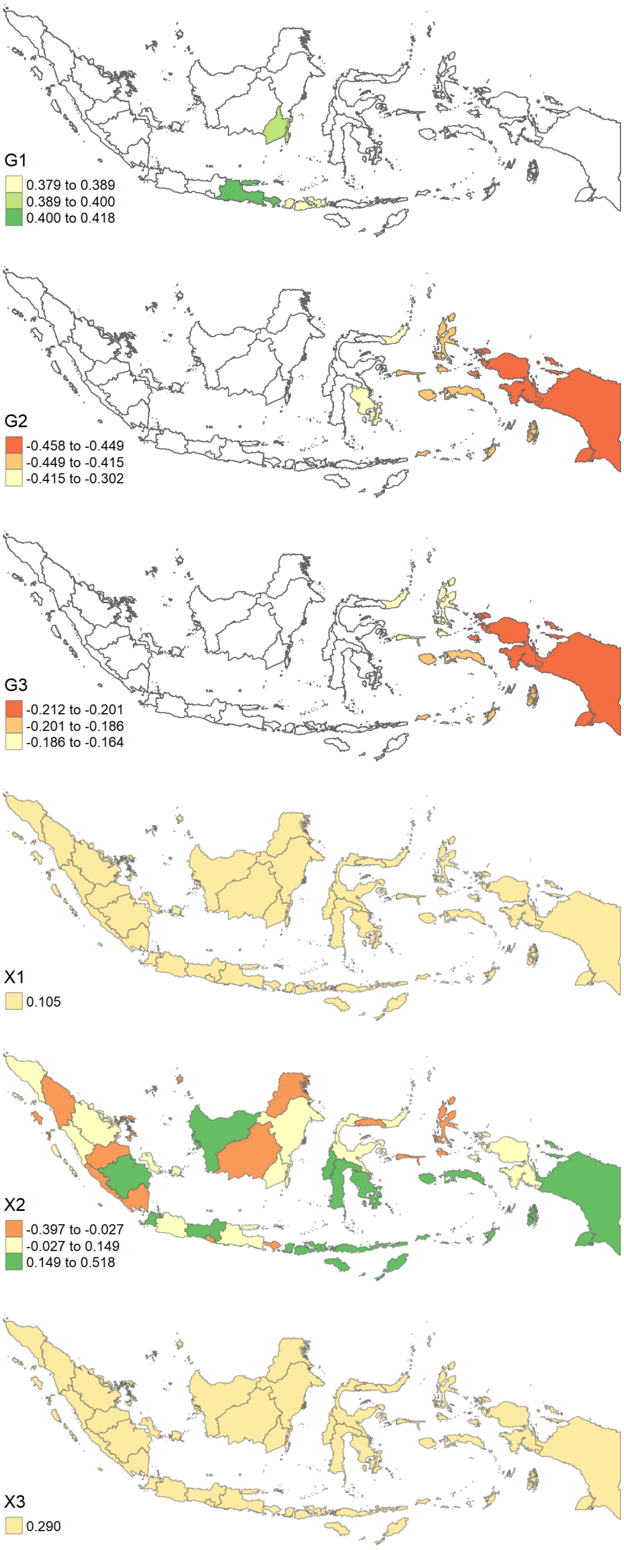


Figure 7: Bandwidth optimization of HGWR model via CV using (a) Adaptive Gaussian Kernel and (b) Adaptive Bisquare Kernel

Figure 8 shows the results of the HGWR coefficient estimates, where most of the local estimates for variables G_1, G_2, and G_3 are significant in provinces located in the eastern region of Indonesia. According to Central Statistics Agency, this might happen because provinces in the western region dominate development in Indonesia, so factors such as G_1, G_2, and G_3 do not have a significant impact directly on the poverty rate in provinces located in the western region of Indonesia. On the other hand, the variables X_1, X_2, and X_3 as sample-level variables are only involved in the early stage, namely in the estimation process using the HLM model. As presented in the HLM results section, variables X_1 and X_3 are estimated with fixed effects, while X_2 is estimated with random effects. Thus, X_1 and X_3 have constant estimates, while X_2 has estimates that vary across groups (provinces). All three estimators are significant based on the significance test results from the HLM model.



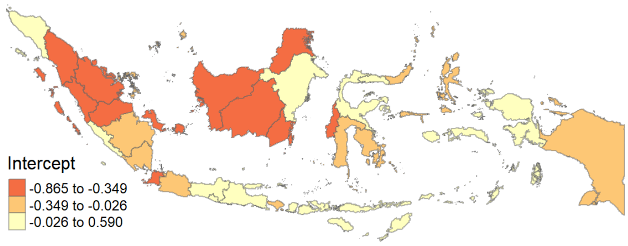


Figure 8: HGWR coefficient estimator with Adaptive Bisquare Kernel

IV.CONCLUSION

Based on the results of the analysis, it can be concluded that HGWR is the best method for modeling poverty rates in Indonesia, which has a spatial hierarchical structure when compared to HLM and GWR. This is evident from the higher R^2 value of the HGWR model, which is 0.8004, compared to HLM and GWR. HGWR is also able to model the spatial hierarchy structure and spatial heterogeneity simultaneously. According to the HGWR model, most of the local estimators for population dependency ratio (G_1), adjusted per capita expenditure (G_2), and economic growth rate (G_3) showed significance in provinces located in eastern Indonesia. In addition, the fixed effects and random effects estimators, namely the percentage of households without access to electricity (X_1), the ratio of per capita normative consumption to net product (X_2), and the percentage of households without access to clean water (X_3) also have a significant influence on the poverty rate in Indonesia.

V. REFERENCES

[1]. A. Adji, T. Hidayat, H. Tuhiman, S. Kurniawan, and A. Maulana, “Measurement of Poverty Line in Indonesia: Theoretical Review and Proposed

Improvements," TNP2K Work. Pap., no. January, pp. 48-e, 2020, [Online]. Available: http://tnp2k.go.id/download/88787WP 48_Measurement of Poverty Line in Indonesia-Theoretical Review and Proposed Improvements.pdf

[2]. C. S. Agency, "Indonesia Poverty Profile in September 2022," no. 07, 2023.

[3]. A. S. Fotheringham, C. Brunsdon, and M. Charlton, Geographically Weighted Regression the Analysis of Spatially Varying Relationships. Chichester, England: Wiley, 2002.

[4]. M. N. Aidi, F. Ernawati, and Siswanto, "Analysis of Factors Affecting the Number of Poor People in Indonesia Using Geographycally Weighted Regression," no. 4, pp. 1538–1553, 2018.

[5]. M. Amiati, A. Adnan, and R. Yendra, "Poverty Data Analysis In Riau Province Using Geographically Weighted Regression Model With Exponential And Tricube Adaptive Kernels," Int. J. Adv. Res. Publ. , vol. 4, no. 1, pp. 1–4, 2020, [Online]. Available: www.ijarp.org

[6]. S. W. Raudenbush and A. S. Bryk, Hierarchical Linear Models. Thousand Oaks, CA SAGE Publications, Inc, 2002.

[7]. H. Goldstein, Multilevel Statistical Models. West Sussex: John Wiley and Sons, Ltd., 2011.

[8]. J. Hox, Multilevel Analysis: Techniques and Applications. New Jersey: Lawrence Erlbaum Associates Publishers, 2002.

[9]. Y. Hu, B. Lu, Y. Ge, and G. Dong, "Uncovering spatial heterogeneity in real estate prices via combined hierarchical linear model and geographically weighted regression," Environ. Plan. B Urban Anal. City Sci., vol. 49, no. 6, pp. 1715–1740, 2022, doi: 10.1177/23998083211063885.

[10]. S. Nakagawa, P. C. D. Johnson, and H. Schielzeth, "The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded," J. R. Soc. Interface, vol. 14, no. 134, 2017, doi: 10.1098/rsif.2017.0213.

[11]. B. Lu, P. Harris, M. Charlton, and C. Brunsdon, "The GWmodel R package: Further topics for exploring spatial heterogeneity using geographically weighted models," Geo-Spatial Inf. Sci., vol. 17, no. 2, pp. 85–101, 2014, doi: 10.1080/10095020.2014.917453.

[12]. D. Octaviani, H. Pau, M. Fatima, and D. Putri, "Application of Geographically Weighted Regression Method on the Human Development Index of Central Java Province," vol. 3, no. 2, pp. 189–201, 2023.

[13]. W. Pribadi and F. Kartiasih, "Enviromental Quality and Poverty in Indonesia," J. Pengelolaan Sumberd. Alam dan Lingkung., vol. 10, no. 1, pp. 89–97, 2020, doi: 10.29244/jpsl.10.1.89-97.

[14]. A. Choirul Huda, A. Az-Zahra, F. Putri Yasmin, I. Wahyu Kusuma Ningrum, and W. Surya Putra, "Spatial Regression Analysis of Poverty Percentage on Eastern Indonesia in 2022," pp. 747–756, 2023.

[15]. J. Garza-Rodriguez, "Poverty and economic growth in Mexico," Soc. Sci., vol. 7, no. 10, 2018, doi: 10.3390/socsci7100183.

[16]. A. R. da Silva and A. S. Fotheringham, "The Multiple Testing Issue in Geographically Weighted Regression," Geogr. Anal., vol. 48, no. 3, pp. 233–247, 2016, doi: 10.1111/gean.12084.

Cite this article as :

Cintia Septemberini, Muhammad Nur Aidi, Anang Kurnia, "Modeling Poverty Data in Indonesia with Spatial Hierarchy Structure Using HLM, GWR, and HGWR Methods", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 11 Issue 1, pp. 260-271, January-February 2024. Available at doi : https://doi.org/10.32628/IJSRSET2411126

Journal URL : https://ijsrset.com/IJSRSET2411126