International Journal of Science: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Analysis of Factors Affecting District/City GRDP in Kalimantan Island

Dikky Wirwana[a*], Muhammad Nur Aidi[b], Anwar Fitrianto[c]

[a,b,c]*IPB University, Jl. Raya Dramaga, Babakan, Dramaga District, Bogor City, West Java, Indonesia.*

[a]*Email: dikkywirwana09@gmail.com*

[b]*Email: muhammadai@apps.ipb.ac.id*

[c]*Email: anwarstat@gmail.com*

Abstract

The Gross Regional Domestic Product (GRDP) is the added value of production obtained from various sectors. The value of GRDP is one of the indicators to see and measure the economic growth of a region. When compared to other islands, Kalimantan Island has a GRDP value that is quite low. Therefore, regression analysis is needed to see what factors affect the value of GRDP. However, the problem that is often found is that the local conditions of each place are different. There are many things behind it, one of which is in terms of geography. This is often referred to as spatial heterogeneity. One of the spatial modeling techniques that overcomes spatial heterogeneity is Geographically Weighted Regression. Because the weighting is based on the location of the observation or the area, it is possible that modeling on more than one explanatory variable has multicolinearity. There are several methods that are able to overcome multicolinearity in the GWR model, including Ridge regression and Least Absolute Shrinkage and Selection Operator (LASSO). In this study, the best model is the Geographically Weighted Regression model with a coefficient of determination (R^2) of 97.63% and an RMSE value of 258711464. The dominant factors affecting the value of GRDP at each location are the Human Development Index (IPM), the number of workers, and the percentage of households using electricity.

Keywords: GRDP; Multicollinearity; Geographically Weighted Regression.

---

* Corresponding author.

1. Introduction

The Gross Regional Domestic Product (GRDP) is the added value of production obtained from various sectors. The GRDP value is obtained from the total added value generated for all business units in a region or is the entire value of final goods and services produced by all economic units in a region. The value of GRDP is one of the indicators to see and measure the economic growth of a region. According to the Central Statistics Agency (2018), the distribution of the value of GRDP per island in Indonesia for the last 4 years from 2014-2018 has not changed significantly [1]. The island of Java is the island with the largest GRDP distribution in Indonesia. When compared to other islands, Kalimantan Island has a GRDP value that is quite low. The value of GRDP from year to year can change. An increase or decrease in the production of goods and services can be indicated through an increase or decrease in the value of GRDP in an area. The increase and decrease in value is due to several factors that influence it. Of course, to find out what causes the value of GRDP from year to year, an analysis is needed that can see what influences this. One of the analyses that can see the relationship between variables is regression analysis.

Regression analysis is used to explain the relationship between the response variable and the explanatory variable. The regression coefficient in the regression model applies globally to each observation location. The regression model is good if there is no spatial variation between locations. However, the problem that is often found is that the location conditions of each place are different. There are many things behind it, one of which is in terms of geography. This is often referred to as spatial heterogeneity.

Spatial heterogeneity is caused by differences in diversity between points of observation. Spatial heterogeneity is reflected in the error in the measurement, which results in heteroscedasticity, meaning that the resulting error variance is not constant [2]. Spatial heterogeneity cannot be ignored in forming a regression model because it will produce a large variance. This results in the hypothesis testing being carried out not giving good results [3]. To overcome the problem of spatial heterogeneity, a method is needed that can overcome this problem. One of the spatial modeling techniques that overcomes spatial heterogeneity is Geographically Weighted Regression .

Geographically Weighted Regression (GWR) is a development of linear regression in which the GWR regression coefficient value applies locally. This model calculates the parameters at each observation location, or in other words, takes into account the location of the observation data. Because each area of the observation location has a different regression parameter value because it is weighted based on the observation location or the area, it is possible that modeling on more than one explanatory variable experiences multicollinearity.

Multicollinearity is a condition in which one or more explanatory variables are correlated with other explanatory variables. The presence of multicollinearity will cause the regression coefficient estimator obtained from the least squares method (MKT) to produce a large variance. If there is multicollinearity between the explanatory variables, then estimating the regression coefficient with MKT will produce an unbiased estimator, but the estimator may have a large variance [4]. Great variety leads to hypothesis testing tending to accept $H_{0,}$ which means that the regression coefficient cannot be estimated with a high degree of accuracy [5]. There are several methods that are able to overcome multicollinearity in the GWR model, including Ridge regression and Least

Absolute Shrinkage and Selection Operator (LASSO).

Several previous studies have examined the problem of overcoming multicolinearity in the GWR model using the Ridge and LASSO regression models. Reference [6] compared Geographically Weighted Ridge Regression (GWRR) and Geographically Weighted Lasso Regression (GWLR) to model GRDP data in 113 districts/cities on the island of Java. The results show that the GWRL model is better than the GWRR model in overcoming the problem of multicolinearity. The results show that the GWLR model is better than the GWR and GWRR models. GWLR has a better prediction accuracy and a better level of stability with a coefficient of determination obtained of 98.37% and a Root Mean Square Error value of 2.3379. GWLR is considered more consistent in dealing with local multicollinearity problems even though the explanatory variables have a high level of multicollinearity.

Reference [5] made a comparison between geo-weighted Gulud regression and geo-weighted Lasso regression on data containing multicolinearity to model GRDP data in 27 districts/cities in West Java province. The results showed that the Lasso regression model showed better performance than the Gulud regression in dealing with multicollinearity, with a coefficient of determination of 99.79% and a Root Mean Square Error value of 0.079. The dominant explanatory variables influencing the value of PAD in districts and cities in West Java province based on the best model are the number of industries, the number of markets, and the number of foreign and domestic tourists.

In this study, GRDP will be modeled with the best model of GWR, GWRR, and GWLR evaluated from the coefficient of determination and Root Mean Square Error. Then they will investigate what factors affect the value of GRDP in 56 districts or cities on the island of Kalimantan.

2. Methodology

2.1 Spatial Heterogeneity Test

Spatial heterogeneity in the model occurs due to differences in characteristics between observation points as well as socio-cultural and geographical conditions. Spatial heterogeneity is reflected in the error in the measurement, which results in heteroscedasticity, meaning that the variance of the resulting error is not constant [2]. To detect the presence or absence of spatial heterogeneity in the model, the Breusch-Pagan test was carried out with the following hypothesis:

$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_n^2 = \sigma^2$ (no spatial heterogeneity)

H_1 : at least one $\sigma_i^2 \neq \sigma^2$ (there is spatial heterogeneity) ; $i = 1, 2, ..., n$

Test Statistics:

$$BP = \left(\frac{1}{2}\right) f^T Z (Z^T Z)^{-1} Z^T f \sim \chi^2_{(p+1)} \quad (1)$$

Where BP is the value of the Breusch-Pagan test , the remainder for the $e_i$ith observation with a matrix of size (nx1), f vector of size (nx1), n the number of observation areas, σ^2= variance of the remainder of e_i, X matrix of size nx (p + 1) which contains a vector **X** with standardized observations. p is the number of explanatory variables.

decision making on the Breusch-Pagan test (BP) rejects H_0if BP > $\chi^2_{(p+1)}$where $\chi^2_{(p+1)}$is the critical point of the test $\chi^2$with a significant level of α.

2.2 Multicolinearity

Multicholinearity is a condition in which there is one or more explanatory variables that correlate with other explanatory variables. Multicholinearity is a condition where there is an almost perfect linear relationship (near dependence) on the columns of the **X matrix** . If there is a perfect linear relationship, $\left|X^T X\right| = 0$ this condition is called exact multicolinearity [7].

Multicolinearity can be seen from the Pearson correlation value between independent variables. If the correlation between independent variables is high, then it is likely that there is an indication that the data has multicolinearity . In addition, another indicator that can detect the presence of multicolinearity is by looking at the value of the VIF (Variance Inflation Factor). The tolerance value, which indicates the presence of multicolinearity, is less than 0.20 or 0.10 and or the VIF value is greater than 5 or 10. The VIF value greater than 10 greatly affects the least squares estimate of the regression coefficient [8]. In GWR modeling, the VIF value is calculated for each of the explanatory variables. The VIF value is expressed as follows:

$$VIF_k(u_i, v_i) = \frac{1}{1 - R_k^2(u_i, v_i)} \tag{2}$$

where $R_k^2(u_i, v_i)$ is the coefficient of determination between X_k the other explanatory variables for each location (u_i, v_i) [9].

The correlation values for each explanatory variable at each location i are as follows:

$$r_{i,k} = \frac{\sum_{j=1}^{k} w_{ij}(x_j - \bar{x}_i)(y_j - \bar{y}_i)}{\sqrt{\sum_{j=1}^{k} w_{ij}(x_j - \bar{x}_i)^2}\sqrt{\sum_{j=1}^{k} w_{ij}(y_j - \bar{y}_i)^2}} \tag{3}$$

2.3 Geographically Weighted Regression (GWR)

The GWR model is a development of linear regression. However, GWR differs from linear regression, which is generally applied at each observation location. GWR produces local parameter estimates for each observation location [10]. In general, the multiple linear regression model is written as follows:

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + \varepsilon_i \; ; \; i = 1, 2, ..., n \qquad (4)$$

where $(\beta_0, ..., \beta_p)$ is the location parameter coefficient and assumed residual $\varepsilon_i \sim N(0, \sigma^2)$. The regression model of equation (4) was then developed into a GWR model. Reference [11] explained that the estimation of local parameter regression coefficients from the GWR model was carried out using the Weighted Least Square (WLS) method, namely by giving different weights for each observation location. The model of GWR to is as follows [10]:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \; ; \; i = 1, 2, ..., n \qquad (5)$$

Where y_i is the value of the response variable to the i-th location, is the value of the x_{ik} k - th explanatory variable at the location (u_i, v_i), β_0 is the intercept value in the GWR model, β_k is the local parameter value for each location (u_i, v_i), and the remainder is assumed $\varepsilon_i \sim N(0, I\sigma^2)$.

2.4 Spatial Weighting Function

Generally, the weight function with a kernel is used for weighted least-squares estimation in the GWR model. Found that spatial weighting greatly affects the results of the GWR estimation [10]. This is because the location around the i-observation greatly affects the parameter estimation at the i-location. Therefore, it is necessary to select the right weighting function to get an accurate model. The weighting used in this study is the Exponential kernel weighting with the following form of function:

$$w_j(u_i, v_i) = \exp\left[\left(-\frac{d_{ij}}{b}\right)\right] \qquad (6)$$

where b is the window width (bandwidth), d_{ij} is the distance from the i-th location to the j-th location obtained from the Euclidean distance as follows:

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2} \qquad (7)$$

Another thing to consider is estimating the bandwidth before doing GWR modeling . A Bandwidth is a circle with a radius of r from the center point of the location that serves as the basis for determining the weight of each observation against the regression model for each location. It is very important to use the bandwidth selection method to estimate the correct function (kernel). An extremely small bandwidth width value will result in an enlarged variance. One way to choose the optimum bandwidth can be done by using the method (Cross Validation , CV). Mathematically, CV can be written as follows:

$$CV(h) = \sum_{i=1}^{n} \left[y_i - y_{\neq i}(h) \right]^2 \quad (8)$$

where $y_{\neq i}$ is the estimated value for y_i by eliminating the observation of the i-th location point in the prediction process and the optimum bandwidth (h) will be obtained by an iterative process until the minimum CV is obtained [10].

2.5 Ridge Regression

Ridge regression was first introduced to control for the instability of the least squares estimator [12]. Ridge regression is an alternative to overcome the multicolinearity between explanatory variables . The Ridge regression solution is obtained by using the least squares method, which is to minimize the number of residual squares of the regression by adding constraints (λ) to the least squares method so that the coefficient shrinks to zero [13]. The Ridge regression coefficient estimator is obtained by minimizing the following equation:

$$\beta_R = \arg\min \left\{ \sum_{i=1}^{n} \left(y_i - \beta_o - \sum_{k=1}^{p} x_{ij} \beta_k \right)^2 + \lambda \sum_{k=1}^{p} \beta_k^2 \right\} \quad (9)$$

with a constraint $\sum_{k=1}^{p} \beta_k^2 \leq \rho$, where ρ is a quantity that controls the amount of depreciation with a value of $\rho \geq 0.$

In the regression coefficient estimation method in ridge regression, there is a value λ that plays a role in controlling the amount of shrinkage. If λ = 0 then the least squares estimate will be obtained. If it is λ increased, then the absolute value of the estimated coefficient becomes smaller, going to zero to λ go to infinity [14]. The solution of the Ridge regression can also be obtained by the equation:

$$\beta_R = \left(X^T X + \lambda I \right)^{-1} X^T Y \quad (10)$$

where I is a pxp sized identity matrix .

The optimal value selection can be obtained by using the λ generalized cross validation (GCV) [15]. The optimal coefficient estimator is obtained from the selection of the value λ that produces the minimum GCV value. The GCV value is formulated as follows:

$$GCV = \frac{\sum_{i=1}^{n} e_{i,\lambda}^2}{\left\{ n - \left[1 + tr(H_\lambda) \right] \right\}^2} \quad (11)$$

with the $e_{i,\lambda}^2$ i-th squared remainder for a given value of c, is the H_λ hat matrix .

2.6 Geographically Weighted Ridge Regression (GWRR)

GWRR is one method that can overcome the problem of multicollinearity in spatial data. GWRR is an extension of the Ridge regression method. The difference between Ridge regression and GWRR is the use of weights as additional information in estimating the regression coefficients [16]. The model of GWRR is as follows:

$$\beta_R = \arg\min \left\{ \sum_{i=1}^{n} \left(y_i - \beta_o(u_i, v_i) - \sum_{k=1}^{p} x_{ij} \beta_k(u_i, v_i) \right)^2 + \lambda \sum_{k=1}^{p} \beta_k^2(u_i, v_i) \right\} \quad (12)$$

The solution of GWRR can also be obtained by the equation:

$$\beta_R(u_i, v_i) = \left(X^T W(u_i, v_i) X + \lambda I \right)^{-1} X^T W(u_i, v_i) Y \quad (13)$$

where **X is the** nxp +1 matrix of explanatory variables, **y** is the response variable column vector of size nx1, **I** is the identity matrix of size pxp , λ is a positive bias constant, and $W(u_i, v_i)$ is a weighted diagonal matrix of size nxn.

2.7 LASSO Regression

Least Absolute Shrinkage and Selection Operator (LASSO) regression was first introduced by [17]. The LASSO parameter coefficient estimator cannot be obtained in closed form as in MKT or Ridge regression, but by using quadratic programming [13]. LASSO is defined as follows:

$$\beta_L = \arg\min \left\{ \sum_{i=1}^{n} \left(y_i - \beta_o - \sum_{k=1}^{p} x_{ij} \beta_k \right)^2 + \lambda \sum_{k=1}^{p} |\beta_k| \right\} \quad (14)$$

with the condition that $\sum_{k=1}^{p} |\beta_k| \le t$, $\sum_{k=1}^{p} |\beta_k| \le t$ is the same as adding a penalty $\lambda \sum_{k=1}^{p} |\beta_k|$ to the sum of the squares of the remainder, so that there is a direct relationship between the t parameter and λ controlling the amount of shrinkage of the regression coefficient [17]

Solved the LASSO problem by modifying the LAR algorithm. The LAR algorithm is as follows [18]:

1. Standardize the explanatory variable so that it has a mean of zero and a variance of one. Start with the remainder $r = y - \bar{y}, \beta_1, \beta_2, ..., \beta_p = 0$.

2. Choose the explanatory variable x_j that is most correlated with r .

3. Change the value β_j from 0 moving towards the least squares coefficient (x_j, r) until the other explanatory variable x_k has a correlation as large as the correlation x_j with the current remainder.

4. Change the value β_j and β_k move in the direction defined by the coefficient of least squares along with the current remainder of the same magnitude.
5. Continue this way until all the explanatory variables have been entered. After min(N-1,p) steps, the full least squares solution is obtained.

Modify the LAR algorithm for the LASSO solution by modifying step 4 in a way that if the non-zero coefficient reaches zero, remove the variable from the active variable group and recalculate the direction of the least squares together.

2.8 Geographically Weighted LASSO Regression (GWLR)

The concept of LASSO which is applied to the GWR model is a spatial method used for heterogeneity and multicollinearity which is then known as Geographically Weighted Regression LASSO. The solution from GWLR is to solve the LASSO formulation with the following constraints:

$$\beta_L = \arg\min \left\{ \sum_{i=1}^{n} \left(y_i - \beta_o(u_i, v_i) - \sum_{k=1}^{p} x_{ij} \beta_k(u_i, v_i) \right)^2 + \lambda \sum_{k=1}^{p} \left| \beta_k(u_i, v_i) \right| \right\} \quad (15)$$

with $\sum_{k=1}^{p} \left| \beta_k(u_i, v_i) \right| \leq s_i$ absolutes. The final estimation of LASSO parameters is carried out simultaneously with the final LASSO solution depending on the previously estimated kernel bandwidth [19].

2.9 Data

In this study, the type of data used is secondary data that can be obtained from the Central Statistics Agency (BPS). The data is Gross Regional Domestic Product (GRDP) at current prices in 2018. The units of observation in this study are 47 districts and 9 cities in the province of the island of Kalimantan. The response variables and explanatory variables can be seen in Table 1 as follows:

Table 1: Research variables

Variable	Variable Name	Unit
Y	GRDP on the basis of the price of the business field	Million Rupiah
X_1	Human Development Index (HDI)	Percent
X_2	Total population	Soul
X_3	Total manpower	Soul
X_4	District/City Regional Minimum Wage	Rupiah
X_5	Locally-generated revenue	Rupiah
X_6	Percentage of Households Using Electricity	Percent
X_7	Percentage of Households Using Gas	Percent
X_8	Number of Hotels and Lodging	Unit

2.10 Data Procedure

The steps in real data analysis are as follows:

1. Exploring data on response variables and explanations to find out the general picture of the data.
2. Pearson correlation analysis (r) was conducted to determine the effect of the explanatory variable on the response variable [20]:

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} \quad (19)$$

where r is the value of the correlation coefficient between the explanatory variable and the response variable. The hypothesis used is as follows:

$H_0 : \rho = 0$
$H_1 : \rho \neq 0$

where the test statistic is $t_0 = \frac{r\sqrt{n-2}}{\left(1-r^2\right)}$, with the critical area H_0 rejected if $t_0 > t_{table}$. If H_0 is rejected , it can be concluded that there is a correlation between two or more variables being compared.

3. Perform linear regression modeling using the least squares method (MKT) with the equation [4]:

$$y_i = b_0 + \sum_{j=1}^{k} x_{ij} b_j + e_i \quad (20)$$

4. Breusch Pagan test to see if there is spatial heterogeneity in the data. With the hypothesis in equation (1).
5. Performing GWR modeling on the data with the following stages:

a. Estimating the bandwidth value with the fixed kernel Exponential function which minimizes the cross validation value in equation (8).
b. Form a weighting matrix in equation (6) for each observation location using the previously obtained bandwidth value.
c. Estimating the GWR model parameter values for each location with the bandwidth and weight values obtained from the previous step.

6. Detecting local multicollinearity in the GWR model with VIF using equation (2) and then looking for the value of the GWR correlation coefficient with equation (3).

7. Conducting GWRR modeling to overcome multicolinearity in the GWR model with equation (13) with the following steps:

a. Estimating the value of the coefficient of bias (λ) and bandwidth with the optimal fixed kernel Exponential function by using cross validation .

b. Form a weighting matrix $W(u_i, v_i)$ for each observation location using the previously obtained bandwidth values.

c. Estimating the estimated coefficient value of the regression parameters for each location based on the weighting matrix and the bias coefficient obtained previously in equation (13).

8. Modeling GWLR using the LARS algorithm by modifying the addition of a weighting matrix on the variables.

9. Mapping the predicted results obtained from the GWR, GWRR, and GWLR models to compare the results visually.

10. Comparing the R^2 and RMSE values obtained from the GWR, GWRR, and GWL models to determine the best model for estimating the value of GRDP. The formulas for R^2 and RMSE are as follows:

$$R2 = \frac{(JKT - JKG)}{JKT} \tag{21}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[y_i - y_i\right]^2} \tag{22}$$

with JKT (Sum of Squares Total) = $\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$, JKG (Sum of Squares of Errors) = $\sum_{i=1}^{n}\left(y_i - y_i\right)^2$ [9].

3. Results and Discussion

3.1 Data Exploration

Data exploration is very necessary to find out and get information about the data. The data used in this study is the district or city Gross Regional Domestic Revenue (GRDP) data on the island of Kalimantan in 2018. The value of GRDP in a district/city is influenced by several factors, one of which is the geographical condition of the area. Regions that become the center of government or the center of the economy tend to produce high GRDP values compared to other regions. The following is the result of mapping GRDP data in regencies or cities on the island of Kalimantan.

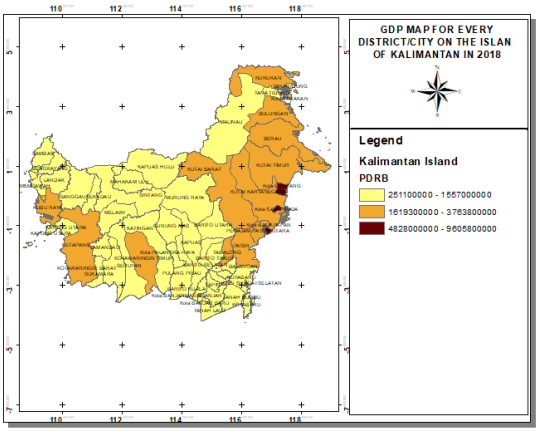

Figure 1: GRDP map of districts/cities in Kalimantan Island in 2018 (not scaled)

Figure 1 explains that the distribution of GRDP values varies. There are 4 districts or cities that have GRDP values in the high category (4828000000 – 9605800000) these districts or cities are Paser, Balikpapan, Samarinda, and Bontang. There are 18 regencies/cities that have GRDP values in the medium category (1619300000 – 3763800000) including Pontianak, Palangkaraya, and Banjarmasin. There are 34 regencies or cities that have GRDP values in the low category (251100000 – 155700000) including Singkawang, Barito Kuala, and Banjar Baru.

3.2 Multicollinearity and Pearson Correlation

Table 2: Value of Variance Inflation Factor (VIF)

Variable	VIF	Variable	VIF
X_1	3.283	X_5	5.299
X_2	25.858	X_6	1.354
X_3	19.882	X_7	2.069
X_4	1.697	X_8	2.030

Based on Table 2, it can be concluded that it shows the detection of multicolinearity between explanatory variables by showing the Variance Inflation Factor (VIF) value. The criteria for testing multicolinearity is by looking at the VIF value, if the VIF value is > 5 then there is multicolinearity. From Table 1, there are VIF values > 5, namely X_2, X_3, and X_5.

The relationship between the response variable and the explanatory variable can be seen from the resulting correlation coefficient. The correlation used is Pearson Correlation with $\alpha = 0.05$. The following is the correlation coefficient generated by the response variable with 8 explanatory variables.

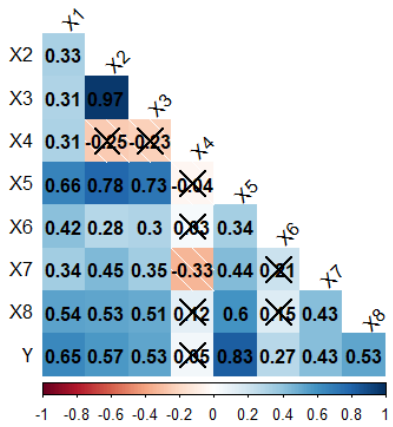

Figure 2: Pearson correlation coefficient value between the response variable and the explanatory variable

The correlation between explanatory variables can also be seen in Figure 2. It can be concluded that there are several explanatory variables that are correlated and real, and some that are not significant are given a cross on the correlation value in Figure 2. Because the variables X_2 and X_3 have very high correlation values. The researchers did not enter the X_2 variable into the model. After further testing, the best model was obtained by removing several variables, including X_2 (Number of Population), X_5 (Regional Original Income), and X_8 (Number of Hotels and Lodging).

3.3 Multiple Linear Regression Modeling

Before using the Geographically Weighted Regression (GWR) model, the linear regression model was first used to determine the initial analysis of the GRDP data and to find out the explanatory variables that had a significant effect on the GRDP data regardless of the location of the observations. The following are the results of parameter estimation, which can be seen in Table 3.

Table 3: Results of Regression Analysis

Variable	**Coefficient**	***p-value***
Intercept	-14253314183.63	0.000
X_1	194506043.142	0.000
X_3	7164,864	0.002
X_4	124.481	0.808
X_6	-11331174.081	0.433
X_7	25391655.794	0.175
R Square		0,561
F		12,800
Sig		0,000

5% alpha level are the Human Development Index (X_1) and the Number of Workers (X_3), while the other explanatory variables have no significant effect on GRDP. Based on the analysis of variance in Table 3, it can be concluded that the simultaneous testing of all explanatory variables performed on the regression model shows that all variables have a significant effect on GRDP with an R value of 56.1%. Simultaneous parameter testing has a *p-value* of 0.000 which means that all explanatory variables affect the response variable and the resulting coefficient of determination is 56.1%.

3.4 Breusch-Pagan test

The Breusch-Pagan test that was carried out resulted in a Chi-square value of 18,113 which was greater than $\chi^2_{5;0.05}$= 11,070 with a p-value of 0.002 which was less than the real level of 0.05, which means that the decision to reject H_0 was obtained, which means that there is an influence of spatial heterogeneity on the GRDP data in the district or city. The cities on the island of Kalimantan in 2018. This spatial heterogeneity indicates that each district/city on the island of Kalimantan has its own characteristics. A modeling method is needed to overcome the spatial heterogeneity. One of the modeling methods that can overcome the presence of spatial heterogeneity is Geographically Weighted Regression.

3.5 Comparison of GWR, GWRR and GWLR Model Models

The comparison of the best models is seen from the largest R^2 and the smallest RMSE value. The R^2 and RMSE values for each model will be presented in Table 4 as follows:

Table 4: Comparison of GWR, GWRR and GWLR models

Model	GWR	GWRR	GWLR
R^2	97.63%	72.18%	52.44%
RMSE	258711464	888169850	1161256178

Based on Table 4 the GWR model produces the largest R^2 value of 97.63% and the smallest RMSE of 258711464. In this case, it means that the GWR model is the best in modeling GRDP data compared to the GWRR and GWLR models.

3.6 Geographically Weighted Regression

The first step to obtain the model at each location required bandwidth (window width). The optimum bandwidth value is obtained by using the cross validation method as in equation 7, the optimum bandwidth is the one that produces the minimum cross validation coefficient value. The bandwidth value obtained is 0.6049869 with a CV value of 20.63374.

The GWR model produces a coefficient of determination (R^2) of 97.63%, this means that the GWR model is

able to explain the diversity of GRDP values of 97.63%, while the rest is explained by other variables outside the model. The map of the estimated value of GRDP in the GWR model is as follows:

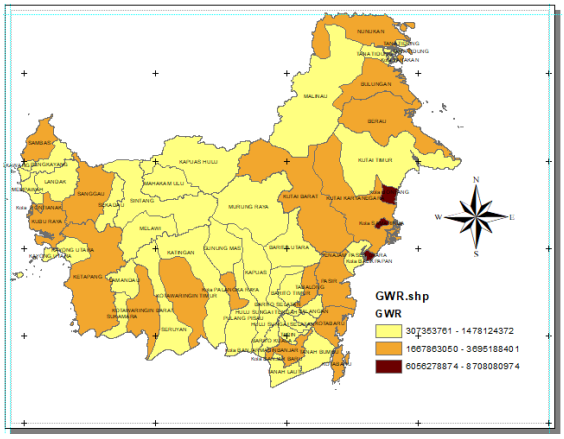

Figure 3: Map of the estimated value of GRDP in the GWR model (not scaled)

Figure 3 is a map of the estimated GRDP values in the GWR model showing that there are several districts/cities that have low, medium and high GRDP values. There are 34 regencies/cities with low GRDP, 19 regencies/cities with moderate GRDP, and 3 regencies/cities with high GRDP.

Three regencies/cities that have a high estimated value of GRDP include the city of Balikpapan, Samarinda City and the City of Bontang, with a range of GRDP values of 6056278874–8708080974. The three regions are influenced by the same explanatory variable, namely the Human Development Index. The Districts or cities that have a high category of estimated GRDP value are areas that are the capital of the province and are the economic centers of the province. In addition to being the provincial capitals and economic centers, these three regions are also central government areas.

The districts or cities that have an estimated value of GRDP in the medium category have a range of estimated values of GRDP from 1667863050–3695188401. In the estimated value of GRDP in the medium category, the estimated value of GRDP is influenced by explanatory variables that differ in each region. Tabalong and Paser Regencies are influenced by the explanatory variables of the Human Development Index (HDI) and the number of workers. The number of workers is influenced by the number of workers because these areas are areas that support Micro, Small and Medium Enterprises (MSMEs). Bulungan Regency is influenced by the explanatory variable, the percentage of households using electricity. While other areas are influenced by the number of workers.

The districts or cities that have a low category of estimated GRDP have a value range of 307353761 – 1478124372. North Barito, Murung Raya and Bulungan regencies are influenced by the explanatory variables of the Human Development Index (HDI) and the number of workers. The percentage of the population using electricity is influenced by the explanatory variable, while other areas are influenced by the number of workers. For simplicity and clarity, local coefficients of positive, negative, and no effect on the value of GRDP in each district or city on the island of Kalimantan will be presented.

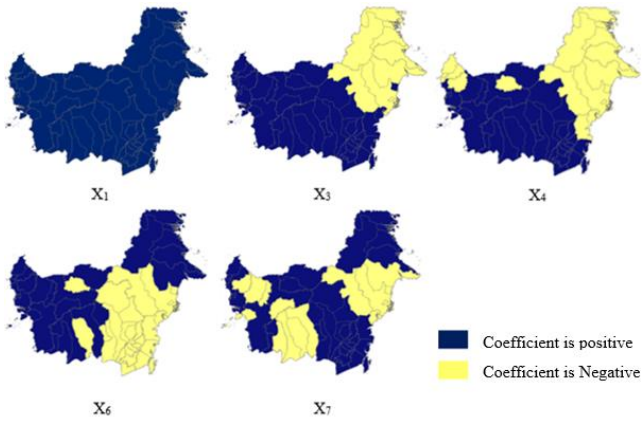

Figure 4: Local coefficient values in the GWR model (not scaled)

Based on Figure 4, the local coefficient values in the GWR model look like they have a pattern. The coefficients of X_3 (number of workers) and X_4 (regency/city regional minimum wages) have almost the same pattern, namely the West, Central, and South Kalimantan regions, the local coefficient is positive. This is presumably because in these areas, the number of workers and the regional minimum wage of districts or cities still greatly influence the value of GRDP. The number of workers and the district/city regional minimum wage have a positive effect on the value of GRDP because the areas of West, Central and South Kalimantan are on average dominated by industry and wholesale and retail trade.

The local coefficient X_1 (human development index) has all positive local coefficients, which means that each additional coefficient in the region makes a positive contribution to GRDP. This means that the human development index has a very influential role on the growth of the GRDP value on the island of Kalimantan.

The local coefficient X_6 (percentage of population using electricity) has a positive effect in 27 districts or cities and in 29 districts or cities has a negative effect on the value of GRDP. This means that the percentage of the population using electricity does not have a significant effect on the overall GRDP value of the island of Kalimantan. There are only a few areas. Like in the areas of West Kalimantan and North Kalimantan, which have a positive coefficient, this means that electricity consumption in West and North Kalimantan is quite high.

The coefficient X_7 (percentage of population using gas) has a positive effect in 40 districts or cities and 16 districts or cities have a negative effect on the value of GRDP. Based on BPS 2018 data, districts or cities that have a positive influence on the value of GRDP are areas that have a high percentage of the population using gas, such as the city of Pontianak, which has a percentage of 97.53%. Further parameter testing will be carried out. Parameter testing is carried out to see which variables are significant or can have a positive influence on the value of GRDP. Based on the parameter test, significant variables were $\alpha = 0.05$ obtained and 4 groups were obtained, namely as follows:

Table 5: Significant variables in each district/city

Group	County/city	Significant variable
1	North Barito, Murung Raya, Tabalong, Balangan and Paser.	Human Development Index (X_1) and Total Labor (X_3).
2	North Barito, Murung Raya, Tabalong, Balangan, Paser, West Kutai, Kutai Kartanegara, East Kutai, Berau, North Penajam Paser, Mahakam Ulu, Balikpapan, Samarinda and Bontang.	Human Development Index (X_1).
3	Hedgehog, Mempawah, Ketapang, Sekadau, Melawi, North Kayong, Kubu Raya, Pontianak, Singkawang, West Kotawaringin, East Kotawaringin, Kapuas, South Barito, North Barito, Sukamara, Lamandau, Seruyan, Pulang Pisau, Gunung Mas, East Barito, Murung Raya, Palangkaraya, Tanah Laut, Kota Baru, Banjar, Barito Kuala, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tabalong, Tanah Bumbu, Balangan, Banjarmasin, Banjar Baru, Paser.	Number of Workers (X_3).
4	Malinau, Bulungan.	Percentage of Households Using Electricity (X_6).

Based on Table 5 can explain the variables that have an influence on the value of GRDP for each district/city in any area, if it is concluded that there are four groups of districts/cities based on significant variables. Group one, namely North Barito, East Barito, Tabalong, Balangan and Paser, the variables of the Human Development Index and the Number of Workers are factors that influence the value of GRDP in the region. In the second group, including West Kutai and Kutai Kartanegara, the variables of the Human Development Index have an influence on the GRDP value in the region. The third group includes the Hedgehog and the Mempawah variable Number of Workers which have an influence on the GRDP value in the region. The last group or the fourth group, namely Malinau and Bulungan, the variable Percentage of Households Using Electricity which has an influence on the value of GRDP in the region. To make it easier and clearer, it will be presented in the form of

pictures of each group of variables that have an influence on the value of GRDP.

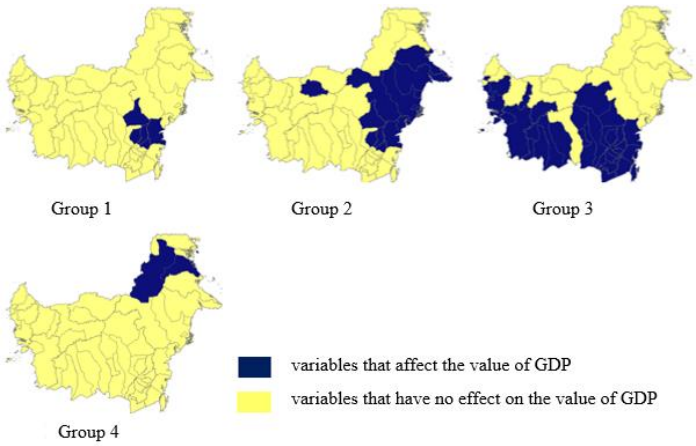

Figure 5: Variables that affect the value of GRDP (not scaled)

Based on Table 5 and Figure 5, it can be concluded that the variables that have an influence on the value of GRDP are like forming a pattern. Regencies or cities that have an influence on the value of GRDP are areas that are close to each other. This means that the regions that have an influence on the GRDP value have similarities in the variables that affect the GRDP value.

4. Conclusions and Suggestions

Based on the research results, the GWR model is best model for modeling GRDP data in 56 districts/cities on the island of Kalimantan. For the factors that influence the value of GRDP in the regency or city of Kalimantan based on the best model, the dominant GWR is the Human Development Index (IPM), the number of workers, and the percentage of households using electricity.

Based on the research that has been done and the conclusions obtained, there are problems that can still be solved in this study, namely checking outlier data and not entering it into the model if there is outlier data, or using Geographically Weighted Robust Regression if outlier data wants to be included in the model.

References

[1]. [BPS]. Central Bureau of Statistics. 2018. Gross Regional Domestic Product (GRDP) of Provinces in Indonesia 2014-2018. Bps. Jakarta

[2]. Anselin L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht(NL): Kluwer Academic.

[3]. Firdaus M. 2011. *Econometrics: An Applicative Approach*. Bumi Aksara Jakarta.

[4]. Myers RH.1990. *Classical and Modern Regression with Applications*. PWS KENT Publishing Company.

[5]. Wulandari R, Saefuddin A, Afendi FM. 2017, Application of Geographically Weighted Gulud Regression and Geographically Weighted LASSO Regression on Data Containing Multicollinearity: The Case of Regional Original Income Data in 27 Regencies/Cities in West Java Province

[Thesis]. Bogor (ID): Bogor Agricultural Institute.

[6]. Yulita, Tiyas., et al., 2016. Geographically Weighted Ridge Regression and Geographically Weighted LASSO Modeling On Spatial Data with Multicollinearity. [Thesis]. Bogor (ID): Bogor Agricultural Institute.

[7]. Draper NR, Smith H. 1998. *Applied Regression Analysis*. 3rd Ed. New York (US): John Wiley & Sons.

[8]. Friday OR, Emenonye C. 2012. The Detention and Correction of Multicollinearity Effects in a Multiple Regression Diagnostics. *Elixir Statistics* 49:10108- 10112.

[9]. Wheeler DC. 2007. Diagnostic Tools and a Remedial Method for Collinearity in Geographically Weighted Regression. *Environment and Planning A* 39: 2464-2481.

[10]. Fotheringham USA, Brunsdon C, Charlton M. 2002. *Geographically Weighted Regression the Analysis of Spatially Varying Relationships*. England (GB): John Wiley and Sons.

[11]. Leung Y, Mei CL, Zhang WX. 2000. Statistical Test for Spatial Nonstationarity Based on The Geographically Weighted Regression Model. *Environment and Planning A* 32: 9-32.

[12]. Hoerl AE, Kennard RW.1970. Ridge Regression: Biased Estimation for Nonthorgonal Problems. *Technometrics* .12: 55-67.

[13]. Hastie T, Tibshirani R, Friendman J. 2009. *The Elements od Statistical Learning Data Mining, Inference, and Prediction*. New York (US): Springer.

[14]. Draper NR, Smith H. 1992. *Applied Regression Analysis*. Ed.2nd. Sumantri B, translator. Jakarta (ID): Gramedia. Translation from: *John Wiley and Sons*.

[15]. Montgomery DC, Peck EA. 1992. *Introduction to Linear Regression Analysis*. Ed 2nd. New York (US): John Wiley & Sons.

[16]. Wheeler DC. 2006. Diagnostic Tools and Remedial Methods for Collinearity in Liniear Regression Models with Spatially Varying Coefficients. The Ohio State University.

[17]. Tibshirani R. 1996. Regression Shrinkage and Selection via The LASSO. *Journal of the Royal Statistical Society Series B*. 58(1): 267-288.

[18]. Hastie T, Thibsirani R, Friedman J. 2008. *The Elements of Statistical Learning Sata Mining, Inference and Prediction*. Ed 2nd. Springe. Stanford University.

[19]. Wheeler DC. 2009. Simultaneous Coefficient Penalization and Model Selection in Geographically Weighted Regression: The Geographically Weighted Lasso. *Journal of Environment and Planning A* 41 (3): 722-742.

[20]. Walpole RE. 1982. *Introduction to Statistics*. Ed 3th. Sumantri B, translator. Jakarta (ID): Gramedia Jakarta.