International Journal of Social Science and Human Research

ISSN (print): 2644-0679, ISSN (online): 2644-0695

Volume 07 Issue 06 June 2024

DOI: 10.47191/ijsshr/v7-i06-07, Impact factor- 7.876

Page No: 3605-3613

Modelling Mean Years of Schooling (MYS) Districts/Cities of West Kalimantan Province

Ketut Ananda Kuvera Witana[1], Made Susilawati[2], Ni Luh Putu Suciptawati[3], Made Ayu Dwi Octavanny[4], Anggun Yuliarum Qur'ani[5], Desak Putu Eka Nilakusmawati[6]

[1,2,3,4,5,6] Mathematics Department, Udayana University, Badung, Bali, Indonesia

ABSTRACT: Mean years of schooling (MYS) is one of the indicators measuring the quality of education in Indonesia, regulated through the nine-year compulsory education policy. West Kalimantan Province is one of the provinces with the lowest MYS rates in Indonesia, with only one out of its 13 districts/cities fulfilling the compulsory education policy. Therefore, research is needed to obtain the best model and understand the factors that influence the MYS districts/cities of West Kalimantan Province. This study uses the Geographically Weighted Panel Regression (GWPR) analysis method, due to the annual changes in MYS numbers indicating time effects in data and differences in MYS numbers in each district/city indicating differences in characteristics between regions. GWPR is a combination of panel data regression model and geographically weighted regression (GWR) model. The results showed that the best model obtained is the GWPR adaptive bisquare model with an AIC value of -147.241 and R^2 value of 80.07%. The variables of per capita income, dependency ratio, literacy rate, and the number of schools significantly influence MYS, and four groups of regions are formed based on significant variables.

KEYWORDS: Panel Data Regression, GWPR, MYS, West Kalimantan

I. INTRODUCTION

Education is a fundamental human right for every individual and plays a crucial role in human development efforts [1]. The role of education is further strengthened by its ability to produce quality human resources. Furthermore, The World Bank (2023) asserts that education is not only the key to poverty alleviation, health improvement, and gender equality achievement, but also the most important factor in achieving equity and inclusivity. In Indonesia, education plays an important role in achieving the nation's goals, namely to educate the nation's life, as stated in the fourth paragraph of the preamble 1945 Constitution [2]. In order to ensure that all Indonesians receive basic education or as an effort to equalize education, and driven by the spirit of Article 31 of the 1945 Constitution which guarantees the right of citizens to receive education and the government's obligation to provide it, the government issued a Nine-Year Compulsory Education Policy through Law Number 20 of 2003 concerning the National Education System which includes Elementary and Junior High Schools [3]. One of the indicators to measure the education level of people in a region is the Mean Years of Schooling (MYS) [4].

MYS is defined as the total years spent by people aged 15 years and over in education and does not include years spent repeating grades. Low MYS reflects the short duration of education in a region, indicating a low level of educational completion of the population. This condition results in the quality of the education system in the region tending to decline, and ultimately results in the low quality of human resources in the region [5]. Based on provincial MYS data in Indonesia by Badan Pusat Statistik (BPS) Indonesia, there has been an increase in MYS from year to year. However, there is a disparity in MYS achievement between provinces in Indonesia, with several provinces not yet meeting the 9-year target according to the compulsory education policy [6]. West Kalimantan Province is one of the provinces that has not yet reached the 9-year MYS target, with an MYS of 7.59 years in 2022. Viewed by its districts/cities, only 1 out of 14 districts/cities, namely Pontianak City, has achieved MYS above 9 years, while the other 13 districts/cities still have an MYS below 9 years [7]. Furthermore, there is a difference in the quality of education reflected in the variation of MYS between districts/cities in West Kalimantan.

Considering the government's efforts to achieve educational equity and to increase the MYS rate to meet the 9-year target according to the compulsory education policy, it is important to understand the factors that significantly influence the MYS of districts/cities in West Kalimantan Province. One way to understand the factors that influence MYS is to model the MYS of districts/cities using a regression approach, which is a statistical method for studying the relationship between independent variables (explanatory) and dependent variables (explained) [8]. In this case, changes in MYS over years indicate the presence of a time effect (temporal), and variations in MYS between districts/cities indicate spatial heterogeneity. Therefore, Geographically Weighted Panel

Regression (GWPR), which is able to accommodate both temporal and spatial effects, becomes the appropriate approach for modelling MYS in this research.

GWPR is a development of regression analysis that combines the concepts of panel data regression (PDR) and geographically weighted regression (GWR) [9]. PDR is a regression analysis that uses panel data, which is a combination of cross-sectional and time-series data, allowing for the analysis of relationships between individuals and their changes over time [10]. Meanwhile, GWR is a development of regression analysis that involves the geographic coordinates of a region, allowing for the estimation of regression parameters that vary in each geographic location [11]. GWR is a development of the regression model that overcomes the limitations of Ordinary Least Squares (OLS) in handling spatial data. OLS produces global regression coefficients that are assumed to be the same for all regions [11]. However, this approach is less accurate when the object of observation are locations and have different characteristics between regions, which can be caused by spatial effects such as spatial dependence (influence between regions) and spatial heterogeneity (differences in characteristics between regions) [12]. To overcome spatial heterogeneity, GWR was developed by involving geographic coordinates into the regression model. This allows the estimation of regression parameters locally for each location, thereby better accommodating spatial heterogeneity. By combining the concepts of PDR and GWR, GWPR is able to overcome cases of spatial heterogeneity in PDR model. Thus, this research aims to obtain the best model and identify the factors that significantly influence the MYS of districts/cities in West Kalimantan Province using the GWPR analysis method, which can accommodate the presence of temporal and spatial effects in the data. To provide more comprehensive information, data is used over five years period from 2018 to 2022.

II. METHODS

The data used in this research are secondary data sourced from BPS Indonesia and BPS of West Kalimantan Province. The data includes MYS and factors suspected to influence MYS from two cities and 12 districts in West Kalimantan Province from 2018 to 2022. The dependent variable (Y) studied in this research is the MYS of districts/cities in West Kalimantan Province with six independent variables, namely per capita income (X_1), dependency ratio (X_2), literacy rate (X_3), number of schools (X_4), student to teacher ratio (X_5), and school participation rate (X_6). Presented in full in Table 1.

Table 1. Research Variables

Code	***Variable***	***Description***
Dependent Variable		
$Y_{i,t}$	Mean Years of Schooling	total years spent by resident aged 15 years and above in completing education
Independent Variable		
$X_{1i,t}$	Per Capita Income	the average income per individual in a region
$X_{2i,t}$	Dependency Ratio	comparison between the number of unproductive individuals and productive individuals
$X_{3i,t}$	Literacy Rate	percentage of the population aged 15 years and above who have the ability to read and write
$X_{4i,t}$	Number of Schools	total number of educational institutions registered and recognized by the government in a region
$X_{5i,t}$	Student to Teacher Ratio	the result of division between the number of students and the number of teachers
$X_{6i,t}$	School Participation Rate	proportion or percentage of school-age children attending their age-appropriate level of education

This analysis is conducted using the GWPR method. GWPR combines the PDR and GWR models, so in the GWPR model, it considers the time dimension in the data and considers the geographic coordinates of each observation area. By combining the concepts of PDR and GWR, GWPR is able to address spatial heterogeneity in the PDR model. This research is conducted using Rstudio Version 4.3.1, Ms. Excel, and ArcGIS 10.8, with the steps for the data analysis method in this research are: (1) Conducting a descriptive analysis to provide a general overview of MYS districts/cities in West Kalimantan Province from 2018 to 2022; (2) Forming a PDR model of common effect (CE), fixed effect (FE) consisting of individual effect and time effect, and random effect (RE) [13]; (3) Selecting the PDR model is carried out using the Chow test and Hausman test [10]. The Chow test uses the hypothesis H_0 there is no difference in intercept between individual units and H_1 there is a difference in intercept between individual units, with the Chow test statistic being the F test with the following formula:

$$F_{calc} = \frac{(RSS_{CEM} - RSS_{FEM})/(N-1)}{RSS_{FEM}/(NT-N-K)}$$

Description:
RSS_{CEM} : residual sums of squares of the CE model
RSS_{FEM} : residual sums of squares of the FE model
N : number of cross section units

T : number of time series units
K : number of independent variables

If H_0 is rejected ($F_{calculated} > F_{N-1,N(T-1)-K}$ or $p_{value} < \alpha$), then the analysis continues to the Hausman test with the hypothesis H_0 error is not correlated with independent variables and H_1 error is correlated with independent variables. The test statistic for the Hausman test uses the following formula:

$$\chi^2_{calc}(K-1) = \left(\hat{\beta}_{FEM} - \hat{\beta}_{REM}\right)^T \left(var\left(\hat{\beta}_{FEM} - \hat{\beta}_{REM}\right)\right)^{-1} \left(\hat{\beta}_{FEM} - \hat{\beta}_{REM}\right)$$

Description:
$\hat{\beta}_{FEM}$: vector of FEM parameter estimates
$\hat{\beta}_{REM}$: vector of REM parameter estimates

If $\chi^2_{calculated}(K-1)$ is less than the critical value or $p_{value} > \alpha$, then there is not enough evidence to reject H_0, meaning the error is not correlated with the independent variables and the selected PDR model is the RE model. If $\chi^2_{calculated}(K-1)$ is more than the critical value or $p_{value} < \alpha$, then H_0 is rejected, meaning the error is correlated with the independent variables and the selected PDR model is the FE model, and continues to step 4. If H_0 in the Chow test cannot be rejected ($F_{calculated} < F_{N-1,N(T-1)-K}$ or $p_{value} > \alpha$), then there is no difference in intercept between individual units with the selected PDR model is the CE model and continues to step 4; (4) Testing the assumptions of the selected PDR model, consisting of multicollinearity test, autocorrelation test, heteroskedasticity test. Multicollinearity test conducted to determine whether there is a relationship between the independent variables in the model with the following test statistic formula [14]:

$$VIF_k = \frac{1}{(1-R_k^2)}$$

Description:
R_k^2 : coefficient determination of the regression equation between X_k and other independent variables

If the VIF value is greater than 5, it means there is multicollinearity, vice versa. Autocorrelation test conducted to determine whether there is autocorrelation between residuals using the Wooldridge test [15]. The hypothesis H_0 there is no autocorrelation and H_1 there is autocorrelation. If $p_{value} < \alpha$, then reject H_0, means there is autocorrelation, vice versa. Heteroskedasticity test conducted to determine whether there is a difference in the variance of the residuals from one observation to other observation, with the hypothesis H_0 there is no inequality of residual variance and H_1 there is inequality of residual variance, using the Breusch-Pagan test with the following test statistic formula [8]:

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

Description:
f : least square residual for the i-th observation
Z : an n × (k + 1) matrix containing normalized vector standardized for each observation.

(5) If there is not enough evidence to reject H_0 ($BP < \chi^2_{(k)}$ atau $p_{value} > \alpha$), it means the heteroskedasticity test is not fulfilled, then the analysis is completed on the PDR model. However, if H_0 is rejected ($BP > \chi^2_{(k)}$ atau $p_{value} < \alpha$), it means the heteroskedasticity test is fulfilled, then the analysis continued to GWPR modelling with the initial step of calculating the Euclidean distance for each district/city [16]; (6) Calculating cross validation (CV) values to determine optimum bandwidth [16]; (7) Calculating the weighting matrix using adaptive kernel functions, consist of adaptive gaussian kernel, adaptive bisquare kernel, and adaptive tricube kernel [11]; (8) Estimating parameters of the GWPR model using the weighted least squares (WLS) method [16]; (9) Determining the best model between the adaptive kernel functions, considering the model with lower AIC value and higher R^2 value [8]; (10) Conducting parameter testing on the GWPR model, includes testing the model's goodness of fit simultaneously and testing the significance of the model's parameters partially [16]. The model's goodness of fit testing is conducted to determine whether the parameters are significantly affected by location, with the hypothesis H_0 there is no β_k affected by location and H_1 there is at least one β_k affected by location (u_{it}, v_{it}). Test statistic

$$F_{calc} = \frac{RSS_{PDR}/df_1}{RSS_{GWPR}/df_2}$$

Description:
RSS_{PDR} : residual sums of squares of the PDR model
RSS_{GWPR} : residual sums of squares of the GWPR model
df : degree of freedom

If the value of $F_{calculated} > F_{\alpha(df_1,df_2)}$ or $p_{value} < \alpha$, then reject H_0, which means there is β_k that is affected by location and there is a difference between the PDR model and the GWPR model. The significance of the model's parameters testing is conducted to determine which independent variables significantly influence Y at location (u_{it}, v_{it}), with the hypothesis H_0 the parameter coefficient of variable X_k does not significantly influence Y and H_1 the parameter coefficient of variable X_k significantly influence Y. Test statistic:

$$t_{calc} = \frac{\hat{\beta}_k(u_{it}, v_{it})}{\hat{\sigma}\sqrt{C_{kk}}}$$

Description:

$\hat{\beta}_k(u_{it}, v_{it})$: vector of parameter estimates at location (u_{it}, v_{it})

C_{kk} : diagonal elements of the matrix $C_i C_i^T$

$\hat{\sigma}$: $\sqrt{RSS_{GWPR}/df1}$

If the value of $|t_{calculated}| > t_{\alpha/2(df)}$ or $p_{value} < \alpha$, then reject H_0, which means the parameter coefficient of variable X_k is significant influence Y; (11) Interpreting the GWPR model and conclusions.

III. RESULTS AND DISCUSSION

The results of the analysis are described as follows.

A. Descriptive Analysis

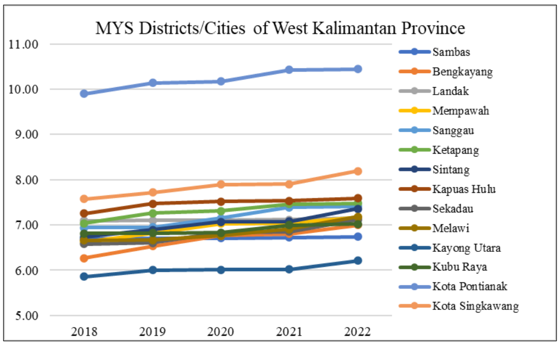

Figure 1. Graph of MYS Districts/Cities of West Kalimantan Province from 2018 to 2022

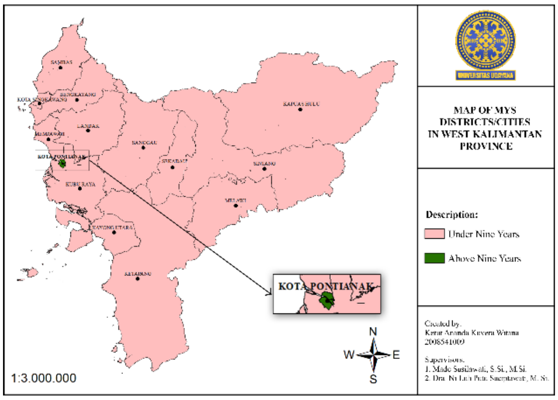

Figure 2. Map of MYS Districts/Cities of West Kalimantan Province from 2018 to 2022

Based on Figure 1 and Figure 2, it is found that the MYS districts/cities of West Kalimantan Province experienced an increase each year during the period from 2018 to 2022, indicating the influence of time series in the data. Although there was an increase, in line with the Nine-Year Compulsory Education Policy, only Pontianak City has an MYS rate above nine years, which is shown by the green area on the map. In contrast, the other 13 districts/cities have MYS rates below nine years, which is shown by the pink area on the map.

B. PDR Model Estimation

PDR model estimation is conducted using three approaches, namely CEM, FEM, and REM.

1) Common Effect Model (CEM)

The estimation results are as follows:

Table 2. Estimation of the common effect model parameters

Variables	*Parameters*	*Estimated Value*	p_{value}
Intercept	α	-0.12497	0.0010570
Per Capita income (X_1)	β_1	0.026657	0.0007492
Dependency Ratio (X_2)	β_2	-0.045535	0.0117136
Literacy Rate (X_3)	β_3	0.16518	1.287×10^{-5}

Number of Schools (X_4)	β_4	0.00041091	0.3650813
Student to teacher ratio (X_5)	β_5	0.082397	0.0153612
School Participation Rate (X_6)	β_6	0.015558	0.0589108
	R^2		**74.83%**
	p_{value}		$< 2.2 \times 10^{-16}$

According to Table 2, the R^2 value for CEM is 74.83%.

2) Fixed Effect Model (FEM)

Two fixed effect models are formed, namely fixed individual effect and fixed time effect. Fixed individual effect model uses 14 dummy variables representing each district/city, while fixed time effect model uses dummy variables representing the range of years from 2018 to 2022. The estimation results for the fixed individual effect model are as follows:

Table 3. Estimation of the fixed individual effect model parameters

Variables	***Parameters***	***Estimated Value***	p_{value}
Sambas	α_1	−0.791135	0.7325
Bengkayang	α_2	0.848385	0.6583
Landak	α_3	−0.533245	0.8221
Mempawah	α_4	2.368569	0.1650
Sanggau	α_5	−1.213983	0.6142
Ketapang	α_6	−2.168574	0.4091
Sintang	α_7	−0.872951	0.7118
Kapuas Hulu	α_8	−0.047317	0.9827
Sekadau	α_9	1.461482	0.4117
Melawi	α_{10}	1.035597	0.5786
Kayong Utara	α_{11}	2.373261	0.1240
Kubu Raya	α_{12}	1.873794	0.2725
Kota Pontianak	α_{13}	1.341616	0.5885
Kota Singkawang	α_{14}	−1.162264	0.6617
Per Capita income (X_1)	β_1	0.0241189	6.865×10^{-5}
Dependency Ratio (X_2)	β_2	−0.0233381	6.861×10^{-7}
Literacy Rate (X_3)	β_3	0.0370120	0.005944
Number of Schools (X_4)	β_4	0.0088783	0.004957
Student to teacher ratio (X_5)	β_5	−0.0025228	0.885800
School Participation Rate (X_6)	β_6	−0.0024414	0.349221
	R^2		**75.95%**
	p_{value}		7.01×10^{-14}

According to Table 3, the R^2 value for fixed individual effect model is 75.95%. The estimation results for the fixed time effect model are as follows:

Table 4. Estimation of the fixed time effect model parameters

Variables	***Parameters***	***Estimated Value***	p_{value}
2018	α_1	−11.2846	0.004247
2019	α_2	−11.2261	0.004769
2020	α_3	−11.4896	0.003590
2021	α_4	−11.6194	0.003132
2022	α_5	−11.4994	0.003736
Per Capita income (X_1)	β_1	0.02397683	0.004120
Dependency Ratio (X_2)	β_2	−0.07510064	0.004616
Literacy Rate (X_3)	β_3	0.15504443	7.587×10^{-5}
Number of Schools (X_4)	β_4	0.00045527	0.323063
Student to teacher ratio (X_5)	β_5	0.09359832	0.013797
School Participation Rate (X_6)	β_6	0.02007894	0.024506

R^2	75.36%
p_{value}	3.1347×10^{-16}

According to Table 4, the R^2 value for fixed time effect model is 75.36%.

3) Random Effect Model (REM)

The estimation results are as follows:

Table 5. Estimation of the random effect model parameters

Variables	*Parameters*	*Estimated Value*	p_{value}
Intercept	α	2.54066876	0.119944
Per Capita income (X_1)	β_1	0.02923342	1.99×10^{-7}
Dependency Ratio (X_2)	β_2	-0.02318035	2.39×10^{-7}
Literacy Rate (X_3)	β_3	0.04202198	0.002513
Number of Schools (X_4)	β_4	0.00212576	0.011662
Student to teacher ratio (X_5)	β_5	-0.00232325	0.892536
School Participation Rate (X_6)	β_6	-0.00101723	0.714615
	R^2		70.05%
	p_{value}		2.22×10^{-16}

According to Table 5, the R^2 value for REM is 70.05%.

C. PDR Model Selection

The selection of the PDR model is conducted using the Chow test and the Hausman test. Two FE models are formed, thus the selection between the two FE models is carried out first, considering a lower AIC value and a higher R^2 value.

Table 6. AIC and R^2 Values for Each FE Model

Model	*AIC*	R^2
Fixed individual effect	−136,8665	75,95%
Fixed time effect	102,431	75,36%

Based on Table 6, it can be concluded that the fixed individual effect model is better and will be used for further analysis.

1) Chow Test

This test was conducted to determine whether there are differences in intercepts between individual units. Based on the analysis results, the p_{value} of chow test was obtained to be $< 2.2 \times 10^{-16}$, which means less than $\alpha = 0.05$. Therefore, rejecting H_0 means that there are differences in intercepts between individual units, and the analysis continued with the Hausman test.

2) Hausman Test

This test was conducted to determine whether there is a correlation between errors and independent variables. Based on the analysis results, the p_{value} of chow test was obtained to be $< 2.2 \times 10^{-16}$, which means less than $\alpha = 0.05$. Therefore, rejecting H_0 means that errors are correlated with independent variables, with the fixed individual effect model selected as the preferred PDR model.

D. Testing the Assumptions of Selected PDR Models

The testing of PDR model assumptions consists of:

1) Multicollinearity Test

This test is conducted to determine whether there is correlation between the independent variables in the model. The result of the VIF value for each independent variable are as follows:

Table 7. The VIF Values for Each Independent Variable

Independent Variable	X_1	X_2	X_3	X_4	X_5	X_6
VIF	1.649059	1.561484	1.910444	1.849764	1.847427	1.517676

Based on Table 7, the VIF value for each independent variable are less than 5. Therefore, it can be concluded that there is no multicollinearity.

2) Autocorrelation Test

This test is conducted to determine whether there is correlation between residuals in the model. Based on the test results using the Wooldridge test, a p_{value} of 1.9×10^{-6} was obtained, which means less than $\alpha = 0.05$. Therefore, it can be concluded that rejecting H_0, which means there is autocorrelation.

3) Heteroscedasticity Test

This test is conducted to determine whether there is a difference in variance between residuals of one observation with another observation. Based on the test results using the Breusch-Pagan test, a p_{value} of 0.0003789 was obtained, means less than $\alpha = 0.05$. Therefore, it can be concluded that rejecting H_0, which means there is a difference in the variance of residuals indicating the presence of heteroscedasticity.

Based on the results of the assumption testing on the selected PDR model, the occurrence of heteroscedasticity indicates the inequality of residual variety between each observation unit. With observation units being district/city, this shows differences in characteristics between regions or the effect of spatial heterogeneity. Therefore, it is important to consider the geographical coordinates of each region. Thus, the analysis will continue with the GWPR approach. Before continuing with the GWPR analysis, data transformation with within transformation will be conducted to accommodate the fixed effects in the selected PDR model.

E. GWPR Modelling

The first step in GWPR modelling is to calculate the Euclidean distance. This involves converting the latitude and longitude coordinates from degrees minutes seconds (DMS) to kilometers (km). After the Euclidean distance has been calculated for each district/city, the next step is to determine the optimum bandwidth using the adaptive kernel weighting function. Therefore, different bandwidth values are obtained for each district/city.

The results of the Euclidean distance calculation and bandwidth are then used to calculate the weighting matrix with adaptive kernel functions, which include adaptive gaussian, adaptive bisquare, and adaptive tricube for each district/city. The result of the weighting matrix is a diagonal matrix with weighting values that have been obtained, and will repeat every year. With a five-year data period, the weighting repeats itself and forms a 70×70 matrix. The next step is to estimate the parameters of the GWPR model using the WLS approach, so that different parameters are generated at each location. After that, the best GWPR model is determined based on the lower AIC value and the higher R^2 of each weighting function. The following are the results of AIC and R^2 for each model:

Table 8. AIC and R^2 Values for Each GWPR Model

Model	*AIC*	R^2
Adaptive Gaussian	$-142{,}9548$	76,78%
Adaptive Bisquare	$-147{,}241$	80,07%
Adaptive Tricube	$-146{,}7966$	80,03%

Based on Table 8, it is obtained that the GWPR adaptive bisquare model is the best model with lower AIC value and higher R^2 value. Next, the goodness-of-fit test for the model was conducted and continued with the significance test of the model parameters.

Simultaneous goodness-of-fit testing is conducted to determine whether the GWPR model is better than the PDR model, and it obtained a calculated F value (4.5243) is greater than the critical F value (1.5678). Therefore, it is concluded that the better model is the GWPR model and the parameters are significantly influenced by location. This shows that there is a difference between the GWPR model and the PDR model. Partial significance testing of model parameters is conducted to determine the independent variable that significantly affects the dependent variable. From the analysis results, it was found that the parameters have varying p_{values}, and the parameters with a $p_{value} < \alpha$ indicate that the independent variable has a significant effect on the dependent variable.

F. GWPR Model Interpretation

From the results of the significance testing of the GWPR model parameters, various models were obtained for MYS for each district/city in West Kalimantan Province. If regional group is formed based on significant independent variables that influenced the dependent variable, then four regional groups are formed as follows.

Table 9. Regional Groups Based on Significant Independent Variables

Significant Independent Variables	*Districts/Cities*
X_1, X_2, X_3, X_4	Sambas, Bengkayang, Landak, Mempawah, Sanggau, Sekadau, Kayong Utara, Kubu Raya, Kota Pontianak, Kota Singkawang
X_1, X_2, X_3	Ketapang
X_1, X_2, X_4	Sintang
X_1, X_2	Kapuas Hulu, Melawi

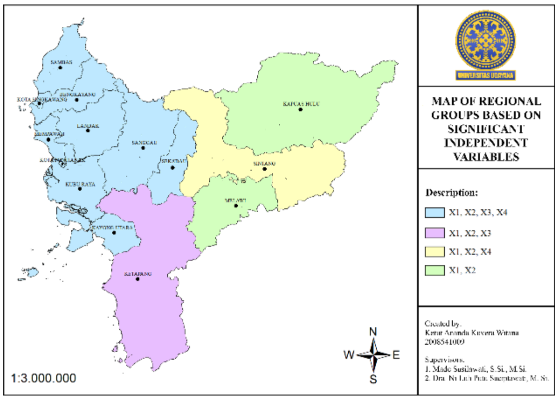

Figure 3. Map of Regional Groups Based on Significant Independent Variables

Based on Table 9 and Figure 3, it can be seen that the independent variables of per capita income (X_1), dependency ratio (X_2), literacy rate (X_3), and number of schools (X_4) significantly influence MYS (Y) in almost all district/city in West Kalimantan Province, across the formed groups. The following are the models formed for each regional group. As an example of the first regional group, the following is the model formed for Kayong Utara District.

$$\hat{Y}_{11t} = (1{,}17 \times 10^{-16}) + 0{,}0215X_{11t1} - 0{,}0211X_{11t2} + 0{,}0451X_{11t3} + 0{,}0082X_{11t4}$$

The model formed for Ketapang District as an example of the second regional group,

$$\hat{Y}_6 t = (6{,}18 \times 10^{-16}) + 0{,}0195\mathrm{X}_{6t1} - 0{,}0212\mathrm{X}_{6t2} + 0{,}0337\mathrm{X}_{6t3}$$

The model formed for Sintang District as an example of the third regional group,

$$\hat{Y}_7 t = (-2{,}17 \times 10^{-16}) + 0{,}0256\mathrm{X}_{7t1} - 0{,}0279\mathrm{X}_{7t2} + 0{,}0093\mathrm{X}_{7t4}$$

And The model formed for Kapuas Hulu Districts as an example of the fourth regional group

$$\hat{Y}_8 t = (-6{,}51 \times 10^{-16}) + 0{,}0219\mathrm{X}_{8t1} - 0{,}0281\mathrm{X}_{8t2}$$

Based on the model formed, it can be seen that X_1 variable has a positive relationship to Y, which means an increase in X_1 will increase Y, in line with research by Wardhana (2020) [17]. X_3 variable has a positive relationship with Y, which means that an increase in X_3 will increase Y, in line with the publication of Education Statistics by BPS (2022) [6]. X_4 variable has a positive relationship to Y, which means an increase in X_4 will increase Y, while X_2 variable has a negative relationship to Y, which means a decrease in X_2 will increase Y, in line with research by Jubaidah & Sukmawaty (2023) [18]. Therefore, in the effort to achieve educational equality and increase the mean years of schooling to fulfill the Nine-Year Compulsory Education Policy, it is necessary to consider per capita income, dependency ratio, literacy rate, and the number of schools.

IV. CONSLUSION

Based on the analysis results, it can be concluded that the best model for modelling MYS districts/cities in West Kalimantan Province from 2018 to 2022 is the GWPR adaptive bisquare model, due to the presence of heteroskedasticity in the PDR model indicating spatial effects in the data. The independent variables that significantly influence MYS are per capita income, dependency ratio, literacy rate, and the number of schools, which are significant in almost all districts/cities, across the formed regional groups. Based on significant independent variables, four regional groups are formed.

From on the findings of the study, some recommendations that can be given are developing the GWPR model with the selected PDR estimation method is random effect model, and the government may form education equalization policies and programs that consider aspects of population income, dependence rate, literacy rate, and the number of educational facilities, especially the number of schools.

V. ACKNOLEDGE

Highest appreciation to the Department of Mathematics, Udayana University, for the invaluable support in completing this research.

REFERENCES

1) The World Bank. (2023). Education Overview: Development News, Research, Data. Retrieved September 22, 2023, from Education website: https://www.worldbank.org/en/topic/education/overview
2) BKN. (2020). Undang-Undang Dasar Negara Republik Indonesia Tahun 1945 dalam Satu Naskah. Retrieved February 6, 2024, from https://www.bkn.go.id/wp-content/uploads/2014/06/UUD-1945.pdf
3) Kemendikbud. (2019). *Percepatan Peningkatan Akses dan Pemerataan Mutu Pendidikan Melalui Kebijakan Zonasi Pendidikan*. Jakarta.

4) UNDP. (2022). Human Development Index. Retrieved September 22, 2023, from https://hdr.undp.org/data-center/human-development-index#/indicies/HDI
5) UNESCO. (2017). Mean years of schooling. Retrieved September 22, 2023, from Glossary website: https://glossary.uis.unesco.org/glossary/en/term/2694/en
6) BPS. (2022). *Statistik Pendidikan 2022* (1st ed.; R. Sinang & I. Maylasari, Eds.). Jakarta: BPS RI.
7) BPS. (2023). *Provinsi Kalimantan Barat dalam Angka 2023* (1st ed.; H. Sucihati, M. Yani, & R. Pertiwi, Eds.). Pontianak: BPS Provinsi Kalimantan Barat.
8) Gujarati, D. N. (2003). *Basic Econometrics* (4th ed.). West Point: McGraw Hill.
9) Yu, D. (2010). Exploring Spatiotemporally Varying Regressed Relationships: The Geographically Weighted Panel Regression Analysis. *The International Archives of the Photogrammetry*, *38*(2), 135–136.
10) Baltagi, B. H. (2005). *Econometric Analysis of Panel Data* (3rd ed.; John Wiley & Sons, Ed.). Houston: John Wiley & Sons Ltd.
11) Fotheringham, A. S. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationship*. Newcastle: John Wiley & Sons Ltd.
12) Anselin, L. (1992). Spatial Data Analysis With GIS : An Introduction To Application In The Social Sciences. In *Geographic Information Systems: A Handbook for the Social Sciences*. California.
13) Greene, W. H. (2018). *Econometric Analysis (Eighth Edition)* (8th ed.; C. Paganelli, Ed.). New York: Pearson Education.
14) O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6
15) Croissant, Y., & Millo, G. (2008). Panel Data Econometrics in R: The plm Package. *Journal of Statistical Software* , *27*(2), 27. https://doi.org/10.18637/jss.v027.i02
16) Caraka, R. E., & Yasin, H. (2017). *Geographically Weighted Regression (GWR) Sebuah Pendekatan Regresi Geografis* (1st ed., Vol. 1; R. E. Caraka & H. Yasin, Eds.). Yogyakarta: Mobius.
17) Wardhana, A., Kharisma, B., & Lingga, A. (2020). Pencapaian Pendidikan dan Kualitas Sumber Daya Manusia Antar Kabupaten dan Kota di Jawa Barat. *Forum Ekonomi*, *22*(2), 5–6. https://doi.org/10.29264/jfor.v22i2.7224
   Jubaidah, S., & Sukmawaty, Y. (2023). Pemodelan Rata-rata Lama Sekolah Kabupaten/Kota di Pulau Kalimantan Dengan Pendekatan Geographically Weighted Regression (GWR). *Repo Mhs ULM*, 12–13