Modeling of School Participation Rate for Senior High School in Indonesia Using Mixed Geographically Weighted Regression Model

Febrina Kusumaningati[1], Jefri E.B.Utomo[1], Maya M.M. Zach[1], Ayu Puspitawati[1] and Nur Chamidah[2]

[1]*Student of Study Program of Statistics, Department of Mathematics,Universitas Airlangga, Surabaya, Indonesia*
[2]*Department of Mathematics, Faculty of Sciences and Technology, Universitas Airlangga, Surabaya, Indonesia*

Keywords: School Participation Rate, Senior High School, Mixed Geographically Weighted Regression.

Abstract: The quality of education can be determined by looking at school participation rate (SPR). The SPR in Indonesia has increased in 2013-2016, but SPR for the age of 16-18 years has the lowest value when compared with the age of 7-12 years and 13-15 years. This study aims to model and analyze the results of SPR modeling and interpret the factors that affect the SPR in every province in Indonesia by using Mixed Geographically Weighted Regression (MGWR) model presented in thematic map. Based on the analysis result, there were 5 provinces where the SPR affected by rate of economic growth; 13 provinces were affected by the number of school and rate of economic growth; 13 provinces were affected by the number of school, rate of economic growth, and Gini Ratio; and 3 provinces were affected by rate of economic growth and rate of labor participation. Also, we obtained through the MGWR model of Papua province that rate of economic growth as a global variable decreased the SPR due to lack of good equity while rate of labor participation reduced the SPR because many students prefer to work than study.

1 INTRODUCTION

The quality of education in Indonesia can be examined by looking at school participation rate according to BPS (2013). School participation rate is the sum of all students who are still in school compared to the whole population within the same age group. From 2013 to 2016, school participation rate in the 7-12 year population increases by 0.67%, 13-15 years increases by 4.07%, and 16-18 years increases 6.99%. School participation rate for 16-18 years (70.83%) has the lowest score compared to 7-12 years (99.09%) and 13-15 years (94.88%).

Rahmatin and Soejoto (2017) studied the influence of poverty rate and number of schools on school participation rate (SPR) in Surabaya city by using multiple linear regression method. The obtained results were the poverty level factor and the number of schools significant to the SPR in Surabaya city.

The vulnerability of dengue hemorrhagic fever disease in Surabaya was studied by Chamidah, et al (2014) by using spatial logistic regression approach, the result showed that every district in Surabaya has different characters or relatively heterogeneous.

This also applies in education. Education in each region is different, for example it can be seen from the luxurious infrastructure that only exist in big cities. Risking those assets, the automatic significant factor for each region is relatively heterogeneous. To discuss the heterogeneity, the researchers used the special method specially mixed geographically weighted regression (MGWR) approach to find out the best SPR model in Indonesia and their significant factors. MGWR is combination between linear regression and Geographically Weighted Regression, so MGWR can be used to identify significant factors on SPR locally and globally.

Kusumaningati, F., Utomo, J., Zach, M., Puspitawati, A. and Chamidah, N.
Modeling of School Participation Rate for Senior High School in Indonesia Using Mixed Geographically Weighted Regression Model.
DOI: 10.5220/0007554409390943
In *Proceedings of the 2nd International Conference Postgraduate School (ICPS 2018)*, pages 939-943
ISBN: 978-989-758-348-3

2 LITERATUR REVIEW

2.1 School Participation Rate

School participation rate is the proportion of all children who are still in school in one particular age group against the population with the appropriate age group (BPS,2017).

2.2 Factors that Influence School Participation Rate

There are some factors which suspected to influence school participation rate:

2.2.1 Number of Schools

Number of schools is a number of SMA/SMK in each province in Indonesia.

2.2.2 The Rate of Economic Growth (PDRB)

The Rate of Economic Growth on BPS calculation is based on production of products and services growth in certain time. The calculated of PDRB :

$$\mathbf{PDRB} = \frac{\mathbf{PDRB_t - PDRB_{t-1}}}{\mathbf{PDRB_{t-1}}} \mathbf{x\ 100\%} \quad (1)$$

2.2.3 Gini Ratio

Gini ratio used to measure the level of overall income inequality. The calculated of gini ratio:

$$\mathbf{GR = 1 - \sum_{i=1}^{n} f_{pi} x (Fc_i + Fc_{i-1})} \quad (2)$$

2.2.4 Percentage of Poor People

Poor people are people who have an average per capita expenditure below the poverty line. Percentage of poor people in BPS calculated :

$$\boldsymbol{P_0 = \frac{1}{n} \sum_{i=1}^{q} \left[\frac{z - y_i}{z}\right]^0} \quad (3)$$

2.2.5 Rate of Labor Force Participation

The labor force participation rate (TPAK) is the percentage of the population aged 15 years and over which is labor forced. The calculation rate of force participation is:

$$\boldsymbol{TPAK = \frac{labor\ participation}{number\ of\ population\ over\ 15\ years}\ x\ 100\ \%} \quad (4)$$

2.3 Geographically Weighted Regression Method

Geographically Weighted Regression (GWR) method is a technique used in spatial regression model that takes the framework of a simple regression model by weighted regression (Fotheringham, *et al.,*). GWR model is expressed as follows:

$$y_i = \beta_0 (u_i, v_i) + \sum_{k=1}^{p} \beta_k (u_i, v_i) x_{ik} + \varepsilon_i \ldots\ldots\ldots (5)$$

2.4 Mixed Geographically Weighted Regression Method

Mixed Geographically Weighted Regression (MGWR) model is a combination model of global regression with GWR that considers a situation where several predictor variables that influence global responses and other predictor variables are local. The MGWR model can be written as follows:

$$y_i = \beta_0 (u_i, v_i) + \sum_{k=1}^{q} \beta_k (u_i, v_i) x_{ik} + \sum_{k=q+1}^{p} \beta_k x_{ik} + \varepsilon_i \ldots\ldots (6)$$

The estimation of this model parameter can use the Weighted Least Square (WLS) approach. The estimated parameters of the MGWR model are as follows:

$$y_i = \mathbf{x_i \beta_i} (u_i, v_i) + \mathbf{x_g \beta_g} + \boldsymbol{\varepsilon} \quad (7)$$

3 MATERIAL AND METHODS

3.1 Data and Data Sources

The data used in this study was secondary data of 2017 obtained from the publication of the Central Bureau of Statistics (BPS). The data obtained consist of a percentage of upper secondary level SPR or age group 16-18 years and the factors that influence it. In this research, there were 35 provinces in Indonesia becoming the unit of observation.

3.2 Research Variables

Variables in this research were a response variable and five predictor variables. The response variable was school participation rate (SPR).

Table 1: Predictor Variables

Predictor Variables	Description	Data Type
X_1	Number of Schools	Discrete
X_2	Economic Growth Rate (PDRB)	Continuous
X_3	Gini Ratio	Continuous
X_4	Percentage of Poor People	Continuous
X_5	Percentage of Work Force on Working Age Population	Continuous

In addition, geographic variables were longitude (u_i) and latitude (v_i).

3.3 Step Analysis

Modelling and analyzing the upper secondary level SPR for each province in Indonesia using the MGWR method with the following steps:

3.3.1 Test Spatial Assumption

Conducting assumption test against response variable. Following steps to test assumptions:

a. Perform normality test against response variable. Data should be normal.
b. Conducting multicollinearity data test. If the value of VIF> 10 then there is a high multicollinearity indication in the data.
c. Conducting heteroscedasticity test with GeoDa software. If the value of p-value> α then there is no heteroscedasticity.
d. Testing the assumption of spatial dependencies in the data using Moran's I test using GeoDa software.

3.3.2 Analyze with GWR Method

Analyze the upper secondary level school participation rate model in each province in Indonesia using GWR method with the following steps:

a. Determine longitude and latitude for each province in Indonesia.
b. Calculate the euclidian distance (d_{ij}) between the i-th location of the j-th location at the coordinates.
c. Determine the optimum bandwidth by using Cross Validation (CV) method.
d. Calculates a weighted matrix by using kernel function for optimum bandwidth.
e. Estimate GWR model parameters by including all predictor variables.
f. Perform partial significance test parameters.

3.3.3 Analyze with MGWR Method

Analyze the upper secondary level school participation rate model in each province in Indonesia using the MGWR method with the following steps:

a. Obtain the model parameter estimator MGWR by using the optimum bandwidth and the same weighting as in the GWR model.
b. Testing the suitability of the MGWR model.
c. Conducting simultaneous testing on global predictor variable parameters in the MGWR model.
d. Perform partial testing on the parameters of local predictor variables on the MGWR model.
e. Compare the GWR model with the MGWR model with AICc criteria

4 RESULTS AND DISCUSSION

4.1 Description Research Variables

The following is a descriptive analysis of research variables used:

Table 2: Description research variables.

Variabel	N	Mean	Minimum	Maximum
X1	34	775.88	85	4225
X2	34	5.2408	0.02	7.67
X3	34	0.3608	0.28	0.44
X4	34	11.135	3.78	27.69
X5	34	68.752	62.63	77.12

4.2 Spatial Assumption

The spatial assumption test aims to determine whether the data to be analyzed has satisfied the

basic assumptions of spatial regression. From normality test, data is normally distributed.

Multicollinearity shows if VIF value<10. The following is VIF value from predictor variables:

Table 3: Variance influence factors value.

Predictor Variables	VIF Value
Number of Scholl (X_1)	1.20
PDRB (X_2)	1.08
Gini Ratio (X_3)	1.27
Percentage of Poor People (X_4)	1.20
Percentage of Work Force on Working Age Population (X_5)	1.15

To test the effect of spatial heterogeneity, it was used Breusch-Pagan test. Based on the test result, it was obtained *p*-value, i.e., 0.01994 less than α=0.05. It means that there was heterogeneity spatial in data. Furthermore, Moran's I test was implemented to find out the effect of spatial dependency and get p-value=0.000. Finally, it was concluded that there was spatial dependency in data.

4.3 Estimation Spatial Regression

This research has fulfilled the basic assumptions of spatial regression thereby the next was step to estimate the parameters of the spatial regression model and to determine the weighted. The weighted used were the kernel function weights while the kernel function used was Fixed-Gaussian. To determine the best model, look at the comparison of AICc values between GWR and MGWR methods. The summary of AICc values is represented in Table 4:

Table 4: Comparison of AICc value between GWR and MGWR methods.

Methods	AIC'c Values
Fixed Gaussian GWR method	242,280491
Fixed Gaussian MGWR method	227,937452

Based on Table 4, the best method is the method that has the smallest AICc value that is Fixed Gaussian method MGWR method with AICc value 227.9374. Therefore, the method used to estimate the best model in this research was Fixed Gaussian MGWR method.

4.4 Partial Test of Local and Global Parameters

After the best model was obtained, the next step was to test the significance of global parameters with GWR 4.0 software. Based on the calculation, the variables that affect global was X_2 with standard residual is 0.667, t-value is -2.078, and the estimation is -.1371.

The other variables are local and the values depend on each region. The summary of significant variables can be seen in Table 5 as follows:

Table 5: Significant variables in each region

Province	Significant Variables
DI Yogyakarta	X1, X2, X3
Papua	X2, X5

MGWR model for province of DIY is as follows:

$$y = -0.002392\ X_1 - 1.371757\ X_2 + 75.174818\ X_3$$

Based on the model above, it can be said that every increase of number of school (X_1) will decrease school participation rate as big as 0.002392, every increase of economic growth rate (X_2) will decrease school participant rate as big as 1.371757, and every increase of gini ratio (X_3) will increase school participation rate as big as 75.174818.

MGWR model for province of Papua is as follow:

$$y = 166.903313 - 1.371757\ X_2 - 1.062467\ X_5$$

Based on the model above, can be said that every increase of economic growth rate (X2) will decrease school participation rate as big as 1.371757 and every increase of percentage of work force on working age population (X5) will decrease school participation rate as big as 1.062467.

Predictor variables estimation that influences both provinces is different. Their economic growth rate is not increasing school participation rate. This is because high economic growth rate is not always causing all of the people to be prosperous, because economic growth rate is often not followed by good equity. Economic growth rate is significant in 34 provinces in Indonesia, but the other predictor variables just influence locally. Hence, modeling of school participation rate for senior high school is

different too. Significant variables for every province in Indonesia, is showed in Table 6.

Table 6: Significant variables in every province in Indonesia.

Province	Significant Variables
Aceh, North Sumatera , North Sulawesi, Gorontalo, North Maluku	X2
West Sumatera, Riau, Jambi, Bengkulu, Kepulauan Riau, NTT, East Kalimantan, North Kalimantan, Central Sulawesi, South Sulawesi, Southeast Sulawesi, West Sulawesi	X1, X2
South Sumatera, Lampung, Kepulauan Bangka Belitung, DKI Jakarta, West Java, Central Java, DI Yogyakarta, East Java, Banten, Bali, NTB, West Kalimantan, Central Kalimantan, South Kalimantan	X1, X2, X3
Maluku, West Papua, Papua	X2, X5

The thematic map of the significant variables of every province in Indonesia is shown in Figure 1.

Figure 1: The thematic map of the significant variablesof every province in Indonesia.

4.5 Conformity Test of MGWR Model

After MGWR model obtained by result of parameter estimation, the next step was conformity testing for MGWR model. Result of conformity test of MGWR model is summarized in table 7.

Table 7: Result of conformity test of MGWR model.

Source	DF	SS	MS	F
Global	28	943.5		
Residual		77		
MGWR Improvement	6.853	418.120	61.014	
MGWR Residual	21.14	525.457	24.848	2.4555

Table 7 shows that F value, i.e., 2.4555 > F table, i.e., 2.36. Therefore, the estimated MGWR model is appropriate.

5 CONCLUSION

Based on the analysis and result which has been done, it is concluded as follows:

Analysis result of school participation rate in Indonesia by using mixed geographically weighted regression (MGWR) model found that global predictor variable was economy growth rate which have a significant effect. Then, local predictor variables are number of schools, gini ratio, percentage of poor people, and percentage of work force on working age population.

By using mixed geographically weighted regression (MGWR) estimation, it was obtained different model for every province because significant predictor variables for every province in Indonesia are the same.

REFERENCES

Badan Pusat Statistik., 2017. https://www.bps.go.id/dynamictable/2015/12/22/1054/angka-partisipasi-sekolah-aps-menurut-provinsi-2011-2017.html. Accessed on 4th April 2018.

Chamidah, N., Syaifudin, T., and Rifada, M., 2014. The Vulnerability of Dengue Hemorrhagic Fever Disease in Surabaya Based on Spatial Logistic Regression Approach. *Journal of Applied Mathematical Sciences*, Vol.8, No.28, pp.1369-1379.

Fotheingham A.S., Brunsdon C., Charlton M., 2002. *Geographically Weighted Regression the Analysis of Spatial Varying Relationships*. United Kingdom: John Wiley & Sons.

Mei, C.L. , He, S.Y. and Fang. K.T., 2004. A Note on the Mixed Geographically Weighted Regression Model. *Journal of Regional Science*, Vol.44, pp. 143-157.

Rahmatin, U.Z, dan Soejoto, A., 2017. Pengaruh Tingkat Kemiskinan dan Jumlah Sekolah Terhadap Angka Partisipasi Sekolah (APS) di Kota Surabaya. *Jurnal Pendidikan Ekonomi Manajemen dan Keuangan*, Vol.1,No. 2, pp. 127-140.