CRIME RATE IN CENTRAL JAVA, INDONESIA 2017: A SPATIAL ANALYSIS USING MIXED GEOGRAPHICALLY WEIGHTED REGRESSION

by Tarno Tarno

Submission date: 10-Aug-2022 04:43PM (UTC+0700)
Submission ID: 1880967986
File name: N_CENTRAL_JAVA,_INDONESIA2017_A_SPATIAL_ANALYSIS_USING_MIXED.pdf (180.04K)
Word count: 1597
Character count: 7801

Far East Journal of Mathematical Sciences (FJMS)

http://www.pphmj.com
http://dx.doi.org/10.17654/MS121010031
Volume 121, Number 1, 2019, Pages 31-37 **ISSN: 0972-0871**

CRIME RATE IN CENTRAL JAVA, INDONESIA 2017: A SPATIAL ANALYSIS USING MIXED GEOGRAPHICALLY WEIGHTED REGRESSION

Nur Azmi Prasetyo and Tarno*

Department of Statistics
Universitas Diponegoro
Semarang 50275, Indonesia
e-mail: tarno.stat@gmail.com

Abstract

Received: September 25, 2019; Accepted: October 28, 2019
2010 Mathematics Subject Classification: 91D25.
Keywords and phrases: crime rate, mixed geographically weighted regression, AICc.
*Corresponding author

1. Introduction

Crimes or violations are acts of someone who can be punished by the KUHP or other laws and regulations that apply in Indonesia (Criminal Statistics, 2016). According to the data on the number of crimes reported in 2016, it was noted that the province of Central Java was ranked eighth out of 34 provinces in Indonesia with 14,353 cases where it was quite high.

The government has endeavored in implementing various policies or programs to deal with this problem, but it has not been optimal. There needs to be the right strategy to solve the problem. The government must know the exact things those are related to the high crime rate in Central Java.

The method that can be used to conduct this strategy is regression analysis. Because the problem of crime is very likely to be influenced by location and neighbor, so the data between observations is difficult to assume to be free from each other. So, an analysis will be done that accommodates spatial problems, namely geographically weighted regression (GWR) and mixed geographically weighted regression (MGWR). It is a regression that geographically weighted. Observations in further locations are weighted with smaller weightings, according to Tobler's first law of geography which states that "*Everything is related to everything else, but near things are more related than distant things*".

MGWR is a combination of global linear regression model with the GWR model. So that the model that will be generated by the MGWR estimator parameters is global in nature and several others are localized according to the observation location. Estimation of parameters in MGWR can be done with WLS method as in the GWR model. Nowadays, we can see

many implementations of this method, and we need to apply this method to model the crime rate in Central Java.

2. Literature Review

MWGR is an advanced form of GWR. The GWR stated here is as explained by Brundson et al. in [2]. It is added by the local variable existence, anyway the formula now becomes:

$$\hat{y} = \sum_{g=1}^{q} a_g x_{ig}(a) + \sum_{l=q+1}^{p} b_l(u_i, v_i) x_{i,l}(b) + \varepsilon_i, \quad i = 1, 2, ..., n, \qquad (1)$$

where a_g is a global parameter, b_l is a local parameter, x_{ig} represents variables of global parameter, while $x_{i,l}$ for local parameter, and ε_i is the residual of MGWR.

The assumption should be met in MGWR is heterogeneity. The test of it is conducted using Breusch-Pagan (BP), where the H_0 states that there is *no heterogeneity spatial* in the data. It is rejected when the value of *BP*-test ≤ *DK*, which is defined as $\{BP \mid BP > \chi^2_{(k-1)}\}$ or when the *p*-value is smaller than α, obviously.

3. Methodology

The data used in this research is the secondary data obtained from the Central Bureau of Statistics. The units of observation are the regencies in Central Java, which consists of 35 regencies. To support the research process, we used *R* studio application.

The variables in this study are the percentage of criminality per regency in Central Java in 2017 (*Y*) as a dependent variable and opened unemployment (*X*1), the number of poor people (*X*2), level of education (*X*3), human development index (*X*4), net enrollment rate (Elementary School) (*X*5), net enrollment rate (Junior High School) (*X*6), net enrollment rate (Senior High School) (*X*7), school enrollment rate (7-12 years old) (*X*8),

school enrollment rate (13-15 years old) ($X9$), school enrollment rate (16-18 years old) ($X10$), the number of workforce employed ($X11$), total workforce ($X12$), the percentage of population ($X13$), population density ($X14$) and population growth rate ($X15$) as the independent variables. Also, latitude (u_i) and longitude (v_i) of the location. To determine the factors affecting the percentage of crime in Central Java 2017, we use the following steps:

1. Find the significant variables and model it using stepwise model forward selection.
2. Test the data for heterogeneity spatial possibility.
3. Find the optimum bandwidth.
4. Compare the AICc and R^2 between global regression and GWR.
5. Model the data using MGWR.
6. Estimate local and global parameters of MGWR.
7. Compare AICc of global regression, GWR and MGWR.

4. Result and Discussion

Before we analyze the data using MGWR, we need to analyze it using global regression and GWR. It is to convince and synchronize with the related theory.

Let us start with the global regression, which is the stepwise model forward selection. This model is used to know the fit regression model with significant variables. It is obtained the model as:

$$\hat{y} = -18.67341253 + 0.09759767X4 + 0.00001147X12 + 0.00069410X14 - 0.01664916X2 + 0.11837098X6. \quad (2)$$

First, we should do the F-test and it is obtained the F-statistics with the value of 17.07 and the p-value of 0.00000007162. It can be concluded that with $\alpha = 5\%$, the model is significant simultaneously.

After that, we need to calculate the heterogeneity test. Using Breusch-Pagan test, it is obtained the *p*-value of 0.01057. It indicates the spatial heterogeneity phenomenon, and it also suggests us to continue this analysis using GWR.

Before we begin the analysis using GWR, we need to know the coordinates (latitude and longitude) of the locations. Then the first to do is to obtain the optimum bandwidth. In this research, we obtain the value of 1.027011.

Table 1. Comparison between GWR and global regression model

Model	R^2	AICc
GWR	0.8037667	141.639
Global	0.7464	142.5304

Table 1 shows the comparison between GWR and global regression model. In the table, it is so clear that the GWR model is better than global regression. It is obvious because the R^2 of GWR is bigger than global regression, and the AICc is smaller.

Then we move to the main topic of this research, which is the use of MGWR. The reason why we need to use this method is because there is a possibility that the best model is by combining the GWR model and the global regression instead. For the first step in this transition of GWR into MGWR, we need to know whether the variable is using global coefficient or the local coefficient. It is tested using the significance of coefficient test, the result is shown in Table 2.

Table 2. Comparison between GWR and global regression

Coefficient	*p*-value
Intercept	0.79
*X*4	0.55
*X*12	0.48
*X*14	0.01
*X*2	0.75
*X*6	0.45

Table 2 shows that there is a variable which is affected by location. It is population density ($X14$), it tells us that the crime rate is affected by the population density of each location with significance level of 5%. We cannot assume this variable statistically is the same for every location, while the other variables are not affected by location.

Table 3. Local variable

Regency	$X14$	Wonogiri	0.00080	Kendal	0.00073
Cilacap	0.00060	Karanganyar	0.00079	Batang	0.00070
Banyumas	0.00064	Sragen	0.00079	Pekalongan City	0.00068
Purbalinga	0.00066	Grbogan	0.00078	Pemalang	0.00065
Banjarnegara	0.00068	Blora	0.00081	Tegal City	0.00062
Kebumen	0.00069	Rembang	0.00081	Brebes	0.00060
Purworejo	0.00072	Pati	0.00079	Magelang Regency	0.00074
Wonosobo	0.00071	Kudus	0.00078	Surakarta	0.00078
Magelang	0.00074	Jepara	0.00077	Salatiga	0.00076
Boyolali	0.00076	Demak	0.00077	Semarang Regency	0.00075
Klaten	0.00077	Semarang City	0.00076	Pekalongan Regency	0.00068
Sukoharjo	0.00078	Temanggung	0.00073	Tegal Regency	0.00061

Then we need to know the value of each coefficient of parameters. It is presented in Table 3 and Table 4.

Table 4. Global variables

Intercept	$X4$	$X12$	$X2$	$X6$
−14.2313	0.071482	0.000011	−0.0168846	0.08412758434

Table 5. Comparison of AICc

Model	AICc
Global regression	142.5304
GWR	141.693
MGWR	141.3

From the models, the main thing is how to convince that the MGWR model is better than the rest of models. In this research, we use the AICc as the criteria. The smaller AICc indicates the better model, anyway the comparison is presented in Table 5. It is clear that MGWR is the best model among the rest, with the AICc value of 141.3.

5. Conclusion

MGWR is the best way to model the percentage of criminality in Central Java 2017. This is because it has the smallest AICc value among the represented models. Anyway the value of AICc of MGWR is 141.3. From Section 3, we can conclude that population density was statistically affected by location. And that we cannot generalize the variable as a global variable. This result of research is expected to be able to help the government to reduce the crime rate of Central Java efficiently.

References

CRIME RATE IN CENTRAL JAVA, INDONESIA 2017: A SPATIAL ANALYSIS USING MIXED GEOGRAPHICALLY WEIGHTED REGRESSION

ORIGINALITY REPORT

12%	9%	6%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	mafiadoc.com Internet Source	5%
2	web.unitn.it Internet Source	1%
3	Ditha Runita, Rohmatul Fajriyah. "Measuring the Indonesian provinces competitiveness by using PCA technique", Journal of Physics: Conference Series, 2017 Publication	1%
4	journal.ugm.ac.id Internet Source	1%
5	www.slideshare.net Internet Source	1%
6	C E Mongi, Y A R Langi, C E J C Montolalu, N Nainggolan. "Comparison of hierarchical clustering methods (case study: data on poverty influence in North Sulawesi)", IOP	1%

Conference Series: Materials Science and Engineering, 2019
Publication

7	E N Manurung, Y Widyaningsih, S Aminah. "Linear regression model as an approach to analyze spatio temporal data to know the influence factors of the number of crimes in North Sumatera", Journal of Physics: Conference Series, 2021 Publication	1 %
8	www.mdpi.com Internet Source	1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

CRIME RATE IN CENTRAL JAVA, INDONESIA 2017: A SPATIAL ANALYSIS USING MIXED GEOGRAPHICALLY WEIGHTED REGRESSION

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7