Geographically and Temporally Weighted Autoregressive to Modeling the Levels of Poverty Population in Java in 2012-2018

Hartinah Djalihu[a]*, Anik Djuraidah[b], Agus Mohamad Soleh[c]

[a,b,c]*Departement of Statistics, Faculty of Mathematics and Natural Science, IPB University, Bogor, Jawa Barat, 16680, Indonesia*
[a]*Email: hartinah_djalihu@apps.ipb.ac.id,* [b]*Email: anikdjuraidah@apps.ipb.ac.id*
[c]*Email: agusms@apps.ipb.ac.id*

Abstract

Geographically and temporally weighted regression (GTWR) is a method applied when there is spatial and temporal diversity in the observation. GTWR model just considers local influences of spatial-temporal response variable on the explanatory variables. The GTWR model can add an autoregressive component of response variable, the resulting model is known as a geographically and temporally weighted autoregressive model (GTWAR). This study aims to perform GTWAR modeling which is applied to the data on the proportion of poor people by districts/cities in Java in 2012-2018. The results showed that GTWAR produced Akaike Information Criterion (AIC) smaller than GTWR, and the coefficient of determination (R^2) is higher than GTWR.

Keywords: poverty; spatial autoregressive; GTWAR.

1. Introduction

Poverty is a global issue faced by many countries in the world, including Indonesia. To measure poverty [1] uses the concept of the ability to meet basic needs (basic needs approach) which is seen as an economic inability to meet basic food and non-food needs measured in terms of expenditure.

----

* Corresponding author.

In the publication of a catalog of poverty profiles in Indonesia in March 2019 [1], it was stated that in terms of the number of most of the poor people still living on Java Island, amounting to 12.72 million people, while the lowest number of poor people was in Kalimantan, namely 0.97 million people. The high number of poor people on the island of Java is a natural thing considering that more than half of Indonesia's population lives on the island of Java. Spatial diversity or also known as heterogeneity is a condition where the measurement of the relationship between variables varies from one location to another [2]. One of the ways to overcome spatial heterogeneity is by using Geographically Weighted Regression (GWR). The GWR model is a development of global regression modeling by adding weights based on the distance from one observation location to another, so that the interpretation for each point of location will be different [2]. However, the GWR method in its analysis only considers spatial without the component of time in its modeling. So the time element in the GWR can be carried out by using the Geographically and Temporal Weighted Regression Method (GTWR). In practice, there are many problems that are not modeled on the GTWR model, including autocorrelation between response variables between locations (spatial lag). Therefore, to deal with the problem of heterogeneity and autocorrelation in an observation involving location and time elements, it can be done by integrating the GTWR model and the spatial autoregressive model which produces a geographically and temporally weighted autoregressive model (GTWAR). Several previous studies related to poverty, including [3] modeled housing prices in Shenzhen, China using the geographically and temporally weighted autoregressive method (GTWAR) [4] used the GTWR approach to estimate house prices with consider the distance between locations based on the route of travel [5] apply the GTWR model to economic growth [6] comparison of GWTR and robust GTWR modelling [7] used robust MGTWR to modeling the percentage of poverty population in java in 2012-2018. Based on this description, this study will model the percentage of poor people in regency/municipality on Java island by developing GTWAR model, selecting the model the best and determine the factors that affect the percentage of poor people in regency/municipality on Java Island in 2012-2018.

2. Materials and Methods

2.1.   *Data*

The data used are in the form of secondary data from BPS-Statistics of Indonesia from 2012 to 2018. The response variable used is the percentage of poor people living in regency/municipality on Java island. The explanatory variables used are shown in Table 1 as follows:

Table 1: Explanatory Variable

	Description of variables	Reference
X1	Gross Regional Domestic Product (Billion Rupiah)	[8]
X2	People with the highest education graduated from elementary school (%)	[9]
X3	Literacy numbers	[10]
X4	Per capita expenditure for food (%)	[9]
X5	Raskin/rasta recipient households (%)	[10]
X6	Population according to age group 15-64 (%)	[9]
X7	Expected years of schooling (EYS) (Year)	[11]
X8	Mean years of schooling (MYS) (Year)	

2.2.   Data Analysis Procedure

The steps of the analysis are as follows:

1. Exploring data on the percentage of poor people in regency/municipality in Java in 2012-2018.
2. Calculating the value of Variance Inflation Factor (VIF) to determine the multicollinearity of the explanatory variables, with criteria if VIF> 10 then multicollinearity occurs between variables.
3. Examination spatial autocorrelation with *Moran's* Index [12].
4. Examination spatial diversity annually on data using the *Breusch Pagan* (BP) test [13].
5. Estimating parameters of GTWAR model:

a.   Determine (W^*y) which will be used as an additional variable in the model, where (W^*) is the spatial dependency weighting matrix (W^*) using exponential distance weights. This matrix is an alternative to the negative exponential function based on distance with the following equation:

$$W_{ij} = \exp(-\alpha \, d_{ij})$$

α is a positive constant.

b.   Determination the spatial-temporal weighted matrix (W):

   i. Calculate temporal-spatial ratio parameters (τ) using the Cross-Validation (*CV*) approach.
   ii. Calculate spatial parameters (λ) and temporal parameters (μ) using the Cross-Validation (*CV*) approach.
   iii. Calculate temporal-spatial distance (d_{ij}^{ST}) and the width of the temporal-spatial window (h_{ST}^2) using the Cross-Validation (*CV*) approach.

$$\left(d_{ij}^{ST}\right)^2 = \lambda\left[\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2\right] + \mu\left(t_i - t_J\right)^2$$

   iv. Determine the weighted matrix (W) of GTWAR model using a measure of spatial-temporal distance with interactions for each observation location based on the Kernel Bisquare function with the following formula:

$$w_j(u_i, v_i) = \begin{cases} \left[\left(1 - \left(\dfrac{d_{ij}^{ST}}{h_{ST}}\right)^2\right)\right]^2 , & jika \; d_{ij} \leq h \\ 0 & , \; jika \; d_{ij} > h \end{cases}$$

c.   Calculating variable estimator GTWAR model.

6. Partial testing of each GTWAR parameter on variables.
7. Making the comparison of the results of the GTWR and GTWAR models based on the *AIC* and pseudo R^2.

3. Result and Discussion

3.1.   Data Exploration

Java is the 13th largest island in the world with an administrative area of 138793.6 km2 with a population density of around 1317 people per km. Provinces in Java Island consist of Banten, the Special Capital Region of Jakarta, West Java, Central Java, East Java and the Special Region of Yogyakarta with a total of 118 regency/municipality. The map of distribution the percentage of poor people in Java for 2012-2018 is presented in Figure 1.
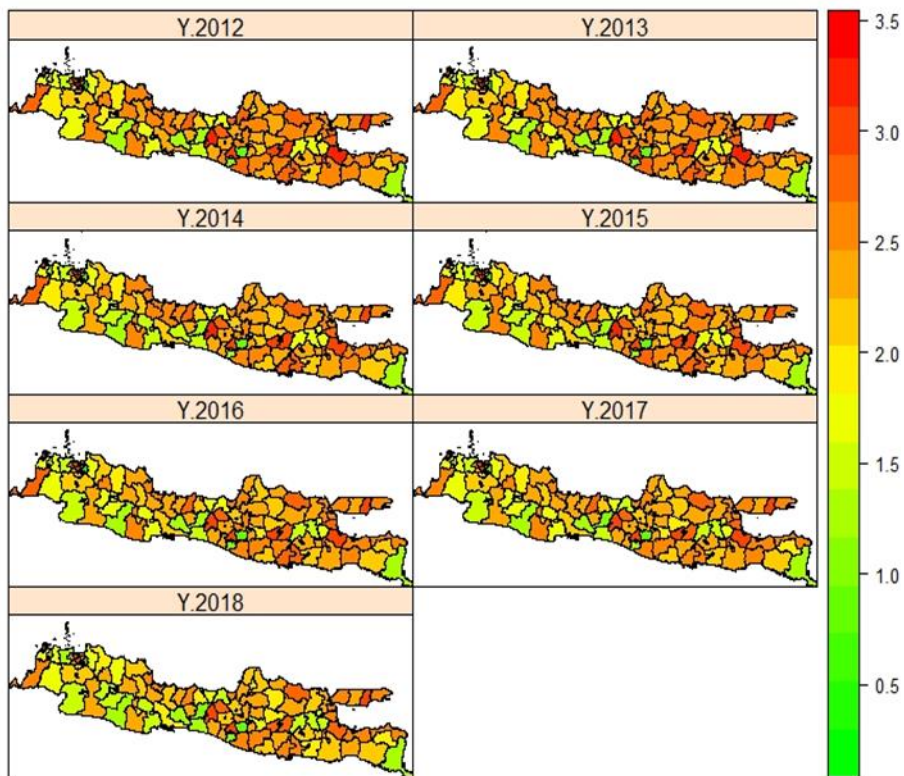

Figure 1: Map of the distribution of the percentage of poor people in Java 2012-2018

It can be seen that regencies/municipalities have similar percentage of poor people tend to be in groups. Most of the adjoining areas have a similar percentage of the poor. This is shown by the similarity in color in Figure 1. Figure 1 also explains that every time change in areas that have a high percentage of poor people tends to decrease every year. The linear relationship between each explanatory variable and the response variable can be seen using the Pearson correlation value. In Table 2 it can be seen that there are explanatory variables that have a positive linear relationship and a negative linear relationship.

Then multicollinearity is checked from the value of VIF (Variance Inflation Factor). Multicollinearity occurs if the VIF value is greater than ten (VIF> 10). The test results showed that in 2012-2018 most explanatory variables had a VIF value of less than 10 (ten), therefore it can be concluded that there is no multicollinearity between variables.

Table 2: Correlation value of response variable with explanatory variable

Variable	Korelasi *Pearson*
Y dengan X_1	-0.57
Y dengan X_2	0.53
Y dengan X_3	0.45
Y dengan X_4	0.37
Y dengan X_5	0.61
Y dengan X_6	-0.49
Y dengan X_7	-0.19
Y dengan X_8	-0.42

3.2. *Geographically and Temporally Weighted Autoregressive (GTWAR)*

The results of the *Moran* Index test and the *Breusch-Pagan* test

Testing with the Moran Moran index on the Y response variable was carried out to see whether or not there was spatial autocorrelation in the data. Based on the results of the Moran index test in Table 3, it can be seen that the Moran index is real at the 5% real level, which indicates that there is a positive Moran index value (I> 0) in the data on the percentage of poor people in regency/municipality of Java Island. These results indicate that the poverty rate in the data has a positive spatial autocorrelation, which means that areas that are close to each other tend to have similar poverty levels (groups).

Table 3: The results of testing the Moran's index on the response variable *y*

Years	Moran's index
2012	0.28
2013	0.28
2014	0.25
2015	0.24
2016	0.25
2017	0.24
2018	0.24
2012-2018	0.28

Testing to determine heterogeneity in data due to spatial influence is done through a variety of homogeneity tests with the Breusch Pagan test for each year from 2012 to 2018 and simultaneously on 118 regency/municipality is presented in Table 4.

Table 4: Test statistics of the Pagan Breusch

Years	P-value
2012	0.0008*
2013	0.0025**
2014	0.1189*
2015	0.2611
2016	0.0268*
2017	0.1864
2018	0.6175
2012-2018	2.654×10^{-9}*

*significant at $\alpha = 5\%$

**significant at $\alpha = 10\%$

Based on Table 4, it is found that observations made from 2012-2018 show that there is spatial diversity in 2012, 2013, 2014 and 2016, but for 2015, 2017 and 2018 there is no spatial diversity in the percentage of poor people on the island of Java, which is thought to be due to differences characteristics at each location. If a simultaneous spatial test is carried out, it is found that there is spatial variation in the percentage of poor people in Java. The existence of spatial diversity over time can be handled by applying temporal geographic weighted regression.

Determination of Weighted Matrix Parameters for the GTWAR Model

The GTWAR model combines spatial-temporal heterogeneity and the effects of spatial autocorrelation. The GTWAR method analysis requires a weighted matrix (W) which is constructed from the euclidean distance matrix. This distance matrix uses the interaction between spatial distance and temporal distance. Therefore, in building a euclidean distance matrix with interactions requires a balancing parameter. The use of balancing parameters is intended because of the difference in units between spatial distance, temporal distance and their interactions. The balancing parameters used in this study are the spatial distance parameter (λ), the temporal distance parameter (μ), the spatial and temporal distance interaction parameter (τ) and the value of the temporal spatial window width (h_{ST}). To calculate these parameters, a cross validation approach is used. The parameter values for each model can be seen in table 5.

Table 5: Parameter values in each models

Models	τ	λ	μ	h_{ST}
GTWR	0.4802	0.9007	0.4325	5.5033
GTWAR	0.1869	0.9547	0.1785	4.7022

Then a comparison of the goodness of the model is performed on the global regression model, GTW and

GTWAR to see which model is better for modeling the percentage of poor people in Java from 2012-2018. The criteria used is to compare the values of AIC dan pseudo R^2 as in Table 6.

Table 6: Comparison of model goodness

	Global Regression	GTWR	GTWAR
AIC	619.908	-1427.851	-1964.819
Pseudo R^2	0.5383	0.6237	0.7265

The comparison of the goodness of several models can be done by looking at the value of AIC and Pseudo R^2. The model goodness indicators presented in Table 5 show that the GTWAR model has a higher Pseudo R^2 value than the other models, which is 72.65% and the smallest AIC is -1964.819. The high Pseudo R^2 value and the small AIC value in the GTWAR model means that the addition of the autoregressive component in y can change the coefficient of GTWR, where the explanatory variable in the model is able to explain the response variable by 72.65%. Based on the AIC and Pseudo R^2 values, it can be concluded that the GTWAR model is better than other comparison models.

Estimation of GTWAR Model Parameters

It can be seen in Figure 2 that the value of $\hat{\beta}_1$ produces an estimated value that varies between regions and years. This means that when the estimation is positive, it will have an impact on the increasing percentage of poor people in certain areas and years when there is an increase in gross regional domestic product (X1), whereas if the estimation is negative it means that it will have an impact on the decreasing percentage of poor people if other independent variables remain for all regencies/municipalities in 2012-2018.
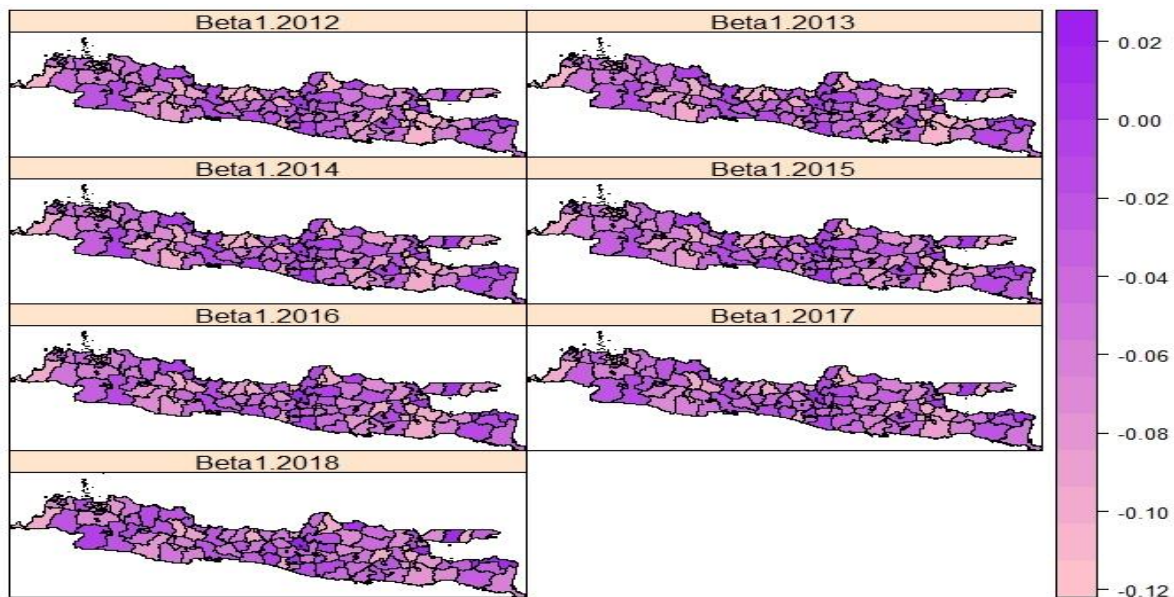

Figure 2: Map of the spatial-temporal distribution of parameter estimator X1

In Figure 3, it can be seen that the value of $\hat{\beta}_2$ also produces an estimated value that varies between regions and

years. This means that when the suspect is positive, it will have an impact on the increasing percentage of poor people in certain areas and years when there is an increase in the percentage of poor people aged 15 years with primary school education (X2), whereas if the suspect is negative it means that it will have an impact on the decline percentage of poor people.
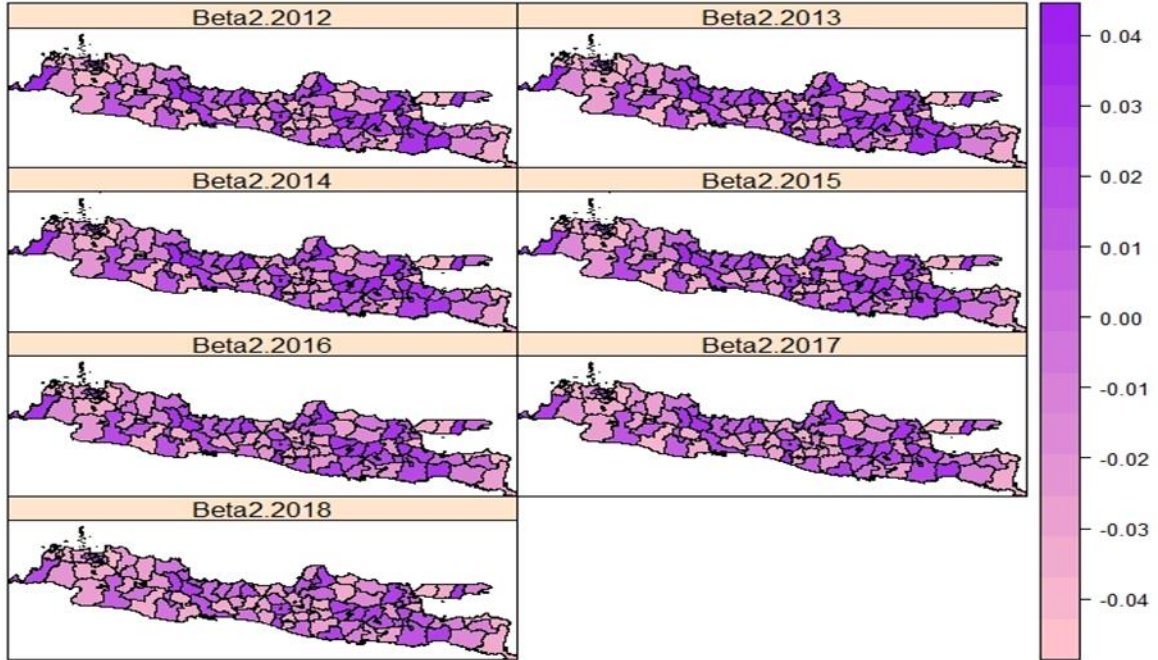

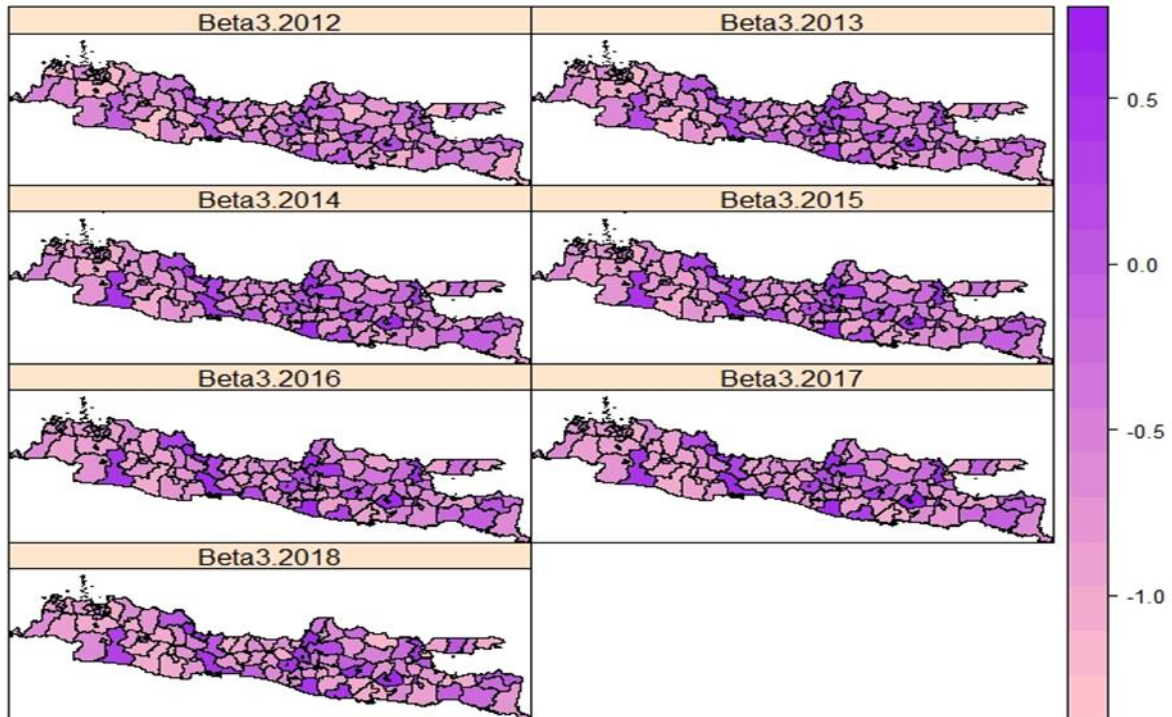**Figure 3:** Map of the spatial-temporal distribution of parameter estimator X2

Figure 4: Map of the spatial-temporal distribution of parameter estimator X3

In Figure 4, you can see the value of $\hat{\beta}_3$, there are also areas that have negative values in 2012-2018. This has an impact on the decrease in the percentage value of poor people in certain areas and years when there is an increase in the literacy rate (X3), while for areas with a positive $\hat{\beta}_3$ value it can be interpreted that the higher the literacy rate will have an impact on the increase in the percentage value of poor people.

Figure 5 shows that the value of $\hat{\beta}_4$ produces an estimated value that varies between regions and years. It means that when the estimation is positive, it will have an impact on the increasing percentage of poor people in certain areas and years when there is an increase in the percentage of per capita expenditure on food (X4), whereas if the estimation is negative, it will have an impact on the decreasing percentage of poor people.
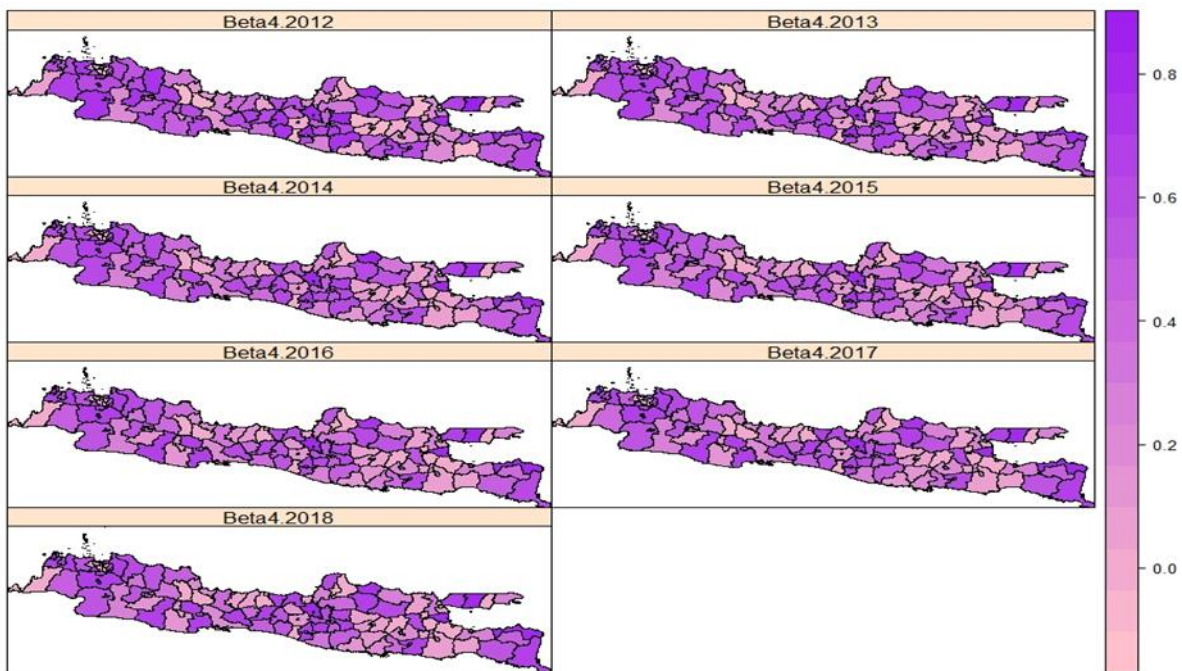

Figure 5: Map of the spatial-temporal distribution of parameter estimator X4

In Figure 6, it can be seen that the value of $\hat{\beta}_5$ produces the same predicted value between regions and years, which is positive. It means that when the estimation is positive, it will have an impact on the increasing percentage of poor people in certain areas and years when there is an increase in the percentage of poor households receiving raskin (X5) if the other independent variables remain for regencies/municipalities in 2012-2018.
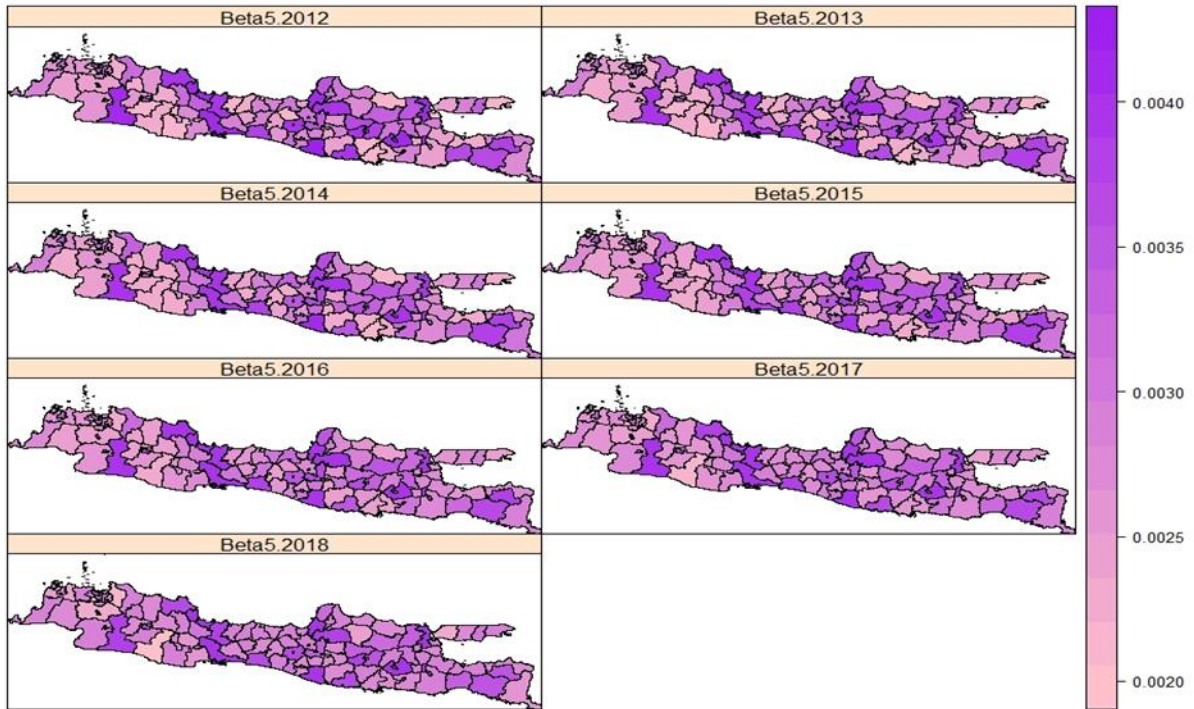
Figure 6: Map of the spatial-temporal distribution of parameter estimator X5

In Figure 7, it can be seen that the value of $\hat{\beta}_6$ produces the same predictive value between regions and years, which is negative. It means that when the estimation is negative, it will have an impact on the decreasing percentage of poor people in certain areas and years when there is an increase in the percentage of population by age group (X6).
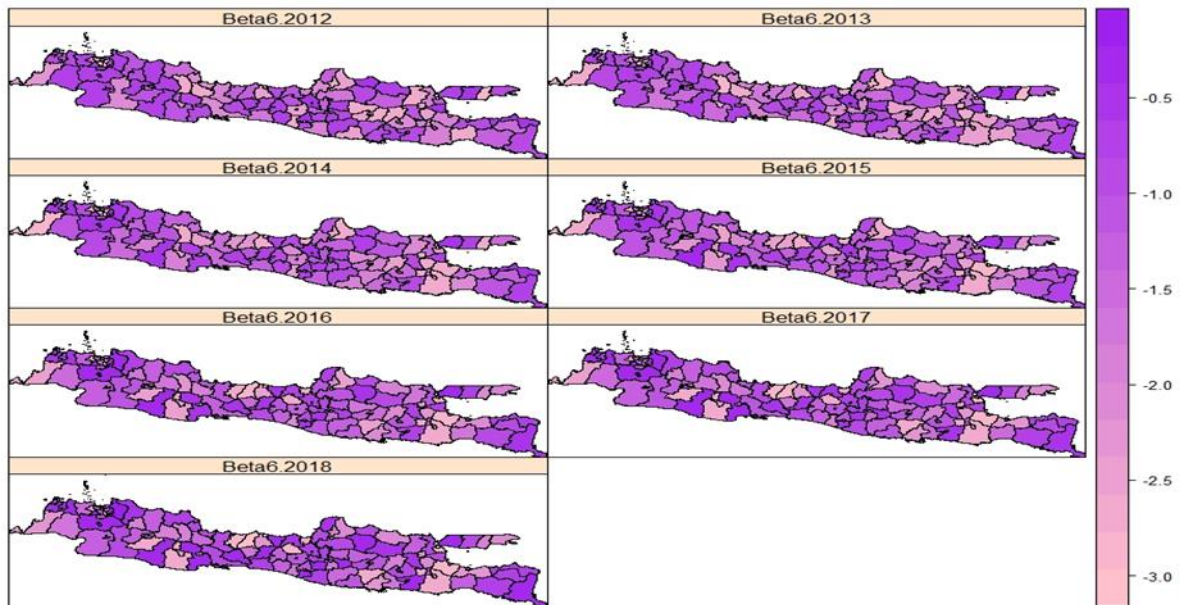

Figure 7: Map of the spatial-temporal distribution of parameter estimator X6

In Figure 8 it is also seen that $\hat{\beta}_7$ produces the same predicted value between regions and years, which is negative. It means that when the estimation is negative, it will have an impact on the decreasing percentage of poor people in certain areas and years when there is an increase in the expected number of years of schooling (X7) if the other independent variables are fixed for all regencies/municipalities in 2012-2018.
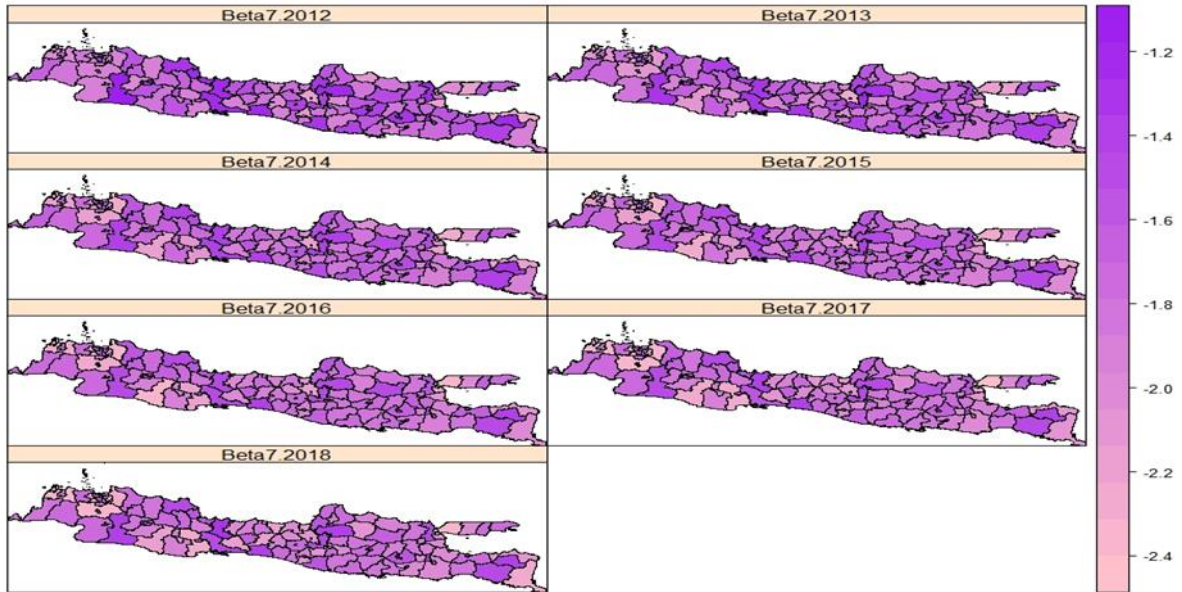

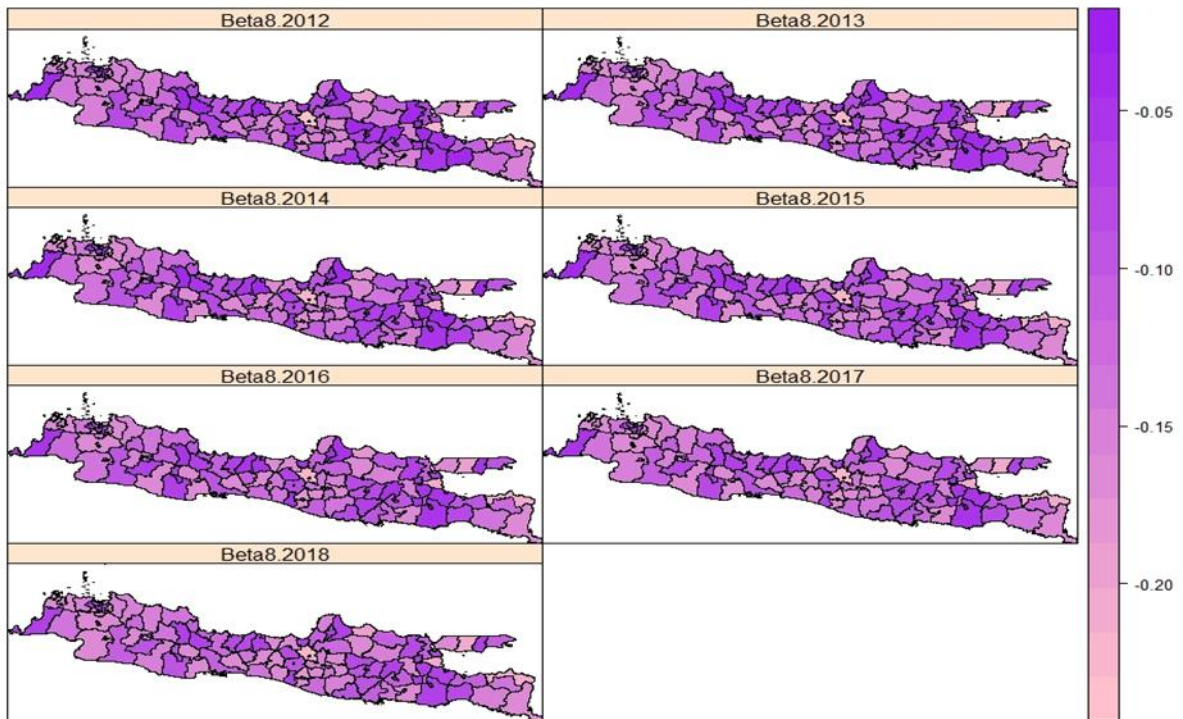**Figure 8:** Map of the spatial-temporal distribution of parameter estimator X7

Figure 9: Map of the spatial-temporal distribution of parameter estimator X8

In Figure 9, it can be seen that the value of $\hat{\beta}_8$ also produces the same predicted value between regions and years, which is negative. It means that when the estimation is negative, it will have an impact on the decreasing percentage of poor people in certain areas and years when there is an increase in the average length of schooling (X8) if the other independent variables are constant for regencies/municipalities in 2012-2018.

4. Conclusion

The percentage of poor people on the island of Java in 2012-2018 was identified as having spatial and temporal diversity so that they could use the GTWR. In addition, the presence of autocorrelation between response variables between locations fulfills the requirements for the GTWAR model to be applied. The high value of Pseudo R^2 and the small value of AIC in the GTWAR model means that it can explain that the addition of an autoregressive component in y can change the coefficient of GTWR. The GTWAR model produces different parameter estimators in each regency/municipality on the island of Java in the 2012-2018 period which causes the factors that affect the percentage of poor people to vary in each regency/municipality on the island of Java.

References

[1]. Badan Pusat Statistik. "Profil Kemiskinan di Indonesia Maret 2019". Jakarta: BPS. 2019.

[2]. A.S. Fotheringham, C. Brunsdon, M. Charlton. Geographically Weighted Regression: The Analysis of Spatially Varying Relationship. England: Jhon Wiley & Sons Ltd. 2002.

[3]. Bo Wu, Rongrong Li, Bo Huang. A Geographically and Temporally Weighted Autoregressive Model with Application to Housing Prices. International Journal of Geographical Information Science, 28(5): 1186-1204, 2014.

[4]. Liu J, Yang Y, Xu S, Zhao Y, Wang Y, Zhang F. A Geographically Temporal Weighted Regression Approach with Travel Distance for House Price Estimation. *Article Entropy MDPI*. 303 (18) : 1-13, 2016.

[5]. M Sholihin, A.M. Soleh, A. Djuraidah. Geographically and Temporally Weighted Regression (GTWR) for Modeling Economic Growth Using R. International Journal of Computer Science and Network. 6(6): 800-805, 2017.

[6]. G. Erda, Indahwati, A. Djuraidah. "A Comparison of GWTR and Robust GTWR Modelling". International Journal of Scientific Research in Science, Engineering, and Technology, vol. 4 (9), pp 453-457, 2018.

[7]. A.H. Asianingrum, A. Djuraidah, Indahwati. robust mixed geographically and temporally weighted regression to modeling the percentage of poverty population in java in 2012-2018. International Journal of Sciences: Basic and Applied Research (IJSBAR), vol. 53 (2), pp 186-197, 2020.

[8]. H. Yasin, Sugito, A. Prakutama. "Analisis Data Kemiskinan di Jawa Tengah Menggunakan Metode Mixed Geographically and Temporally Weighted Regressions (MGTWR)". Biostatistics, vol. 1(1), pp. 15-23, 2015.

[9]. R. Affifah, Y. Andriyana, I.M. Jaya. "Robust Geographically Weighted Regression with Least Absolute Deviation Method". American Institute of Physics, pp. 1-9, 2017.

[10]. Nuramaliyah, A. Saefuddin, M.N. Aidi. "The Best Global and Local Variables of the Mixed

Geographically and Temporally Weighted Regression Model. Indonesian Journal of Statistics and Its Applications, vol. 3(3), pp. 320-330, 2019.

[11]. K.Y. Widiyanti, H. Yasin, Sugito. "Pemodelan Proporsi Penduduk Miskin Kabupaten dan Kota di Provinsi Jawa Tengah Menggunakan Geographically and Temporally Weighted Regression". Jurnal Gaussian, vol. 3(4), pp. 691-700, 2014.

[12]. Anselin L. A Note on Small Sample Properties of Estimators in a FirstOrder Spatial Autoregressive Model. *Environ. Plan. A Econ. Sp.* 14(8) 1023–1030, 1982.

[13]. Anselin L. Spatial Econometrics: Methods and Models. Dordrecht: Academic Publishers. 1988.

[14]. H. Sugi, M.N. Aidi, A. Djuraidah. Analysis of the Geographically and Temporally Weighted Regression (GTWR) of the GRDP the Construction Sector in Java Island. Forum Geografi, 33(1), 2019.