

The Step Construction of Geographically Weighted Panel Regression in Air Polluter Standard Index (APSI) Data

Budi Warsito^{1,*}, Hasbi Yasin², Dwi Ispriyanti³ and Arief Rachman Hakim⁴

^{1,2,3,4}Department of Statistics, Faculty of Science and Mathematics, Diponegoro University, Semarang – Indonesia

Abstract. Geographically Weighted Panel Regression or GWPR is a local linear regression model that combines GWR model and panel data regression model with considering spatial effect, especially spatial heterogeneity problem. This article is focused on the soft computation of GWPR model using Fixed Effect Model (FEM). Parameter estimation in GWPR is obtain by Weighted Least Squares (WLS) methods and the resulting model for each location will be different from one to another. This study will compare the fixed-effect GWPR model with several weighting functions. The best model is determined based on the biggest coefficient of determination (R^2) value. In this study, the model is applied in the Air Polluter Standard Index (APSI) in Surabaya City, East Java. The results of this study indicate that Fixed Effect GWPR model with a fixed exponential kernel weighting function is the best model to describe the APSI because it has the smallest AIC.

Keywords: **APSI; Fixed Effect, GWPR; Panel Regression.**

1 Introduction

Regression method is a statistical method used to model dependence of one dependent variable on one or more independent variables [1]. Regression methods that can be used to model data involving cross section elements and time series elements are called panel data regression methods. However, in panel data regression cannot overcome the existence of spatial effects. Meanwhile, spatial regression can be used to overcome the existence of spatial effects. Spatial effects are divided into two parts, namely spatial dependence and spatial heterogeneity [2]. Spatial dependence is developed through Spatial Autoregressive (SAR) model and Spatial Error Model (SEM) [3]. While spatial heterogeneity was developed through Geographically Weighted Regression (GWR) [4].

The model construct of panel-type data and spatial effects, particularly spatial heterogeneity problems, can be developed through spatial panel regression analysis by combining Geographically Weighted Regression (GWR) models with panel data regression models to form Geographically Weighted Panel Regression (GWPR) models. GWPR is a merger between the GWR model and the panel data regression model [5]. GWR is defined as the development of the regression by adding geographical location where the data is obtained so that the estimated parameters generated will be local in accordance with the number of observation locations [4]. In the GWPR model, the time element is involved in the GWR model [5]. GWPR which is a combination of

GWR model with panel data regression will produce GWPR fixed effect model and GWPR random effect [6].

Air pollution is one of the real problems that can threaten human life. Air quality degradation in big cities is one indicator of problems due to air being polluted by air pollutant elements. The main influential factors which are the main sources of air pollution in big cities are transportation machinery, exhaust gas industry, population density, the number of shopping centre, air humidity [7, 8]. The main elements of air pollutants according to the Air Pollutant Standard Index (APSI) are Carbon Monoxide (CO), Particulate Matter (PM), Sulfur Dioxide (SO₂) Nitrogen Dioxide (NO₂) and Ozone (O₃) [8]. Factors Causes and effects of air pollution have been closely related to the location of the observation, meaning that between locations will have spatially different causes and effects. Therefore, demographically, the potential impacts and causes of air pollution will vary between regions [9, 10]. Spatial regression method is one approach that can be used for air pollution data modeling. The model is a Geographically Weighted Regression (GWR) model, which is a regression method that involves location effects into predictors [4]. Because the APSI data in this study are spatio temporal, in this study the GWR model is used based on the panel regression model. This model is known as the FE-GWPR model.

The parameter estimation used in the GWPR model is obtained by assigning weight to each location. There are several ways in which to determine the weight of each location of observation, one of which uses kernel

* Corresponding author: budiwrst2@gmail.com

functions, which can be divided into two, fixed and adaptive kernels. In this study, FE-GWPR will be applied to APSI data modeling in Surabaya City, East Java.

2 Literature Review

2.1 Fixed Effect Panel Regression Model

The fixed effect model is also known as Least Square Dummy Variable (LSDV), therefore the observed value on the coefficient μ_i is a dummy variable having different values for each i-th individual [11]. The parameter estimation on the fixed influence model one uses in transformation [11], ie equation (1) is averaged for the whole time to obtain the equation:

$$\bar{y}_i = \alpha + \mu_i + \bar{\mathbf{X}}_i \boldsymbol{\beta} + \bar{\varepsilon}_i \quad (1)$$

There are several approaches to estimate the fixed effect model, one of which is the fixed effect within transformation model [12]. The effect of the fixed effect within transformation model is done by eliminating the effect of the cross section unit (and then the value of the dependent and independent variables of each cross section unit is averaged over time.) Continue to estimate the fixed effect within transformation model can be done by Ordinary Least Squares (OLS).

2.2 Fixed Effect Geographically Weighted Panel Regression Model

The main idea of Geographically Weighted Panel Regression is similar to GWR analysis for cross-sectional data [6]. In the GWPR model, it is assumed that the time series of observations in the geographic location is the realization of a smooth spatiotemporal process [6]. GWPR model is a combination of the GWR model and the panel data regression model involving time elements in the GWR model [5]. The GWPR Fixed Effect model equation is:

$$y_{it} = \sum_{k=1}^p \beta_k(i) x_{itk} + \varepsilon_{it} \quad (2)$$

where $i = 1, 2, \dots, n$ (number of location) and $t = 1, 2, \dots, T$ (number of time periods)

Equation (2) is obtained from the transformation of within transformation [13]. This transformation consists of reducing the fixed effect model equations with the mean model equations. The fixed effect model is expressed in the following equation:

$$y_{it} = \beta_0(i) + \mu_i + \sum_{k=1}^p \beta_k(i) x_{itk} + \varepsilon_{it} \quad (3)$$

Whereas, the average equation is expressed in the following equation:

$$\bar{y}_i = \beta_0(i) + \mu_i + \sum_{k=1}^p \beta_k(i) \bar{x}_{ik} + \bar{\varepsilon}_i \quad (4)$$

The result of the transformation from the reduction of equation (3) to equation (4) is expressed as follows:

$$y_{it}^* = \sum_{k=1}^p \beta_k(i) x_{itk}^* + \varepsilon_{it}^* \quad (5)$$

which:

$$y_{it}^* = y_{it} - \bar{y}_i; \quad x_{itk}^* = x_{itk} - \bar{x}_{ik}; \quad \text{and} \\ \varepsilon_{it}^* = \varepsilon_{it} - \bar{\varepsilon}_i$$

The estimation of Fixed Effect GWPR model parameters uses the Weighted Least Square (WLS) approach as estimated in the Geographically Weighted Regression (GWR) model [5, 14] so that the estimator of the model parameters is as follows:

$$\hat{\boldsymbol{\beta}}(i) = [\mathbf{X}^T \mathbf{W}(i) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(i) \mathbf{y}$$

\mathbf{W} is (nt x nt) matrix, which is as follows:

$$\mathbf{W}(i) = \text{diag}(w_1(i), w_2(i), \dots, w_n(i))$$

The weighting function in the GWPR model is the same as the weight in the GWR model, which depends on the distance between points of observation. Observations in the local sampling location are given weight based on the kernel function in GWPR as well as in GWR then the weight is applied to all time periods [6]. One to determine the weighting matrix by using kernel functions [15] and one type of kernel function, namely function adaptive kernel with an exponential kernel function [13] which is stated in the following formula:

$$w_j(i) = \exp(-d_{ij}/h_i)$$

where d_{ij} denotes the distance between the location (u_i, v_i) to location (u_j, v_j) and h_i are positive parameters are known and are usually called smoothing parameter (*bandwidth*) for location (u_i, v_i) .

In the kernel weighting function, there is a bandwidth parameter. Bandwidth is analogous to a radius of a circle, so that an observation location within the circle radius is still considered influential in forming parameters at one point of observation location [11]. There are several methods that can be used to choose the optimum bandwidth and one of them uses Cross Validation (CV) [4]. The calculation of CV in GWPR is equal to GWR which is calculated based on the average dependent and independent variables for the whole time [6] and is defined as follows:

$$CV(h_i) = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h_i))^2$$

where $\hat{y}_{\neq i}(h_i)$: the fitted value of y_i with the observation at location omitted from fitting process. Testing of GWPR model parameters was carried out using the analogy of the F test and t test as discussed in Leung [17] and Purhadi [18].

3 Methodology

This research aims to develop the Fixed Effect Geographically Weighted Panel Regression (FE-GWPR) model, which is intended to see the influence of the predictor variables to the Air Polluter Standard Index (APSI) locally. This data was collected from five location:

- (1) Centre Surabaya (Jalan Ketabang kali//SUF1)
- (2) South Surabaya (Jalan Masjid Al Akbar Gayungan/SUF4),
- (3) West Surabaya (Jalan Simomulyo/SUF3),
- (4) East Surabaya (Jalan Arief Rahman Hakim/SUF5), and
- (5) Wonorejo Surabaya (Jalan Kendal Sari 117/SUF6).

The respon variable is the APSI data, and the predictor variables are:

1. The traffic velocity,
2. The population density,
3. The business center aspect,
4. The air humidity,

In this research R is used to analyse the Fixed Effect GWPR model. The following steps are taken to the data analysis in this study:

1. Obtain APSI data and the predictor variables that it influence.
2. Perform a fixed effect model estimation.
3. Test the assumption of panel data regression, namely normality test, non-autocorrelation test, homoskedasticity and non-multicollinearity.
4. Test the spatial heterogeneity and local non-multicollinearity assumptions.
5. Calculate the euclidean distance between one location which is located in the (u_i, v_i) coordinates and other location located at (u_j, v_j) coordinates.
6. Calculate the optimum bandwidth with local sample data (average overall time data) using the CV method and the fixed exponential kernel weighting matrix.
7. Estimating Fixed Effect GWPR model parameters uses deviation between data with average data for each location unit against time units using a fixed exponential kernel weighting matrix.
8. Testing the Fixed Effect GWPR model.
9. Obtain the final model and then interpret the model based on the smallest value of AIC.

4 Result and Discussion

4.1 Panel Regression Model

The following is the estimation of the fixed effect panel regression model for APSI data in Surabaya City, East Java Province:

$$\hat{y}_{it} = -0.7767 X_{1it} - 2.2147 X_{2it} + 0.0307 X_{3it} + 0.0003 X_{4it}$$

In testing the panel data regression assumptions, it was concluded that the residual normality test was normally distributed, the variance of residual homoskedasticity test was not constant, in the non-multicollinearity test there was no linear relationship between the independent variables and there was no residual autocorrelation in the non-autocorrelation test.

Based on the assumption of panel data regression models, there are assumptions that are not met, namely the existence of heterogeneity. Based on the cross section unit which is an area, the possibility of heterogeneity that occurs due to the condition of the observation area that is not homogeneous so that spatial heterogeneity occurs. Based on testing of spatial heterogeneity obtained Breusch-Pagan [1] BP values of 124.340 and p-value of 0.0000, which means that there is spatial heterogeneity. Whereas, the assumption of local non-multicollinearity is fulfilled because the VIF value of each independent variable for each location less than 10 which indicates that there is no local multicollinearity.

4.2 Fixed Effect GWPR of APSI data

The first step taken in Fixed Effect GWPR modeling is to determine the geographical coordinates of each sample that will be used. After that, calculate the average dependent and independent variables for the entire time at each location to get the optimum bandwidth value using cross validation (CV) criteria and weighting values. The weighting matrix is the diagonal value of the weight that has been obtained and repeats as much as the time unit to get the parameter estimation. The model that is formed will be different at each location.

Firstly, select the best weighting function which is used to develop the Fixed Effect GWPR model. Table 1 shows that the Fixed Bandwidth Exponential Kernel is the best weighting scheme of APSI data because it has the smallest value of AIC.

Table 1. Selection the best model using AIC

Model	AIC
Fixed Effect-GWPR	2437.841
Fixed Effect Panel Regression	2448.047

Table 2 shows the goodness of fits of Fixed Effect GWPR uses the fixed bandwidth exponential weighting kernel based on Leung [13] and Purhadi [14]. The result conclude that the Fixed Effect GWPR is the not

significantly better than Fixed Effect Panel Regression Model without spatial effect.

Table 2. Goodness of Fits of Fixed Effect GWPR

Results	Decision
F = 1.0178 and Sig. = 0.4416	H ₀ not rejected, which means that Fixed Effect GWPR Model are not significantly difference with Fixed Effect Panel Regression

The final model in each location is shown in Table 3. The results of parameter estimation at each location have different values. This shows that the FE-GWPR model can be used to detect the influence of spatial predictor variables based on the location of the observation.

Table 3. Model Parameters at Each Location

Location	β_1	β_2	β_3	β_4
SUF1	-0.4799	-2.2859	-0.0199	-0.0001
SUF3	-1.0252	-2.8311	0.0220	0.0001
SUF4	-0.8428	-1.9087	0.1152	0.0003
SUF5	-0.3111	-2.3039	0.0155	0.0006
SUF6	-1.0464	-1.7288	0.0297	0.0006

4 Conclusion

The results of this study indicate that Fixed Effect GWPR model with a fixed exponential kernel weighting function is the better model than the Fixed Effect Panel Regressions to describe the APSI data in Surabaya City, East Java Province because it has the smallest AIC.

We would like to give thanks to Directorate of Research and Public Services, The Ministry of Research, Technology and Higher Education Republic of Indonesia for their support. This research was funded by "PTUPT" Research Grant 2018.

References

1. N. Draper, H. Smith, Analisis Regresi Terapan. PT. Gramedia Pustaka Utama. Jakarta (1992)
2. L. Anselin, *Spatial Econometrics: Methods and Models*, Dordrecht: Kluwer Academic Publishers (1988)
3. L. Anselin, *Handbook of Spatial Analysis*, London: Sage Publications (2009)
4. A.S. Fotheringham, C. Brunson, M. Charlton, *Geographically Weighted Regression*, England: John Wiley & Sons Ltd (2002)
5. A.Y. Qur'ani, *Jurnal Mahasiswa Statistik* **2**, 3 (2014)
6. D. Yu, *Remote Sensing and Spatial Information Science* **38**, Part II (2010)
7. F. Atash, *Cities*, **24**, 6 p. 399-409 (2007)
8. M. Fahimi, B. Dharma, D. Fetararayani, A. Baskoro, G. Sugiarto, C. Effendi, *Jurnal Penyakit Dalam*, **12**, 1, p. 1-9 (2012)
9. A. Gilbert, J. Chakraborty, *Social Science Research*, **40**, p. 273-286 (2011)
10. D. Robinson, J. M. Lloyd, *International Journal of Applied Earth Observation and Geoinformation*, **21**, p. 374-383 (2013)
11. B.H. Baltagi, *Econometrics Analysis of Panel Data (3rd ed.)*, Chicester, England: John Wiley & Sons Ltd (2005)
12. J.M. Wooldridge, *Econometric Analysis of and Panel Data*, London, England: The MIT Cross Section Press (2002)
13. F. Bruna, D. Yu, *Geographically Weighted Panel Regression. XI Congreso Galego de Estatística e Investigación de Operacións*, (A Coruna 24-26 de outubro de 2013)
14. R. Cai, D. Yu, M. Oppenheimer, *Journal of Agricultural and Resource Economics* **39**, 2 (2014)
15. C. Chasco, I. Garcia, J. Vicens, J. Munich Personal RePEc Archive (MPRA) , Working Paper, No. 1682 (2007)
16. D.C. Wheeler, P. Antonio, *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, Berlin: Springer (2010)
17. Y. Leung, C.L. Mei, W.X Zhang, W. X. *Environment and Planning A*, **32**, p. 9-32 (2000)
18. Purhadi, H. Yasin, *European Journal of Scientific Research* **69**, 2 (2012)