Geographically Weighted Regression Model with Kernel Bisquare and Tricube Weighted Function on Poverty Percentage Data in Central Java Province.

N F T A Nugroho¹, I Slamet²
¹Department of Mathematics, ²Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Sebelas Maret Jl. Ir. Sutami 36A, Kenitingan, Surakarta, 57126, Indonesia
E-mail: nungkifauzi@gmail.com

Abstract. Poverty is a socio-economic condition of a person or group of people who can not fulfil their basic need to maintain and develop a dignified life. This problem still cannot be solved completely in Central Java Province. Currently, the percentage of poverty in Central Java is 13.32% which is higher than the national poverty rate which is 11.13%. In this research, data of percentage of poor people in Central Java Province has been analyzed through geographically weighted regression (GWR). The aim of this research is therefore to model poverty percentage data in Central Java Province using GWR with weighted function of kernel bisquare, and tricube. As the results, we obtained GWR model with bisquare and tricube kernel weighted function on poverty percentage data in Central Java province. From the GWR model, there are three categories of region which are influenced by different of significance factors.

Keywords: Percentage of poverty, GWR, Bisquare and Tricube kernel function.

1. Introduction
Poverty problem has been investigated by numerous researchers from different point of views. We refer to Fotheringham [4], Nugroho [8], and Slamet et al. [10] as examples. One of the categories of poverty is people who has low income ability to fulfil their basic needs (Slamet et al. [10]). Chaudry and Wimer [2] explained that poverty is an important indicator of community and children welfare. Poverty and low family income seriously affect in children development, especially in cognitive and educational aspects.

According to Guting [6], spatial data is data related to space occupied by objects that can be dots, lines and spaces. In the regression model, spatial data analysis in the form of points can be modelled through geographically weighted regression. Furthermore, estimation of model parameters also can be derived without any difficulties (Fotheringham [4]). This regression model is part of spatial analysis in which weight is taken based on position or distance of one observation location with other observation location. The location point is the coordinate of longitude and latitude of an area on the map.

In this research, we present the best global regression model on poverty percentage data in Central Java Province. Furthermore, the analysis is developed through geographically weighted regression model with the weighting function of kernel bisquare and tricube model.
2. Literature Review

In this section we review multiple linear regression, spatial heterogeneity, geographically weighted regression, and mapping models.

**Multiple Linear Regression**

Multiple linear regression model with $p$-1 independent variables given as follow.

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$  \hspace{1cm} (1)

In matrix form the equation (1) can be written as

$$Y = X\hat{\beta} + \varepsilon$$  \hspace{1cm} (2)

where $Y$ is the dependent variable vector ($n \times 1$), $X$ is the independent variable matrix ($n \times p$), $\hat{\beta}$ is vector parameter ($p \times 1$), $\varepsilon$ is an error vector ($p \times 1$) mean zero and variance $\sigma^2 I$, $I$ is the identity matrix. Parameter estimation is obtained by minimizing the sum of squared errors (SSE) as below

$$\varepsilon^T \varepsilon = (Y - X\hat{\beta})^T (Y - X\hat{\beta})$$

$$= Y^T Y - 2\hat{\beta}^T X^TY + \hat{\beta}^T X^TX\hat{\beta}$$ \hspace{1cm} (3)

The next step to minimize the sum of squared errors is decrease equation (3) against $\hat{\beta}$.

$$0 = X^TX\hat{\beta} - X^TY$$

$$X^TX\hat{\beta} = X^TY$$

$$(X^TX)^{-1}X^TX\hat{\beta} = (X^TX)^{-1}X^TY$$ \hspace{1cm} (4)

Because of $(X^TX)^{-1}X^TX$ equal to $I$, then the parameter estimation of $\hat{\beta}$ in can be expressed by

$$\hat{\beta} = (X^TX)^{-1}X^TY.$$ \hspace{1cm} (5)

The testing statistic $F$ of regression model is

$$F = \frac{SSR}{SSE}.$$ \hspace{1cm} (6)

Partial parameter test of regression model is

$$t = \frac{b_k}{s(b_k)}.$$ \hspace{1cm} (7)

Normality and multicollinearity assumption tests are as follow. (Gujarati [5]).

i. The testing statistic of normality using Kolmogorov Smirnov is

$$D = \max \{|F_n(\varepsilon_i) - F_0(\varepsilon_i)|\}$$ \hspace{1cm} (8)

ii. The testing statistic of multicollinearity is

$$VIF = \frac{1}{1-R^2}$$ \hspace{1cm} (9)

If the VIF value is more than 10, then it indicates the presence of multicollinearity among independent variables. These two assumptions must be fulfilled for a global regression model.

**Spatial heterogeneity.** After finishing the assumption test, the next testing is spatial heterogeneity. Spatial heterogeneity occurs when the independent variable gives different effects for each location. Spatial heterogeneity testing can be done with Breusch-Pagan test (Anselin [1]).

$$BP = \frac{1}{2} [h^T Z(Z^TZ)^{-1}Z^T h] \sim \chi^2_k$$ \hspace{1cm} (10)

**Geographically weighted regression.** Geographically weighted regression (GWR) is an expansion of the linear regression model. The difference with the linear regression model is the GWR model taking into t the location aspect. The GWR model defined as

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p-1} \beta_k(u_i, v_i) x_{i,k} + \varepsilon_i, i = 1, 2, \ldots, n,$$ \hspace{1cm} (11)
dependent variable matrix. Bisquare and tricube kernel functions will be used to obtain the value of the sum of the minimum error squares can be obtained by deriving the equation (13) to equal to zero so that

\[ w_{ij} \sum_{j=1}^{n} w_{ij}^2 = \sum_{j=1}^{n} w_{ij} (y_i - (\beta_0(u_i, v_i) + \sum_{k=1}^{p-1} \beta_k(u_i, v_i) x_{i,k}))^2 \]  

Then equation (12) can be expressed in matrix form as follow

\[ e^T W(u_i, v_i) e = Y^T W(u_i, v_i) Y - 2 \beta^T X^T W(u_i, v_i) Y + \beta^T X^T W(u_i, v_i) \beta \]  

The spatial weights are used to estimate the parameters of the GWR model. The following bisquare and tricube kernel functions will be used.

\[
  w_{ij} = \begin{cases} 
    1 - \left( \frac{d_{ij}}{h_i} \right)^2, & d_{ij} < h_i \\
    0, & d_{ij} \geq h_i 
  \end{cases}
\]  

(16)

\[
  w_{ij} = \begin{cases} 
    1 - \left( \frac{d_{ij}}{h_i} \right)^3, & d_{ij} < h_i \\
    0, & d_{ij} \geq h_i 
  \end{cases}
\]  

(17)

where \( d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \). \( d_{ij} \) is the Euclidean distance. Bandwidth is the radius of a point to the extent to which the point influences that form a circle. The best bandwidth calculations can use cross validation (CV) as follow

\[ CV = \sum_{i=1}^{n} \sum_{j=1}^{n} (y_i - \hat{y}_{j\neq i}(b))^2 \]  

(18)

where \( \hat{y}_{j\neq i}(b) \) is the estimated value of the dependent variable with location \( j \neq i \) (Fotheringham et al. [4]).

**Mapping models.** In this research, mapping of the region in Central Java Province is based on factors affecting poverty. After determining the GWR model, then we classify regions based on factors that affect poverty significantly by using partial test parameters. The partial test of \( \beta(u_i, v_i) \) on the GWR model is as follow.

\[
  t = \frac{\hat{\beta}_k(u_i, v_i)}{SE(\hat{\beta}_k(u_i, v_i))}
\]  

(19)
where $\hat{\beta}_k(u_i, v_i)$ is an estimated parameter of the GWR model, and $SE(\hat{\beta}_k(u_i, v_i))$ is a standard GWR model error. This test statistic will be rejected if $|t| > t_{(\frac{\alpha}{2}, n-p)}$.

3. Research Method

For achieving the research goal, there are two steps taken. Those are collecting data and determining the model.

For the first step, data of percentage poverty and factors affecting poverty in Central Java was collected from Badan Pusat Statistik (Central Bureau of Statistics) (BPS[2]). This data consists of $X_1$ (labor force participation rate as), $X_2$ (percentage of population working as farmers), $X_3$ (HDI), and $X_4$ (population growth rate).

The next step that needs to be done in this research is to determine the global regression model using the least squares method. Having obtained the best model then it needs to test spatial heterogeneity. If there is spatial heterogeneity in the variance error, then the data can be modelled by geographically weighted regression. When GWR model was obtained, then we classify the region in the province of Central Java based on variables that affect poverty.

4. Results

We obtained global regression model for poverty percentage by using the least square method as follow.

$$Y = 3.736 + 0.084X_1 - 11.622X_2 + 0.120X_3 + 0.090X_4$$

(20)

From the model, it can be seen that there are four independent variables that influence significantly. Model (20) yields a coefficient of determination $R^2 = 67.85\%$.

The normality and multicollinearity assumptions in this model have been fulfilled. From the test of spatial heterogeneity, we obtained BP value which is 11.438 which is greater than the value of $X^2_{(0.1; 4)} = 7.779$. this means there is spatial heterogeneity. For that, we use geographically weighted regression to determine the model. The geographically weighted regression model is formed based on four independent variables that have been obtained from the global regression model i.e. $(X_1)$, $(X_2)$, $(X_3)$, $(X_4)$. In determining the estimation value of GWR model, parameters depend on the weighting of each location. For that, it needs to determine first optimum.

Table 1 presents the bandwidth values with the bisquare and tricube kernel weighted function.

| Table 1. Bandwidth Values and Minimum CVs. |
|------------------------------------------|
| Bandwidth | CV         |
| Bisquare  | 2.794016  | 291.1594  |
| Tricube   | 2.794004  | 290.0084  |

Then the bandwidth value is used to determine the weighting matrix of each GWR. Based on the equation $W(u_i, v_i) = diag[w_{i1}, w_{i2}, ..., w_{in}]$ we have a weighted value for each location. The weighted value is then used to determine the estimated parameters of the GWR model.

From geographically weighted regression model of Central Java Province it can be arranged classification of area in Central Java Province based on factors that affect poverty as shown in Fig. 1 and Fig. 2.
Figure 1. The classification of regions using GWR bisquare.
Figure 2. The classification of regions using GWR tricube.

Fig. 1 shows classification of all regions in Central Java Province based on geographically weighted regression model with bisquare kernel weighted function while Figure 2 shows classification of the regions based on geographically weighted regression model with tricube kernel weighted function. The difference of color indicates the independent variables that affect the percentage of poverty in the area. From Figure 1 and Figure 2, it can be noticed that bisquare model and bisquare model provide different classification for Banjarnegara district and Kudus district. From Table 2, it shows that the determinant coefficient of bisquare model is almost same as tricube model.

5. Conclusion

Based on the discussion, it can be stated that the percentage of poverty in Central Java can be modelled using geographically weighted regression model using bisquare and tricube weighting functions. Based on the model obtained, regions can be classified according to the level of poverty. This classification will assist the government on their effort of decreasing the level of poverty in Central Java Province. Furthermore, this attempt of modelling provides significant step on application of bisquare and tricube kernels function on spatial regression.

References

[1] Anselin, L., Spatial Econometrics : Methods and Models, Kluwer Academic Publishers, Dordrecht, 1998.
[2] Badan Pusat Statistik, Provinsi jawa tengah dalam angka tahun 2016, Jakarta, 2016.
[3] Chaudry, A., and C. Wimer., Poverty is Not Just an Indicator: The Relationship between Income, Poverty, and Child Well-Being, Academic Pediatrics 16 (2016), no. 3, S23–S29.
[4] Fotheringham, S., C. Brunsdon, and M. Charlton, Geographically Weighted Regression - The Analysis of Spatially Varying Relationships, Wiley, Chichester, 2002.
[5] Gujarati, N. D., Essential of Econometrics, McGraw-Hill, Inc, New York, 2006.
[6] Guting, R.H., An Introduction to Spatial Database Systems, The International Journal on Very Large Data Bases 3 (1994), no. 4, 357–399.
[7] Kutner, M.H., C.J. Nachtsheim., and J. Neter., Applied Linear Regression Models 4th ed, McGraw-Hill Companies, Inc., New York, 2004.
[8] Nugroho, H., Kemiskinan, Ketimpangan, dan Kesenjangan, Aditya Media, Yogyakarta, 1995.
[9] Sinaga, K.P., Poverty Data Modeling in North Sumatera Province Using Geographically Weighted Regression (GWR) Method, International Journal of Science and Research Vol. 4 (2015), no. 2, 1738–1742.
[10] Slamat, I., Nugroho, N.F.T.A., and Muslich., Geographically Weighted Regression on Poverty Indicator, Paper in publication process in IOP AD-INTERCOMME.