International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 5, Issue 3, 2019 Special Edition: *Science, Applied Science, Teaching and Education*

A Goodness of Fit Test of Geographically Weighted Polynomial Regression Models and Its Application on Life Expectancy Modelling

Toha Saifudin*[1], Fatmawati[2], Nur Chamidah[3]

Department of Mathematics, Faculty of Science and Technology, Airlangga University

Email: tohasaifudin@fst.unair.ac.id; fatmawati@fst.unair.ac.id; nur-c@fst.unair.ac.id

Abstract

Geographically weighted polynomial regression (GWPolR) is a spatial model with varying coefficients and polynomial relationships between response and its predictors. It is a generalisation of geographically weighted regression (GWR) models. By this generalisation, it has more parameters and better goodness of fit measures than the GWR does. Nevertheless, it is important to decide statistically whether the GWPolR model describes a given data set significantly better than a GWR model does. So, to carry out the work this paper aims to derive an ANOVA type test statistic and provide a guideline for performing the test in practice. Then, two simulated data sets were used to evaluate test performance. Those examples have shown that the test procedure has performed well and has provided a feasible way to choose an appropriate model for a given data set. In Human Development Index modelling, the GWPolR model was not significantly better than GWR model.

Keywords: Geographically weighted polynomial regression, Goodness of fit test, Human Development Index

Introduction

The ordinary linear regression (OLR) model has been one of the useful methods in analysing the relationships among variables. However, its uniformity assumption over observations may be unrealistic in spatial data sets (Fotheringham et al., 1996; Fotheringham, 1997). Some approaches have been proposed. One of them is the Geographically Weighted Regression (GWR) model (Brunsdon et al., 1996; Fotheringham et al., 1997). In the GWR model, the parameters are assumed to be functions of the locations.

The GWR model has been one of the useful methods in spatial analysis (Fotheringham et al., 2002) and many authors have studied the scope of its theory (Brunsdon et al., 1999; Fotheringham et al., 1998 and 2002). In application, it has been also widely applied to different areas, for example: in climatology (Al-Ahmadi & Al-Ahmadi, 2013; Brunsdon et al., 2001; and Wang et al., 2012), in econometrics (Mittal et al., 2004; Lu et al., 2014), and in the social field (Fotheringham et al., 2001; Han & Gorman, 2013). The GWR model is robust from multicollinearity (Fotheringham and Oshan, 2016).

The GWR model is an extension of the OLR model. Even though the GWR coefficients are spatially varying, the response variable in each location is modeled as a linear function of a set of explanatory variables. However, not all explanatory variables have a linear relationship with the response. Non-linearity in the relationships of variables commonly exists in many real-life situations. In spatial research, some of them are suspected to need non-linear relationships (Chamidah et al., 2014; Chiang et al., 2015). As the nonlinear relationships are present in the real situation, the model based on the linear approach may be unrealistic. Therefore, some approach models which accommodate the real data pattern are required to improve the basic GWR model.

To overcome the problem, a generalisation of the GWR model using a polynomial function approach has proposed (Saifudin et al., 2017; 2018; 2019). The model was called geographically weighted polynomial regression (GWPolR). In those studies, the GWPolR model was compared with the GWR model through a sample data set based on residual sum of squares (RSS) and determination coefficients (R^2). Based on the sample used, the GWPolR model yielded better goodness of fit indicators than the GWR model did. However, it could not be ascertained statistically whether GWPolR was significantly better than GWR in that case. Thus, for regression problems, a goodness of fit test is needed (Saifudin et al., 2018; 2019). It was an open problem to follow up.

As a generalisation of the GWR model, the GWPolR model has larger number of parameters than the GWR model does. The models with more parameters commonly have higher goodness of fit indicator values. Conversely, models with fewer parameters have greater ease in use and interpretation. As the improvement of the GWPolR model is significant, the model should be selected to use. On the other hand, the GWR is still reasonable to use when the improvement is not significant. So, we need to make sure the GWPolR model describes a data set significantly better than the basic GWR model does. It seems that there has not been a formal way to do this work. Therefore, this paper aims to derive a goodness of fit test and provide a guideline for performing the test in practice. Furthermore, we evaluate the performance of the test procedure based on some simulated data sets.

Research Method

The GWR model has explored in the form of

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{p} \beta_j(u_i, v_i)x_{ij} + \varepsilon_i, \; i = 1, 2, \dots, n, \tag{1}$$

where $\beta_j(u_i, v_i), j = 0, 1, 2, \dots, p$ are unknown parameters at location (u_i, v_i), and ε_i is normally distributed error with a zero mean and variance σ^2 for all $i = 1, 2, \dots, n$ (Brunsdon et al.,1996 and 1999; Fotheringham et al., 1998 and 2002). The weighted least square (WLS) estimator for the GWR coefficients at location (u_i, v_i) is

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = [\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_i, v_i)\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_i, v_i)\boldsymbol{y}, \tag{2}$$

where

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \text{ and } \mathbf{W}(u_i, v_i) =$$

$$\text{diag}[K_h(d_{i1}), K_h(d_{i2}), \dots, K_h(d_{in})]. \tag{3}$$

Furthermore, $K_h(\cdot) = K\left(\frac{\cdot}{h}\right)$ with $K(\cdot)$ is a kernel function, h is the bandwidth, and d_{ij} is the distance between location (u_i, v_i) and (u_j, v_j) (Fotheringham et al., 2002). Then, the RSS of GWR model is

$$\text{RSS}_{gwr} = \hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y}^{\mathrm{T}}(\mathbf{I} - \mathbf{L})^{\mathrm{T}}(\mathbf{I} - \mathbf{L})\boldsymbol{y}\,, \tag{4}$$

where

$$\mathbf{L} = \begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}}[\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_1, v_1)\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_1, v_1) \\ \boldsymbol{x}_2^{\mathrm{T}}[\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_2, v_2)\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_2, v_2) \\ \vdots \\ \boldsymbol{x}_n^{\mathrm{T}}[\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_n, v_n)\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_n, v_n) \end{bmatrix} \tag{5}$$

is called a hat matrix, $\mathbf{I}$ is an identity matrix of order *n*, and $\boldsymbol{x}_i^{\mathrm{T}} = (1, x_{i1}, x_{i2}, \dots, x_{ip})$ is the i^{th}-row of the matrix $\mathbf{X}$ (Fotheringham et al., 2002).

A generalisation of model (1) has proposed, namely GWPolR model in the form of

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p}\sum_{j=1}^{d_k}\beta_{k,j}(u_i, v_i){x_{ik}}^{j} + \varepsilon_i. \tag{6}$$

The WLS estimator for the GWPolR model at a given location (u_i, v_i) can be expressed as

$$\hat{\boldsymbol{\beta}}_{pol}(u_i, v_i) = \left[\boldsymbol{X}_{pol}^{\mathrm{T}}\boldsymbol{W}(u_i, v_i)\boldsymbol{X}_{pol}\right]^{-1}\boldsymbol{X}_{pol}^{\mathrm{T}}\boldsymbol{W}(u_i, v_i)\boldsymbol{y} \tag{7}$$

where

$$\boldsymbol{X}_{pol} = \begin{bmatrix} 1 x_{11} x_{11}^2 & \cdots & x_{11}^{d_1} & \cdots & x_{1p} x_{1p}^2 & \cdots & x_{1p}^{d_p} \\ 1 x_{21} x_{21}^2 & \cdots & x_{21}^{d_1} & \cdots & x_{2p} x_{2p}^2 & \cdots & x_{2p}^{d_p} \\ & & & \vdots & & & \\ 1 x_{n1} x_{n1}^2 & \cdots & x_{n1}^{d_1} & \cdots & x_{np} x_{np}^2 & \cdots & x_{np}^{d_p} \end{bmatrix}, \tag{8}$$

and $\boldsymbol{W}(u_i, v_i)$ and $\boldsymbol{y}$ are defined as in equation(3) (Saifudin et al., 2018; 2019). Then, the RSS of GWPolR model is

$$\mathrm{RSS}_{Pol} = \sum_{i=1}^{n} \hat{\varepsilon}_i^{*2} = \hat{\boldsymbol{\varepsilon}}_{Pol}^{\mathrm{T}} \hat{\boldsymbol{\varepsilon}}_{Pol} = \boldsymbol{y}^{\mathrm{T}}(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\boldsymbol{y}. \quad (9)$$

where

$$\mathbf{G} = \begin{bmatrix} \boldsymbol{x}_1^{*\mathrm{T}}\left(\mathbf{X}_{Pol}^{\mathrm{T}}\mathbf{W}(u_1,v_1)\mathbf{X}_{Pol}\right)^{-1}\mathbf{X}_{Pol}^{\mathrm{T}}\mathbf{W}(u_1,v_1) \\ \boldsymbol{x}_2^{*\mathrm{T}}\left(\mathbf{X}_{Pol}^{\mathrm{T}}\mathbf{W}(u_2,v_2)\mathbf{X}_{Pol}\right)^{-1}\mathbf{X}_{Pol}^{\mathrm{T}}\mathbf{W}(u_2,v_2) \\ \vdots \\ \boldsymbol{x}_n^{*\mathrm{T}}\left(\mathbf{X}_{Pol}^{\mathrm{T}}\mathbf{W}(u_n,v_n)\mathbf{X}_{Pol}\right)^{-1}\mathbf{X}_{Pol}^{\mathrm{T}}\mathbf{W}(u_n,v_n) \end{bmatrix} \quad (10)$$

is an $n \times n$ *hat matrix* of the GWPolR model, $\mathbf{I}$ is an identity matrix of order *n*, and $\boldsymbol{x}_i^{*\mathrm{T}}$ is the i^{th}-row of the matrix $\boldsymbol{X}_{pol}$[18, 19].

In this research, we will construct a goodness of fit test of the GWPolR model. This test evaluates the improvement of GWPolR from GWR. Suppose that RSS_{gwr} and RSS_{Pol} are RSS of GWR and GWPolR model, respectively. Then, the improvement of GWPolR from GWR is notated by $\Delta\mathrm{RSS} = \mathrm{RSS}_{gwr} - \mathrm{RSS}_{Pol}$. Here, a test statistic will be constructed by comparing the $\Delta\mathrm{RSS}$ with the RSS of initial model, i.e., RSS_{gwr}. To conclude whether the improvement is significant or not, the distribution of the test statistic will be searched. Furthermore, the performance of the test procedure will be evaluated by using some simulated datasets based on the test guidelines.

Results and Discussion

A Goodness of Fit Test Statistic

We assume that the following two assumptions hold on the GWR and GWPolR models:

Assumption 1. The error terms $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ are distributed as a Normal distribution with zero means and constant variance σ^2.

Assumption 2. Suppose that $\hat{y}_i$ and $\hat{y}_i^*$ is the fitted value of y_i at location *i* for GWR and GWPolR models, respectively. For all $i = 1, 2, \dots, n$, $\hat{y}_i$ and $\hat{y}_i^*$ are unbiased estimates of

$E(y_i)$ based on GWR and GWPolR models, respectively, i.e., $E(\hat{y}_i) = E(y_i)$ and $E(\hat{y}_i^*) = E(y_i)$ for all *i*.

Then, we state the following hypothesis:

H$_0$: A GWPolR model is not significantly better than a basic GWR model in describing the given data set

H$_1$: A GWPolR model is significantly better than a basic GWR model in describing the given data set

To test the hypothesis, a test statistic and its approximated distribution is constructed in the following theorem.

Theorem 1. Let $\text{RSS}_{gwr} = \boldsymbol{y}^{\text{T}}(\mathbf{I}-\mathbf{L})^{\text{T}}(\mathbf{I}-\mathbf{L})\boldsymbol{y}$ be the residual sum of squares of GWR model, where $\mathbf{L}$ is the *hat matrix* of the GWR model. Let $\text{RSS}_{Pol} = \boldsymbol{y}^{\text{T}}(\mathbf{I}-\mathbf{G})^{\text{T}}(\mathbf{I}-\mathbf{G})\boldsymbol{y}$ be the residual sum of squares of GWPolR model, where $\mathbf{G}$ is the *hat matrix* of the GWPolR model. Let ΔRSS be the difference between the residual sum of squares of the GWR model and that of the GWPolR model, i.e., $\Delta\text{RSS} = \text{RSS}_{gwr} - \text{RSS}_{Pol}$, then the goodness of fit test statistic

$$F_{gof} = \frac{\Delta\text{RSS}/\varphi_1}{\text{RSS}_{gwr}/\delta_1} \tag{11}$$

is approximately distributed F with $\frac{\varphi_1^2}{\varphi_2}$ degrees of freedom in the numerator and $\frac{\delta_1^2}{\delta_2}$ degrees of freedom in the denominator, where $\varphi_i = tr(\mathbf{A}^i)$ and $\delta_i = tr\left(\left((\mathbf{I}-\mathbf{L})^{\text{T}}(\mathbf{I}-\mathbf{L})\right)^i\right)$ for $i = 1, 2$ and $\mathbf{A} = (\mathbf{I}-\mathbf{L})^{\text{T}}(\mathbf{I}-\mathbf{L}) - (\mathbf{I}-\mathbf{G})^{\text{T}}(\mathbf{I}-\mathbf{G})$.

Proof. Based on equations (4) and (9), the ΔRSS can be expressed as

$$\Delta\text{RSS} = \text{RSS}_{gwr} - \text{RSS}_{Pol} = \boldsymbol{y}^{\text{T}}\mathbf{A}\boldsymbol{y}, \tag{12}$$

where $\mathbf{A} = (\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L}) - (\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})$ is a positive semidefinite matrix since $\Delta\mathrm{RSS} \geq 0$ for any $\boldsymbol{y}$. Under GWR model and assumptions 1 and 2,we have

$$E(\hat{\boldsymbol{\varepsilon}}) = E(\boldsymbol{y}) - E(\hat{\boldsymbol{y}}) = \mathbf{0}, \text{ and } E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) = \sigma^2\mathbf{I}. \tag{13}$$

Then, RSS_{gwr} can be expressed as

$$\mathrm{RSS}_{gwr} = \left(\hat{\boldsymbol{\varepsilon}} - E(\hat{\boldsymbol{\varepsilon}})\right)^{\mathrm{T}}\left(\hat{\boldsymbol{\varepsilon}} - E(\hat{\boldsymbol{\varepsilon}})\right) = \boldsymbol{\varepsilon}^{\mathrm{T}}(\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L})\boldsymbol{\varepsilon}. \tag{14}$$

So,

$$E(\mathrm{RSS}_{gwr}) = E\left(tr(\boldsymbol{\varepsilon}^{\mathrm{T}}(\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L})\boldsymbol{\varepsilon})\right) = tr\left((\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L})E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}})\right) = \sigma^2\delta_1 \quad , \tag{15}$$

where $\delta_1 = tr\left((\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L})\right)$.

On the other hand, under GWPolR model and assumptions 1 and 2, we have

$$E(\hat{\boldsymbol{\varepsilon}}_{Pol}) = E(\boldsymbol{y}) - E(\hat{\boldsymbol{y}}_{Pol}) = \mathbf{0}, \text{ and } E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) = \sigma^2\mathbf{I}. \tag{16}$$

Then, RSS_{Pol} can also be expressed as

$$\mathrm{RSS}_{Pol} = \left(\hat{\boldsymbol{\varepsilon}}_{Pol} - E(\hat{\boldsymbol{\varepsilon}}_{Pol})\right)^{\mathrm{T}}\left(\hat{\boldsymbol{\varepsilon}}_{Pol} - E(\hat{\boldsymbol{\varepsilon}}_{Pol})\right) = \boldsymbol{\varepsilon}^{\mathrm{T}}(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\boldsymbol{\varepsilon}. \tag{17}$$

Therefore, by following equation (15) then $E(\mathrm{RSS}_{Pol}) = \sigma^2\gamma_1$ where $\gamma_1 = tr\left((\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\right)$.

According to equations (14) and (17), the $\Delta\mathrm{RSS}$ can also be elaborated as

$$\Delta\mathrm{RSS} = \mathrm{RSS}_{gwr} - \mathrm{RSS}_{Pol} = \boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{A}\,\boldsymbol{\varepsilon}. \tag{18}$$

Hence, we have

$$E(\Delta\mathrm{RSS}) = E(\boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{A}\,\boldsymbol{\varepsilon}) = E\left(tr(\boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{A}\,\boldsymbol{\varepsilon})\right) = E\left(tr(\mathbf{A}\,\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}})\right) = tr(\mathbf{A})E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}] = \sigma^2\varphi_1, \tag{19}$$

where $\varphi_1 = tr(\mathbf{A})$.

From equation (18), we know that the $\Delta\mathrm{RSS}$ can be expressed as a quadratic form of normal variable with a symmetric and positive semidefinite matrix **A**. From the distribution theory, a quadratic form of standardised normal variables, i.e., $\boldsymbol{\xi}^{\mathrm{T}}\mathbf{A}\boldsymbol{\xi}$ where $\boldsymbol{\xi} \sim N(\mathbf{0},\mathbf{I})$ and **A** is symmetric, is distributed as χ^2 distribution if and only if **A** is idempotent (Rencher & Schaalje, 2008; Hogg et al., 2013). For the random variable

$$\frac{\Delta \text{RSS}}{\sigma^2} = \left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right)^{\text{T}} \mathbf{A} \left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right), \tag{20}$$

we know that $\frac{\boldsymbol{\varepsilon}}{\sigma} \sim N(\mathbf{0}, \mathbf{I})$, but the matrix **A** is generally not idempotent due to the complexity of the weighted matrix $\mathbf{W}(u_i, v_i)$ which is different at each location (u_i, v_i). So, the quantity $\frac{\Delta \text{RSS}}{\sigma^2}$ is generally not distributed as an exact χ^2 distribution. But, there are several approximation for the distribution of the quadratic form (Yuan & Bentler, 2010). A simpler method has proposed to approximate the distribution of this quadratic form by multiplying a constant *c* with a χ^2 variable with *r* degrees of freedom, i.e., writed as $c\chi_r^2$, if the matrix **A** is symmetric and positive semidefinite (Leung et al.,2000). Then, the constant *c* and *r* are choosen in such a way so the mean and variance of $c\chi_r^2$ and those of the quadratic form $\frac{\Delta \text{RSS}}{\sigma^2}$ are made to match each other. For the random variable χ_r^2, we know that its mean and variance are *r* and 2*r*, respectively. So, the mean and variance of $c\chi_r^2$ are *cr* and $2c^2r$, respectively.

For the quadratic form $\frac{\Delta \text{RSS}}{\sigma^2}$, we know from equation (19) that its mean is φ_1. Its variance is derived by the following explanation. Since the matrix **A** is symmetric and positive semidefinite, there is an orthogonal matrix **P** of order *n* such that

$$\mathbf{P}^{\text{T}}\mathbf{A}\mathbf{P} = \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n), \tag{21}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix which have the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ of the matrix **A** in its main diagonal. Suppose that

$$\boldsymbol{\eta} = (\eta_1, \eta_2, \cdots, \eta_n)^{\text{T}} = \mathbf{P}^{\text{T}} \frac{\boldsymbol{\varepsilon}}{\sigma}. \tag{22}$$

According to the properties of multivariate normal distribution, then $\eta_1, \eta_2, \cdots, \eta_n$ are independent and identically distributed (iid) N(0,1). On the other hand, from equation (22), we have $\frac{\boldsymbol{\varepsilon}}{\sigma} = \mathbf{P}\boldsymbol{\eta}$. So, we obtain

$$\frac{\Delta \text{RSS}}{\sigma^2} = \boldsymbol{\eta}^{\text{T}}\mathbf{P}^{\text{T}}\mathbf{A}\mathbf{P}\boldsymbol{\eta} = \boldsymbol{\eta}^{\text{T}}\boldsymbol{\Lambda}\boldsymbol{\eta} = \sum_{i=1}^{n} \lambda_i \eta_i^2. \tag{23}$$

Because of the fact $\eta_i \sim iid\ \text{N}(0,1)$ then $\eta_i^2 \sim iid\ \chi^2_{(1)}$ for i =1, 2, ⋯, n. Therefore, $\text{var}(\eta_i^2) = 2$ and

$$\text{var}\left(\frac{\Delta \text{RSS}}{\sigma^2}\right) = \sum_{i=1}^{n} \lambda_i^2 \,\text{var}(\eta_i^2) = 2\sum_{i=1}^{n} \lambda_i^2. \tag{24}$$

If $\lambda_1, \lambda_2, \cdots, \lambda_n$ are eigenvalues of $\mathbf{A}$ then $\lambda_1^2, \lambda_2^2, \cdots, \lambda_n^2$ are eigenvalues of the matrix $\mathbf{A}^2$. So,

$$\text{var}\left(\frac{\Delta \text{RSS}}{\sigma^2}\right) = 2\, tr(\mathbf{A}^2) = 2\varphi_2, \tag{25}$$

where $\varphi_2 = tr(\mathbf{A}^2)$.

Based on the approach rule above, by equalising each mean and variance of $c\chi_r^2$ and $\frac{\Delta \text{RSS}}{\sigma^2}$, it can be written the following equation system

$$\begin{cases} cr = \varphi_1, \\ 2c^2 r = 2\varphi_2. \end{cases} \tag{26}$$

By solving the equation system (26), it is obtained $c = \frac{\varphi_2}{\varphi_1}$, and $r = \frac{\varphi_1^2}{\varphi_2}$. So, the distribution of $\frac{\Delta \text{RSS}}{c\sigma^2} = \frac{\varphi_1 \Delta \text{RSS}}{\varphi_2 \sigma^2}$ can be approximated by a χ_r^2 distribution with $r = \frac{\varphi_1^2}{\varphi_2}$ degrees of freedom, where $\varphi_i = tr(\boldsymbol{A}^i),\ i = 1, 2$ with $\mathbf{A} = (\mathbf{I} - \mathbf{L})^{\text{T}}(\mathbf{I} - \mathbf{L}) - (\mathbf{I} - \mathbf{G})^{\text{T}}(\mathbf{I} - \mathbf{G})$.

If a basic GWR is used to fit the data and satisfies assumptions 1 and 2, the residual sum of squares can be expressed as $\text{RSS}_{gwr} = \boldsymbol{\varepsilon}^{\text{T}}(\mathbf{I} - \mathbf{L})^{\text{T}}(\mathbf{I} - \mathbf{L})\boldsymbol{\varepsilon}$, where $\mathbf{L}$ is the hat matrix of the basic GWR model as stated in equation (5) [25]. Approximated distribution of $\frac{\delta_1 \text{RSS}_{gwr}}{\delta_2 \sigma^2}$ is a χ^2 distribution with $\frac{\delta_1^2}{\delta_2}$ degrees of freedom, where $\delta_i = tr\left(\left((\mathbf{I} - \mathbf{L})^{\text{T}}(\mathbf{I} - \mathbf{L})\right)^i\right)$ for i = 1, 2.

Let the statistic F_{gof} be defined as

$$F_{gof} = \frac{\dfrac{\varphi_1 \Delta \mathrm{RSS}}{\varphi_2 \sigma^2} \Big/ \left(\dfrac{\varphi_1^2}{\varphi_2}\right)}{\dfrac{\delta_1 \mathrm{RSS}_{gwr}}{\delta_2 \sigma^2} \Big/ \left(\dfrac{\delta_1^2}{\delta_2}\right)}. \tag{27}$$

Then, the distribution of F_{gof} may reasonably be approximated by an F-distribution with $\frac{\varphi_1^2}{\varphi_2}$ degrees of freedom in the numerator and $\frac{\delta_1^2}{\delta_2}$ degrees of freedom in the denominator. If we simplify equation (27), we obtain equation (11).■

No significant difference between GWR and GWPolR models for the given data leads to the fact that the quantity ΔRSS is close to zero. It means that the quantity F_{gof} is sufficiently small. Intuitively, a small value of F_{gof} supports the null hypothesis. Otherwise, a large value of F_{gof} indicates that the null hypothesis should be rejected. Hence, we reject the null hypothesis and conclude that the GWPolR describes a given data set significantly better than the basic GWR does if $F_{gof} > F_\alpha(\varphi_1^2/\varphi_2, \delta_1^2/\delta_2)$, where $F_\alpha(\varphi_1^2/\varphi_2, \delta_1^2/\delta_2)$ is the upper 100α percentage point of the F-distribution for a given α.

A Guideline for Performing the Test

The calculation of F_{gof} test statistic can be constructed by using Table 1. Suppose that $F_\alpha(\varphi_1^2/\varphi_2, \delta_1^2/\delta_2)$ is the upper 100α percentage point of F-distribution with degree of freedom φ_1^2/φ_2 in the numerator and δ_1^2/δ_2in the denominator for a given α. Then, we reject the null hypothesis if $F_{gof} > F_\alpha(\varphi_1^2/\varphi_2, \delta_1^2/\delta_2)$. We can also use a p-value

$$p = P\left(F_{gof} \geq f_{gof}\right), \tag{28}$$

where f_{gof} is an observed value of the test statistic F_{gof}. If the p-value is less than a given significance level α, we reject the null hypothesis. We accept it otherwise.

Table 1. An ANOVA table for performing the test

Source of Variation	Degrees of freedom	Sum of squares	Mean Squares	F_{gof}
GWPolR Residuals	γ_1	RSS_{Pol}		
GWPolR Improvement	φ_1	ΔRSS	$\frac{\Delta\text{RSS}}{\varphi_1}$	$\frac{\Delta\text{RSS}/\varphi_1}{\text{RSS}_{gwr}/\delta_1}$
GWR Residuals	δ_1	RSS_{gwr}	$\frac{\text{RSS}_{gwr}}{\delta_1}$	

Application

Here, we used three data sets. The first two data sets were simulated data sets. Each data set was used to see whether the test conclusion was suitable with the true model or not. The last, it was applied to real data for modeling life expectancy based on human development index and per capita expenditure.

The First Simulated Data Set

For the first example, we generated a data set according to the following GWR model

$$y_i = \beta_1(u_i, v_i) + \beta_2(u_i, v_i)x_i + \varepsilon_i. \tag{29}$$

Here, we used a sample size of 12. The spatial locations were randomly located on a cartesian coordinate system with random points in the form of (u_i, v_i), $i = 1, 2, \dots, 12$. An example of generated data sets is listed in Table 2. Its scatter plot tends to follow linear trend (Figure 1).

Table 2. The first data set

Number of observation	y	x	u	v
1	15.93	1.90	4.76	0.68
2	19.97	3.70	4.06	0.92
3	10.60	2.48	0.57	1.86
4	10.38	1.34	1.41	1.35
5	15.92	1.55	5.49	4.10
6	20.68	4.12	5.43	1.55
7	12.88	0.39	4.51	4.00
8	15.05	3.15	1.07	3.23
9	24.69	4.01	3.08	2.11
10	8.76	0.60	1.81	5.43
11	16.10	1.96	4.26	0.87
12	14.98	0.78	5.97	1.49

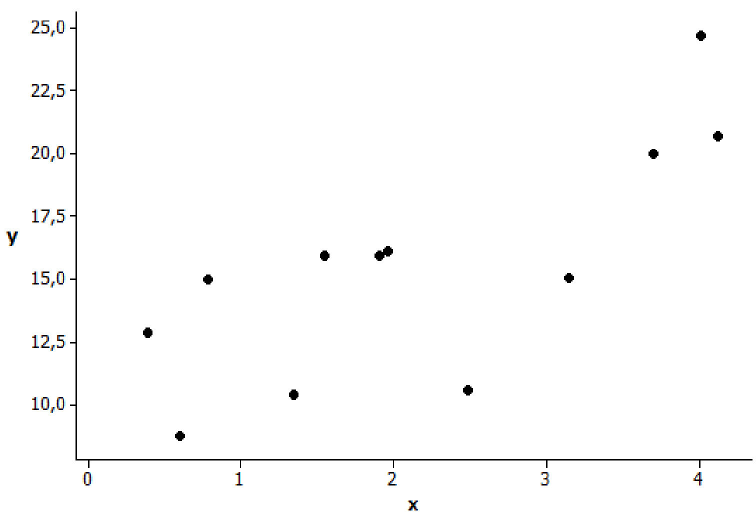


Figure 1. Scatter plot of the first data

The data set was firstly modeled by using equation (29). Then, it was also modeled by the following GWPolR model

$$y_i = \beta_1(u_i, v_i) + \beta_{2,1}(u_i, v_i)x_i + \beta_{2,2}(u_i, v_i)x_i^2 + \varepsilon_i. \qquad (30)$$

Based on Cross Validation with Gaussian kernel [5] we found that the optimal bandwidth for GWR and GWPolR were 1.632766 and 1.270955 units, respectively. The performance indicators for both models are presented in Table 3.

Table 3. Performance indicators for the first example

Model	RSS	R^2
GWR	21.30691	90.93%
GWPolR	2.838471	98.79%

From Table 3, the performance indicators of GWPolR are better than those of GWR. It seems that the GWPolR gives improvement from GWR. However, we have not known whether the improvement is statistically significant or not. Hence, the goodness of fit test procedure described above was conducted. Its results are presented in Table 4. For this sample, we found φ_1^2/φ_2= 5.36923 and $\delta_1^2/\delta_2 = 6.94807$. By using significance level of 0.05, the value of $F_{0.05}$(5.36923, 6.94807) was 3.94638. Furthermore, the *p*-value of this test was 0.29928. So, we can not reject the null hypothesis. Here, the GWPolR model is not significantly better than GWR model in describing the given data set. It means that the conclusion is according to the data condition which is generated by GWR model.

Table 4. An ANOVA table for performing the test on the first example

Source of Variation	Degrees of freedom	Sum of squares	Mean Squares	F_{gof}
GWPolR Residuals	2.21064	2.83847		
GWPolR Improvement	2.96974	18.46844	6.21887	
GWR Residuals	5.18038	21.30691	4.11300	1.512

The Second Simulated Data Set

Here, a data set based on the GWPolR model in equation (30) was generated. The sample size and the determination of spatial coordinates are similar to those in the first data set. The data set is listed in Table 5. Its scatter plot tends to follow nonlinear trend (Figure 2). Then, the data set is modeled by using both models (29) and (30). Based on the CV criterion with Gaussian kernel weighting function, we found that the optimal bandwidth for GWR and GWPolR were 0.9156273 and 1.100645, respectively. The goodness of fit indicators for both models are presented in Table 6.

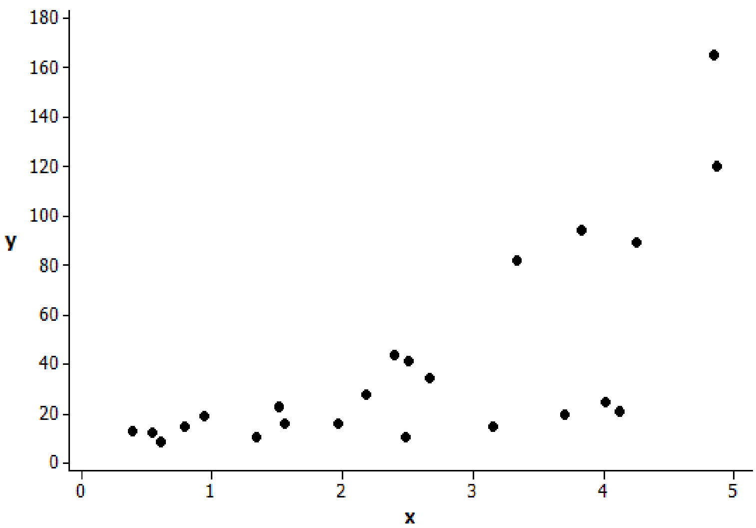


Figure 2. Scatter plot of the second data

Table 5. The second data set

Number of observation	y	x	u	v
1	34.51	2.67	2.12	1.24
2	89.42	4.25	3.38	1.76
3	165.35	4.85	5.68	4.67
4	22.50	1.51	2.02	3.88
5	41.32	2.50	3.55	1.64
6	81.95	3.34	5.61	3.72
7	94.39	3.83	5.51	2.59
8	43.73	2.40	4.75	1.97
9	12.55	0.54	3.03	4.48
10	19.35	0.94	4.59	4.47
11	27.74	2.18	1.65	2.02
12	120.29	4.87	3.59	2.19

From Table 6, the performance indicators of GWPolR model are better than those of GWR model. It means that the GWPolR gives goodness improvement from GWR model. However, we have not known whether the improvement is statistically significant or not. Hence, the goodness of fit test procedure described above was performed and presented in Table 7. For this sample, we found $\varphi_1^2/\varphi_2 = 0.01525$and $\delta_1^2/\delta_2 = 5.52262$. For a significance level of 0.05, the value of $F_{0.05}(0.01525, 5.52262)$ was 0.10741. The *p*-value of this test was 0.00896. Hence, we reject the null hypothesis and conclude that GWPolR model is significantly better than GWR model in describing the given data set. This conclusion is according to the true data set which is generated from GWPolR model.

Table 6. Performance indicators for the second example

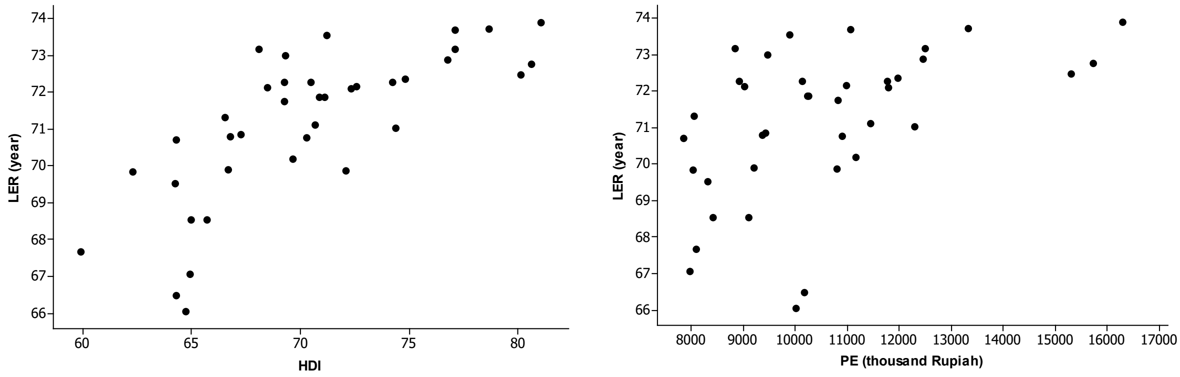
Model	RSS	R^2
GWR	259.16520	98.95%
GWPolR	42.39748	99.83%

Table 7. An ANOVA table for performing the test on the second example

Source of Variation	Degrees of freedom	Sum of squares	Mean Squares	F_{gof}
GWPolR Residuals	3.27040	42.39748		
GWPolR Improvement	0.07975	216.76772	2,718.09053	
GWR Residuals	3.35015	259.16520	77.35928	35.135

The Real Data: Life Expectancy Data

Life Expectancy data in this research were obtained from The Statistics of East Java, Indonesia. The data involved 38 observation units consisting of 29 districts and 9 cities in East Java in 2017. The observed attributes of each district or city are Life Expectancy Rate (LER) in years, Human Development Index (HDI) without units, and Percapita Expenditure (PE) in thousands of Rupiah. In this study, the dependent variable is LER. Whereas, the independent variables are HDI and PE. The trend of relationships between LER and each independent variable can be seen on Figure 3. It seems that there are nonlinear trends.



(a) (b)

Figure 3. Scatter plot of LER versus (a) HDI and (b) PE

Table 8. Performance indicators for the Life Expectancy data

Model	RSS	R^2
GWR	30.3514	80.31%
GWPolR	21.7375	85.90%

Based on the CV criterion with Gaussian kernel weighting function, we found that the optimal bandwidth for GWR and GWPolR were 0.8702024 and 0.7367324, respectively. The goodness of fit indicators for both models are presented in Table 8. From Table 8, the GWPolR model gave better performance than GWR model. In addition, the GWPolR model has reduced the RSS value by 8.6139 from GWR. Also, it increases R^2 by 5.59% from GWR. Furthermore, the goodness of fit test procedure described above was performed and presented in Table 9. For this sample, we found $\varphi_1^2/\varphi_2 = 3.21623$ and $\delta_1^2/\delta_2 = 4.51826$. By using a significance level of 0.05, the value of $F_{0.05}(3.21623, 4.51826)$ was 5.82602. Furthermore, the *p*-value of this test was 0.07538. Hence, we couldn't reject the null hypothesis and conclude that the GWPolR model was not significantly better than the GWR model in describing the real data.

Table 9. An ANOVA table for performing the test on the Life Expectancy data

Source of Variation	Degrees of freedom	Sum of squares	Mean Squares	F_{gof}
GWPolR Residuals	4.51465	21.7375		
GWPolR Improvement	0.29890	8.6139	28.81867	
GWR Residuals	4.81355	30.3514	6.30541	4.57047

Conclusion

A test statistic of an ANOVA type can be built on the residual sum of squares of the models. The test statistic approximately follows *F*-distribution. Based on the generated data sets, the goodness of fit test procedure empirically performs well for testing the related models. In other words, the test can correctly support the true models.

Acknowledgments

The authors thank to Airlangga University for the financial support of this publication. Then the authors thank to Statistics of East Java, Indonesia for kindly providing the data. The authors also give high appreciation for all people who contributed to the completion of this paper.

References

Al-Ahmadi, K., and Al-Ahmadi, S. (2013). Rainfall-Altidude Relationship in Saudi Arabia. *Advances in Meteorology*. doi: 10.1155/2013/363029 (accesed: January 25th, 2018).

Brunsdon, C., Fotheringham A. S., and Charlton, M. (1996). Geographically Weighted Regression: AMethod For Exploring Spatial Nonstationarity. *Geographical Analysis,* **28**(4): 281–298.

Brunsdon, C., Fotheringham A. S., and Charlton, M. (1999). Some Notes on Parametric Significance Tests for Geographically Weighted Regression. *Journal of Regional Science*, **38**(3): 497–524.

Brunsdon, C., McClatchey, J., and Unwin, D. J. (2001). Spatial Variations In The Average Rainfall–Altitude Relationship In Great Britain: An Approach Using Geographically Weighted Regression. *International Journal of Climatology*, **21**: 455–466.

Chamidah, N., Saifuddin, T., and Rifada, M. (2014). The Vulnerability Modeling of Dengue Hemorrhagic Fever Disease in Surabaya Based on Spatial Logistic Regression Approach. *Applied Mathematical Sciences,* **8**(28): 1369 – 1379.

Chiang, Y-H., Peng, T-C., and Chang, C-O. (2015). The nonlinear effect of convenience stores on residential property prices: A case study of Taipei, Taiwan. *Habitat International*, **46**: 82–90.

Fotheringham, A. S. (1997). Trends in quantitative methods I: stressing the local. *Progress in Human Geography*, **21**: 88-96.

Fotheringham, A. S., Charlton, M., and Brunsdon, C. (1996). The geography of parameter space: an investigation into spatial non-stationarity. *International Journal of Geographical Information Systems*, **10**: 605 – 627.

Fotheringham, A. S., Charlton, M., and Brunsdon, C. (1997). Two techniques for exploring non-stationarity in geographical data. *Geographical Systems*, **4**: 59-82.

Fotheringham, A. S., Charlton, M., and Brunsdon, C. (1998). Geographically Weighted Regression: A natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, **30**: 1905–1927.

Fotheringham, A. S., Charlton, M., and Brunsdon, C. (2001). Spatial Variation in School Performance: a Local Analysis Using Geographically Weighted Regression. *Geographical & Environmental Modelling,* **5**(1): 43–66.

Fotheringham, A. S., Brunsdon, C., and Chartlon, M. (2002). *Geographically Weighted Regression: The Analysis Of Spatially Varying Relationships*. USA: John Wiley & Sons.

Han, D., and Gorman, D. M. (2013). Exploring Spatial Associations Between On-Sale Alcohol Availability, Neighborhood Population Characteristic, and Violent Crime in a Geographically Isolated City. *Journal of Addiction*. doi: 10.1155/2013/356152 (accessed May 20th, 2018).

Hogg, R.V., McKean, J.W., & Craig, A.T. 2013. *Introduction to Mathematical Statistics*, Seventh Edition. Pearson Education, USA.

Leung, Y., Mei, C. L., and Zhang, W. X. (2000). Statistical Tests for Spatial Nonstationarity based on The Geographically Weighted Regression Model. *Environment and Planning A*, **32**: 9–32.

Lu, B., Charlton, M., Harris, P., and Fotheringham, A. S. (2014). Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. *International journal of Geographical Information Science*. doi: 10.1080/13658816.2013.865739 (accessed: May 30th, 2018).

Mittal, V., Kamakura, W. A., and Govind, R. (2004). Geographic Pattern in Customer Service and Satisfaction: An Empirical Investigation. *Journal of Marketing*, **68**: 48–62.

Nakaya T, Fotheringham AS, Charlton, M., & Brunsdon C, 2005, Geographically weighted Poisson regression for disease associative mapping, *Statistics in Medicine*, 24, 2695-2717.

Rencher, A. C., and Schaalje, G. G. (2008). *Linear Models in Statistics* (2nd ed.). New York, USA: John Willey & Sons.

Saifudin, T., Fatmawati, and Chamidah, N. (2017). Perluasan Geographically Weighted Regression Menggunakan Fungsi Polinomial (Expansion of Geographically Weighted Regression Using Polynomial Functions). *Prosiding Seminar Nasional Integrasi Matematika Dan Nilai Islami*, **1**(1): 15 – 20.

Saifudin, T., Fatmawati, & Chamidah, N. (2018). Geographically Weighted Polynomial Regression: Application to Poverty Modeling in East Java Province, Indonesia. *International Journal of Academic and Applied Research*, **2**(11): 4 – 11.

Saifudin, T., Suliyanto, and Ana, E. Development of Geographically Weighted Regression Using Polynomial Function Approach and Its Application on Life Expectancy Data. *International Journal of Innovation, Creativity and Change*. 2019; 5(3): 271 – 289.

Wang, C., Zhang, J., and Yan, X. (2012). The Use of Geographically Weighted Regression for the Relationship among Extreme Climate Indices in China. *Mathematical Problems in Engineering*. doi: 10.1155/2012/369539 (accessed: May 25th, 2018).

Yuan, K. H., and Bentler, P. M. (2010). Two Simple Approximations to the Distribution of Quadratic Forms. *British Journal of Mathematical and Statistical Psychology*, **63**(2): 273–291.