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

Toha Saifudin*1, Fatmawati2, Nur Chamidah3

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 (GWPoIIR). In those studies, the GWPoIIR 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 GWPoIIR 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

\[ y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{p} \beta_j(u_i, v_i)x_{ij} + \varepsilon_i, \quad i = 1, 2, \ldots, n, \]  

where \( \beta_j(u_i, v_i), j = 0, 1, 2, \ldots, p \) are unknown parameters at location \((u_i, v_i)\), and \( \varepsilon_i \) is normally distributed error with a zero mean and variance \( \sigma^2 \) for all \( i = 1, 2, \ldots, n \) (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

\[ \hat{\beta}(u_i, v_i) = [X^T W(u_i, v_i) X]^{-1} X^T W(u_i, v_i) y, \]  

where
\[
X = \begin{bmatrix}
1 & x_{11} & \ldots & x_{1p} \\
1 & x_{21} & \ldots & x_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
1 & x_{n1} & \ldots & x_{np}
\end{bmatrix},
\quad
y = \begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix},
\quad
\text{and } W(u_i, v_i) = \text{diag}[K_h(d_{i1}), K_h(d_{i2}), \ldots, K_h(d_{im})].
\]

Furthermore, \(K_h(\cdot) = K(\cdot) / h\) 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

\[
\text{RSS}_{gwr} = \hat{e}^T \hat{e} = y^T (I - L)^T (I - L) y,
\]

where

\[
L = \begin{bmatrix}
x_1^T [X^T W(u_1, v_1) X]^{-1} X^T W(u_1, v_1) \\
x_2^T [X^T W(u_2, v_2) X]^{-1} X^T W(u_2, v_2) \\
\vdots \\
x_n^T [X^T W(u_n, v_n) X]^{-1} X^T W(u_n, v_n)
\end{bmatrix}
\]

is called a hat matrix, \(I\) is an identity matrix of order \(n\), and \(x_i^T = (1, x_{i1}, x_{i2}, \ldots, x_{ip})\) is the \(i^{th}\)-row of the matrix \(X\) (Fotheringham et al., 2002).

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

\[
y_i = \beta_0(u_i, v_i) + \sum_{k=1}^P \sum_{j=1}^{d_k} \beta_{k,j}(u_i, v_i) x_{ik}^j + \epsilon_i.
\]

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

\[
\hat{\beta}_{pol}(u_i, v_i) = \left[X_{pol}^T W(u_i, v_i) X_{pol}\right]^{-1} X_{pol}^T W(u_i, v_i) y
\]

where

\[
X_{pol} = \begin{bmatrix}
1 x_{11}^2 & \ldots & x_{11}^{d_1} & \ldots & x_{1p} x_{1p}^2 & \ldots & x_{1p}^{d_p} \\
1 x_{21}^2 & \ldots & x_{21}^{d_1} & \ldots & x_{2p} x_{2p}^2 & \ldots & x_{2p}^{d_p} \\
\vdots \\
1 x_{n1}^2 & \ldots & x_{n1}^{d_1} & \ldots & x_{np} x_{np}^2 & \ldots & x_{np}^{d_p}
\end{bmatrix},
\]

and \(W(u_i, v_i)\) and \(y\) are defined as in equation (3) (Saifudin et al., 2018; 2019). Then, the RSS of GWPolR model is
RSS_{Pol} = \sum_{i=1}^{n} \hat{\epsilon}_{i}^2 = \hat{\mathbf{e}}_{Pol}^T \hat{\mathbf{e}}_{Pol} = \mathbf{y}^T (\mathbf{I} - \mathbf{G})^T (\mathbf{I} - \mathbf{G}) \mathbf{y}. \quad (9)

where

\[
\mathbf{G} = \begin{bmatrix}
    \mathbf{x}_1^T (\mathbf{X}_{Pol}^T \mathbf{W}(u_1, v_1) \mathbf{X}_{Pol})^{-1} \mathbf{X}_{Pol}^T \mathbf{W}(u_1, v_1) \\
    \mathbf{x}_2^T (\mathbf{X}_{Pol}^T \mathbf{W}(u_2, v_2) \mathbf{X}_{Pol})^{-1} \mathbf{X}_{Pol}^T \mathbf{W}(u_2, v_2) \\
    \vdots \\
    \mathbf{x}_n^T (\mathbf{X}_{Pol}^T \mathbf{W}(u_n, v_n) \mathbf{X}_{Pol})^{-1} \mathbf{X}_{Pol}^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 \(\mathbf{x}_i^T\) is the \(i^{th}\)-row of the matrix \(\mathbf{X}_{Pol}\).[{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 \(\Delta\text{RSS} = \text{RSS}_{gwr} - \text{RSS}_{Pol}\). Here, a test statistic will be constructed by comparing the \(\Delta\text{RSS}\) with the RSS of initial model, i.e., \(\text{RSS}_{gwr}\). 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 \(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\) are distributed as a Normal distribution with zero means and constant variance \(\sigma^2\).

**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 \(i = 1, 2, \ldots, n\), \(\hat{y}_i\) 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 $RSS_{gwr} = y^T(I - L)^T(I - L)y$ be the residual sum of squares of GWR model, where $L$ is the hat matrix of the GWR model. Let $RSS_{pol} = y^T(I - G)^T(I - G)y$ be the residual sum of squares of GWPolR model, where $G$ is the hat matrix of the GWPolR model. Let $\Delta RSS$ be the difference between the residual sum of squares of the GWR model and that of the GWPolR model, i.e., $\Delta RSS = RSS_{gwr} - RSS_{pol}$, then the goodness of fit test statistic

$$F_{gof} = \frac{\Delta RSS/\varphi_1}{RSS_{gwr}/\delta_1}$$

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

and $A = (I - L)^T(I - L) - (I - G)^T(I - G)$.

**Proof.** Based on equations (4) and (9), the $\Delta RSS$ can be expressed as

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

\[
E(\hat{\varepsilon}) = E(y) - E(\hat{y}) = 0, \text{ and } E(\varepsilon \varepsilon^T) = \sigma^2 I.
\]

Then, RSS\(_{gwr} \) can be expressed as

\[
RSS_{gwr} = (\hat{\varepsilon} - E(\hat{\varepsilon}))^T(\hat{\varepsilon} - E(\hat{\varepsilon})) = \varepsilon^T(I - L)^T(I - L)\varepsilon. \quad (14)
\]

So,

\[
E(RSS_{gwr}) = E\left(tr(\varepsilon^T(I - L)^T(I - L)\varepsilon)\right) = tr\left((I - L)^T(I - L)E(\varepsilon \varepsilon^T)\right) = \sigma^2 \delta_1 ,
\]

where \( \delta_1 = tr\left((I - L)^T(I - L)\right) \).

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

\[
E(\hat{\varepsilon}_{Pol}) = E(y) - E(\hat{y}_{Pol}) = 0, \text{ and } E(\varepsilon \varepsilon^T) = \sigma^2 I.
\]

Then, RSS\(_{pol} \) can also be expressed as

\[
RSS_{pol} = (\hat{\varepsilon}_{pol} - E(\hat{\varepsilon}_{pol}))^T(\hat{\varepsilon}_{pol} - E(\hat{\varepsilon}_{pol})) = \varepsilon^T(I - G)^T(I - G)\varepsilon. \quad (17)
\]

Therefore, by following equation \( 15 \) then \( E(RSS_{pol}) = \sigma^2 \gamma_1 \) where \( \gamma_1 = tr\left((I - G)^T(I - G)\right) \).

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

\[
\Delta RSS = RSS_{gwr} - RSS_{pol} = \varepsilon^T A \varepsilon. \quad (18)
\]

Hence, we have

\[
E(\Delta RSS) = E(\varepsilon^T A \varepsilon) = E\left(tr(\varepsilon^T A \varepsilon)\right) = E\left(tr(A \varepsilon \varepsilon^T)\right) = tr(A)E[\varepsilon \varepsilon^T] = \sigma^2 \varphi_1 ,
\]

where \( \varphi_1 = tr(A) \).

From equation \( 18 \), we know that the \( \Delta RSS \) 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., \( \xi \sim N(0, I) \) and \( A \) is symmetric, is
distributed as $\chi^2$ distribution if and only if $A$ is idempotent (Rencher & Schaalje, 2008; Hogg et al., 2013). For the random variable

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

(20)

we know that $\frac{\varepsilon}{\sigma} \sim N(0,1)$, but the matrix $A$ is generally not idempotent due to the complexity of the weighted matrix $W(u_i, v_i)$ 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 $\chi^2$ 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 $\chi^2$ variable with $r$ degrees of freedom, i.e., writed as $c\chi^2_r$, 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 $c\chi^2_r$ and those of the quadratic form $\frac{\Delta\text{RSS}}{\sigma^2}$ are made to match each other. For the random variable $\chi^2_r$, we know that its mean and variance are $r$ and $2r$, respectively. So, the mean and variance of $c\chi^2_r$ are $cr$ and $2cr$, respectively.

For the quadratic form $\frac{\Delta\text{RSS}}{\sigma^2}$, we know from equation (19) that its mean is $\varphi_1$. 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

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

(21)

where $\Lambda$ 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

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

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

\[
\frac{\Delta RSS}{\sigma^2} = \eta^T P^T A P \eta = \eta^T A \eta = \sum_{i=1}^{n} \lambda_i \eta_i^2. \tag{23}
\]

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

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

If \( \lambda_1, \lambda_2, \ldots, \lambda_n \) are eigenvalues of \( A \) then \( \lambda_1^2, \lambda_2^2, \ldots, \lambda_n^2 \) are eigenvalues of the matrix \( A^2 \). So,

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

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

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

\[
\begin{align*}
\frac{cr}{\sigma^2} &= \varphi_1, \\
2c^2r &= 2\varphi_2.
\end{align*} \tag{26}
\]

By solving the equation system (26), it is obtained \( e = \frac{\varphi_2}{\varphi_1} \), and \( r = \frac{\varphi_1^2}{\varphi_2} \). So, the distribution of \( \frac{\Delta RSS}{\sigma^2} = \frac{\varphi_2 \Delta RSS}{\varphi_1 \sigma^2} \) can be approximated by a \( \chi^2_r \) distribution with \( r = \frac{\varphi_1^2}{\varphi_2} \) degrees of freedom, where

\[
\varphi_i = \text{tr}(A^i), \quad i = 1, 2 \text{ with } A = (I - L)^T (I - L) - (I - G)^T (I - G).
\]

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 \( \text{RSS}_{gwr} = \varepsilon^T (I - L)^T (I - L) \varepsilon \), where \( L \) 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 \( \chi^2 \) distribution with \( \delta_2 \) degrees of freedom, where \( \delta_i = \text{tr}\left(\left((I - L)^T (I - L)\right)^i\right) \) for \( i = 1, 2. \)
Let the statistic $F_{gof}$ be defined as

$$
F_{gof} = \frac{\frac{\varphi_1^2 \Delta RSS}{\varphi_2 \sigma^2}}{\left(\frac{\varphi_1^2}{\delta_1^2} + \frac{\varphi_2^2}{\delta_2^2}\right)}.
$$

(27)

Then, the distribution of $F_{gof}$ may reasonably be approximated by an $F$-distribution with $\varphi_1^2 \varphi_2$ degrees of freedom in the numerator and $\delta_1^2 \delta_2^2$ 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 $\Delta RSS$ 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 100$\alpha$ percentage point of the $F$-distribution for a given $\alpha$.

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 100$\alpha$ percentage point of $F$-distribution with degree of freedom $\varphi_1^2 / \varphi_2$ in the numerator and $\delta_1^2 / \delta_2$ in the denominator for a given $\alpha$. 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

$$p = P(F_{gof} \geq f_{gof}),$$

(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 \( \alpha \), we reject the null hypothesis. We accept it otherwise.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Mean Squares</th>
<th>( F_{gof} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWPolR Residuals</td>
<td>( \gamma_1 )</td>
<td>( RSS_{pol} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GWPolR Improvement</td>
<td>( \varphi_1 )</td>
<td>( \Delta RSS )</td>
<td>( \frac{\Delta RSS}{\varphi_1} )</td>
<td>( \frac{RSS_{gwr}}{\delta_1} )</td>
</tr>
<tr>
<td>GWR Residuals</td>
<td>( \delta_1 )</td>
<td>( RSS_{gwr} )</td>
<td>( \frac{RSS_{gwr}}{\delta_1} )</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1.** An ANOVA table for performing the test

**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

\[
y_i = \beta_1(u_i, v_i) + \beta_2(u_i, v_i)x_i + \epsilon_i.
\]

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 \((u_i, v_i), i = 1, 2, \ldots, 12\). 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

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

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

\[ y_i = \beta_1(u_i, v_i) + \beta_2(u_i, v_i)x_i + \beta_3(u_i, v_i)x_i^2 + \epsilon_i. \]  

(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>
<thead>
<tr>
<th>Model</th>
<th>RSS</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWR</td>
<td>21.30691</td>
<td>90.93%</td>
</tr>
<tr>
<td>GWPolR</td>
<td>2.838471</td>
<td>98.79%</td>
</tr>
</tbody>
</table>

Table 3. Performance indicators for the first example

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 \( \varphi_1^2/\varphi_2 = 5.36923 \) 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

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Mean Squares</th>
<th>$F_{gof}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWPolR Residuals</td>
<td>2.21064</td>
<td>2.83847</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GWPolR Improvement</td>
<td>2.96974</td>
<td>18.46844</td>
<td>6.21887</td>
<td></td>
</tr>
<tr>
<td>GWR Residuals</td>
<td>5.18038</td>
<td>21.30691</td>
<td>4.11300</td>
<td>1.512</td>
</tr>
</tbody>
</table>

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](image)
Table 5. The second data set

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

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

<table>
<thead>
<tr>
<th>Model</th>
<th>RSS</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWR</td>
<td>259.16520</td>
<td>98.95%</td>
</tr>
<tr>
<td>GWPolR</td>
<td>42.39748</td>
<td>99.83%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Mean Squares</th>
<th>Fgof</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWPolR Residuals</td>
<td>3.27040</td>
<td>42.39748</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GWPolR Improvement</td>
<td>0.07975</td>
<td>216.76772</td>
<td>2,718.09053</td>
<td></td>
</tr>
<tr>
<td>GWR Residuals</td>
<td>3.35015</td>
<td>259.16520</td>
<td>77.35928</td>
<td>35.135</td>
</tr>
</tbody>
</table>

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

Table 8. Performance indicators for the Life Expectancy data

<table>
<thead>
<tr>
<th>Model</th>
<th>RSS</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWR</td>
<td>30.3514</td>
<td>80.31%</td>
</tr>
<tr>
<td>GWPolR</td>
<td>21.7375</td>
<td>85.90%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Mean Squares</th>
<th>$F_{gof}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWPolR Residuals</td>
<td>4.5165</td>
<td>21.7375</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GWPolR Improvement</td>
<td>0.29890</td>
<td>8.6139</td>
<td>8.6139</td>
<td>28.81867</td>
</tr>
<tr>
<td>GWR Residuals</td>
<td>4.81355</td>
<td>30.3514</td>
<td>6.30541</td>
<td>4.57047</td>
</tr>
</tbody>
</table>

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


