J M S K
Jurnal Matematika, Statistika, & Komputasi
http://journal.unhas.ac.id/index.php/jmsk
Vol .15, No. 2, 25-32, January 2019

e-ISSN: 2614-8811

KONSTRUKSI UJI KESESUAIAN MODEL *GEOGRAPHICALLY WEIGHTED POLYNOMIAL REGRESSION*

Toha Saifudin, Fatmawati, dan Nur Chamidah*

Abstract

Geographically Weighted Polynomial Regression (GWPolR) is a generalization of *Geographically Weighted Regression* (GWR) model. By using the generalization, GWPolR has parameters much more than GWR model. In general, excess of the number of parameter will have a higher appropriate value. However, the model which has less parameter will have the excess for easing in application and its interpretation. Nevertheless, when the model has more the parameters, then the model will be better significantly to be used. Therefore, the aim of this paper is to construct the conformity between hypothesis test with respect to the GWPolR model.

Keywords: *Geographically weighted polynomial regression*, *Geographically Weighted Regression*, uji kesesuaian model

Abstrak

Geographically Weighted Polynomial Regression (GWPolR) merupakan perumuman dari model *Geographically Weighted Regression* (GWR). Dengan perumuman tersebut, model GWPolR memiliki jumlah parameter yang lebih banyak daripada model GWR. Umumnya, kelebihan model dengan jumlah parameter lebih banyak adalah memiliki nilai kesesuaian lebih tinggi. Sebaliknya, model dengan jumlah parameter yang sedikit memiliki kelebihan berupa kemudahan dalam aplikasi dan interpretasinya. Namun demikian, jika model dengan jumlah parameter yang lebih banyak ternyata secara signifikan lebih baik maka sudah seharusnya model tersebut dipilih untuk digunakan. Oleh karena itu, tujuan paper ini adalah mengkonstruksi uji hipotesis kesesuaian model GWPolR.

Kata Kunci: *Geographically weighted polynomial regression*, *Geographically Weighted Regression*, uji kesesuaian model

1. Pendahuluan

Metode GWR merupakan metode analisis data spasial yang populer digunakan dalam sejumlah bidang penerapan. Namun demikian, beberapa upaya perluasan untuk perbaikan model telah dilakukan oleh beberapa peneliti. Sebagai contoh, misalnya: Fotheringham dkk (2002) memberikan beberapa perluasan GWR untuk model *generalized linear*, dan Brunsdon dkk (1999) membahas model GWR semiparametrik.

Menurut beberapa referensi, seperti Fotheringham dkk (2002) dan Leung dkk (2000), model GWR dapat dinyatakan dalam bentuk

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

dengan $(y_i, x_{i1}, x_{i2}, \dots, x_{ip})$ adalah nilai pengamatan variabel respon y dan variabel-variabel prediktor x_1, x_2, …, dan x_p di lokasi (u_i, v_i), $\beta_j(u_i, v_i)$ untuk $j = 1, 2, \dots, p$ adalah parameter atau koefisien regresi di lokasi (u_i, v_i), dan ε_i , $i = 1, 2, \dots, n$, adalah galat model yang

*Universitas Airlangga
nur-c@fst.unair.ac.id

diasumsikan berdistribusi Normal dengan rata-rata 0 dan variansi σ^2 yang biasanya dinyatakan dengan $\varepsilon_i \sim N(0, \sigma^2)$.

Model GWR merupakan model dengan koefisien atau parameter regresi yang bervariasi berdasarkan lokasi. Namun jika dilihat model di masing-masing lokasi, semua variabel prediktor dihubungkan terhadap respon menggunakan fungsi linier. Faktanya, tidak semua variabel prediktor memiliki hubungan linier dengan respon. Mungkin diantara variabel prediktor yang terlibat dalam model, ada beberapa darinya yang memiliki pola non linier. Untuk mengatasi hal tersebut, Saifudin dkk (2017) telah membuat model perluasan atau perumuman dengan menggunakan pendekatan fungsi polinomial. Dalam Saifudin dkk (2017), model tersebut dinamakan *Geographically Weighted Polynomial Regression* (GWPolR).

Model GWPolR merupakan perumuman dari model GWR dengan memperumum fungsi linier menjadi fungsi polinomial. Tujuan penggunaan fungsi polinomial dalam perumuman ini adalah agar dapat lebih mengakomodasi perilaku sampel yang kemungkinan ada satu atau lebih variabel prediktor memiliki hubungan non linier dengan respon. Konsekuensi dari perumuman tersebut, model GWPolR memiliki jumlah parameter yang lebih banyak daripada model GWR. Umumnya, kelebihan model dengan jumlah parameter yang lebih banyak adalah memiliki nilai kesesuaian lebih tinggi. Sebaliknya, model dengan jumlah parameter yang lebih sedikit memiliki kelebihan berupa kemudahan dalam aplikasi dan interpretasinya. Namun demikian, jika model dengan jumlah parameter yang lebih banyak ternyata secara signifikan lebih baik maka sudah seharusnya model tersebut dipilih untuk digunakan. Oleh karena itu, diperlukan bahasan tentang uji signifikansi kesesuaian model GWPolR. Berdasarkan hal tersebut, maka tujuan paper ini adalah mengkonstruksi uji hipotesis kesesuaian model GWPolR.

2. Model *Geographically Weighted Polynomial Regression*

Menurut Saifudin dkk (2017), model GWPolR sebagai perumuman dari model GWR pada persamaan (1) dapat dinyatakan sebagai

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \sum_{j=1}^{d_k} \beta_{k,j}(u_i, v_i)\, {x_{ik}}^{j} + \varepsilon_i\,, \quad i = 1, 2, \dots, n. \tag{2}$$

Dalam format matriks, model (2) dapat ditulis sebagai

$$y_i = \boldsymbol{x}_i^{*\mathrm{T}}\, \boldsymbol{\beta}_{pol}(u_i, v_i) + \varepsilon_i\,, \quad i = 1, 2, \dots, n \tag{3}$$

dengan

$$\boldsymbol{x}_i^{*\mathrm{T}} = \begin{pmatrix} 1 & x_{i1} & x_{i1}^2 & \cdots & x_{i1}^{d_1} & \cdots & x_{ip} & x_{ip}^2 & \cdots & x_{ip}^{d_p} \end{pmatrix},$$

dan

$$\boldsymbol{\beta}^{\mathrm{T}}{}_{pol}(u_i, v_i) = (\ \beta_0(u_i, v_i)\ \beta_{1,1}(u_i, v_i)\ \beta_{1,2}(u_i, v_i) \cdots \beta_{1,d_1}(u_i, v_i) \cdots \\ \beta_{p,1}(u_i, v_i)\ \beta_{p,2}(u_i, v_i) \cdots \beta_{p,d_p}(u_i, v_i)\ ).$$

Menurut Saifudin (2017), estimator *weighted least square* untuk parameter model GWPolR pada persamaan (3) di lokasi (u_i, v_i) dapat dinyatakan sebagai

$$\hat{\boldsymbol{\beta}}^{\mathrm{T}}{}_{pol}(u_i, v_i) = \left[\mathbf{X}_{pol}^{\mathrm{T}} \mathbf{W}(u_i, v_i) \mathbf{X}_{pol}\right]^{-1} \mathbf{X}_{pol}^{\mathrm{T}} \mathbf{W}(u_i, v_i) \boldsymbol{y}\,, \tag{4}$$

dengan

$$\boldsymbol{y} = (y_1, y_2, \dots, y_n)^T\,,$$

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

$$\mathbf{W}(u_i, v_i) = \mathrm{diag}[K_h(d_{i1}), K_h(d_{i2}), \dots, K_h(d_{in})\ ],$$

$K_h(\cdot) = K\left(\frac{\cdot}{h}\right)$ untuk $K(\cdot)$ adalah fungsi kernel yang diberikan, dan h adalah *bandwidth*.

Berdasarkan persamaan (4), vektor nilai estimasi model GWPolR untuk variabel respon y di n lokasi dapat dinyatakan sebagai

$$\hat{\boldsymbol{y}}_{pol} = (\hat{y}_1^*, \hat{y}_2^*, \dots, \hat{y}_n^*)^{\mathrm{T}} = \mathbf{C}\,\mathbf{y}\,, \tag{5}$$

dengan

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

dinamakan *hat matrix* dari model GWPolR. Berdasarkan persamaan (5), vektor residual dapat ditulis sebagai

$$\hat{\boldsymbol{\varepsilon}}_{pol} = \boldsymbol{y} - \hat{\boldsymbol{y}}_{pol} = (\mathbf{I} - \mathbf{C})\boldsymbol{y}\,. \tag{6}$$

Selanjutnya, dengan persamaan (6) Jumlah Kuadrat Galat (JKG) model GWPolR adalah

$$\mathrm{JKG}_{\mathrm{Pol}} = \hat{\boldsymbol{\varepsilon}}_{pol}^{\mathrm{T}} \hat{\boldsymbol{\varepsilon}}_{pol} = \boldsymbol{y}^{\mathrm{T}} (\mathbf{I} - \mathbf{C})^{\mathrm{T}} (\mathbf{I} - \mathbf{C}) \boldsymbol{y}\,, \tag{7}$$

dengan **I** adalah matriks identitas orde n.

3. Uji Kesesuaian Model *Geographically Weighted Polynomial Regression*

Dalam bagian ini akan dikonstruksi statistik uji kesesuaian model GWPolR berdasarkan JKG model. Distribusi pendekatan dari statistik ujitersebut kemudian akan diinvestigasi untuk menguji apakah model GWPolR dapat menggambarkan data sampel secara signifikan lebih baikdaripada model GWR. Dalam bahasan ini, diberikan dua asumsi sebagai berikut:

Asumsi 1: galat model yaitu $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ diasumsikan berdistribusi Normal dengan mean 0 dan variansi σ^2

Asumsi 2: Misalkan bahwa $\hat{y}_i^*$ adalah nilai estimasi dari y_i di lokasi (u_i, v_i). Untuk $i = 1, 2, \dots, n$, $\hat{y}_i^*$ adalah estimator unbias dari $E(y_i)$, yaitu $E(\hat{y}_i^*) = E(y_i)$.

3.1 Distribusi Jumlah Kuadrat Galat Model GWPolR

Misalkan $\boldsymbol{x}_i^{*\mathrm{T}} = (1\ \ x_{i1}\ \ x_{i1}^2\ \dots\ x_{i1}^{d_1}\ \dots\ \ x_{ip}\ \ x_{ip}^2\ \dots\ x_{ip}^{d_p})$ adalah baris ke-i dari matriks $\mathbf{X}_{Pol}$, $i = 1, 2, \dots, n$ dan $\hat{\boldsymbol{\beta}}_{Pol}(u_i, v_i)$ adalah vektor estimator parameter di lokasi (u_i, v_i). Nilai dugaan bagi y_i dapat dinyatakan sebagai

$$\begin{aligned} \hat{y}_i^* &= \boldsymbol{x}_i^{*\mathrm{T}}\, \hat{\boldsymbol{\beta}}_{Pol}(u_i, v_i) \\ &= \boldsymbol{x}_i^{*\mathrm{T}} \left[\mathbf{X}_{Pol}^{\mathrm{T}} \mathbf{W}(u_i, v_i) \mathbf{X}_{Pol}\right]^{-1} \mathbf{X}_{Pol}^{\mathrm{T}} \mathbf{W}(u_i, v_i) \boldsymbol{y}, \end{aligned} \tag{8}$$

dan residual model sebagai $\hat{\varepsilon}_i^* = y_i - \hat{y}_i^*$, untuk $i = 1, 2, \dots, n$.

Misalkan $\hat{\boldsymbol{y}}_{Pol} = (\hat{y}_1^*\ \ \hat{y}_2^*\ \dots\ \hat{y}_n^*)^{\mathrm{T}}$ menyatakan vektor nilai dugaan bagi y_i dan $\hat{\boldsymbol{\varepsilon}}_{Pol} = (\hat{\varepsilon}_1^*\ \ \hat{\varepsilon}_2^*\ \dots\ \hat{\varepsilon}_n^*)^{\mathrm{T}}$ menyatakan vektor residual, maka

$$\hat{\boldsymbol{y}}_{Pol} = \mathbf{G}\boldsymbol{y},\ \text{dan}\ \ \hat{\boldsymbol{\varepsilon}}_{Pol} = \boldsymbol{y} - \hat{\boldsymbol{y}}_{Pol} = (\mathbf{I} - \mathbf{G})\boldsymbol{y} \tag{9}$$

dengan

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

merupakan *hat matrix* dari model GWPolR dan **I** adalah matriks identitas berorde *n*.

Oleh karena itu,

$$\hat{\boldsymbol{\varepsilon}}_{Pol} = \boldsymbol{y} - \hat{\boldsymbol{y}}_{Pol} = (\mathbf{I}-\mathbf{G})\boldsymbol{y} \tag{11}$$

dengan **I** adalah matriks identitas orde *n*.

Misalkan jumlah kuadrat galat model GWR Polinomial dinotasikan dengan JKG_{Pol}, maka

$$\text{JKG}_{Pol} = \sum_{i=1}^{n} \hat{\varepsilon}_i^{*2} = \hat{\boldsymbol{\varepsilon}}_{Pol}^{\mathrm{T}}\hat{\boldsymbol{\varepsilon}}_{Pol} = \boldsymbol{y}^{\mathrm{T}}(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\boldsymbol{y}. \tag{12}$$

Selanjutnya, Estimator tak bias bagi variansi galat diberikan dalam Teorema 1 berikut.

Teorema 1. Misalkan model GWPolR memenuhi Asumsi 1 dan Asumsi 2, dan $\hat{\boldsymbol{\beta}}_{Pol}(u_i,v_i)$ sebagai estimator parameter di lokasi (u_i,v_i), maka estimator tak bias untuk variansi galat (σ^2) diberikan sebagai berikut:

$$\hat{\sigma}^2 = \frac{\text{JKG}_{Pol}}{\gamma_1}, \tag{13}$$

dengan $\gamma_1 = tr\left((\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\right) = n - \left(2\,tr(\mathbf{G}) - tr\left(\mathbf{G}^{\mathrm{T}}\mathbf{G}\right)\right)$.

Bukti:
Berdasarkan Asumsi 1 dan Asumsi 2, maka

$$E(\hat{\boldsymbol{\varepsilon}}_{Pol}) = E(\boldsymbol{y}) - E(\hat{\boldsymbol{y}}_{Pol}) = \mathbf{0},$$
$$E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\right) = \sigma^2\mathbf{I},$$

dengan $\boldsymbol{\varepsilon} = (\varepsilon_1\ \varepsilon_2\ \ldots\ \varepsilon_n)^{\mathrm{T}}$ adalah vektor galat model. Dengan demikian, JKG_{Pol} dapat dinyatakan sebagai berikut

$$\begin{aligned}\text{JKG}_{Pol} &= \left(\hat{\boldsymbol{\varepsilon}}_{Pol} - E(\hat{\boldsymbol{\varepsilon}}_{Pol})\right)^{\mathrm{T}}\left(\hat{\boldsymbol{\varepsilon}}_{Pol} - E(\hat{\boldsymbol{\varepsilon}}_{Pol})\right) \\ &= \boldsymbol{\varepsilon}^{\mathrm{T}}(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\boldsymbol{\varepsilon}\ .\end{aligned}$$

Oleh karena itu, maka

$$\begin{aligned}E(\text{JKG}_{Pol}) &= E\left(\boldsymbol{\varepsilon}^{\mathrm{T}}(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\boldsymbol{\varepsilon}\right) \\ &= E\left(tr\left(\boldsymbol{\varepsilon}^{\mathrm{T}}(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\boldsymbol{\varepsilon}\right)\right) \\ &= tr\left((\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\right)\right) \\ &= \sigma^2\gamma_1\ ,\end{aligned} \tag{14}$$

dengan $\gamma_1 = tr\left((\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\right)$. Berdasarkan persamaan (14) dapat diambil bahwa estimator tak bias bagi σ^2 adalah $\hat{\sigma}^2 = \frac{\text{JKG}_{Pol}}{\gamma_1}$. Secara operasional, γ_1 dapat dinyatakan dengan rumus berikut

$$\begin{aligned}\gamma_1 &= tr\left((\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\right) \\ &= tr(\mathbf{I}) - tr(\mathbf{G}) - tr(\mathbf{G}^{\mathrm{T}}) + tr(\mathbf{G}^{\mathrm{T}}\mathbf{G}) \\ &= n - \left(2\,tr(\mathbf{G}) - tr\left(\mathbf{G}^{\mathrm{T}}\mathbf{G}\right)\right).\ \blacksquare\end{aligned} \tag{15}$$

Berdasarkan Teorema 1, besaran JKG_{Pol} dapat digunakan untuk mengestimasi variansi galat σ^2. Besaran ini dapat digunakan untuk mengukur kesesuaian model GWR Polinomial terhadap data sampel. Semakin kecil nilai besaran ini, menunjukkan semakin sesuai model tersebut diterapkan pada data sampel. Namun untuk kepentingan uji kesesuaian model, diperlukan pengetahuan tentang distribusi dari besaran JKG_{Pol}. Oleh karena itu diberikan pendekatan distribusi untuk JKG_{Pol} yang dapat dilihat pada Teorema 2 berikut.

Teorema 2. Misalkan model GWPolR memenuhi asumsi 1 dan 2. Misalkan $\hat{\boldsymbol{\beta}}_{Pol}(u_i, v_i)$ adalah estimator parameter model di lokasi (u_i, v_i). Distribusi untuk variabel acak

$$\frac{\gamma_1 \text{JKG}_{Pol}}{\gamma_2 \sigma^2} \tag{16}$$

dapat didekati menggunakan distribusi χ_r^2 dengan derajat bebas $r = \frac{\gamma_1^2}{\gamma_2}$, dengan

$$\gamma_i = tr\left(\left((\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\right)^i\right), i = 1, 2. \tag{17}$$

Bukti:

Dari persamaan (18) dapat dilihat bahwa JKG_{Pol} dapat dinyatakan sebagai bentuk kuadrat dari variabel Normal dengan $(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})$ merupakan matriks simetri dan semidefinit positif. Berdasarkan teori distribusi bentuk kuadrat (1) diketahui bahwa suatu bentuk kuadrat dari variabel Normal Baku, yaitu $\boldsymbol{\xi}^{\mathrm{T}}\mathbf{A}\boldsymbol{\xi}$ dengan $\boldsymbol{\xi} \sim N(\mathbf{0},\mathbf{I})$ dan $\mathbf{A}$ matriks simetri, berdistribusi χ^2 jika dan hanya jika $\mathbf{A}$ matriks idempoten (sebagai contoh, lihat Rencher & Schaalje (2008), *corollary* 1 dari Teorema 5.5, hal. 118). Untuk variabel

$$\frac{\text{JKG}_{Pol}}{\sigma^2} = \left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right)^{\mathrm{T}} (\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G}) \left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right), \tag{18}$$

diketahui bahwa $\frac{\boldsymbol{\varepsilon}}{\sigma} \sim N(\mathbf{0}, \mathbf{I})$, tetapi matriks $(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})$ umumnya tidak idempoten, karena elemen pembentuknya yaitu matriks pembobot $\mathbf{W}(u_i, v_i)$ berbeda pada setiap lokasi *i*. Karena itu $\frac{\text{JKG}_{Pol}}{\sigma^2}$ tidak berdistribusi χ^2 secara eksak, namun dapat ditentukan distribusi χ^2 secara pendekatan (*approximated* χ^2). Berdasarkan teori distribusi bentuk kuadrat (Yuan & Bentler, 2010), pendekatan distribusi χ^2 terhadap bentuk kuadrat (18) dapat dilakukan dengan cara mengalikan variabel χ_r^2 dengan konstanta *c*, yang dapat ditulis $c\chi_r^2$. Selanjutnya *c* dan *r* ditentukan sedemikian hingga rata-rata dan variansi dari $c\chi_r^2$ dan bentuk kuadrat yang didekati saling bersesuaian (sama). Untuk variabel χ_r^2, berdasarkan teori dapat diketahui bahwa rata-rata dan variansinya masing-masing adalah *r* dan 2*r*. Oleh karena itu mean dan variansi dari variabel $c\chi_r^2$ masing-masing adalah *cr* dan $2c^2r$.

Untuk variabel bentuk kuadrat $\frac{\text{JKG}_{Pol}}{\sigma^2}$, berdasarkan persamaan (14) diketahui mempunyai mean γ_1. Adapun variansinya diuraikan berikut. Karena matriks $(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})$ simetri dan semidefinit positif, maka terdapat matriks ortogonal $\mathbf{P}$ orde *n* sedemikian hingga

$$\mathbf{P}^{\mathrm{T}}(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\mathbf{P} = \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \tag{19}$$

dengan $\boldsymbol{\Lambda}$ adalah matriks diagonal yang elemen diagonal utamanya merupakan nilai-nilai eigen dari matriks $(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})$. misalkan bahwa

$$\boldsymbol{\eta} = (\eta_1, \eta_2, \cdots, \eta_n)^{\mathrm{T}} = \mathbf{P}^{\mathrm{T}} \frac{\boldsymbol{\varepsilon}}{\sigma}. \tag{20}$$

Sesuai dengan sifat distribusi Normal multivariat, maka $\eta_1, \eta_2, \cdots, \eta_n$ merupakan variabel acak independen dan identik berdistribusi Normal baku, yang ditulis $\eta_i \sim iid\ \text{N}(0,1)$. Di sisi lain, dari persamaan (20) diperoleh $\frac{\boldsymbol{\varepsilon}}{\sigma} = \mathbf{P}\boldsymbol{\eta}$, dan dari persamaan (18) diperoleh

$$\begin{aligned}\frac{\text{JKG}_{Pol}}{\sigma^2} &= \boldsymbol{\eta}^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}(\mathbf{I}-\mathbf{G})^{T}(\mathbf{I}-\mathbf{G})\mathbf{P}\boldsymbol{\eta} \\ &= \boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{\eta}\end{aligned}$$

$$= \sum_{i=1}^{n} \lambda_i \eta_i^2. \tag{21}$$

Sesuai dengan teori distribusi, oleh karena $\eta_i \sim iid\ \mathrm{N}(0,1)$ maka $\eta_i^2 \sim iid\ \chi^2_{(1)}$ untuk i =1, 2, ⋯, n. oleh karena itu $\mathrm{var}(\eta_i^2) = 2$, dan

$$\mathrm{var}\left(\frac{\mathrm{JKG}_{Pol}}{\sigma^2}\right) = \sum_{i=1}^{n} \lambda_i^2\, \mathrm{var}(\eta_i^2) = 2\sum_{i=1}^{n} \lambda_i^2. \tag{22}$$

Jika $\lambda_1, \lambda_2, \cdots, \lambda_n$ adalah nilai-nilai eigen dari matriks $(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})$ maka $\lambda_1^2, \lambda_2^2, \cdots, \lambda_n^2$ merupakan nilai-nilai eigen dari matriks $\left((\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\right)^2$. Oleh karena itu

$$\begin{aligned} \mathrm{var}\left(\frac{\mathrm{JKG}_{Pol}}{\sigma^2}\right) &= 2\,tr\left(\left((\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\right)^2\right) \\ &= 2\gamma_2, \end{aligned} \tag{23}$$

dengan $\gamma_2 = tr\left(\left((\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\right)^2\right)$.

Berdasarkan persamaan (14) and (23), dapat dibentuk sistem persamaan berikut

$$\begin{cases} cr = \gamma_1, \\ 2c^2 r = 2\gamma_2. \end{cases} \tag{24}$$

Penyelesaian dari sistem persamaan (24) adalah $c = \frac{\gamma_2}{\gamma_1}$, dan $r = \frac{\gamma_1^2}{\gamma_2}$. Dengan demikian, distribusi dari $\frac{\mathrm{JKG}_{Pol}}{c\sigma^2}$ atau $\frac{\gamma_1 \mathrm{JKG}_{Pol}}{\gamma_2 \sigma^2}$ dapat didekati menggunakan distribusi χ^2 dengan derajat bebas $\frac{\gamma_1^2}{\gamma_2}$, dengan $\gamma_i = tr\left(\left((\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\right)^i\right), i = 1, 2$. ■

3.2 Statistik Uji Kesesuaian Model *Geographically Weighted Polynomial Regression* dan Distribusinya

Jumlah Kuadrat Galat dan distribusi pendekatannya yang telah diperoleh pada bagian 3.1 di atas selanjutnya akan digunakan untuk menguji kesesuaian model, apakah GWPolR dapat memodelkan data secara signifikan lebih baik daripada GWR. Dalam uji kesesuaian model ini dihipotesiskan sebagai berikut:

H_0 : Model GWPolR tidak berbeda dengan model GWR,

H_1 : Model GWPolR signifikan lebih sesuai daripada model GWR.

Untuk menguji hipotesis di atas, dalam paper ini dikonstruksi suatu statistik uji dan distribusi pendekatannya.melalui Teorema 3 berikut.

Teorema 3. Misalkan $\mathrm{JKG}_{gwr} = \boldsymbol{y}^{\mathrm{T}}(\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L})\boldsymbol{y}$ adalah jumlah kuadrat galat model GWR dengan $\mathbf{L}$ adalah *hat matrix* model GWR, dan $\mathrm{JKG}_{Pol} = \boldsymbol{y}^{\mathrm{T}}(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\boldsymbol{y}$ adalah jumlah kuadrat galat model GWPolR dengan $\mathbf{G}$ adalah *hat matrix* model GWPolR. Distribusi dari statistik

$$F_1 = \frac{\mathrm{JKG}_{Pol}/\gamma_1}{\mathrm{JKG}_{gwr}/\delta_1} \tag{25}$$

mendekati distribusi F dengan derajat bebas pembilang $db_1 = \frac{\gamma_1^2}{\gamma_2}$ dan derajat bebas penyebut $db_2 = \frac{\delta_1^2}{\delta_2}$, dengan $\gamma_i = tr\left(\left((\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\right)^i\right)$ dan $\delta_i = tr\left(\left((\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L})\right)^i\right)$ untuk $i = 1, 2$.

Bukti:

Jika model GWPolR digunakan untuk memodelkan data sampel, maka jumlah kuadrat galat dapat dinyatakan dengan $\text{JKG}_{Pol} = \boldsymbol{y}^{\mathrm{T}}(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\boldsymbol{y}$. Berdasarkan Teorema 2, distribusi dari variabel $\frac{\gamma_1 \text{JKG}_{Pol}}{\gamma_2 \sigma^2}$ dapat didekati dengan distribusi χ_r^2 dengan derajat bebas $r = \frac{\gamma_1^2}{\gamma_2}$, untuk $\gamma_i = tr\left(\left((\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\right)^i\right), i = 1, 2.$

Jika model GWR digunakan untuk memodelkan data sampel, maka jumlah kuadrat galat dapat dinyatakan dengan $\text{JKG}_{gwr} = \boldsymbol{y}^{\mathrm{T}}(\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L})\boldsymbol{y}$, dengan $\mathbf{L}$ adalah *hat matrix* model GWR. Menurut Leung dkk (2000), distribusi dari variabel $\frac{\delta_1 \text{JKG}_{gwr}}{\delta_2 \sigma^2}$ dapat didekati dengan distribusi χ_d^2 dengan derajat bebas $d = \frac{\delta_1^2}{\delta_2}$, untuk

$$\delta_1 = tr\left((\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L})\right), \text{ dan} \tag{26}$$

$$\delta_2 = tr\left(\left((\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L})\right)^2\right). \tag{27}$$

Dengan cara yang sama seperti dalam mendapatkan persamaan (15), secara operasional δ_1 dapat dinyatakan dalam bentuk

$$\delta_1 = n - \left(2\, tr(\mathbf{L}) - tr\left(\mathbf{L}^{\mathrm{T}}\mathbf{L}\right)\right). \tag{28}$$

Secara teori distribusi, diketahui bahwa variabel $\frac{\chi_r^2 / r}{\chi_d^2 / d}$ berdistribusi *F* dengan derajat bebas pembilang *r* dan derajat bebas penyebut *d*. Oleh karena itu, distribusi dari statistik

$$F_1 = \frac{\frac{\gamma_1 \text{JKG}_{Pol}}{\gamma_2 \sigma^2} \Big/ \left(\frac{\gamma_1^2}{\gamma_2}\right)}{\frac{\delta_1 \text{JKG}_{gwr}}{\delta_2 \sigma^2} \Big/ \left(\frac{\delta_1^2}{\delta_2}\right)}, \tag{29}$$

mendekati distribusi *F* dengan derajat bebas pembilang $db_1 = \frac{\gamma_1^2}{\gamma_2}$ dan derajat bebas penyebut $d_2 = \frac{\delta_1^2}{\delta_2}$. Jika disederhanakan, maka persamaan (29) menjadi persamaan (25), dengan $\gamma_i = tr\left(\left((\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\right)^i\right)$ dan $\delta_i = tr\left(\left((\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L})\right)^i\right)$ untuk $i = 1, 2$. ■

Jika H_0 benar, maka secara logika besaran $\frac{\text{JKG}_{Pol}}{\text{JKG}_{gwr}}$ mendekati 1. Jika H_0 salah, maka besaran tersebut cenderung kecil mendekati 0. Berdasarkan logika tersebut, semakin kecil nilai F_1 maka semakin mendukung untuk menolak H_0 yang berarti bahwa model GWPolR signifikan lebih sesuai daripada model GWR. Oleh karena itu, pada tingkat signifikansi α yang diberikan, tolak H_0 jika $F_1 < F_{1-\alpha}\left(\gamma_1^2/\gamma_2, \delta_1^2/\delta_2\right)$ dan simpulkan bahwa model GWPolR signifikan lebih sesuai daripada model GWR. Jika tidak demikian, maka dapat dikatakan bahwa model GWPolR tidak dapat memberikan peningkatan ketepatan yang signifikan dari model GWR.

4. Kesimpulan

Uji kesesuaian model GWPolR dibangun untuk memeriksa apakah model GWPolR menggambarkan kumpulan data tertentu secara signifikan lebih baik daripada GWR atau tidak. Uji ini bisa dibangun berdasarkan jumlah kuadrat galat dari model. Statistik uji kesesuaian model GWPolR dapat didekati menggunakan distribusi F.

Daftar Pustaka

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

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

[3]. Leung, Y., Mei, CL., & Zhang, WX., 2000. Statistical Tests for Spatial Nonstationarity based on The Geographically Weighted Regression Model. *Environment and Planning A*. 32: 9–32.

[4]. Rencher, A.C., & Schaalje, G.G., 2008. *Linear Models in Statistics,* 2^{nd} Edition. John Willey & Sons, New York, USA.

[5]. Saifudin, T., Fatmawati, dan Chamidah, N., 2017. *Perluasan Geographically Weighted Regression Menggunakan Fungsi Polinomial*, Prosiding Seminar Nasional Integrasi Matematika dan Nilai Islami, 1[1], Juli 2017, 15-20.

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