βeta

Vol. 5 No. 1 (Mei) 2012, Hal. 40-56

p-ISSN: 2085-5893 / e-ISSN: 2541-0458
http://jurnalbeta.ac.id

ESTIMASI DAN PENGUJIAN HIPOTESIS GEOGRAPHICALLY WEIGHTED REGRESSION

Alfira Mulya Astuti[1]

Abstrak: Salah satu analisis statistika yang menghubungkan variabel respon dengan variabel bebas yaitu metode regresi. Hasil keluaran (output) dari metode ini adalah estimasi dari parameter yang menghubungkan variabel bebas dan variabel respon. Masalah utama dari metode ini adalah jika metode ini diterapkan pada data spatial. Untuk mengatasi permasalahan pada data spatial maka metode statistik yang akan digunakan adalah Geographically Weighted Regression (GWR), yaitu model yang menggunakan faktor geografis sebagai variabel bebas yang dapat mempengaruhi variabel respon. Estimasi pada model GWR dengan pendekatan MLE menghasilkan estimator yang sama dengan pendekatan Weighted Least Square (WLS) yang sudah umum digunakan. Namun pendekatan ini tidak dapat digunakan secara langsung untuk menaksir estimator varians.

Kata Kunci: *Metode Regresi; Geographically Weighted Regression (GWR)*

A. PENDAHULUAN

Metode regresi merupakan metode statistik yang paling umum digunakan. Metode regresi yaitu metode yang menghubungkan variabel respon dengan variabel bebas dengan hasil keluaran (output) utamanya adalah estimasi dari parameter yang membentuk suatu model tertentu (Draper dan Smith, 1992). Masalah utama dari metode ini adalah jika metode ini diterapkan pada data spatial, dimana metode *Ordinary Least Square* (OLS) untuk estimasi parameter model regresi dengan asumsi *error* identik independen dan berdistribusi normal yang harus dipenuhi, maka akan diperoleh satu model taksiran untuk semua data. Hal inilah yang menyebabkan ketidaksesuaian model pada data spatial.

[1] IAIN Mataram, Mataram, Indonesia,alfiramulyastuti@yahoo.co.id

Data spatial merupakan data pengamatan bila melibatkan informasi koordinat lokasi pengambilan data di samping data mengenai peubah-peubah yang sedang diamati. Analisis terhadap data spatial memerlukan perhatian lebih dibandingkan dengan analisis data nonspatial, khususnya ketika menggunakan regresi. Salah satu hal yang harus mendapat perhatian pada penanganan data spatial adalah kemungkinan munculnya heterogenitas spatial. Heterogenitas spatial muncul karena kondisi data di lokasi yang satu dengan lokasi yang lain tidak sama, baik dari segi geografis, keadaan sosial-budaya maupun hal-hal lain yang melatarbelakanginya. Salah satu dampak yang ditimbulkan dari munculnya heterogenitas spasial adalah parameter regresi bervariasi secara spasial atau disebut juga terjadi nonstasioneritas spasial pada parameter regresi.

Pada regresi OLS (*Ordinary Least Square*) diasumsikan bahwa nilai duga parameter regresi akan tetap (konstan), artinya parameter regresi bernilai sama untuk setiap titik di dalam wilayah penelitian (parameter global). Bila terjadi heterogenitas spasial pada parameter regresi, maka informasi yang tidak dapat ditangani oleh metode regresi OLS akan ditampung sebagai galat. Bila kasus semacam itu terjadi, regresi OLS menjadi kurang mampu dalam menjelaskan fenomena data yang sebenarnya. Untuk mengantisipasi munculnya heterogenitas spasial pada parameter regresi, regresi OLS dikembangkan menjadi *Geographically Weighted Regression* (GWR). Pada GWR, parameter regresi diasumsikan bervariasi secara spasial. Melalui penggunaan GWR akan dapat diketahui variasi spasial dalam nilai duga parameter, sehingga interpretasi yang berbeda dan berharga dapat diperoleh untuk setiap titik lokasi yang diteliti.

Banyaknya variabel prediktor memungkinkan akan terjadinya pelanggaran asumsi tentang tidak adanya multikolinieritas dalam data. Oleh karena itu diperlukan suatu metode untuk mereduksi beberapa variabel yang tidak signifikan terhadap responnya. Salah satu metode yang dapat digunakan adalah metode stepwise GWR.

B. TINJAUAN PUSTAKA

Metode regresi yang merupakan metode yang memodelkan hubungan antara variabel respon (y) dan variabel bebas ($x_1, x_2, \ldots, x_p$). Model regresi linier secara umum dinyatakan dengan

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon$$

Jika diambill sebanyak *n* pengamatan, maka model di atas dapat ditulis sebagai:

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + \varepsilon_i \qquad (2.1)$$

dengan *i* = 1, 2, ... , *n* ; *k* = 1, 2, ... , *p* ; $\beta_0, \beta_1, \ldots, \beta_p$ adalah parameter model dan $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ adalah error yang diasumsikan identik, independen, dan berdistribusi Normal dengan mean nol dan varians konstan σ^2. Pada model ini, hubungan antara variabel bebas dan variabel respon dianggap konstan pada setiap lokasi geografis. Estimator dari parameter model didapat dari persamaan

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{y} \qquad (2.2)$$

dengan

$\boldsymbol{\beta}$: vektor dari parameter yang ditaksir berukuran *n* x (*p*+1)

X : matrik data berukuran *n* x (*p*+1) dari variabel bebas yang elemen pada kolom pertama bernilai 1

y : vektor observasi dari variabel respon berukuran (*n* x 1)

k : banyaknya variabel bebas (*k* = 1, 2, ... , *p*)

2.1 Model GWR

Metode GWR adalah suatu teknik yang membawa kerangka dari model regresi sederhana menjadi model regresi yang terboboti (Fotheringham, et al., 2002)

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \,, \qquad (2.3)$$

dengan

y_i : pengamatan pada lokasi ke-*i* (*i* = 1, 2, ... , *n*)

(u_i, v_i) : koordinat *longitude latitude* dari titik ke-*i* pada suatu lokasi geografis.

$\beta_k(u_i, v_i)$: realisasi dari fungsi kontinyu $\beta_k(u, v)$ pada titik ke-*i*

ε_i : *error* yang diasumsikan identik, independen, dan berdistribusi Normal dengan mean nol dan varians konstan σ^2

Dengan demikian setiap nilai parameter dihitung pada setiap titik lokasi georafis. Jadi setiap titik lokasi geografis mempunyai nilai parameter regresi yang berbeda-beda. Hal ini menghasilkaan variasi pada nilai parameter regresi di suatu kumpulan wilayah geografis. Jika nilai parameter regresi konstan pada tiap-tiap wilayah geografis, maka model GWR adalah model global. Artinya tiap-tiap wilayah geografis mempunyai model yang sama. Hal ini merupakan kasus spesial dari GWR.

2.2 Estimasi parameter model GWR

Pada model GWR diasumsikan bahwa data observasi yang dekat dengan titik ke-i mempunyai pengaruh yang besar pada estimasi dari βk (ui,vi) dari pada data yang berada jauh dari titik ke-i. Esensi yang bisa diambil dari hal tersebut adalah persamaan diatas mengukur hubungan model pada semua titik ke-i. Lokal parameter βk (ui,vi) diestimasi menggunakan WLS (Leung, 2000). Pada GWR sebuah observasi diboboti dengan nilai yang berhubungan dengan titik ke-i. Bobot wij, untuk j = 1,2, . . . ,n, pada tiap lokasi (ui,vi) diperoleh sebagai fungsi yang kontinyu dari jarak antara titik ke-i dan titik data lainnya. Misal matriks berikut merupakan matriks dari lokal parameter

$$\mathbf{B} = \begin{bmatrix} \beta_0(u_1,v_1) & \beta_1(u_1,v_1) & \ldots & \beta_p(u_1,v_1) \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ \beta_0(u_n,v_n) & \beta_1(u_n,v_n) & \ldots & \beta_p(u_n,v_n) \end{bmatrix} \quad (2.4)$$

Estimasi tiap baris adalah dengan persamaan berikut:

$$\hat{\beta}(i) = \left(\mathbf{X}^T \mathbf{W}(i)\mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W}(i)\mathbf{y} \quad (2.5)$$

dengan

X = matrik data dari variabel bebas

y = vektor variabel respon

W(*i*) = matriks pembobot

$= diag\ [w_{i1}, w_{i2}, \ldots, w_{in}]$ (2.6)

Estimasi dari (2.5) merupakan estimasi dari *least square* tetapi matriks pembobot tidak konstan, sehingga **W**(*i*) dihitung untuk tiap *i* dan w_{ij} mengindikasikan kedekatan atau bobot tiap titik data dengan lokasi *i*. Hal ini yang membedakan GWR dengan tradisional WLS yang mempunyai matrik bobot yang konstan.

Selain menghasilkan estimasi parameter lokal untuk tiap-tiap lokasi geografis, GWR juga menghasilkan versi lokal untuk seluruh standar regression pada seluruh lokasi geografis misalnya ukuran *goodness of fit*. Hal ini dapat memberikan informasi pada pemahaman aplikasi dari model dan untuk penelitian lebih lanjut apakah diperlukan penambahan variabel independen pada model GWR. Hal yang penting lainnya adalah titik dimana parameter lokal diestimasi dengan model GWR tidak memerlukan titik dimana data diambil. Estimasi dari parameter dapat didapat dari semua lokasi geografis. Dengan demikian, pada sistem dengan data titik lokasi geografis yang besar, estimasi model GWR dari lokal parameter.

2.3 Pembobotan model GWR

Peran pembobot pada model GWR sangat penting karena nilai pembobot ini mewakili letak data observasi satu dengan lainnya. Oleh karena itu, sangat dibutuhkan ketepatan cara pembobotan (Chasco, et al., 2007). Skema pembobotan pada GWR dapat menggunakan beberapa metode yang berbeda, salah satu metode pembobotan yang biasa digunakan adalah kernel Gaussian dan fungsi pembobotan *bisquare* (Bocci, et al., 2006). Fungsi *bisquare* untuk menghitung titik ke-*M* yang terdekat adalah:

$$w_{ij} = \begin{cases} [1 - (d_{ij} / b)^2]^2, & \text{jika } j \text{ adalah salah satu titik ke-}M \text{ yang ekat dari itik ke-}i \\ 0, & \text{untuk lainnya} \end{cases} \tag{2.7}$$

dengan *b* adalah jarak titik ke-*M* yang terdekat dan d_{ij} merupakan jarak Euclidean $d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$.

Kriteria untuk penentuan nilai *M* yang tepat dapat diperoleh dengan pendekatan *least square* yaitu dengan menggunakan kriteria *cross-validation*

$$CV = \sum_{i=1}^{n} \left[y_i - \hat{y}_{i^*}(b) \right]^2$$

(2.8)

dengan i* ≠ i dan $\hat{y}_{i^*}(b)$ adalah nilai dugaan untuk y_i dengan pengamatan pada titik ke-*i* diabaikan dalam proses kalibrasinya.

2.4 *Maximum Likelihood Estimator* dan Statistik Uji

Sejauh ini metode *Maximum Likelihood Estimator* (MLE) merupakan metode estimasi yang umum digunakan (Casella dan Berger, 1990). Jika kita memiliki sampel $x_1, x_2, \ldots, x_n$ yang iid dari populasi yang berdistribusi $f(x \mid \theta_1, \ldots, \theta_k)$, fungsi *likelihood*-nya didefinisikan sebagai:

$$L(\theta \mid x) = L(\theta_1, \ldots, \theta_k \mid x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i \mid \theta_1, \ldots, \theta_k)$$

(2.9)

Jika fungsi likelihood dapat diturunkan terhadap θ_i, maka akan diperoleh penyelesaian atau estimasi parameter-parameter $(\theta_1, \ldots, \theta_k)$ dengan memaksimumkan fungsi:

$$\frac{\partial}{\partial \theta_i} L(\theta \mid x) = 0, \quad i = 1, \ldots, k$$

(2.10)

MLE berhubungan dengan metode *Likelihood Ratio Test* (LRT) dalam penentuan statistik uji. LRT Λ dapat diperoleh dari proses pembagian:

$$\Lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})}$$

(2.11)

$L(\hat{\omega}) = \max_{\omega} L(\omega)$ dan $L(\hat{\Omega}) = \max_{\Omega} L(\Omega)$

dengan

ω : himpunan parameter di bawah hipotesis nol (H_0)

Ω : himpunan parameter di bawah populasi

L(ω) : fungsi *likelihood* di bawah H_0

L(Ω) : fungsi *likelihood* di bawah populasi

Keputusan tolak H_0 jika $\Lambda < \Lambda_0 < 1$.

C. PEMBAHASAN

1.1 Estimasi parameter GWR

Metode GWR adalah suatu teknik yang membawa kerangka dari model regresi sederhana menjadi model regresi yang terboboti (persamaan 2.3). Asumsi yang digunakan pada model ini adalah *error* berdistribusi normal dengan *mean* nol dan variansi konstan. Leung, 2000 menggunakan pendekatan *Weighted Least Square* (WLS) dalam mengestimasi parameter.

Pada penulisan ini akan digunakan pendekatan MLE dalam estimasi parameternya. Langkah awal dari pendekatan tersebut adalah dengan membentuk *likelihood* sebagai berikut:

$$
\begin{aligned}
L(\boldsymbol{\beta}_k, \sigma^2 \mid y) &= \prod_{i=1}^{n} f(y_i \mid \boldsymbol{\beta}_k, \sigma^2) \\
&= \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \left(y_i - \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} \right)^2 \right) \right) \qquad (3.1) \\
&= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} \right)^2 \right)
\end{aligned}
$$

Faktor letak geografis merupakan faktor pembobot pada model GWR. Faktor ini memiliki nilai yang berbeda untuk setiap daerah yang menunjukkan sifat lokal pada model. Oleh karena itu pembobot diberikan pada bentuk *likelihood* untuk mendapatkan model lokal GWR.

$$
\begin{aligned}
\exp\left(w_{i(j)}\right) L(\boldsymbol{\beta}_k, \sigma^2 \mid y) &= w_{i(j)} (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp\left(w_{i(j)}\right) \\
&\quad \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} \right)^2 \right)
\end{aligned}
$$

$$
\begin{aligned}
L^*(\beta_k, \sigma^2 \mid y) &= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \\
&\quad \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} w_{i(j)} \left(y_i - \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} \right)^2 \right)
\end{aligned}
$$

(3.2)

Setelah diperoleh bentuk *likelihood* yang terboboti (*L*)* kemudian dilakukan operasi logaritma natural pada model tersebut untuk memudahkan operasi matematik sehingga diperoleh estimasi parameternya.

$$\begin{aligned}
\ln L^*(\beta_k,\sigma^2 \mid y) &= \ln\left((2\pi)^{-n/2}\left(\sigma^2\right)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} w_{i(j)}\left(y_i-\beta_0(u_i,v_i)+\sum_{k=1}^{p}\beta_k(u_i,v_i)x_{ik}\right)^2\right)\right)\\
&= -\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln(\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n} w_{i(j)}\left(y_i-\beta_0(u_i,v_i)+\sum_{k=1}^{p}\beta_k(u_i,v_i)x_{ik}\right)^2\\
&= -\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln(\sigma^2)-\frac{1}{2\sigma^2}\mathrm{Q} \qquad (3.3)
\end{aligned}$$

dengan

$$\begin{aligned}
\mathrm{Q} &= \sum_{i=1}^{n} w_{i(j)}\left(y_i-\beta_0(u_i,v_i)+\sum_{k=1}^{p}\beta_k(u_i,v_i)x_{ik}\right)^2\\
&= \left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}(i)\right)^T\mathbf{W}(i)\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}(i)\right)
\end{aligned}$$

Estimasi parameter diperoleh dengan memaksimalkan bentuk ln*L** yaitu men*differensial*kan terhadap setiap parameternya.

1.1.1 Estimasi parameter $\boldsymbol{\beta}(i)$

Estimasi parameter $\boldsymbol{\beta}(i)$ diperoleh dengan men*differensial*kan persamaan (3.3) terhadap $\boldsymbol{\beta}(i)$ dengan penjabaran sebagai berikut:

$$\begin{aligned}
\frac{\partial\left(\ln L^*(\beta_k,\sigma^2\mid y)\right)}{\partial\left(\beta(i)\right)^T} &= \frac{\partial(Q)}{\partial\left(\beta(i)\right)^T}\\
0 &= \frac{\partial\left(\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}(i)\right)^T\mathbf{W}(i)\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}(i)\right)\right)}{\partial\left(\beta(i)\right)^T}\\
0 &= \frac{\partial\left(\mathbf{y}^T\mathbf{W}(i)\mathbf{y}-\mathbf{y}^T\mathbf{W}(i)\mathbf{X}\boldsymbol{\beta}(i)-\boldsymbol{\beta}(i)^T\mathbf{X}^T\mathbf{W}(i)\mathbf{y}+\boldsymbol{\beta}(i)^T\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\boldsymbol{\beta}(i)\right)}{\partial\left(\beta(i)\right)^T}\\
0 &= -\mathbf{X}^T\mathbf{W}(i)\mathbf{y}-\mathbf{X}^T\mathbf{W}(i)\mathbf{y}+2\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\boldsymbol{\beta}(i)\\
0 &= -2\mathbf{X}^T\mathbf{W}(i)\mathbf{y}+2\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\boldsymbol{\beta}(i)\\
-2\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\boldsymbol{\beta}(i) &= -2\mathbf{X}^T\mathbf{W}(i)\mathbf{y}
\end{aligned}$$

sehingga dari persamaan di atas diperoleh estimasi parameter $\beta(i)$ adalah $\hat{\boldsymbol{\beta}}(i)=\left(\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(i)\mathbf{y}$

(3.4)

Setelah diperoleh estimator $\boldsymbol{\beta}(i)$ maka akan dicari sifat-sifat dari estimator tersebut. Untuk menunjukkan sifat ke-takbias-an dari estimator diperoleh dengan cara berikut:

$$
\begin{aligned}
E\left(\hat{\boldsymbol{\beta}}(i)\right) &= E\left(\left(\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(i)\mathbf{y}\right)\\
&= \left(\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(i)E(\mathbf{y})\\
&= \left(\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\boldsymbol{\beta}(i)\\
&= \boldsymbol{\beta}(i)
\end{aligned}
$$

Pembuktian di atas menunjukkan bahwa estimator $\hat{\boldsymbol{\beta}}(i)$ merupakan estimator tak bias untuk $\boldsymbol{\beta}(i)$.

$$
\begin{aligned}
Var\left(\hat{\boldsymbol{\beta}}(i)\right) &= Var\left(\left(\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(i)\mathbf{y}\right)\\
&= \left(\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(i)Var(\mathbf{y})\left(\left(\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(i)\right)^T\\
&= \left(\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(i)\sigma^2\left(\left(\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(i)\right)^T\\
&= \left(\left(\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(i)\right)\left(\left(\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(i)\right)^T\sigma^2\\
&= \mathbf{C}\mathbf{C}^T\sigma^2
\end{aligned}
$$

dengan $\mathbf{C} = \left(\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(i)$

$$
\begin{aligned}
\left|Var\left(\hat{\boldsymbol{\beta}}(i)\right)\right| &= \left|\mathbf{C}\mathbf{C}^T\sigma^2\right|\\
&= \sigma^2\left|\mathbf{C}\mathbf{C}^T\right|
\end{aligned}
$$

Maka $\left|\mathbf{C}\mathbf{C}^T\right|$ harus sekecil mungkin agar $\hat{\boldsymbol{\beta}}(i)$ penaksir yang efisien.

$$
\lim_{n\to\sim} Var\left(\hat{\boldsymbol{\beta}}(i)\right) = \lim_{n\to\sim}\sigma^2\left|\mathbf{C}\mathbf{C}^T\right| = \lim_{n\to\sim}\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{n-(p+1)}\left|\mathbf{C}\mathbf{C}^T\right| = 0
$$

Karena $\hat{\boldsymbol{\beta}}(i)$ merupakan penaksir tak bias $\boldsymbol{\beta}(i)$ dan $\lim_{n\to\sim} Var\left(\hat{\boldsymbol{\beta}}(i)\right) = 0$ sehingga dapat dikatakan bahwa $\hat{\boldsymbol{\beta}}(i)$ merupakan penaksir yang konsisten.

1.1.2 Estimasi parameter σ^2

Estimasi parameter σ^2 diperoleh dengan men*differensial*kan persamaan (3.3) terhadap σ^2 dengan penjabaran sebagai berikut:

$$\frac{\partial\left(\ln L^*(\beta_k,\sigma^2 \mid y)\right)}{\partial\left(\sigma^2\right)} = \frac{\partial\left(-\frac{n}{2}\ln(\sigma^2)-\frac{1}{2\sigma^2}\mathrm{Q}\right)}{\partial\left(\sigma^2\right)}$$

$$0 = -\frac{n}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}\mathrm{Q}$$

$$0 = \frac{-n\sigma^2+\mathrm{Q}}{2(\sigma^2)^2}$$

$$0 = -n\sigma^2+\mathrm{Q}$$

$$n\sigma^2 = \mathrm{Q}$$

sehingga dari persamaan diatas diperoleh estimasi parameter σ^2 adalah: $\hat{\sigma}^2 = \frac{\mathrm{Q}}{n}$. Namun penaksir tersebut merupakan penaksir σ^2 yang bersifat global, sehingga hal ini tidak akan memiliki arti pada model GWR. Oleh sebab itu, penaksir σ^2 dihitung dengan sifat kelokalan model GWR.

Pada model GWR estimasi untuk tiap pengamatan ke-*i* ($\hat{y}_i$) adalah sebagai berikut:

$$\begin{aligned}\hat{y}_i &= x_i^T\hat{\boldsymbol{\beta}}(i)\\ &= x_i^T\left[\left(\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(i)\mathbf{y}\right]\end{aligned}$$

Secara umum untuk semua pengamatan dapat ditulis dalam bentuk:

$$\hat{\mathbf{y}} = \mathbf{S}\mathbf{y} \qquad (3.5)$$

dengan $\mathbf{S} = \begin{bmatrix} x_i^T\left(\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(i) \\ \vdots \\ x_n^T\left(\mathbf{X}^T\mathbf{W}(n)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(n)\end{bmatrix}$

Dari persamaan (12) dapat diperoleh

$$\begin{aligned}\hat{\boldsymbol{\varepsilon}} &= \mathbf{y}-\hat{\mathbf{y}}\\ &= \mathbf{y}-\mathbf{S}\mathbf{y}\\ &= \left(\mathbf{I}-\mathbf{S}\right)\mathbf{y}\end{aligned}$$

Nilai jumlah kuadrat *error*-nya adalah:

$$
\begin{aligned}
SSE &= \hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} \\
&= \left[(\mathbf{I}-\mathbf{S})\mathbf{y}\right]^T \left[(\mathbf{I}-\mathbf{S})\mathbf{y}\right] \\
&= \mathbf{y}^T (\mathbf{I}-\mathbf{S})^T (\mathbf{I}-\mathbf{S})\mathbf{y}
\end{aligned}
$$

(3.6)

Persamaan (3.6) juga dapat dimodifikasi dengan memperhatikan asumsi-asumsi pada model GWR yaitu:

$$
\begin{aligned}
E(\hat{\boldsymbol{\varepsilon}}) &= E(\mathbf{y}-\hat{\mathbf{y}}) \\
&= E(\mathbf{y}) - E(\hat{\mathbf{y}}) \\
&= X\boldsymbol{\beta}(i) - X\hat{\boldsymbol{\beta}}(i) \\
&= 0
\end{aligned}
$$

(3.7)

Varians *error*-nya yaitu:

$$
\begin{aligned}
Var(\boldsymbol{\varepsilon}) &= E\left[(\boldsymbol{\varepsilon}-E(\boldsymbol{\varepsilon}))^T (\boldsymbol{\varepsilon}-E(\boldsymbol{\varepsilon}))\right] \\
\sigma^2 &= E\left[(\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} - 0 - 0 + 0)\right] \\
\sigma^2 &= E(\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon})
\end{aligned}
$$

(3.8)

Dari persamaan (3.7) dan (3.8) maka persamaan (3.6) dapat dimodifikasi menjadi:

$$
\begin{aligned}
SSE &= \hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} \\
&= (\hat{\boldsymbol{\varepsilon}} - E(\hat{\boldsymbol{\varepsilon}}))^T (\hat{\boldsymbol{\varepsilon}} - E(\hat{\boldsymbol{\varepsilon}})) \\
&= \left[(\mathbf{I}-\mathbf{S})\mathbf{y} - E((\mathbf{I}-\mathbf{S})\mathbf{y})\right]^T \left[(\mathbf{I}-\mathbf{S})\mathbf{y} - E((\mathbf{I}-\mathbf{S})\mathbf{y})\right] \\
&= \left[(\mathbf{I}-\mathbf{S})(\mathbf{y}-E(\mathbf{y}))\right]^T \left[(\mathbf{I}-\mathbf{S})(\mathbf{y}-E(\mathbf{y}))\right] \\
&= (\mathbf{y}-E(\mathbf{y}))^T (\mathbf{I}-\mathbf{S})^T (\mathbf{I}-\mathbf{S})(\mathbf{y}-E(\mathbf{y})) \\
&= \boldsymbol{\varepsilon}^T (\mathbf{I}-\mathbf{S})^T (\mathbf{I}-\mathbf{S})\boldsymbol{\varepsilon}
\end{aligned}
$$

(3.9)

$$\begin{aligned} E(SSE) &= E\left(\boldsymbol{\varepsilon}^T (\mathbf{I}-\mathbf{S})^T (\mathbf{I}-\mathbf{S})\boldsymbol{\varepsilon}\right) \\ &= E\left(tr\left(\boldsymbol{\varepsilon}^T (\mathbf{I}-\mathbf{S})^T (\mathbf{I}-\mathbf{S})\boldsymbol{\varepsilon}\right)\right) \\ &= E\left(tr\left((\mathbf{I}-\mathbf{S})^T (\mathbf{I}-\mathbf{S})\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}\right)\right) \\ &= tr\left(\mathbf{I}^\mathbf{T}\mathbf{I} - \mathbf{I}^\mathbf{T}\mathbf{S} - \mathbf{S}^\mathbf{T}\mathbf{I} + \mathbf{S}^\mathbf{T}\mathbf{S}\right)E(\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}) \\ &= tr\left(\mathbf{I} - \mathbf{S} - \mathbf{S}^\mathbf{T} + \mathbf{S}^\mathbf{T}\mathbf{S}\right)\sigma^2 \\ &= \left(n - 2tr(\mathbf{S}) + tr(\mathbf{S}^\mathbf{T}\mathbf{S})\right)\sigma^2 \end{aligned}$$

(3.10)

Dari persamaan (3.9) diperoleh bahwa

$E\left(\frac{SSE}{\left(n-2tr(\mathbf{S})+tr(\mathbf{S}^\mathbf{T}\mathbf{S})\right)}\right) = \sigma^2$ dan diperoleh penaksir σ^2 yang tak bias adalah:

$$\hat{\sigma}*^2 = \frac{SSE}{\left(n-2tr(\mathbf{S})+tr(\mathbf{S}^\mathbf{T}\mathbf{S})\right)}$$

(3.11)

1.2 Pengujian Hipotesis

Pada tahap pengujian hipotesis akan digunakan metode LRT dengan menggunakan hasil yang telah diperoleh pada sub bab sebelumnya. Uji hipotesis yang pertama dilakukan adalah pengujian model secara serentak untuk menguji signifikansi dari faktor geografis yang merupakan inti dari model GWR. Bentuk hipotesisnya adalah:

H_0 : $\beta_k(u_i,v_i) = \beta_k \qquad k = 1,2,...,p$

(tidak ada pengaruh faktor geografis pada model)

H_1 : paling tidak ada satu $\beta_k(u_i,v_i) \neq \beta_k$

(ada pengaruh faktor geografis pada model)

Setelah terbentuk hipotesis maka langkah selanjutnya adalah menentukan himpunan parameter dibawah H_0 (ω):ω = $\{\boldsymbol{\beta}_k, \sigma^2\}$ dan membentuk fungsi *likelihood* dibawah H_0 (L(ω)) yang sama dengan persamaan (4.1). Kemudian dilanjutkan pada langkah penentuan himpunan parameter dibawah H_1 (Ω): Ω = $\{\boldsymbol{\beta}_k(u_i,v_i), \sigma^2\}$ dan pembentukan fungsi *likelihood* dibawah H_1 (L(Ω)) sesuai dengan persamaan (3.2).

Pada sub bab sebelumnya telah diperoleh estimasi untuk tiap parameter baik di bawah H_0 maupun H_1. Dari hasil estimasi tersebut maka disubtitusikan yang memaksimumkan fungsi likelihood baik di bawah H_0 maupun H_1. Langkah selanjutnya adalah menentukan statistik uji model GWR dengan pendekatan LRT, berdasar pada uji *F,* yang dapat digunakan untuk membandingkan model GWR dan model regresi global. Uji ini berdasarkan hasil SSE dibagi dengan banyak derajat bebas yang efektif (persamaan 3.10) yang mendekati distribusi χ^2 dengan derajat bebas banyak derajat bebas yang efektif. Secara matematis dituliskan sebagai berikut:

$$F = \frac{SSE_0 / n-(p+1)}{SSE_1 / \left(n - 2tr(\mathbf{S}) + tr(\mathbf{S}^{\mathbf{T}}\mathbf{S})\right)} = \frac{SSE_0 / df_1}{SSE_1 / df_2} \qquad (3.12)$$

Tolak H_0 jika $F_{hit} > F_{\alpha;(df1;df2)}$ atau P_value < α.

Jika pada pengujian model secara serentak diperoleh keputusan tolak H_0, maka langkah selanjutnya adalah melakukan uji parsial dengan hipotesis:

H_0 : $\beta_{1k} = \beta_{2k} = ... = \beta_{nk}$ untuk *k* yang diberikan (*k* = 1, 2, ..., *p*)

(tidak ada perbedaan yang signifikan antara satu daerah dengan daerah lainnya)

H_1 : paling tidak ada satu $\beta_{ik}\,(i = 1,2,...,n)$ yang berbeda

(ada perbedaan yang signifikan antara satu daerah dengan daerah lainnya)

Untuk melakukan pengujian di atas maka ditentukan terlebih dahulu varians sampel yang dinotasikan sebagai

$$V_k = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_{ik} - \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_{ik}\right)^2 \qquad (3.13)$$

dengan $\hat{\beta}_{ik}$ (*i* = 1, 2, ..., *n*) diperoleh dari persamaan (3.4).

Langkah selanjutnya adalah menentukan distribusi dari V_k di bawah H_0. Misalkan $\hat{\boldsymbol{\beta}}_k = \left(\hat{\beta}_{1k} \quad \hat{\beta}_{2k} \quad \cdots \quad \hat{\beta}_{nk}\right)^T$ dan **J** merupakan matriks *n* x *n* yang semua elemennya adalah 1, maka persamaan (3.12) dapat ditulis dengan:

$$V_k = \frac{1}{n}\hat{\boldsymbol{\beta}}_k{}^T\left[\mathbf{I} - \frac{1}{n}\mathbf{J}\right]\hat{\boldsymbol{\beta}}_k$$

(3.14)

Di bawah H_0 semua β_{ik} (i = 1, 2, ..., n) sama, sehingga dapat diasumsikan bahwa:

$$E(\hat{\beta}_{1k}) = E(\hat{\beta}_{2k}) = \cdots = E(\hat{\beta}_{nk}) = \mu_k$$

(3.15)

Secara umum dapat ditulis $E(\hat{\beta}_k) = \mu_k \mathbf{1}$

(3.16)

Dari persamaan (3.16) dan $\mathbf{1}^T\left[\mathbf{I} - \frac{1}{n}\mathbf{J}\right] = \mathbf{0}$ serta $\left[\mathbf{I} - \frac{1}{n}\mathbf{J}\right]\mathbf{1} = \mathbf{0}$, maka V_k dapat ditulis sebagai:

$$V_k = \frac{1}{n}\left(\hat{\boldsymbol{\beta}}_k - E\left(\hat{\boldsymbol{\beta}}_k\right)\right)^T\left[\mathbf{I} - \frac{1}{n}\mathbf{J}\right]\left(\hat{\boldsymbol{\beta}}_k - E\left(\hat{\boldsymbol{\beta}}_k\right)\right)$$

(3.17)

Misalkan $\mathbf{e}_k$ merupakan vektor yang elemennya bernilai 1 untuk elemen ke-k dan 0 untuk lainnya, maka diperoleh:

$$\beta_{ik} = e_k^T\hat{\boldsymbol{\beta}}(i) = e_k^T\left(\mathbf{X}^T\mathbf{W}(i)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(i)\mathbf{y}$$

$$\hat{\boldsymbol{\beta}}_k = \left(\hat{\beta}_{1k} \quad \hat{\beta}_{2k} \quad \cdots \quad \hat{\beta}_{nk}\right)^T = \mathbf{B}\mathbf{y}$$

(3.18)

dengan

$$\mathbf{B}^* = \begin{pmatrix} e_k^T\left(\mathbf{X}^T\mathbf{W}(1)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(1) \\ \vdots \\ e_k^T\left(\mathbf{X}^T\mathbf{W}(n)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(n) \end{pmatrix}$$

(3.19)

Subtitusi persamaan (3.18) ke persamaan (3.17) maka akan diperoleh:

$$V_k = \frac{1}{n}(\mathbf{y} - E(\mathbf{y}))^T \mathbf{B}^{*T} \left[\mathbf{I} - \frac{1}{n}\mathbf{J}\right]\mathbf{B}^*(\mathbf{y} - E(\mathbf{y}))$$
$$= \boldsymbol{\varepsilon}^T \left(\frac{1}{n}\mathbf{B}^{*T}\left[\mathbf{I} - \frac{1}{n}\mathbf{J}\right]\mathbf{B}^*\right)\boldsymbol{\varepsilon}$$

(3.20)

dengan $\boldsymbol{\varepsilon}$ berdistribusi Normal dengan mean nol dan varians σ^2 dan $\frac{1}{n}\mathbf{B}^T\left[\mathbf{I} - \frac{1}{n}\mathbf{J}\right]\mathbf{B}$ merupakan matriks semi definit positif.

Kemudian dengan menggunakan pendekatan yang sama pada pembahasan sebelumnya, maka akan ditentukan:

$$E(V_k) = E\left(\boldsymbol{\varepsilon}^T \left(\frac{1}{n}\mathbf{B}^T\left[\mathbf{I} - \frac{1}{n}\mathbf{J}\right]\mathbf{B}\right)\boldsymbol{\varepsilon}\right)$$
$$= E\left(tr\left(\boldsymbol{\varepsilon}^T \left(\frac{1}{n}\mathbf{B}^T\left[\mathbf{I} - \frac{1}{n}\mathbf{J}\right]\mathbf{B}\right)\boldsymbol{\varepsilon}\right)\right)$$
$$= E\left(tr\left(\frac{1}{n}\mathbf{B}^T\left[\mathbf{I} - \frac{1}{n}\mathbf{J}\right]\mathbf{B}\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\right)\right)$$
$$= tr\left(\frac{1}{n}\mathbf{B}^T\left[\mathbf{I} - \frac{1}{n}\mathbf{J}\right]\mathbf{B}\right)E(\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon})$$
$$= tr\left(\frac{1}{n}\mathbf{B}^T\left[\mathbf{I} - \frac{1}{n}\mathbf{J}\right]\mathbf{B}\right)\sigma^2$$

(3.21)

Distribusi dari E(V_k)/ σ^2 akan mendekati χ^2 dengan derajat bebas $tr\left(\frac{1}{n}\mathbf{B}^T\left[\mathbf{I} - \frac{1}{n}\mathbf{J}\right]\mathbf{B}\right)$.

Selanjutnya untuk menghitung nilai statistik uji adalah berdasar pada uji F yang merupakan rasio antara dua distribusi χ^2 yaitu

$$F_1 = \frac{V_k / tr\left(\frac{1}{n}\mathbf{B}^T\left[\mathbf{I} - \frac{1}{n}\mathbf{J}\right]\mathbf{B}\right)}{\hat{\sigma}^{*2}}$$

$$F_1 = \frac{V_k / df *_1}{\hat{\sigma} *^2} \quad (3.22)$$

dengan $\hat{\sigma} *^2$ sesuai dengan persamaan (3.11). Tolak H_0 jika $F_{1\text{-}hit} > F_{\alpha;(df*1;df2)}$ atau P_value < α.

D. KESIMPULAN

Estimasi parameter model GWR dengan pendekatan MLE menghasilkan estimator yang sama dengan pendekatan WLS yang sudah umum digunakan. Namun untuk estimasi σ^2 tidak dapat secara langsung menggunakan pendekatan MLE. Hal ini dikarenakan jika digunakan MLE secara langsung maka estimator akan bersifat global dan tidak memiliki arti pada model GWR yang bersifat lokal, sehingga digunakan pendekatan atas sifat kelokalan model GWR. Statistik uji yang digunakan baik untuk model serentak maupun parsial adalah dengan pendekatan LRT berdasar pada uji F.

DAFTAR PUSTAKA

Bocci, C., Petrucci, A., dan Rocco, E. (2006), “An Application of Geographically Weighted Regression to Agricultural Data for Small Area Estimates”, *Dipartimento di Statistica “G. Parenti”*, Universita degli Studi di Firenze, Italy.

Charlton, M., Fotheringham, S., dan Brunsdon, C. (2006), “Geographically Weighted Regression”, *Document ESRC National Centre for Research Methods*, NCRM Methods Review Papers.

Chasco, C., Garcia, I., dan Vicens, J. (2007), “Modeling spatial variations in household disposable income with Geographically Weighted Regression”, *Munich Personal RePEc Archive Paper No. 1682.*

Cressie, N.A.C. (1991), *Statistics For Spatial Data,* John Wiley & Sons, Inc. United States of America.

Draper, N.R. dan Smith, H. (1992), *Analisis Regresi Terapan,* Edisi Kedua, PT Gramedia Pustaka Utama, Jakarta.

Fotheringham, A.S., Brundson, C., dan Charlton, M. (2002) “*Geographically Weighted Regression: the analysis of spatially varying relationships*”, John Wiley & Sons Ltd, England.

Leung, Y. (2000), "Statistical Tests for Spatial Non-Stationarity Based on the Geographically Weighted Regression Model", *Journal*, The Chinese University of Hong Kong, Hong Kong.

Mittal, V., Kamakura, W.A., dan Govind, R. (2004), "Geographic Patterns in Customer Service and Satisfaction: An Empirical Investigation", *Journal of Marketing,* Vol. 68, Hal. 48-62.

Zhang, L. dan Gove, J.H. (2005), "Spatial Assessment of Model Errors from Four Regression Techniques", *Forest Science,*Vol. 51, No. 4, hal. 334-346.

Zhuang, D. (2006), "Spatial Dependence and Neighborhood Effects in Mortgage Lending: A Geographically Weighted Regression Approach", *Paper,*University of Southern California, Los Angeles.