BAB III

MIXED GEOGRAPHICALLY WEIGHTED REGRESSION (MGWR)

3.1 *Mixed Geographically Weighted Regression* (MGWR)

Model *Mixed Geographically Weighted Regression* (MGWR) merupakan model kombinasi atau gabungan antara regresi global dengan GWR yang mempertimbangkan situasi dimana beberapa variabel independen yang mempengaruhi variabel dependen bersifat global dan variabel independen yang lainnya bersifat lokal. Pada model MGWR beberapa koefisien pada model GWR diasumsikan konstan untuk seluruh titik pengamatan, sedangkan yang lain bervariasi sesuai lokasi pengamatan data. Model MGWR dengan p variabel independen dan q variabel independen diantaranya bersifat lokal, dengan mengasumsikan bahwa intersep model bersifat lokal (Purhadi dan Yasin, 2012). Model MGWR dapat dituliskan sebagai berikut :

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{q} \beta_k (u_i, v_i) x_{ik} + \sum_{k=q+1}^{p} \beta_k x_{ik} + \varepsilon_i \qquad (3.1)$$

dengan

y_i = nilai observasi variabel dependen ke-i

x_{ik} = nilai observasi variabel independen ke-k pada lokasi pengamatan ke-i

$\beta_0(u_i, v_i)$ = konstanta atau intersep pada pengamatan ke-i

(u_i, v_i) = koordinat letak geografis (*longitude, latitude*) dari lokasi pengamatan ke-i

$\beta_k(u_i, v_i)$ = koefisien regresi observasi variabel independen ke-k pada lokasi pengamatan ke-i

β_k = koefisien regresi observasi variabel independen ke-k

ε_i = error pengamatan ke-i diasumsikan identik, independen dan berdistribusi normal dengan mean nol dan varian konstan ó^2.

i = 1 , 2 , ... , n

Model MGWR adalah model regresi yang beberapa koefisien dari peubah independennya bersifat konstan, sedangkan yang lainnya bervariasi secara spasial. Penggabungan dari model GWR dan model MGWR tersebut didapat setelah dilakukan pengujian variabilitas spasial. Salah satu prosedur dalam analisis model MGWR adalah pengujian variabilitas spasial untuk menentukan koefisien global dan koefisien lokal. Pengujian ini dilakukan menggunakan statistik uji F, dengan langkah-langkahnya seperti berikut :

1) Perumusan hipotesis :

   $H_0 :\ \beta_k(u_i, v_i) = \beta_k$

   Untuk i = 1,2,..,n dan k = indeks koefisien yang diasumsikan global

   $H_1 : \beta_k\,(u_i, v_i) \neq \beta_k$

   Untuk k = indeks koefisien yang diasumsikan global

2) Statistik uji :

$$F = \frac{(JKR_{MGWR} - JKR_{GWR})/_{df_1}}{JKR_{GWR}/_{df_2}} \quad = \left[\frac{(Y^T R_2 Y - Y^T R_1 Y)}{df_1}\right]\left[\frac{(Y^T R_1 Y}{df_2}\right]^{-1} \qquad (3.2)$$

   Dimana :

   JKR_{MGWR} adalah JKR model dengan koefisien ke-k global dan koefisien lain bervariasi spasial ;

   JKR_{GWR} adalah JKR model GWR awal dengan $R_2 = \left((I - S_2)^T (I - S_2)\right)$ dan $R_1 = \left((I - S_1)^T (I - S_1)\right)$;

   Derajat bebas $df_1 = tr[(I - S_2) - (I - S_1)]$;

   Derajat bebas $df_2 = tr[(I - S_1)]$ (Chang Lin Mei, 2005 : 4-12).

3) Kriteria pengujian :

   Tolak H_0 jika $F \geq F_{\alpha, df_1, df_2}$ atau dapat dikatakan suatu koefisien memiliki pengaruh yang nyata.

Untuk menguji model MGWR, maka dilakukan uji hipotesis kesesuaian model regresi global dan MGWR, dengan langkah-langkah sbb:

1) Perumusan hipotesis :

$H_0 : \beta_k(u_i, v_i) = \beta_k$ untuk i = 1,2,..,n dan k = 0,1,2,...,q

(Model MGWR tidak berbeda dengan Model Regresi Global).

$H_1 : Minimal\ ada\ satu\ \beta_k\ (u_i, v_i) \neq \beta_k$ dengan k = 0,1,2,...,p dan i = 1 ,2,...,n

(Model MGWR berbeda dengan Model Regresi Global)

2) Statistik uji :

$$F(1) = \left[\frac{y^T[(I-H)-((I-S)^T(I-S))]^y/v_1}{y^T(I-S)^T(I-S)y/_{u_1}}\right] \tag{3.3}$$

Dimana :

$S = S_l + (I - S_l)X_g[X_g^T(I - S_l)^T(I - S_l)X_g]^{-1}X_g^T(I - S_l)^T(I - S_l)$;

$v_i = tr\left([(I - H) - (I - S)^T(I - S)]^i\right)$;

$u_i = tr([(I - S)^T(I - S)]^i), i = 1,2$;

$H = X(X^TX)^{-1}X^T$;

$df_1 = \left[\frac{V_1{}^2}{V_2{}^2}\right]$ dan $df_2 = \left[\frac{u_1{}^2}{u_2{}^2}\right]$.

3) Kriteria pengujian :

Tolak H_0 jika $F(1) \geq F_{\alpha,df_1,df_2}$

Selanjutnya dilakukan pengujian parsial untuk mengetahui parameter variabel independen global yang berpengaruh secara signifikan terhadap variabel dependen, dengan langkah-langkahnya sbb :

1) Perumusan hipotesis :

$H_0 : \beta_k(u_i, v_i) = 0$ (variabel global x_k tidak signifikan)

$H_1 : \beta_k(u_i, v_i) \neq 0$ (variabel global x_k signifikan)

Untuk k = 1,2,...,p

2) Statistik uji :

$$T_g = \frac{\widehat{\beta}_k}{\widehat{\sigma}\sqrt{g_{kk}}} \tag{3.4}$$

dimana :

g_{kk} adalah elemen diagonal ke-k dari matriks GG^T signifikansi sebesar α ,

$G = \left[X_g{}^T(I - S_l)^T(I - S_l)X_g\right]^{-1} X_g^T(I - S_l)^T(I - S_l)$;

$\widehat{\sigma}^2 = \frac{y^T(I-S)^T(I-S)y}{tr((I-S)^T(I-S))}$.

3) Kriteria pengujian :

Tolak H_o jika $\left|T_{ghit}\right| > t_{\alpha/2}, df$, dimana $df \quad = \left[\frac{u_1{}^2}{u_2}\right]$

Untuk mengetahui pengaruh signifikan parameter variabel independen yang bersifat lokal dilakukan uji parsial, dengan langkah-langkah sbb :

1) Perumusan hipotesis :

$H_0 :\ \beta_k(u_i, v_i) = 0$

(variabel global x_k pada lokasi ke-i tidak signifikan)

$H_1 :\ \beta_k(u_i, v_i) \neq 0$

(variabel global x_k pada lokasi ke-i signifikan)

Untuk k = 1,2,...,q dan i = 1 , 2,....,n

2) Statistik uji :

$$T_{lhit} = \frac{\widehat{\beta}_k(u_i,v_i)}{\widehat{\sigma}\sqrt{m_{kk}}} \tag{3.5}$$

Dimana :

m_{kk} adalah elemen diagonal ke-k dari matriks MM^T.

dengan $M = \left[X_l{}^T W(u_i, v_i)X_l\right]^{-1} X_l^T W(u_i, v_i)(I - X_g G)$

3) Kriteria Pengujian :

Tolak H_o jika $|T_{lhit}| > t_{\alpha/2}, df$, dimana $df \quad = \left[\frac{u_1{}^2}{u_2}\right]$

3.2 *Akaike Information Criterion Corrected* (AIC_C)

Akaike Information Criterion Corrected (AIC_c) merupakan pengembangan dari *Akaike Information Criterion* (AIC). Pengukuran untuk kualitas relatif dari model statistik berdasarkan data yang diberikan untuk pemodelan model terbaik dari beberapa model yang ada dinyatakan dengan AIC, dengan rumus :

$$AIC = 2k - 2\ln(likelihood) \qquad (3.6)$$

Dimana :

k = banyak parameter yang akan di taksir

$\ln(likelihood)$ = nilai maksimum *likelihood* model

Ukuran yang digunakan untuk mengukur kebaikan model (*goodnees-of-fit*) dan mempertimbangkan prinsip *parsimony* adalah AIC. Namun ukuran ini dinilai bias pada sampel kecil, sehingga ukuran ini dikoreksi dengan AIC *Corrected* (AICc). Rumusnya adalah :

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1} \qquad (3.7)$$

Dimana :

n = ukuran sampel

Jika nilai k semakin besar atau variabel yang ditaksirnya semakin banyak, maka untuk penggunaan nilai AIC_c akan lebih baik daripada dengan nilai AIC.

Alasan digunakannya AICc adalah berawal dari prinsip *parsimony* yang menyatakan bahwa model terbaik diharapkan terbentuk dari koefisien/parameter regresi yang tidak banyak tapi mampu menjelaskan model secara keseluruhan. (*Multimodel Inference Understanding AIC and BIC Model Selection*, Burnham & Anderson, 2004 : 270).

3.3 Pembobotan

Nilai pembobot pada model GWR sangat penting, karena mewakili letak data observasi satu dengan lainnya. Masing-masing observasi memiliki nilai pembobotan sebesar 1 pada model regresi global tanpa pembobotan geografis. Pada model GWR, pembobotan bervariasi sesuai lokasi pada titik regresi ke-i, dimana $0 \leq w_{ij} \leq 1$ dan w_{ij} semakin kecil ketika jarak d_{ij} bertambah. Berarti jika observasi dekat dengan titik regresi, maka akan memberikan bobot yang besar dibandingkan dengan yang jauh dari titik regresi. Fungsi yang digunakan adalah fungsi Kernel untuk mengestimasi paramater dalam model GWR. Pembobot yang terbentuk dengan menggunakan fungsi Kernel ini adalah

1) Fungsi *Fixed Gaussian* :

$$w_j(u_i,v_i) = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right) \tag{3.8}$$

2) Fungsi *Fixed Bisquare* :

$$w_j(u_i,v_i) = \begin{cases} \left(1-(d_{ij}/h)^2\right)^2, & jika \quad d_{ij} \leq h \\ 0, & jika \quad d_{ij} > h \end{cases} \tag{3.9}$$

3) Fungsi *Adaptive Bisquare* :

$$w_j(u_i,v_i) = \begin{cases} \left(1-(d_{ij}/h_i)^2\right)^2, & jika \quad d_{ij} \leq h \\ 0, & jika \quad d_{ij} > h \end{cases} \tag{3.10}$$

4) Fungsi *Adaptive Gaussian* :

$$w_j(u_i,v_i) = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h_{i(p)}}\right)^2\right) \tag{3.11}$$

Dengan d_{ij} jarak antara lokasi (u_i, v_i) dan h adalah parameter penghalus (*bandwith*) dengan rumus :

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + (v_i - v_j)^2} \tag{3.12}$$

Dimana :

u_i = *Longitude* pada lokasi i ;

u_j = *Longitude* pada lokasi j ;

v_i = *Latitude* pada lokasi i ;

v_j = *Latitude* pada lokasi j ;

h = parameter non negatif yang diketahui dan biasanya disebut parameter penghalus *(bandwidth)*.

Bandwidth merupakan radius dari suatu lingkaran, sehingga jika sebuah titik lokasi berada di dalam radius lingkaran tersebut, maka masih dianggap memiliki pengaruh terhadap penaksiran koefisien regresi pada titik lokasi i tersebut. Dalam pembentukan model GWR, *bandwidth* berperan penting karena akan berpengaruh pada ketepatan model terhadap data, yaitu mengatur varians dan bias dari model.

Fungsi dari *bandwidth* adalah untuk menentukan bobot dari suatu lokasi terhadap lokasi lain yang digunakan sebagai pusat. Semakin dekat wilayah dengan daerah pusat, akan semakin besar pula pengaruh yang diberikan. Ada beberapa metode yang digunakan untuk memilih *bandwidth* optimum, salah satu diantaranya adalah metode *Cross Validation (*CV*)*. Rumusnya adalah :

$$CV(h) = \sum_{i=1}^{n} (y_i - \hat{y}_{i \neq i}(h))^2 \qquad (3.13)$$

Dimana :

$\hat{y}_{i \neq i}(h)$ merupakan nilai penaksir (y_i) dimana pengamatan di lokasi (u_i, v_i) dihilangkan dari proses estimasi. Untuk mendapatkan nilai h yang optimal, maka diperoleh dari h yang menghasilkan nilai CV yang minimum.

Proses untuk mendapatkan *bandwidth* yang meminimumkan nilai CV bisa dilakukan juga dengan metode *golden section search* pada *software* GWR4.

3.4 Estimasi Parameter

Model MGWR dalam mengestimasi parameternya dapat menggunakan pendekatan *Weighted Least Square (WLS)* (Mei, dkk. 2004). Langkah awal yaitu dengan membentuk matriks pembobot untuk setiap lokasi pengamatan. Estimasi parameter untuk model MGWR adalah sebagai berikut :

$$y = X_l \beta_l(u_i, v_i) + X_g \beta_g + \varepsilon \qquad (3.14)$$

Dimana :

X_g : matriks variabel independen global

X_l : matriks variabel independen lokal

β_g : vektor parameter variabel independen global

$\beta_l(u_i, v_i)$: matriks parameter variabel independen lokal

$$X_l = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1q} \\ 1 & x_{21} & \cdots & x_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nq} \end{bmatrix}, \quad X_g = \begin{bmatrix} x_{1,(q+1)} & x_{1,(q+2)} & \cdots & x_{1p} \\ x_{2,(q+1)} & x_{2,(q+2)} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,(q+1)} & x_{n,(q+2)} & \cdots & x_{np} \end{bmatrix}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \; dan \; \beta_l(u_i, v_i) = \begin{bmatrix} \beta_0(u_i, v_i) \\ \beta_1(u_i, v_i) \\ \vdots \\ \beta_q(u_i, v_i) \end{bmatrix}, \beta_g = \begin{bmatrix} \beta_{(q+1)} \\ \beta_{(q+2)} \\ \vdots \\ \beta_p \end{bmatrix}, i = 1,2, \dots, n$$

(Chang-Lin Mei , GWR *Technique for Spatial Data Analysis* Halaman 9)

$$\tilde{y}_i = y_i - \sum_{k=1}^{q} \beta_k (u_i, v_i) x_{ik} = \sum_{k=q+1}^{p} \beta_k x_{ik} + \varepsilon_i \; , i = 1 \, , 2 \, , \dots, n \qquad (3.15)$$

Estimator parameter untuk model GWR adalah

$$\hat{\beta}_l(u_i, v_i) = \left[X_l{}^T W(u_i, v_i) X_l \right]^{-1} X_l^T W(u_i, v_i) \tilde{y} \qquad (3.16)$$

Dengan W adalah diagonal matrik pembobot berukuran n x n .

$$W(u_i, v_i) = diag(w_1(u_i, v_i), w_2(u_i, v_i), \dots, w_n(u_i, v_i)) \qquad (3.17)$$

(*Journal of regional science : A note on the MGWR Model*, Changlin Mei, 2004 :145).

Misalkan $x_{li}^T = (1, x_{i1}, x_{i2}, \dots, x_{iq})$ adalah elemen baris ke-i dari matriks X_l. Maka nilai prediksi untuk $\tilde{y}$ pada (u_i, v_i) untuk seluruh pengamatan dapat dituliskan sebagai berikut :

$$\hat{\tilde{y}} = (\hat{\tilde{y}}_1, \hat{\tilde{y}}_2, \dots \hat{\tilde{y}}_n)^T = S_l \tilde{y} \qquad (3.18)$$

Dimana :

$$S_l = \begin{pmatrix} x_{l1}^T (X_l{}^T W(u_1, v_1) X_l)^{-1} X_l{}^T W(u_1, v_1) \\ x_{l2}^T (X_l{}^T W(u_2, v_2) X_l)^{-1} X_l{}^T W(u_2, v_2) \\ \vdots \\ x_{ln}^T (X_l{}^T W(u_n, v_n) X_l)^{-1} X_l{}^T W(u_n, v_n) \end{pmatrix} \qquad (3.19)$$

Setelah itu disubtitusikan elemen dari $\hat{\beta}_l(u_i, v_i)$ kedalam model MGWR pada persamaan berikut :

$$\tilde{y} = y - X_g \beta_g = X_l \beta_l(u_i, v_i) + \varepsilon \qquad (3.20)$$

Menurut metode *Ordinary Least Squares*(OLS) diperoleh estimasi koefisien konstan, sehingga diperoleh untuk estimasi parameter bersifat global adalah

$$\hat{\beta}_g = \left[X_g{}^T (I - S_l)^T (I - S_l) X_g \right]^{-1} X_g^T (I - S_l)^T (I - S_l) y \qquad (3.21)$$

Dengan mensubstitusikan $\widehat{\beta}_g$ kedalam persamaan (3.16) maka akan diperoleh estimasi untuk koefisien lokal pada lokasi (u_i, v_i) adalah

$$\widehat{\beta}_l(u_i, v_i) = \left[X_l{}^T W(u_i, v_i) X_l\right]^{-1} X_l^T W(u_i, v_i)(y - X_g \widehat{\beta}_g) \qquad (3.22)$$

Untuk i = 1 , 2 , ... , n dengan $S_g = X_g \left(X_g^T X_g\right)^{-1} X_g^T$

$$\widehat{\widehat{y}} = (S_l \widetilde{y}) = S_l(y - X_g \widehat{\beta}_g) \qquad (3.33)$$

Oleh karena itu, nilai *fitted-value* dari dependen untuk n lokasi pengamatan adalah $\hat{y} = S_y$, dengan $S = S_l + (I - S_l) X_g [X_g^T \left(I - S_l\right)^T (I - S_l) X_g\big]^{-1} X_g^T (I - S_l)^T (I - S_l)$.

Estimator $\widehat{\beta}_g$ merupakan estimator tak bias dan efisien untuk β_g. Sedangkan estimator $\widehat{\beta}_l(u_i, v_i)$ merupakan estimator tak bias dan efisien untuk $\beta_l(u_i, v_i)$. (Purhadi dan Yasin, 2012 : 531).