BAB II

TINJAUAN PUSTAKA

1.1 Analisis Regresi Berganda

Analisis regresi adalah salah satu metode statistika yang mempelajari pola hubungan secara matematis antara satu variabel *endogenus* (Y) dengan satu atau lebih variabel *eksogenus* (X). Menurut Drapper dan Smith dalam Safitri (2014) hubungan antara satu variabel *endogenus* dengan satu atau lebih variabel *eksogenus* dapat di nyatakan dalam model regresi linier. Secara umum hubungan tersebut dapat dinyatakan sebagai berikut :

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \qquad (2.1)$$

dengan:

Y_i = variabel *endogenus* untuk pengamatan ke-i, untuk i = 1, 2, …, n.

$\beta_0, \beta_1, \beta_2, \ldots, \beta_{p-1}$ = parameter.

$X_{i,1}, X_{i,2}, \ldots, X_{i,p-1}$ = variabel *eksogenus*.

ε_i = sisa (*error*) untuk pengamatan ke-i yang diasumsikan berdistribusi normal yang saling bebas dan identik dengan rata-rata 0 (nol) dan varians σ^2 .

Dalam notasi matriks persamaan di atas dapat ditulis menjadi :

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} X = \begin{bmatrix} 1 & X_{1.1} & X_{1.2} & \cdots & X_{1,p-1} \\ 1 & X_{2.1} & X_{2.2} & \cdots & X_{2.p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n.1} & X_{n.2} & \cdots & X_{n.p-1} \end{bmatrix} \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad (2.2)$$

dimana :

$\mathbf{Y}$ = vektor variabel tidak bebas berukuran n x 1.

$\mathbf{X}$ = matriks variabel bebas berukuran n x (p – 1).

$\boldsymbol{\beta}$ = vektor parameter berukuran p x 1.

$\boldsymbol{\varepsilon}$ = vektor *error* berukuran n x 1.

1.2 Pemodelan *Spatial*

Tobler (1970) mengemukakan hukum pertama tentang geografi, yaitu kondisi pada salah satu titik atau area berhubungan dengan kondisi pada salah satu titik atau area yang berdekatan. Hukum tersebut merupakan dasar pengkajian permasalahan berdasarkan efek lokasi atau *spatial*. Model regresi klasik jika digunakan sebagai alat analisis pada permodelan data *spatial*, dapat menyebabkan kesimpulan yang kurang tepat karena asumsi *error* saling bebas dan asumsi homogenitas tidak terpenuhi.

Anselin (1988) menjelaskan dua efek spatial dalam ekonometrika meliputi efek *spatial dependence* dan *spatial heterogenity*. *Spatial dependence* menunjukkan adanya keterkaitan (*autocorrelation*) antar lokasi obyek penelitian (*cross sectional data set*). *Spatial heterogenity* mengacu pada keragaman bentuk fungsional dan parameter pada setiap lokasi, lokasi-lokasi kajian menunjukkan ketidak homogenan dalam data.

Menurut LeSage (1999) dan Anselin (1988), secara umum model *spatial* dapat dinyatakan dalam bentuk persamaan (2.3) dan (2.4)

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \qquad (2.3)$$

dengan

$$\mathbf{u} = \lambda\mathbf{W}\mathbf{u} + \varepsilon$$

$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}) \tag{2.4}$$

dimana

$\mathbf{y}$	: vektor variabel *endogenus*, berukuran $n \times 1$
$\boldsymbol{X}$	: matriks variabel *eksogenus*, berukuran $n \times (k+1)$
$\boldsymbol{\beta}$	: vektor parameter koefisien regresi, berukuran $(k+1) \times 1$
ρ	: parameter koefisien *spatial lag* variabel *endogenus*
λ	: parameter koefisien *spatial lag* pada *error*
$\mathbf{u}$	: vektor *error* pada persamaan (2.3) berukuran $n \times 1$
ε	: vektor *error* pada persamaan (2.4) berukuran $n \times 1$, yang berdistribusi normal dengan mean nol dan varians $\sigma^2 \mathbf{I}$
W	: Matriks pembobot, berukuran n x n
I	: matriks identitas, berukuran $n \times n$
n	: banyaknya amatan atau lokasi (i = 1, 2, 3, …, n)
k	: banyaknya variabel independen (k = 1, 2, 3, ..., l)

Error regresi (**u**) yang diasumsikan memiliki efek lokasi random dan mempunyai autokorelasi secara spatial. W_1 dan W_2 merupakan pembobot yang menunjukan hubungan *continguity* atau fungsi jarak antar lokasi dan diagonalnya bernilai nol. Berikut ini adalah bentuk matriks persamaan (2.3).

$$\mathbf{y} = [y_1 \quad y_2 \quad \ldots \quad y_n]^T \quad \mathbf{u} = [u_1 \quad u_2 \quad \cdots \quad u_n]^T \quad \boldsymbol{\varepsilon} = [\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_n]^T$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & \cdots & w_{1\mathrm{n}} \\ w_{21} & w_{13} & w_{23} & \cdots & w_{2\mathrm{n}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{\mathrm{n}1} & w_{\mathrm{n}2} & w_{\mathrm{n}3} & \cdots & w_{\mathrm{nn}} \end{bmatrix}$$

Pemodelan *spatial* diantaranya yaitu:

Spatial Autoregressive Model (SAR) terjadi apabila λ = 0, seperti pada persamaan (2.5)

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= (0)\,\mathbf{W}\mathbf{u} + \varepsilon \\ \mathbf{u} &= \varepsilon \end{aligned} \tag{2.5}$$

model persamaan di atas mengasumsikan bahwa proses *autoregressive* hanya pada variabel *endogenus*.

1.3 Uji Dependensi *Spatial*

Dependensi *spatial* menunjukkan bahwa pengamatan di suatu lokasi bergantung pada pengamatan di lokasi lain yang letaknya berdekatan. Pengukuran dependensi *spatial* bisa menggunakan Moran's I. Hipotesis yang digunakan adalah :

H_o : I_M = 0 (tidak ada autokorelasi antar lokasi)

H_1 : I_M ≠ 0 (ada autokorelasi antar lokasi)

Statistik uji (Lee dan Wong, 2001) disajikan pada persamaan berikut.

$$Z_{hitung} = \frac{\mathrm{I_M} - \mathrm{I_{Mo}}}{\sqrt{\mathrm{var}(\mathrm{I_M})}} \tag{2.6}$$

dimana

$$I_M = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$E(I_M) = -\frac{1}{n-1}$$

$$\mathrm{var}(I_M) = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_o^2}{(n+1)(n-1)S_o^2}$$

$$S_1 = \frac{1}{2}\sum_{i\neq j}^{n}(w_{ij} + w_{ij})^2 \qquad S_2 = \sum_{i=1}^{n}(w_{io} + w_{oi})^2$$

$$S_o = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \qquad w_{io} = \sum_{j=1}^{n} w_{ij} \qquad w_{oi} = \sum_{j=1}^{n} w_{ji}$$

keterangan :

x_i = data ke-i (i = 1, 2, ..., n)

x_j = data ke-j (j = 1, 2, ..., n)

$\bar{x}$ = rata-rata data

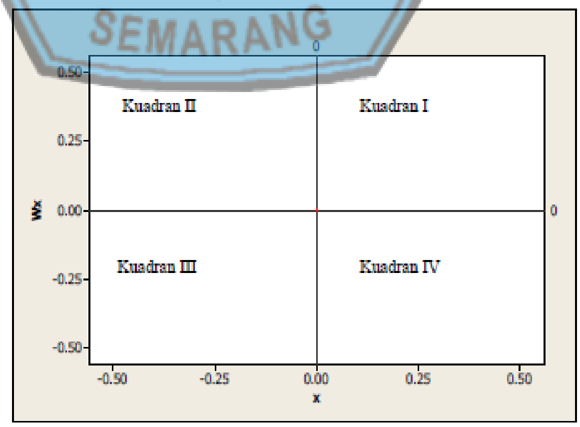
w_{ij} = elemen matriks bobot spatial

var (I_M) = varians Moran's I

E(I_M) = *expected value* Moran's I

Pengambilan keputusannya adalah Ho ditolak jika $|Z_{hitung}| > Z_{\alpha/2}$. Nilai dari indeks I adalah antara -1 dan 1. Apabila I > Io maka data memiliki autokorelasi positif, jika I < Io maka data memiliki autokorelasi negatif. Pola pengelompokan dan penyebaran antar lokasi dapat juga disajikan dengan Moran's *Scatterplot.* Moran's *Scatterplot* menunjukkan hubungan antara nilai amatan pada suatu lokasi (distandarisasi) dengan rata-rata nilai amatan dari lokasi-lokasi yang bertetanggaan dengan lokasi yang bersangkutan (Lee dan Wong, 2001). *Scatterplot* tersebut terdiri atas empat kuadran, yaitu kuadran I, II, III, dan IV. Lokasi-lokasi yang banyak berada di kuadran I dan III cenderung memiliki autokorelasi positif, sedangkan lokasi-lokasi yang banyak berada di kuadran II dan IV cenderung memiliki autokorelasi negatif. Berikut adalah penjelasan dari masing-masing kuadran (Perobelli dan Haddad, 2003).

- Kuadran I (*High-High*), menunjukkan lokasi yang mempunyai nilai amatan tinggi dikelilingi oleh lokasi yang mempunyai nilai amatan tinggi.
- Kuadran II (*Low-High*), menunjukkan lokasi yang mempunyai nilai amatan rendah dikelilingi oleh lokasi yang mempunyai nilai amatan tinggi.
- Kuadran III (*Low-Low*), menunjukkan lokasi yang mempunyai nilai amatan rendah dikelilingi oleh lokasi yang mempunyai nilai amatan rendah.
- Kuadran IV (*High-Low*), menunjukkan lokasi yang mempunyai nilai amatan tinggi dikelilingi oleh lokasi yang mempunyai nilai amatan rendah.



Gambar 2.1 *Moran's Scatterplot*

1.4 Uji Model *Spatial*

Untuk mengetahui adanya efek *spatial* pada observasi model SAR, SEM dan SAC maka dapat diuji dengan menggunakan statistik *Lagrange Multiplier test* (Anselin, 1988). Hipotesis yang digunakan adalah sebagai berikut:

H_0 : $\rho = 0$ (tidak adanya dependensi *spatial autoregressive* dalam model)

H_1 : $\rho \neq 0$ (ada dependensi *spatial autoregressive* dalam model)

Statistik uji yang digunakan adalah *Lagrange Multiplier test* :

$$LM = E^{-1}\left[(R_y)^2 T_{22} - 2R_y R_e T_{12} + (R_e)^2 (D + T_{11})\right] \quad (2.7)$$

Pengambilan keputusan adalah H_0 ditolak jika $LM > \chi^2_{(\alpha,m)}$. Hal ini karena nilai LM hitung secara *asymptotically* mengikui distribusi *Chi-Square*, dengan m = jumlah parameter *spatial* (SAR = 1, SEM = 1, SAC = 2)

$$R_y = \frac{\mathbf{e}^T \mathbf{W}\mathbf{e}}{\sigma^2}$$

$$R_e = \frac{\mathbf{e}^T \mathbf{W}\mathbf{e}}{\sigma^2}$$

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$$

$$T_{fg} = tr\left\{\mathbf{W}_f \mathbf{W}_g + \mathbf{W}_f^T \mathbf{W}_g\right\} \text{ dimana } f,g = 1,2$$

$$D = \sigma^{-2}(\mathbf{WX\beta})^T \mathbf{M}(\mathbf{WX\beta})$$

$$E = (D + T_{11})T_{22} - (T_{12})^2$$

e adalah *least square residual* untuk observasi. Jika matriks pembobot *spatial* adalah $\mathbf{W}$ maka $T_{11} = T_{12} = T_{22} = T = tr\left\{(\mathbf{W}^T + \mathbf{W})\mathbf{W}\right\}$

Nilai statistik uji mengikuti distribusi asymtotik *Chi-Square* dengan derajat bebas m. Jika nilai $LM > \chi^2_{(\alpha,m)}$ maka Ho akan ditolak. Sehinga terdapat dependensi *spatial*.

1.5 Matriks Pembobot *Spatial* (*Spatial Weighting Matrix*)

Matriks pembobot *spatial* (**W**) dapat diperoleh dari ketersinggungan antar wilayah dan jarak dari ketetanggaan (*neighborhood*) atau jarak antara satu *region* dengan *region* yang lain. Menurut LeSage (1999), ada beberapa metode untuk mendefinisikan hubungan persinggungan (*contiguity*) antar wilayah, antaralain sebagai berikut :

1. *Linear Contiguity* (Persinggungan tepi); mendefinisikan W_{ij} = 1 untuk *region* yang berada di tepi (*edge*) kiri maupun kanan *region* yang menjadi perhatian, W_{ij} = 0 untuk *region* lainnya.
2. *Rook Contiguity* (Persinggungan sisi); mendefinisikan W_{ij} = 1 untuk *region* yang bersisian (*common side*) dengan *region* yang menjadi perhatian, W_{ij} = 0 untuk *region* lainnya.
3. *Bhisop Contiguity* (Persinggungan sudut); mendefinisikan W_{ij} = 1 untuk *region* yang titik sudutnya (*common vertex*) bertemu dengan sudut *region* yang menjadi perhatian, W_{ij} = 0 untuk *region* lainnya.
4. *Double Linear Contiguity* (Persinggungan dua tepi); mendefinisikan W_{ij} = 1 untuk dua *entity* yang berada di sisi (*edge*) kiri dan kanan *region* yang menjadi perhatian, W_{ij} = 0 untuk *region* lainnya.

5. *Double Rook Contiguity* (Persinggungan dua sisi); mendefinisikan W_{ij} = 1 untuk dua *entity* di kiri, kanan, utara dan selatan *region* yang menjadi perhatian, W_{ij} = 0 untuk *region* lainnya.
6. *Queen Contiguity* (persinggungan sisi-sudut); mendefinisikan W_{ij} = 1 untuk *entity* yang bersisian (*common side*) atau titik sudutnya (*common vertex*) bertemu dengan *region* yang menjadi perhatian, W_{ij} = 0 untuk *region* lainnya.

Dalam Penelitian ini menggunakan pembobot *Queen customize* karena matriks pembobot *spatial* ini tidak hanya mempertimbangkan faktor persinggungan dan kedekatan antar lokasi wilayah akan tetapi faktor-faktor lainnya yang disesuaikan dengan karakteristik masalahnya. Karakteristik yang dimaksud adalah adanya hubungan saling mempengaruhi antar wilayah karena memiliki hubungan timbal balik. Dimana W=1 untuk wilayah yang bersisian (*common size*) atau titik sudutnya (*common vertex*) bertemu dengan wilayah yang menjadi perhatian, Wij=0 untuk wilayah lainnya.

1.6 Model SAR (*Spatial Autoregressive Model*)

Menurut Anselin (1988), Model Spatial Autoregresive adalah model yang mengkombinasikan model regresi sederhana dengan lag spasial pada variabel dependen dengan menggunakan data cross section. Model spasial autoregressive terbentuk apabila W2 = 0 dan ρ = 0 , sehingga model ini mengasumsikan bahwa proses autoregressive hanya pada variabel respon (Lee dan Yu, 2010). Model umum SAR panel ditunjukan oleh persamaan sebagai berikut:

$$Y_{it} = \rho \sum_{j=1}^{N} W_{ij} Y_{it} + \alpha + X_{it}\beta + \varepsilon_{it}$$

Menurut Abdul Karim, et al. dimana Yit merupakan variabel respon pada unit observasi ke-i dan waktu ke-t, ρ adalah koefisien spasial autoregressive dan Wij adalah elemen matriks pembobot spasial, Xit adalah variabel prediktor pada unit observasi ke-i dan waktu ke-t, β adalah koefisien slope, α adalah intersep model regresi, εit adalah komponen error pada unit observasi ke-i dan waktu ke-t.

1.7 Demam Berdarah *Dengue* (DBD)

Demam berdarah dengue (DBD) adalah suatu penyakit yang disebabkan oleh infeksi virus dengue yang ditularkan melalui gigitan nyamuk *aedes aegypty* dan *aedes albopictus*. Tingginya angka kesakitan DBD disebabkan karena adanya iklim tidak stabil dan curah hujan cukup banyak pada musim penghujan yang merupakan sarang perkembangbiakan nyamuk *aedes aegypty* yang cukup potensial. Angka kematian DBD di Jawa Tengah tahun 2015 sebesar 1,6 persen, sedikit menurun bila dibandingkan CFR tahun 2014 yaitu 1,7 persen.