

## PEMODELAN ANGKA HARAPAN HIDUP PROVINSI JAWA TENGAH MENGUNAKAN *ROBUST SPATIAL DURBIN MODEL*

Maghfiroh Hadadiah Mukrom<sup>1</sup>, Hasbi Yasin<sup>2</sup>, Arief Rachman Hakim<sup>3</sup>

<sup>1,2,3</sup> Departemen Statistika FSM Universitas Diponegoro

email: maghfirohhadadiah98@gmail.com

### ABSTRACT

Spatial regression is a model used to determine relationship between response variables and predictor variables that gets spatial influence. If there are spatial influences on both variables, the model that will be formed is Spatial Durbin Model. One reason for the inaccuracy of the spatial regression model in predicting is the existence of outlier observations. Removing outliers in spatial analysis can change the composition of spatial effects on data. One way to overcome of outliers in the spatial regression model is by using robust spatial regression. The application of M-estimator is carried out in estimating the spatial regression parameter coefficients that are robust against outliers. The aim of this research is obtaining model of number of life expectancy in Central Java Province in 2017 that contain outliers. The results by applying M-estimator to estimating robust spatial durbin model regression parameters can accommodate the existence of outliers in the spatial regression model. This is indicated by the change in the estimating coefficient value of the robust spatial durbin model regression parameter which can increase adjusted  $R^2$  value becomes 93,69% and decrease MSE value becomes 0,12551.

**Keywords:** Outliers, M-estimator, Spatial Durbin Model, Number of Life Expectancy.

### 1. PENDAHULUAN

Analisis regresi merupakan suatu teknik analisis statistik yang bertujuan untuk melihat hubungan antara variabel respon dan variabel prediktor sehingga mampu memprediksikan variabel respon jika variabel prediktornya diketahui. Ketika ditemukan adanya efek spasial (lokasi) pada data yang biasa disebut dengan data spasial, maka analisis regresi yang digunakan adalah regresi spasial. Pemodelan dengan menggunakan regresi spasial salah satunya adalah model *Spatial Durbin Model* (SDM). Model SDM adalah model regresi spasial yang menunjukkan adanya efek spasial dalam variabel respon maupun variabel prediktor<sup>[1]</sup>.

Dalam analisis data, terkadang ditemukan adanya pencilan (*outlier*). Adanya pencilan membuat estimasi parameter menjadi bias. Untuk menganalisis data yang terkontaminasi oleh pencilan dapat digunakan suatu metode analisis statistik yaitu regresi *robust*. Salah satu metode estimasi pada regresi *robust* adalah *M-estimator* yang merupakan metode regresi *robust* yang paling sederhana baik secara teoritis maupun secara komputasi<sup>[4]</sup>. Salah satu metode yang dapat digunakan untuk deteksi pencilan pada data spasial adalah *Moran's Scatterplot*. Karena data yang akan digunakan adalah data spasial, maka analisis regresi *robust* yang digunakan adalah regresi *robust spatial*.

Salah satu aplikasi dari model regresi spasial yaitu pada kasus Angka Harapan Hidup (AHH). Pembentukan model regresi spasial pada penelitian ini dilakukan menggunakan *software R* yang dapat memberikan hasil estimasi yang lebih efektif dan efisien karena tidak perlu melakukan perhitungan manual secara berkali-kali. Penelitian ini difokuskan pada pembentukan model SDM dan *robust* SDM dengan menggunakan *software R* pada kasus AHH menurut kabupaten/kota di Provinsi Jawa Tengah tahun 2017.

### 2. TINJAUAN PUSTAKA

#### 2.1 Regresi Linier Berganda

Regresi linier berganda merupakan model regresi yang mengandung lebih dari satu variabel prediktor<sup>[14]</sup>. Persamaan regresi linier berganda dengan k variabel prediktor adalah sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (1)$$

## 2.2 Matriks Pembobot Spasial

Matriks pembobot spasial digunakan untuk menentukan bobot antar lokasi yang diamati berdasarkan hubungan ketetangaan antar lokasi. Ketetangaan dapat didefinisikan dalam beberapa cara, yaitu<sup>[17]</sup>:

1. *Rook Contiguity* (Persinggungan Sisi)  
Daerah pengamatannya ditentukan berdasarkan sisi-sisi yang saling bersinggungan dan sudut tidak diperhitungkan.
2. *Bishop Contiguity* (Persinggungan Sudut)  
Daerah pengamatannya ditentukan berdasarkan sudut-sudut yang saling bersinggungan dan sisi tidak diperhitungkan.
3. *Queen Contiguity* (Persinggungan Sisi-Sudut)  
Daerah pengamatannya ditentukan berdasarkan sisi-sisi yang saling bersinggungan dan sudut juga diperhitungkan.

## 2.3 Uji Moran's I

Untuk mengetahui apakah ada autokorelasi spasial antar lokasi dapat dilakukan uji autokorelasi spasial dengan menggunakan uji *Moran's I*<sup>[9]</sup>, dengan hipotesis sebagai berikut:

H<sub>0</sub>: Tidak ada autokorelasi spasial antar lokasi

H<sub>1</sub>: Ada autokorelasi spasial antar lokasi

Statistik Uji:  $Z_{hitung} = \frac{I - E(I)}{\sqrt{\text{var}(I)}}$

dengan:

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

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$$

$$E(I) = I_0 = -\frac{1}{n-1}$$

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

$$c_{ij} = (x_i - \bar{x})(x_j - \bar{x})$$

$$S_2 = \sum_{i \neq j}^n (w_{i0} + w_{0i})^2$$

$$\text{var}(I) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{(n^2 - 1) S_0^2} - [E(I)]^2$$

$$w_{i0} = \sum_{i=1}^n w_{ij} ; w_{0i} = \sum_{j=1}^n w_{ji}$$

Keputusan: H<sub>0</sub> ditolak jika

## 2.4 Regresi Spasials

Model umum regresi spasial dapat dituliskan sebagai berikut<sup>[11]</sup>:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (2)$$

$$\mathbf{u} = \lambda \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim \mathbf{N}(0, \sigma_\varepsilon^2 \mathbf{I}_n)$$

- dengan:  $\mathbf{y}$  = Vektor variabel dependen berukuran n x 1  
 $\rho$  = Koefisien parameter spasial lag dari variabel dependen  
 $\mathbf{W}$  = Matriks pembobot spasial yang berukuran n x n  
 $\mathbf{X}$  = Matriks variabel independen berukuran n x (k+1)  
 $\boldsymbol{\beta}$  = Vektor koefisien parameter regresi berukuran (k+1) x 1  
 $\lambda$  = Koefisien parameter spasial *error*  
 $\mathbf{u}$  = Vektor *error* yang mempunyai efek spasial dengan ukuran nx1  
 $\boldsymbol{\varepsilon}$  = Vektor *error* dengan ukuran nx1

Dari persamaan model umum regresi spasial (2), dapat dibentuk beberapa model lain sebagai berikut<sup>[1]</sup>:

1. Jika  $\rho=0$  dan  $\lambda=0$  maka disebut model regresi linier klasik dengan persamaan yang terbentuk adalah:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3)$$

2. Jika  $\rho \neq 0$  dan  $\lambda=0$  disebut regresi *Spatial Autoregressive Model* (SAR) dengan persamaan yang terbentuk adalah:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (4)$$

3. Jika  $\rho=0$  dan  $\lambda \neq 0$  disebut regresi *Spatial Error Model* (SEM) dengan persamaan yang terbentuk adalah:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (5)$$

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

4. Jika  $\rho \neq 0$  dan  $\lambda \neq 0$  disebut *Spatial Autoregressive Moving Average* (SARMA) dengan persamaan yang terbentuk adalah:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (6)$$

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

## 2.5 *Spatial Durbin Model* (SDM)

*Spatial Durbin Model* (SDM) merupakan model regresi spasial yang memiliki bentuk seperti *Spatial Autoregressive Model* (SAR) yang memiliki spasial lag pada variabel respon ( $\mathbf{Y}$ ) seperti pada persamaan (4). Hanya saja, SDM memiliki ciri khas adanya spasial lag pada variabel prediktor ( $\mathbf{X}$ )<sup>[1]</sup>. Model SDM memiliki bentuk persamaan sebagai berikut<sup>[12]</sup>:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \alpha \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathbf{N}(0, \sigma_\varepsilon^2 \mathbf{I}_n) \quad (7)$$

Atau dapat dituliskan sebagai berikut:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \quad (8)$$

dengan:  $\mathbf{Z} = [\mathbf{1}_n \quad \mathbf{X} \quad \mathbf{W}\mathbf{X}]$ ;  $\boldsymbol{\delta} = \begin{bmatrix} \alpha \\ \boldsymbol{\beta} \\ \boldsymbol{\theta} \end{bmatrix}$

$\alpha$  = Parameter konstan

$\boldsymbol{\theta}$  = Vektor parameter lag spasial variabel prediktor berukuran  $k \times 1$

$\mathbf{1}_n$  = Vektor yang berisi angka 1 berukuran  $n \times 1$

Parameter model SDM dapat diestimasi dengan menggunakan metode *Maximum Likelihood Estimation* (MLE), diperoleh estimasi parameter model SDM sebagai berikut:

$$\frac{1}{\lambda_{\min}} < \rho < \frac{1}{\lambda_{\max}}$$

$$\hat{\boldsymbol{\delta}} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{y} - \hat{\rho} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{W}\mathbf{y} = \hat{\boldsymbol{\delta}}_0 - \hat{\rho} \hat{\boldsymbol{\delta}}_a \quad (9)$$

$$\hat{\sigma}^2 = \frac{(\mathbf{e}_0 - \hat{\rho} \mathbf{e}_a)^T (\mathbf{e}_0 - \hat{\rho} \mathbf{e}_a)}{n} \quad (10)$$

## 2.6 Pengujian Model Regresi

### 2.6.1 Uji Kecocokan Model

Untuk menguji kecocokan model SDM digunakan prosedur uji dengan hipotesis sebagai berikut<sup>[15]</sup>:

$H_0 : \rho = \beta_j = \theta_j = 0$ , dengan  $j = 1, 2, \dots, k$

$H_1 : \rho \neq 0$  atau paling tidak ada satu  $\beta_j \neq 0, \theta_j \neq 0$ , dengan  $j = 1, 2, \dots, k$

Statistik Uji:

$$F_{hitung} = \frac{SS_R/k}{SS_E/(n-k-1)} = \frac{MSR}{MSE} \quad (11)$$

Keputusan:  $H_0$  ditolak jika nilai  $F_{hitung} > F_{\alpha, k, n-k-1}$  atau  $p\text{-value} < \alpha$

### 2.6.2 Uji Signifikansi Parameter

Pengujian signifikansi parameter pemodelan spasial menggunakan uji Wald<sup>[1]</sup>, untuk menguji parameter  $\rho$  digunakan hipotesis sebagai berikut<sup>[15]</sup>:

$$H_0 : \rho = 0 ; \quad H_1 : \rho \neq 0$$

$$\text{Statistik Uji: } Wald_{\rho} = \frac{\hat{\rho}^2}{\text{var}(\hat{\rho})} \quad (12)$$

Untuk menguji parameter  $\beta$ :

$$H_0 : \beta_j = 0 ; \quad H_1 : \beta_j \neq 0, j = 1, 2, \dots, k$$

$$\text{Statistik Uji: } Wald_{\beta} = \frac{\hat{\beta}_j^2}{\text{var}(\hat{\beta}_j)} \quad (13)$$

Untuk parameter  $\theta$  menggunakan hipotesis sebagai berikut:

$$H_0 : \theta_j = 0 ; \quad H_1 : \theta_j \neq 0, j = 1, 2, \dots, k$$

$$\text{Statistik Uji: } Wald_{\theta} = \frac{\hat{\theta}_j^2}{\text{var}(\hat{\theta}_j)} \quad (14)$$

Kriteria pengambilan keputusan adalah  $H_0$  ditolak jika nilai  $Wald > \chi_{\alpha,1}^2$ .

### 2.6.3 Uji Asumsi Regresi

#### a. Uji Normalitas

Untuk menguji residual berdistribusi normal dapat digunakan uji *Kolmogorov-Smirnov*. Tahapan uji *Kolmogorov-Smirnov* sebagai berikut<sup>[5]</sup>:

$H_0$ : Residual berdistribusi normal

$H_1$ : Residual tidak berdistribusi normal

$$\text{Statistik Uji: } D = \sup_e |F_0(e) - S(e)| \quad (15)$$

Kriteria uji:  $H_0$  ditolak jika  $D \geq d_{(1-\alpha/2)}$  atau  $p\text{-value} < \alpha$

dengan  $d_{(1-\alpha/2)}$  adalah nilai kritis yang diperoleh dari tabel *Kolmogorov-Smirnov*.

#### b. Uji Multikolinieritas

Deteksi multikolinieritas bertujuan untuk menguji apakah dalam model regresi ditemukan adanya korelasi antar variabel prediktor. Salah satu cara untuk mengetahui adanya multikolinieritas dengan nilai Faktor Inflasi Varian (VIF), didefinisikan dengan rumus<sup>[13]</sup>:

$$VIF(x_j) = \frac{1}{(1-R_j^2)}, \quad i = 1, 2, \dots, n \quad (16)$$

Nilai  $VIF > 10$  menunjukkan multikolinieritas yang kuat<sup>[13]</sup>.

#### c. Uji Autokorelasi

Uji ini digunakan untuk mendeteksi data yang ada apakah terjadi autokorelasi atau tidak. Dengan menggunakan lambang :  $E(\varepsilon_i \varepsilon_j) = 0 \quad i \neq j$

Salah satu cara untuk mendeteksi adanya autokorelasi dapat menggunakan metode *Durbin-Watson* dengan hipotesis sebagai berikut:

$H_0$ : Tidak ada autokorelasi positif atau negatif

$H_1$ : Ada autokorelasi antar positif atau negatif

$$\text{Statistik Uji: } d = \frac{\sum_{i=2}^{i=n} (e_i - e_{i-1})^2}{\sum_{i=1}^{i=n} e_i^2} \quad (17)$$

dengan:

$d$  = nilai *Durbin – Watson* dari hasil perhitungan data

$e_i$  = residual ke  $i$  ;  $e_{i-1}$  = residual ke  $i-1$

#### d. Uji Heteroskedastisitas

Uji ini untuk melihat apakah variansi error satu pengamatan ke pengamatan yang lain konstan atau tidak. Lambang homoskedastisitas:  $E(\varepsilon_i^2) = \sigma^2 \quad i = 1, 2, \dots, n$

Heteroskedastisitas dapat diperiksa dengan menggunakan uji Glejser<sup>[10]</sup>. Uji Glejser dilakukan dengan meregresikan nilai absolut residual  $|e_i|$  terhadap variabel-variabel bebas yang diperkirakan mempunyai hubungan erat dengan  $\sigma^2$ .

$$|e_i| = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + v \quad (18)$$

## 2.7 Ukuran Kecocokan Model

### a. Koefisien Determinasi ( $R^2$ )

Koefisien determinasi ( $R^2$ ) dikenal untuk mengetahui proporsi variasi dalam  $y$  yang dapat dijelaskan oleh beberapa variabel  $x$  secara bersama-sama (gabungan)<sup>[10]</sup>. Nilai  $R^2$  terletak antara nol dan satu. Kecocokan model “lebih baik” jika  $R^2$  semakin dekat dengan satu<sup>[10]</sup>. Kebanyakan peneliti lebih suka menggunakan koefisien determinasi yang disesuaikan (*Adjusted  $R^2$* )<sup>[14]</sup>, rumusnya sebagai berikut:

$$R_{adj,k}^2 = 1 - \frac{SS_E / (n - k - 1)}{SS_T / (n - 1)} \quad (19)$$

### b. Mean Square Error (MSE)

Dalam analisis regresi, istilah *Mean Square Error* (MSE) digunakan untuk merujuk pada estimasi tak bias dari variansi *error*<sup>[13]</sup>. MSE diperoleh berdasarkan rumus<sup>[13]</sup>:

$$MSE = \frac{SS_E}{n - k - 1} \quad (20)$$

Jika nilai MSE semakin kecil hingga mendekati nol, dapat dikatakan bahwa model regresi semakin baik.

## 2.8 Pencilan

Pencilan adalah pengamatan yang tampak berbeda dengan pengamatan lainnya pada sekumpulan data yang ada<sup>[3]</sup>. Salah satu metode yang dapat digunakan untuk mendeteksi pencilan adalah *Moran’s Scatterplot* yang merupakan salah satu metode yang dapat digunakan untuk mendeteksi adanya pencilan spasial<sup>[16]</sup>. *Moran’s Scatterplot* merupakan grafik plot normal dengan nilai atribut  $\left( Z | f(i) = \frac{f(i) - \mu_f}{\sigma_f} \right)$  yaitu rata-rata nilai ketetanggaan dari

nilai atribut yang telah dinormalisasi. Grafik dibagi menjadi empat kuadran, plot data yang berada di kuadran bagian kiri atas dan kanan bawah merupakan pencilan spasial.

Selain dilihat dari grafik, deteksi pencilan spasial menggunakan *Moran’s Scatterplot* juga dapat diidentifikasi secara matematis menggunakan rumus:

$$\left( Z[f(i)]_x \left( \sum_j (\mathbf{w}_{ij} Z[f(j)]) \right) \right)$$

dengan  $\mathbf{W}$  adalah matriks ketetanggaan,  $Z_i = \frac{f(i) - \mu_f}{\sigma_f}$ ,  $I_i = \left( \sum_j (\mathbf{W}_{ij} \mathbf{Z}[f(j)]) \right)$ ,  $\mu_f$  dan  $\sigma_f$  adalah rata-rata dan standar deviasi dari fungsi  $f(i)$ . Jika nilainya kurang dari 0 maka termasuk pencilan spasial, dan jika tidak maka sebaliknya.

## 2.9 Regresi Robust

Regresi *robust* adalah metode regresi yang digunakan ketika distribusi dari residual tidak normal atau ada beberapa pencilan yang mempengaruhi model<sup>[7]</sup>. Metode ini adalah alat yang penting untuk menganalisis data yang terkontaminasi oleh pencilan dan dapat memberikan hasil yang resisten akan kehadiran pencilan. Salah satu metode estimasi dalam regresi *robust* adalah *Robust M-estimator*.

### 2.9.1 Robust M-Estimator

*Robust M-estimator* sebenarnya meminimumkan fungsi obyektif<sup>[7]</sup>:

$$\min_{\beta} \sum_{i=1}^n \rho(u_i) = \min_{\beta} \sum_{i=1}^n \rho\left(\frac{e_i}{s}\right) = \min_{\beta} \sum_{i=1}^n \rho\left(\frac{y_i - \sum_{j=0}^k x_{ij}\beta_j}{s}\right) \quad (21)$$

dengan  $s$  adalah skala estimasi robust. Estimasi  $s$  yang digunakan adalah:

$$s = \frac{MAD}{0,6745} = \frac{\text{median}|e_i - \text{median}(e_i)|}{0,6745} \quad (22)$$

Nilai 0,6745 menjadikan  $s$  sebagai estimator tak bias dari  $\sigma$  jika  $n$  besar dan error berdistribusi normal<sup>[13]</sup>.

Untuk mendapatkan estimasi parameter dengan meminimumkan persamaan (21). Dengan turunan parsial pertama dari  $\rho$  terhadap  $\beta_j$  ( $j = 0, 1, \dots, k$ ) disamakan dengan 0, sehingga:

$$\sum_{i=1}^n x_{ij} \psi\left(\frac{y_i - \sum_{j=0}^k x_{ij}\beta_j}{s}\right) = 0, \quad j = 0, 1, \dots, k \quad (23)$$

Diberikan solusi dengan mendefinisikan fungsi pembobot:

$$w(u_i) = \frac{\psi\left(\frac{y_i - \sum_{j=0}^k x_{ij}\beta_j}{s}\right)}{\left(\frac{y_i - \sum_{j=0}^k x_{ij}\beta_j}{s}\right)} \quad (24)$$

dan  $w_i = w(u_i)$ . Kemudian estimasi persamaan (23) dapat ditulis:

$$\sum_{i=1}^n x_{ij} w_i \left( y_i - \sum_{j=0}^k x_{ij}\beta_j \right) = 0, \quad j = 0, 1, \dots, k \quad (25)$$

Persamaan (25) diselesaikan dengan *Iteratively Reweighted Least Square* (IRLS). Pada notasi matriks, persamaan (25) dapat ditulis:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y} \quad (26)$$

Iterasi akan berhenti jika  $\hat{\beta}_j$  konvergen yaitu selisih nilai  $\hat{\beta}_j^{(m+1)}$  dan  $\hat{\beta}_j^{(m)}$  mendekati 0.

### 2.9.2 Fungsi Obyektif

Fungsi yang digunakan untuk mencari fungsi pembobot pada regresi *robust* adalah fungsi obyektif<sup>[8]</sup>. Salah satu fungsi pembobot yang dapat digunakan adalah fungsi pembobot *Tukey Bisquare* sebagai berikut:

$$w(u_i) = \begin{cases} \left[ 1 - \left( \frac{u_i}{c} \right)^2 \right]^2, & \text{untuk } |u_i| \leq c \\ 0, & \text{untuk } |u_i| > c \end{cases}$$



Nilai  $c$  disebut *tunning constant*, dan *tunning constant* untuk fungsi pembobot *Tukey Bisquare* pada metode estimasi *M-estimator* adalah  $c = 4,685^{[8]}$ .

## 2.10 Angka Harapan Hidup

Menurut Badan Pusat Statistik Indonesia, Angka Harapan Hidup (AHH) pada suatu umur  $x$  adalah rata-rata tahun hidup yang masih akan dijalani oleh seseorang yang telah berhasil mencapai umur  $x$  pada suatu tahun tertentu dan dalam situasi mortalitas yang berlaku di lingkungan masyarakatnya. Terdapat beberapa faktor yang berpengaruh signifikan terhadap AHH diantaranya adalah faktor pendidikan, faktor kesehatan dan faktor ekonomi. Variabel rata-rata lama sekolah digunakan sebagai variabel untuk menjelaskan faktor pendidikan. Variabel persentase rumah tangga berperilaku hidup bersih dan sehat dan jumlah posyandu digunakan sebagai variabel untuk menjelaskan faktor kesehatan. Variabel persentase penduduk miskin dan pengeluaran perkapita disesuaikan digunakan sebagai variabel untuk menjelaskan faktor ekonomi.

## 3. METODOLOGI PENELITIAN

### 3.1. Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari katalog Provinsi Jawa Tengah Dalam Angka 2018 yang dikeluarkan oleh Badan Pusat Statistik (BPS) Provinsi Jawa Tengah dan buku Profil Kesehatan Provinsi Jawa Tengah Tahun 2017 yang dikeluarkan oleh Dinas Kesehatan Provinsi Jawa Tengah. Unit observasi dalam penelitian ini adalah 35 kabupaten dan kota di Provinsi Jawa Tengah. Variabel yang digunakan dalam penelitian ini adalah data AHH sebagai variabel respon dan rata-rata lama sekolah ( $x_1$ ), persentase rumah tangga berperilaku hidup bersih dan sehat ( $x_2$ ), jumlah posyandu ( $x_3$ ), persentase penduduk miskin ( $x_4$ ) dan pengeluaran perkapita disesuaikan ( $x_5$ ) sebagai variabel prediktor.

### 3.2. Langkah-langkah Analisis

Langkah-langkah analisis yang akan dilakukan pada penelitian ini sebagai berikut:

1. Mendapatkan data AHH beserta lima faktor yang diduga mempengaruhinya.
2. Menentukan matriks pembobot spasial berdasarkan *queen contiguity*.
3. Melakukan uji autokorelasi spasial dengan uji *Moran's I*.
4. Mengestimasi parameter dengan regresi *Spatial Durbin Model*.
5. Melakukan uji kecocokan model dan uji signifikansi parameter.
6. Menghitung koefisien determinasi yang disesuaikan (*Adjusted R<sup>2</sup>*) dan MSE.
7. Melakukan uji asumsi pada model.
8. Melakukan pendeteksian pencilan dengan *Moran's Scatterplot*.
9. Mengestimasi parameter dengan regresi *Robust Spatial Durbin Model M-estimator*.
10. Melakukan uji kecocokan model dan uji signifikansi parameter.
11. Menghitung koefisien determinasi yang disesuaikan (*Adjusted R<sup>2</sup>*) dan MSE.
12. Menentukan model regresi terbaik dan menganalisisnya.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Estimasi Parameter Regresi *Robust Spatial Durbin Model*

Untuk memperoleh estimasi parameter model *robust spatial durbin*, maka dilakukan estimasi menggunakan metode kuadrat terkecil. Estimasi parameter dalam metode kuadrat terkecil dapat diperoleh dengan meminimumkan jumlah kuadrat residual sebagai berikut:

$$\sum e_i^2 = \sum [(\mathbf{I} - \rho \mathbf{W})\mathbf{y}_i - \mathbf{Z}_1 \delta_{1i} - \mathbf{Z}_2 \delta_{2i} - \dots - \mathbf{Z}_k \delta_{ki}]^2 = \mathbf{e}^T \mathbf{e} \quad (27)$$

$$\text{sehingga diperoleh } \hat{\delta}_{OLS} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T (\mathbf{I} - \rho \mathbf{W}) \mathbf{y} \quad (28)$$

Untuk mendapatkan fungsi pengaruh, maka persamaan (28) dapat ditulis menjadi:

$$\hat{\delta}_{OLS} = (\mathbf{Z}^T \Psi \mathbf{Z})^{-1} \mathbf{Z}^T \Psi (\mathbf{I} - \rho \mathbf{W}) \mathbf{y} \quad (29)$$

Dari fungsi pengaruh dapat didefinisikan fungsi pembobot :  $b_i = b(u_i) = \frac{\psi(u_i)}{u_i}$

Maka dari itu dari persamaan (29) dapat ditulis:

$$\hat{\delta}_{OLS} = (\mathbf{Z}^T \mathbf{B} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{B} (\mathbf{I} - \rho \mathbf{W}) \mathbf{y} \quad (30)$$

dengan  $\mathbf{B}$  adalah matriks pembobot yang berukuran  $n \times n$  dengan elemen-elemen diagonal yang berisi pembobot  $b_1, b_2, \dots, b_n$ .

#### 4.2 Uji Moran's I

Diperoleh nilai Z-hitung beserta nilai *Moran's I* sebagai berikut:

**Tabel 1.** Uji *Moran's I*

Variabel	Nilai <i>Moran's I</i>	Z-hitung	Kesimpulan
$y$	0,5752232	5,2978	Tolak $H_0$
$x_1$	0,19489697	1,9503	Terima $H_0$
$x_2$	-0,12105565	-0,7779	Terima $H_0$
$x_3$	0,06377452	0,8069	Terima $H_0$
$x_4$	0,35389376	3,2687	Tolak $H_0$
$x_5$	0,17116800	1,7432	Terima $H_0$

Dari Tabel 1 diperoleh hasil bahwa variabel  $y$  dan  $x_1$  memiliki nilai  $|Z_{hitung}| > 1,96$ , yang berarti  $H_0$  ditolak. Hal ini menunjukkan bahwa terdapat autokorelasi spasial antar lokasi pada variabel respon dan variabel prediktor. Karena pada variabel respon dan minimal satu dari variabel prediktor menunjukkan adanya autokorelasi spasial antar lokasi, maka model SDM dapat digunakan dalam penelitian.

#### 4.3 Regresi *Spatial Durbin Model*

Model SDM yang diperoleh sebagai berikut:

$$\begin{aligned} \hat{y}_i = & 0,63839 \sum_{j=1}^n w_{ij} y_j + 23,70977 + 0,61799 x_{i1} + 0,02843 x_{i2} - 2,6744 \times 10^{-6} x_{i3} \\ & - 0,19592 x_{i4} - 2,6278 \times 10^{-4} x_{i5} - 0,84814 \sum_{j=1}^n w_{ij} x_{j1} - 0,01467 \sum_{j=1}^n w_{ij} x_{j2} \\ & + 0,00138 \sum_{j=1}^n w_{ij} x_{j3} + 0,11522 \sum_{j=1}^n w_{ij} x_{j4} + 6,111 \times 10^{-4} \sum_{j=1}^n w_{ij} x_{j5} \end{aligned}$$

##### 4.3.1 Uji Hipotesis

- **Uji Kecocokan Model Regresi SDM**

Berdasarkan hasil komputasi dengan menggunakan *software R*, diperoleh nilai  $F_{hitung} = 8,7936$  dimana nilai tersebut lebih besar dari  $F_{tabel} = 2,49$ . Kemudian diperoleh nilai *p-value* sebesar  $0,0000 < \alpha = 5\%$ . Sehingga diperoleh keputusan bahwa  $H_0$  ditolak pada taraf signifikansi 5%. Jadi, dapat disimpulkan bahwa model SDM telah sesuai untuk menggambarkan hubungan antara variabel respon dan variabel prediktor.

- **Uji Signifikansi Parameter Model SDM**

Dapat dilihat bahwa parameter lag spasial ( $\rho$ ) dan parameter  $\beta_4$  berpengaruh signifikan terhadap model karena nilai wald yang diperoleh  $> \chi_{0,05,1}^2 = 3,841$ , sedangkan parameter  $\beta_1, \beta_2, \beta_3, \beta_5, \theta_1, \theta_2, \theta_3, \theta_4$  dan  $\theta_5$  tidak berpengaruh signifikan terhadap model karena nilai wald yang diperoleh  $< \chi_{0,05,1}^2 = 3,841$ .



**Tabel 2.** Uji Signifikansi Parameter SDM

Parameter	Wald	Keputusan	Parameter	Wald	Keputusan
Rho	5,59767	H <sub>0</sub> ditolak	$\beta_5$	1,44171	H <sub>0</sub> diterima
<i>Intercept</i>	1,80201	H <sub>0</sub> diterima	$\theta_1$	0,70375	H <sub>0</sub> diterima
$\beta_1$	2,84664	H <sub>0</sub> diterima	$\theta_2$	0,07086	H <sub>0</sub> diterima
$\beta_2$	1,32565	H <sub>0</sub> diterima	$\theta_3$	1,38198	H <sub>0</sub> diterima
$\beta_3$	$2,868 \times 10^{-5}$	H <sub>0</sub> diterima	$\theta_4$	0,4234	H <sub>0</sub> diterima
$\beta_4$	7,05707	H <sub>0</sub> ditolak	$\theta_5$	1,54393	H <sub>0</sub> diterima

### 4.3.2 Ukuran Kecocokan Model

#### a. Adjusted R<sup>2</sup>

$$R^2_{Adj,k} = 0,6963$$

Artinya angka harapan hidup dipengaruhi oleh rata-rata lama sekolah, persentase rumah tangga berperilaku hidup bersih dan sehat, jumlah posyandu, persentase penduduk miskin dan pengeluaran per kapita disesuaikan sebanyak 69,63% dan sisanya sebanyak 30,37% dipengaruhi oleh faktor yang lain.

#### b. MSE

$$MSE = 1,15641$$

### 4.3.3 Uji Asumsi Model

- **Asumsi Normalitas**

Berdasarkan hasil komputasi dengan menggunakan *software R*, diperoleh nilai hitung  $D = 0,16089 < \text{nilai tabel } d = 0,224$  yang diperoleh dari tabel *Kolmogorov-Smirnov* dan nilai probabilitas yang didapat adalah sebesar  $0,2928 > \alpha = 0,05$  yang menunjukkan bahwa H<sub>0</sub> diterima atau residual model berdistribusi normal.

- **Asumsi Multikolinieritas**

Berdasarkan hasil komputasi dengan menggunakan *software R*, diperoleh hasil yaitu nilai  $VIF(x_1) = 5,78165$ ,  $VIF(x_2) = 1,71299$ ,  $VIF(x_3) = 1,86858$ ,  $VIF(x_4) = 2,3402$ ,  $VIF(x_5) = 4,05949$ ,  $VIF(x_6) = 7,59969$ ,  $VIF(x_7) = 2,46578$ ,  $VIF(x_8) = 2,39775$ ,  $VIF(x_9) = 5,1459$  serta  $VIF(x_{10}) = 4,89374$ . Dapat dilihat bahwa nilai  $VIF < 10$  pada semua variabel prediktor, maka dapat disimpulkan tidak terjadi multikolinieritas pada variabel prediktor.

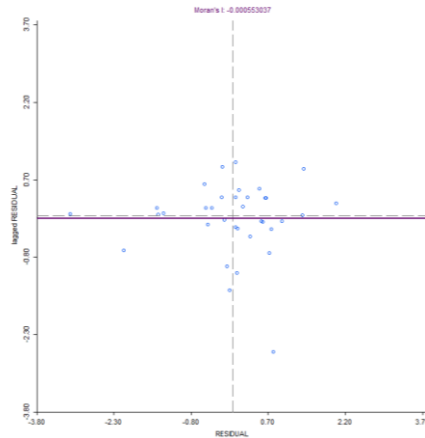
- **Asumsi Non – Autokorelasi**

Berdasarkan hasil komputasi dengan menggunakan *software R*, diperoleh nilai hitung  $d = 2,252943$ . Nilai  $4-d_U$  sebesar 2,1971 dan nilai  $4-d_L$  sebesar 2,8399. Dapat dilihat bahwa nilai  $4-d_U = 2,1971 < d < 4-d_L = 2,8399$  yang menunjukkan bahwa tidak ada keputusan.

- **Asumsi Homoskedastisitas**

Berdasarkan hasil komputasi dengan menggunakan *software R*, diperoleh hasil yaitu nilai  $\text{Prob}(x_1) = 0,90241$ ,  $\text{Prob}(x_2) = 0,21038$ ,  $\text{Prob}(x_3) = 0,23907$ ,  $\text{Prob}(x_4) = 0,11063$ ,  $\text{Prob}(x_5) = 0,36929$ ,  $\text{Prob}(x_6) = 0,49575$ ,  $\text{Prob}(x_7) = 0,45537$ ,  $\text{Prob}(x_8) = 0,18547$ ,  $\text{Prob}(x_9) = 0,21109$  serta  $\text{Prob}(x_{10}) = 0,28783$ . Hasil pengujian menunjukkan bahwa semua variabel prediktor tidak signifikan pada tingkat signifikansi 5%, karena semua nilai probabilitasnya  $> \alpha = 0,05$ . Sehingga dapat disimpulkan bahwa tidak terjadi kasus heteroskedastisitas pada model SDM dengan 5 variabel prediktor.

#### 4.4 Plot Residual



**Gambar 1.** Plot Residual

Pada Gambar 1, secara visual terdapat beberapa pencilan spasial dalam data residual. Untuk memastikan data mana saja yang termasuk pencilan, maka dilakukan deteksi pencilan secara formal dengan rumus  $(z[f(i)]x(\sum_j(w_{ij}z[f(j)]))) < 0$ .

#### 4.5 Pendeteksian Pencilan

Dari hasil *output*, dapat diketahui bahwa ada sebanyak 19 pencilan spasial terdeteksi, yaitu wilayah ke 1, 2, 9, 10, 11, 14, 16, 17, 18, 20, 23, 24, 27, 29, 30, 31, 32, 33 dan 34.

#### 4.6 Regresi Robust

##### 4.6.1 Regresi Robust M-Estimator

Model regresi *robust M-estimator* dengan pembobot *Tukey Bisquare* yang didapatkan setelah proses 18 kali iterasi adalah sebagai berikut :

$$\hat{y}_i = 0,63839 \sum_{j=1}^n w_{ij} y_j + 28,76139 + 0,67309 x_{i1} + 0,03036 x_{i2} + 0,00045 x_{i3} \\ - 0,11198 x_{i4} - 0,00025 x_{i5} - 1,70228 \sum_{j=1}^n w_{ij} x_{j1} + 0,05127 \sum_{j=1}^n w_{ij} x_{j2} \\ + 0,00214 \sum_{j=1}^n w_{ij} x_{j3} - 0,09736 \sum_{j=1}^n w_{ij} x_{j4} + 0,00025 \sum_{j=1}^n w_{ij} x_{j5}$$

#### 4.7 Pemilihan Model Regresi Terbaik

**Tabel 3.** Perbandingan Regresi SDM dan Regresi Robust SDM

	<i>Adjusted R<sup>2</sup></i>	MSE
<b>Regresi SDM</b>	0,6963	1,15641
<b>Regresi Robust SDM</b>	0,9369	0,12551

Kriteria pemilihan model regresi terbaik yaitu mempunyai  $R^2_{Adj,k}$  terbesar dan nilai MSE terkecil. Diperoleh hasil bahwa model *robust* SDM memiliki nilai *Adjusted R<sup>2</sup>* lebih besar yaitu sebesar 0,9369 dan nilai MSE yang lebih kecil yaitu sebesar 0,12551. Maka dari itu dapat disimpulkan bahwa model *robust* SDM lebih baik daripada model SDM.

## 5. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dilakukan, maka dapat ditarik beberapa kesimpulan sebagai berikut:

1. Model regresi spasial terbaik yang diperoleh adalah

$$\hat{y}_i = 0,63839 \sum_{j=1}^n w_{ij}y_j + 28,76139 + 0,67309x_{i1} + 0,03036x_{i2} + 0,00045x_{i3} - 0,11198x_{i4} - 0,00025x_{i5} - 1,70228 \sum_{j=1}^n w_{ij}x_{j1} + 0,05127 \sum_{j=1}^n w_{ij}x_{j2} + 0,00214 \sum_{j=1}^n w_{ij}x_{j3} - 0,09736 \sum_{j=1}^n w_{ij}x_{j4} + 0,00025 \sum_{j=1}^n w_{ij}x_{j5}$$

Model yang lebih baik digunakan untuk menjelaskan angka harapan hidup di Provinsi Jawa Tengah tahun 2017 adalah model *robust spatial durbin*, karena nilai *Adjusted R<sup>2</sup>* yang diperoleh lebih besar dan nilai MSE yang diperoleh lebih kecil.

2. Faktor-faktor yang berpengaruh signifikan terhadap angka harapan hidup di Provinsi Jawa Tengah pada tahun 2017 adalah rata-rata lama sekolah, persentase rumah tangga berperilaku hidup bersih dan sehat, jumlah posyandu, persentase penduduk miskin dan pengeluaran per kapita disesuaikan.

## DAFTAR PUSTAKA

- [1] Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.
- [2] Badan Pusat Statistik (BPS). 2018. *Jawa Tengah Dalam Angka 2018*. Semarang: Badan Pusat Statistik Jawa Tengah.
- [3] Barnett, V., dan Lewis, T. 1994. *Outliers in Statistical Data (3rd ed.)*. New York: John Wiley and Sons, Inc.
- [4] Chen, C. 2002. *Robust Regression and Outlier Detection with The ROBUSTREG Procedure*, pp 265-27. SAS Institute Inc., Lary, NC.
- [5] Conover, W. J. 1980. *Practical Nonparametric Statistics (2nd ed)*. New York: John Wiley and Sons, Inc.
- [6] Dinas Kesehatan. 2018. *Profil Kesehatan Provinsi Jawa Tengah Tahun 2017*. Semarang: Dinas Kesehatan Provinsi Jawa Tengah.
- [7] Draper, N. R., dan Smith, H. 1998. *Applied Regression Analysis (3rd ed.)*. New York: John Wiley and Sons.
- [8] Fox, J. 2002. *Robust Regression : Appendix to An R and S-Plus Companion to Applied Regression*.
- [9] Goodchild, M. F. 1986. *Spatial Autocorrelation*. Norwich: Geo Books.
- [10] Gujarati, D. N. 2004. *Basic Econometrics 4th Edition*. New York: The McGraw-Hill Companies.
- [11] LeSage, J. P. 1999. *The Theory and Practice of Spatial Econometrics*. Ohio: Department of Economics, University of Toledo.
- [12] LeSage, J., dan Pace, R. K. 2009. *Introduction to Spatial Econometrics*. New York: Taylor & Francis Group.
- [13] Montgomery, D. C., dan Peck, E. A. 1992. *Introduction To Linier Regression Analysis*. New York: John Wiley and Sons, Inc.
- [14] Montgomery, D. C., dan Runger, G. C. 2003. *Applied Statistics and Probability for Engineers*. New York: John Wiley & Sons, Inc.
- [15] Ramadani, I. R., Rahmawati, R., Hoyyi, A. 2013. "Analisis Faktor-faktor yang Mempengaruhi Gizi Buruk Balita di Jawa Tengah Dengan Metode Spatial Durbin Model". *Jurnal Gaussian*. Vol. 2, No. 4 : Hal. 333-342. Universitas Diponegoro, Semarang.
- [16] Shekhar, S., Lu, C. T., and Zhang, P. 2003. "A Unified Approach to Detecting Spatial Outliers". *GeoInformatica* 7:2, pp. 139-166.
- [17] Wuryandari, T., Hoyyi, A., Kusumawardani, D., Rahmawati, D. 2014. "Identifikasi Autokorelasi Spasial Pada Jumlah Pengangguran di Jawa Tengah Menggunakan Indeks Moran". *Jurnal Media Statistika*. Vol. 7, No. 1 : Hal. 1-10. Universitas Diponegoro.