

BAB I PENDAHULUAN

1.1 Latar Belakang

Regresi linier merupakan metode statistik untuk memodelkan hubungan antara variabel dependen dan variabel independen. Metode ini banyak diterapkan untuk memprediksi dan menganalisis hubungan antar variabel, termasuk dalam mengkaji pengaruh faktor pendidikan, kesehatan, dan ekonomi. Regresi linier berasumsi bahwa hubungan antar variabel bersifat tetap di seluruh unit pengamatan (Safitri & Amaliana, 2021). Asumsi ini kurang relevan ketika data yang dianalisis memiliki dimensi waktu dan spasial, karena hubungan antar variabel dapat bervariasi antar waktu maupun antar lokasi. Oleh karena itu, diperlukan pengembangan model lanjutan yang mampu mengakomodasi dinamika temporal dan heterogenitas spasial antar unit pengamatan secara lebih akurat.

Analisis data dengan dimensi waktu (*time series*) dan lintas individu (*cross section*) memerlukan pendekatan yang mengakomodasi keduanya secara bersamaan. Regresi data panel tepat digunakan karena dapat menganalisis perubahan dan perbedaan antar unit observasi, serta meningkatkan akurasi estimasi dengan mempertimbangkan variasi tiap unit (Chairunnisa & Fauzan, 2023). Regresi data panel memiliki tiga pendekatan yaitu *Common Effect Model* (CEM) yang mengasumsikan homogenitas antar unit dan waktu, *Fixed Effect Model* (FEM) mengakomodasi perbedaan antar unit dengan *intercept* berbeda, dan *Random Effect Model* (REM) yang menyatakan perbedaan antar unit bersifat acak dan tidak berkorelasi dengan variabel independen (Alviani, 2021). Namun, dalam penerapannya, regresi data panel sering digunakan pada data antar wilayah yang berdekatan secara geografis, sehingga *error* yang dihasilkan berpotensi bersifat heterogen akibat adanya keterkaitan antar wilayah. Oleh karena itu, regresi data panel perlu dikembangkan menjadi regresi panel spasial dengan menambahkan efek spasial untuk menangkap keterkaitan antar wilayah (Paranoan dkk., 2025).

Analisis data spasial merupakan metode statistik untuk mengidentifikasi pola hubungan antar wilayah berdasarkan karakteristik dan lokasi geografisnya (Paramita dkk., 2021). Regresi spasial dibedakan menjadi dua pendekatan berdasarkan tipe data, yaitu pendekatan titik dan pendekatan area (Mahading dkk., 2020). Pendekatan titik biasanya digunakan untuk data dengan lokasi spesifik, sedangkan pendekatan area digunakan untuk mengatasi dependensi spasial, yaitu ketergantungan antar wilayah yang berdekatan (Pratiwi dkk., 2018). Pendekatan area sesuai untuk menganalisis data Indeks Pembangunan Manusia (IPM), karena nilai IPM suatu wilayah dapat dipengaruhi oleh kondisi pembangunan wilayah disekitarnya, sehingga membentuk pola spasial. Beberapa model analisis spasial berbasis area yang umum digunakan antara lain *Spatial Autoregressive Model* (SAR) dan *Spatial Error Model* (SEM). Model SAR menggabungkan regresi linier dengan *lag* spasial pada variabel independen, yaitu saat nilai variabel dependen suatu wilayah berkorelasi dengan wilayah sekitarnya (Rahmawati & Bimanto, 2021). Sedangkan, model SEM digunakan ketika terdapat korelasi spasial pada komponen *error* akibat faktor yang tidak teramati secara langsung (Mar'ah dkk., 2025).

Pengembangan *Spatial Error Model* dengan *Random Effect* (SEM-RE) digunakan dalam penelitian ini untuk mengakomodasi dimensi waktu sekaligus perbedaan antar wilayah. Model ini mempertimbangkan ketergantungan spasial dan efek acak, sehingga sesuai untuk menganalisis IPM yang dipengaruhi oleh variasi wilayah dan perubahan antar waktu. Penelitian Andini & Wachidah (2023) menunjukkan bahwa SEM-RE memberikan hasil optimal dengan nilai *Aikake Information Criterion* (AIC) dan *Bayessian Information Criterion* (BIC) terkecil pada data kemiskinan di provinsi Jawa Tengah tahun 2011-2020.

Pada pemodelan regresi spasial data panel seperti SEM-RE, keberadaan pencilan (*outlier*) dapat memberikan pengaruh berlebihan terhadap *error* dan estimasi parameter, sehingga menurunkan akurasi model. Deteksi *outlier* dalam konteks dimensi spasial dan temporal dapat dilakukan menggunakan statistik *Difference in Fitted Value* (DFFITS) (Said dkk., 2024). Untuk mengatasinya, diperlukan pendekatan regresi yang tahan terhadap *outlier*, salah satunya regresi *robust*. Maharani (2024) menunjukkan bahwa model *Robust Spatial Autoregressive* (RSAR) dengan *MM-Estimator* mampu menangani *outlier* lebih akurat dibandingkan model klasik pada data tuberkulosis di Sulawesi Selatan. Regresi *robust* bertujuan menghasilkan estimasi yang tetap akurat meskipun terdapat *outlier* (Chen, 2002), salah satunya dengan pendekatan *Generalized M-Estimator* (GM-Estimator) yang merupakan pengembangan dari *M-Estimator* (Said dkk., 2024). GM-Estimator menggunakan fungsi pembobot untuk mengurangi pengaruh observasi ekstrem berleverage tinggi, sehingga SEM-RE menjadi lebih stabil meskipun data mengandung *outlier* spasial dan temporal. Penelitian Said dkk. (2024) membandingkan GM-Estimator dan MM-Estimator dalam pemodelan harga saham sektor teknologi tahun 2023, dan menunjukkan bahwa GM-Estimator memberikan estimasi yang lebih akurat.

Penelitian terkait SEM-RE dan GM-Estimator umumnya berfokus pada kasus sosial ekonomi sebagai faktor pembangunan manusia. Berdasarkan hal tersebut, penelitian ini mengkaji IPM di Indonesia tahun 2018-2022. Pemilihan periode ini memiliki nilai strategis karena mencerminkan tiga fase penting pembangunan nasional: periode stabil pada 2018–2019 sebelum pandemi, tahun krisis 2020 saat pandemi COVID-19 mengganggu pendidikan, kesehatan, dan daya beli, serta fase pemulihan 2021–2022 melalui intervensi masif pemerintah. Dinamika ini menjadikan periode 2018–2022 sangat representatif untuk menangkap variasi spasial dan temporal IPM secara komprehensif. IPM merupakan indikator komprehensif yang digunakan untuk mengukur kualitas hidup penduduk berdasarkan tiga dimensi utama, yaitu umur panjang dan hidup sehat (diukur melalui angka harapan hidup saat lahir), pengetahuan (diukur melalui angka melek huruf dan rata-rata lama sekolah), serta standar hidup layak (diukur melalui pendapatan per kapita yang disesuaikan) (Nasruddin & Azizah, 2022).

Berdasarkan data Badan Pusat Statistik (BPS), IPM Indonesia meningkat dari 71,39 pada tahun 2018 menjadi 72,91 pada tahun 2022, namun kenaikannya tidak merata. Provinsi seperti DKI Jakarta, mencatat IPM di atas 80, sedangkan Papua masih berada di kisaran 60 (BPS, 2024). Ketimpangan ini mencerminkan adanya dependensi spasial dan efek panel akibat perbedaan karakteristik antar provinsi serta fluktuasi antar waktu. Variasi besar ini berpotensi menghasilkan *outlier*,

yaitu observasi yang menyimpang jauh dari pola umum, yang dapat menurunkan akurasi estimasi model. Oleh karena itu, penelitian ini bertujuan membentuk model *Robust Spatial Error Model-Random Effect* (RSEM-RE) dengan *GM-Estimator* untuk mengidentifikasi faktor-faktor yang secara signifikan memengaruhi IPM di Indonesia selama periode 2018-2022. Hasil penelitian ini dapat menjadi referensi dan acuan bagi instansi terkait dalam merumuskan kebijakan pembangunan yang lebih efektif.

1.2 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data IPM di Indonesia tahun 2018-2022.
2. Fungsi pembobot yang digunakan pada regresi robust adalah *Schweppe*.
3. Matriks pembobot spasial yang digunakan berbasis *K-Nearest Neighbors* (K-NN) yang ditentukan berdasarkan jarak *haversine*.
4. Penentuan model regresi terbaik berdasarkan nilai R^2 dan RMSE.

1.3 Tujuan dan Manfaat Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Mendapatkan model SEM-RE dan RSEM-RE untuk data IPM di Indonesia tahun 2018-2022.
2. Mengidentifikasi faktor-faktor yang memengaruhi IPM di Indonesia tahun 2018-2022 berdasarkan model RSEM-RE.

Berdasarkan tujuan tersebut, hasil penelitian ini diharapkan dapat memberikan manfaat sebagai berikut.

1. Menambah informasi mengenai pemodelan RSEM-RE menggunakan *GM-Estimator* dan pembobot spasial berbasis K-NN dalam menganalisis IPM di Indonesia tahun 2018-2022.
2. Menambah wawasan pembaca dalam mengembangkan analisis regresi panel spasial menggunakan model RSEM-RE pada berbagai studi kasus, khususnya pada data IPM.
3. Memberikan informasi berbasis data yang dapat menjadi referensi bagi instansi terkait dan masyarakat mengenai faktor-faktor yang memengaruhi IPM pada tahun 2018-2022.

1.4 Landasan Teori

1.4.1 Regresi Data Panel

Regresi data panel merupakan metode statistik yang digunakan untuk menganalisis pengaruh variabel independen terhadap variabel dependen pada data yang menggabungkan dimensi *time series* dan *cross section* pada unit yang diamati berulang kali dalam periode berbeda (Rahayu dkk., 2023). Model umum regresi data panel ditunjukkan pada Persamaan (1).

$$y_{it} = \beta_0 + \sum_{k=1}^K \beta_k x_{kit} + \varepsilon_{it} \quad (1)$$

dengan,

- y_{it} : nilai variabel dependen individu ke- i untuk periode waktu ke- t
 β_0 : parameter *intercept*
 β_k : parameter yang ditaksir dari koefisien variabel independen
 x_{kit} : nilai variabel independen ke- k untuk individu ke- i periode waktu ke- t
 ε_{it} : *error* pada unit observasi ke- i dan waktu ke- t
 i : unit *cross section* (1,2,3, ..., N)
 t : unit *time series* (1,2,3, ..., T)

Model regresi data panel dapat dibagi menjadi tiga jenis berdasarkan asumsi pengaruh yang digunakan, yaitu (Alamsyah dkk., 2022):

1. Jika model tidak memperhitungkan perbedaan perilaku individu dan waktu, sehingga nilai *intercept* (β_0) yang sama untuk seluruh observasi, maka model disebut *Common Effect Model* (CEM) yang ditunjukkan pada Persamaan (1).
2. Jika diasumsikan bahwa *intercept* (β_0) berbeda antar individu tetapi tetap sepanjang waktu, serta *slope* (β_k) sama untuk semua individu dan waktu, maka digunakan *Fixed Effect Model* (FEM) untuk mengatasi keterbatasan ini, Model FEM ditunjukkan pada Persamaan (2).

$$y_{it} = \beta_{0i} + \sum_{k=1}^K \beta_k x_{kit} + \varepsilon_{it} \quad (2)$$

3. *Random Effect Model* (REM) mempertimbangkan perbedaan *intercept* (β_0) yang mungkin terjadi antar individu dan waktu yang mungkin terjadi. Bentuk umum model REM ditunjukkan pada Persamaan (3).

$$y_{it} = \beta_0 + \sum_{k=1}^K \beta_k x_{kit} + v_{it}; v_{it} = u_i + \varepsilon_{it} \quad (3)$$

1.4.2 Pemilihan Model Data Panel

Pemilihan model estimasi regresi data panel bertujuan menentukan model terbaik di antara CEM, FEM, dan REM. Uji spesifikasi model yang digunakan, sebagai berikut (Indrasetianingsih & Wasik, 2020):

1. Uji Chow

Uji Chow digunakan untuk menentukan pilihan antara FEM atau CEM yang tepat digunakan. Hipotesis yang digunakan:

$H_0: \beta_{01} = \beta_{02} = \dots = \beta_{0N} = 0$ (CEM)

H_1 : minimal ada satu $\beta_{0i} \neq 0$ (FEM); $i = 1, 2, \dots, N$

Statistik uji yang digunakan:

$$F_{hit} = \frac{\left(\frac{SSE_1 - SSE_2}{N - 1} \right)}{\left(\frac{SSE_2}{NT - N - K} \right)} \sim F_{(N-1), (NT-N-K)} \quad (4)$$

dengan,

SSE_1 : *Sum Square Error* CEM

SSE_2 : *Sum Square Error* FEM

N : jumlah unit *cross section*

T : jumlah unit *time series*

K : jumlah variabel independen

Kriteria penolakan H_0 adalah jika $F_{hit} \geq F_{(N-1),(NT-N-K)}$ dengan keputusan bahwa model yang paling tepat adalah FEM, sebaliknya terima H_0 yang berarti model yang tepat adalah CEM (Alamsyah dkk., 2022).

2. Uji Hausman

Uji Hausman digunakan untuk menentukan model terbaik antara *Fixed Effect Model* atau *Random Effect Model* yang tepat digunakan. Hipotesis yang digunakan:

$H_0: corr(X_{it}, \varepsilon_{it}) = 0$ (REM)

$H_1: corr(X_{it}, \varepsilon_{it}) \neq 0$ (FEM); $i = 1, 2, \dots, N$; $t = 1, 2, \dots, T$

Statistik uji yang digunakan:

$$\chi^2 = (\hat{\beta}_{FEM} - \hat{\beta}_{REM})^t [var(\hat{\beta}_{FEM}) - var(\hat{\beta}_{REM})]^{-1} (\hat{\beta}_{FEM} - \hat{\beta}_{REM}) \quad (5)$$

Kriteria penolakan H_0 adalah jika $\chi^2 \geq \chi^2_{(\alpha, K)}$ atau $p - value < \alpha$, sebaliknya terima H_0 yang berarti model yang tepat adalah REM (Alamsyah dkk., 2022).

1.4.3 Random Effect Model

Random Effect Model (REM) digunakan untuk mengestimasi data panel dengan variabel gangguan yang dapat saling berhubungan antar waktu atau antar individu. *Fixed Effect Model* (FEM) memiliki kelemahan, salah satunya adalah berkurangnya nilai derajat kebebasan (*degree of freedom*) akibat penambahan banyak variabel *dummy* untuk setiap unit individu atau waktu, yang berdampak pada penurunan efisiensi estimasi parameter. REM mengatasi masalah ini dengan menganggap perbedaan antar unit atau waktu sebagai bagian dari komponen *error* yang acak, bukan sebagai bagian dari intersep tetap (Septianingsih, 2022). Model REM dapat dituliskan seperti pada persamaan (3).

Keterangan:

y_{it} : nilai variabel dependen individu ke- i untuk periode waktu ke- t

β_0 : parameter *intercept*

β_k : parameter yang ditaksir dari koefisien variabel independen ($k = 1, 2, \dots, K$)

x_{kit} : nilai variabel independen ke- k untuk individu ke- i periode waktu ke- t

u_i : *error* pada unit *cross section*

ε_{it} : *error* pada unit *time series*

Dengan demikian, matriks dari persamaan (3) ditunjukkan seperti persamaan (6):

$$y = X\beta + v \quad (6)$$

dengan,

y : vektor observasi dengan ukuran $NT \times 1$

X : matriks variabel independen dengan ukuran $(NT \times (K + 1))$

β : vektor koefisien dengan ukuran $((K + 1) \times 1)$

v : vektor *error* komponen dengan ukuran $NT \times 1$

Berdasarkan struktur *error* pada persamaan (3), maka variansi dari *error* total v_{it} ditunjukkan pada persamaan (7).

$$Var(v_{it}) = Var(u_i + \varepsilon_{it}) = Var(u_i) + Var(\varepsilon_{it}) = \hat{\sigma}_u^2 + \hat{\sigma}_\varepsilon^2 \quad (7)$$

Selain itu, efek individu u_i bersifat tetap untuk setiap waktu t , maka terdapat kovariansi antar waktu pada unit yang sama, yaitu

$$Cov(v_{it}, v_{is}) = Cov(u_i + \varepsilon_{it}, u_i + \varepsilon_{is}) = Cov(u_i, u_i) = \hat{\sigma}_u^2 \text{ untuk } t \neq s \quad (8)$$

Berdasarkan persamaan (8), struktur matriks variansi-kovariansi v_{it} untuk REM dituliskan pada persamaan (9) (Greene, 2002).

$$\Omega = Var(v_i) = \begin{bmatrix} \hat{\sigma}_u^2 + \hat{\sigma}_\varepsilon^2 & \hat{\sigma}_u^2 & \cdots & \hat{\sigma}_u^2 \\ \hat{\sigma}_u^2 & \hat{\sigma}_u^2 + \hat{\sigma}_\varepsilon^2 & \cdots & \hat{\sigma}_u^2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_u^2 & \hat{\sigma}_u^2 & \cdots & \hat{\sigma}_u^2 + \hat{\sigma}_\varepsilon^2 \end{bmatrix} = \hat{\sigma}_u^2 (\mathbf{I}_N \otimes \mathbf{J}_T) + \hat{\sigma}_\varepsilon^2 \mathbf{I}_{NT} \quad (9)$$

Model REM memiliki *error* yang bersifat heteroskedastik dan menunjukkan autokorelasi, sehingga diperlukan transformasi untuk memastikan *error* menjadi homoskedastik dan tidak berkorelasi. Transformasi dilakukan dengan mengalikan kedua sisi dengan $\Omega^{-\frac{1}{2}}$. Sehingga diperoleh persamaan (10).

$$\Omega^{-\frac{1}{2}} \mathbf{y} = \Omega^{-\frac{1}{2}} (\mathbf{X}\boldsymbol{\beta} + \mathbf{v}) \quad (10)$$

Persamaan (10) kemudian disederhanakan menjadi persamaan (11).

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\mathbf{v}} \quad (11)$$

Karena $\tilde{\mathbf{v}}$ telah memenuhi asumsi *Ordinary Least Squares* (OLS), maka prinsip OLS dapat digunakan untuk meminimalkan *Sum Square Error* (SSE). Berdasarkan persamaan (11) SSE dapat diperoleh seperti pada persamaan (12).

$$SSE = \tilde{\mathbf{v}}^T \tilde{\mathbf{v}} = (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})^T (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}) \quad (12)$$

Untuk mendapatkan penduga $\hat{\boldsymbol{\beta}}$ dapat dilakukan dengan meminimalkan SSE, yaitu melakukan turunan parsial SSE terhadap penduga $\hat{\boldsymbol{\beta}}$ lalu disamakan dengan nol seperti pada persamaan (13).

$$\frac{\partial SSE}{\partial \boldsymbol{\beta}} = 0$$

$$\frac{\partial \left((\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})^T (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}) \right)}{\partial \boldsymbol{\beta}} = 0$$

$$-2\tilde{\mathbf{X}}^T \tilde{\mathbf{y}} + 2\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}\hat{\boldsymbol{\beta}} = 0$$

$$\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}\hat{\boldsymbol{\beta}} = \tilde{\mathbf{X}}^T \tilde{\mathbf{y}} \quad (13)$$

Berdasarkan persamaan (13) diperoleh persamaan (14)

$$\tilde{\mathbf{X}} = \Omega^{-\frac{1}{2}} \mathbf{X}, \tilde{\mathbf{y}} = \Omega^{-\frac{1}{2}} \mathbf{y} \quad (14)$$

Kemudian terbentuk persamaan (15) (Greene, 2002)

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}^T \Omega^{-1} \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \Omega^{-1} \tilde{\mathbf{y}} = (\mathbf{X}^T \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Omega^{-1} \mathbf{y} \quad (15)$$

1.4.4 Regresi Spasial

Regresi spasial merupakan pengembangan dari regresi linier berganda yang memperhitungkan faktor lokasi geografis dalam analisis hubungan antara variabel independen dan dependen. Metode ini mengakomodasi dua jenis efek spasial, yaitu dependensi spasial dan heterogenitas spasial (Anselin, 1988). Model umum regresi spasial ditunjukkan pada persamaan (16) (LeSage, 2008):

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \text{ dimana } \mathbf{u} = \lambda \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma_{\varepsilon}^2 \mathbf{I}_N) \quad (16)$$

dengan,

- $\mathbf{y}$: vektor variabel dependen berukuran $NT \times 1$
- $\boldsymbol{\beta}$: vektor parameter regresi berukuran $(K + 1) \times 1$
- ρ : parameter koefisien spasial *lag* variabel dependen
- $\mathbf{X}$: matriks variabel dependen berukuran $NT \times (K + 1)$
- $\mathbf{W}$: matriks pembobot dengan ukuran $N \times N$
- λ : parameter koefisien spasial pada *error*
- $\mathbf{u}$: vektor *error* yang mempunyai efek spasial dengan ukuran $NT \times 1$
- $\boldsymbol{\varepsilon}$: vektor *error* dengan ukuran $NT \times 1$

Menurut Anselin (1988) dari persamaan umum model regresi spasial pada persamaan (16) dapat dibentuk beberapa model lain, seperti berikut:

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

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (17)$$

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

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \lambda \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon} \end{aligned} \quad (18)$$

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

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \lambda \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon} \end{aligned} \quad (19)$$

1.4.5 Matriks Pembobot Spasial

Matriks pembobot spasial *K-Nearest Neighbors* (K-NN) adalah matriks yang digunakan untuk mengetahui efek spasial berdasarkan K tetangga terdekat (Agustina dkk., 2022). Dalam analisis spasial geografis, penentuan tetangga ini tidak cukup hanya menggunakan jarak *eulidean* karena mengabaikan kelengkungan bumi. Oleh karena itu, jarak *haversine* digunakan sebagai alternatif yang lebih akurat, karena menghitung jarak lengkung di permukaan bumi berdasarkan koordinat lintang (*longitude*) dan bujur (*latitude*). Pembobot K-NN dengan perhitungan jarak *haversine* ditunjukkan pada persamaan (20) (Yulianto dkk., 2018).

$$d_{ij} = R \cdot c \quad (20)$$

dengan,

$$c = 2 \cdot \arcsin(\sqrt{a})$$

$$a = \sin^2\left(\frac{\Delta lat}{2}\right) + \cos(v_i) \cdot \cos(v_j) \cdot \sin^2\left(\frac{\Delta long}{2}\right)$$

$$R = 6,371 \text{ km}$$

Matriks W pembobot spasial K-NN memberikan bobot 1 pada K lokasi terdekat dari setiap lokasi i (pada kolom j yang sesuai), sementara kolom-kolom lainnya bernilai 0 (Permana dkk., 2024). Rumus pembobot spasial K-NN didefinisikan pada persamaan (21).

$$W = \begin{cases} 1, j \text{ adalah salah satu dari } K \text{ tetangga terdekat dari lokasi } i \\ 0, \text{lainnya} \end{cases} \quad (21)$$

Pada pembobot spasial K-NN nilai K lokasi terdekat ditentukan oleh peneliti. Standarisasi matriks dilakukan untuk membuat jumlah elemen pada setiap baris menjadi sama dengan satu menggunakan persamaan (22) (Murdani dkk., 2022):

$$w_{ij} = \frac{c_{ij}}{\sum c_i}, i = 1, 2, \dots, n; j = 1, 2, \dots, k \quad (22)$$

Keterangan:

w_{ij} : nilai matriks pembobot spasial pada baris ke- i dan kolom ke- j

c_{ij} : nilai matriks K-NN pada baris ke- i kolom ke- j

$\sum c_i$: total nilai matriks K-NN pada baris ke- i

1.4.6 Uji Dependensi Spasial

Dependensi spasial terjadi ketika nilai observasi pada suatu wilayah dipengaruhi oleh nilai observasi di wilayah lain yang berdekatan secara geografis (Agustina dkk., 2022). Salah satu uji yang digunakan untuk mengidentifikasi adanya dependensi spasial yaitu *Morans' I* (Maharani, 2024). Hipotesis yang digunakan:

$H_0: I = 0$ (tidak terdapat dependensi spasial)

$H_1: I \neq 0$ (terdapat dependensi spasial)

Perhitungan uji *Moran's I* dilakukan menggunakan persamaan (23) (Iffa dkk., 2023):

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (23)$$

Statistik uji *Moran's I* (Anselin, 1988) yaitu:

$$Z_{hitung} = \frac{I - E(I)}{\sqrt{Var(I)}} \sim N(0,1) \quad (24)$$

dengan,

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

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

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}; S_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2; S_2 = \sum_{i=1}^n \left(\sum_{j=1}^n w_{ij} + \sum_{j=1}^n w_{ji} \right)^2$$

Keterangan:

- I : indeks Moran
 n : jumlah lokasi amatan
 x_i : nilai pengamatan pada lokasi ke- i
 x_j : nilai pengamatan pada lokasi ke- j
 $\bar{x}$: rata-rata nilai pengamatan
 w_{ij} : elemen matriks pembobot spasial terstandarisasi antara lokasi i dan j
 $E(I)$: nilai ekspektasi dari I
 $Var(I)$: nilai variansi dari I

Pengambilan keputusan pada uji ini dilakukan dengan menolak H_0 jika nilai $Z_{hitung} \geq Z_{\alpha/2}$ atau $p - value < \alpha$ yang menunjukkan terdapat dependensi spasial. Adapun kriteria uji *Moran's I*:

- Jika $I > E(I)$ maka terjadi autokorelasi positif, artinya lokasi-lokasi yang berdekatan memiliki nilai efek spasial yang serupa.
- Jika $I = E(I)$ maka tidak terjadi autokorelasi, artinya data tidak berkelompok berdasarkan wilayahnya.
- Jika $I < E(I)$ maka terjadi autokorelasi negatif, artinya lokasi-lokasi yang berdekatan memiliki nilai efek spasial yang berbeda (Juliani & Nasution, 2024).

1.4.7 Pengujian *Lagrange Multiplier*

Uji *Lagrange Multiplier* (LM) digunakan untuk mengidentifikasi efek dependensi spasial dalam data, sehingga membantu menentukan model regresi spasial yang tepat (Usali dkk., 2021). Jika LM_{lag} signifikan, model SAR adalah model yang sesuai. Jika LM_{error} signifikan, model SEM adalah model yang sesuai (Anselin, 2010). Apabila kedua uji LM_{lag} dan LM_{error} tidak menunjukkan signifikansi, maka model yang paling sesuai adalah regresi berganda dengan metode *Ordinary Least Square* (OLS) (Agustina dkk., 2022).

1. *Lagrange Multiplier Lag*

Uji *lagrange multiplier lag* bertujuan mendeteksi ketergantungan spasial pada *lag*, dan jika hasilnya signifikan, maka penelitian dapat dilanjutkan dengan menggunakan model regresi spasial SAR (Tumanggor & Simamora, 2023). Hipotesis pemodelan spasial *lag*:

$H_0: \rho = 0$ (tidak ada ketergantungan spasial *lag*)

$H_1: \rho \neq 0$ (terdapat ketergantungan spasial *lag*)

Statistik uji LM_{lag} dihitung dengan menggunakan persamaan (25).

$$LM_{SAR} = \frac{[e^t(I_T \otimes W)y/\hat{\sigma}_e^2]^2}{J} \quad (25)$$

dengan,

$$J = \frac{1}{\hat{\sigma}_e^2} \left[\left((I_T \otimes W)X\hat{\beta} \right)^t (I_{NT} - X(X^tX)^{-1}X^t(I_T \otimes W)X\hat{\beta} + TT_W - \hat{\sigma}_e^2 \right]$$

$$T_W = tr(W^2 + W^tW)$$

Keterangan:

- e : vektor *error* model
 NT : jumlah lokasi observasi
 W : matriks pembobot spasial terstandarisasi berukuran $N \times N$
 X : matriks variabel independen dengan ukuran $NT \times (K + 1)$
 β : vektor koefisien parameter regresi dengan ukuran $(K + 1) \times 1$
 I : matriks identitas berukuran $T \times 1$
 y : vektor variabel dependen berukuran $NT \times 1$

Kriteria pengambilan keputusan adalah menolak H_0 jika $LM_{SAR} \geq \chi^2_{(\alpha;1)}$ atau $p - value < \alpha$, artinya terdapat dependensi spasial pada *lag*, maka model SAR yang sesuai untuk digunakan.

2. Lagrange Multiplier Error

Uji *lagrange multiplier error* bertujuan mendeteksi adanya ketergantungan spasial pada *error*, dan jika hasilnya signifikan, maka penelitian dilanjutkan menggunakan model regresi spasial SEM (Tumanggor & Simamora, 2023). Hipotesis pemodelan spasial *error*:

$H_0: \lambda = 0$ (tidak ada kebergantungan spasial *error*)

$H_1: \lambda \neq 0$ (terdapat kebergantungan spasial *error*)

Statistik uji LM_{error} dihitung dengan menggunakan persamaan (26).

$$LM_{SEM} = \frac{[e^t(I_T \otimes W)e/\hat{\sigma}_e^2]^2}{TT_W} \quad (26)$$

Kriteria pengambilan keputusan adalah menolak H_0 jika $LM_{SEM} \geq \chi^2_{(\alpha;1)}$ atau $p - value < \alpha$, artinya terdapat dependensi spasial pada *error*, maka model SEM yang sesuai untuk digunakan.

1.4.8 Regresi Spasial Data Panel

Menurut Rizki dan Taqiyyuddin (2021), regresi data panel pada beberapa wilayah dapat menimbulkan heterogenitas *error* akibat ketergantungan antar unit *cross section* (Rizki & Taqiyyuddin, 2021). Oleh karena itu, aspek spasial perlu dipertimbangkan dalam pemodelan data yang melibatkan wilayah dan waktu. Persamaan (27) menunjukkan model regresi linier gabungan dengan efek spesifik spasial tanpa efek interaksi spasial (Elhorst, 2014).

$$y_{it} = X_{it}\beta + \mu_i + \varepsilon_{it} \quad (27)$$

dengan,

- i : unit *cross section* (1,2,3, ..., N)
 t : unit *time series* (1,2,3, ..., T)
 y_{it} : variabel dependen pada unit ke- i dan waktu ke- t
 X_{it} : matriks variabel independen pada unit ke- i dan waktu ke- t
 β : vektor $((K + 1) \times 1)$ untuk parameter dari variabel independen
 μ_i : efek spesifik spasial pada unit ke- i
 ε_{it} : *error* pada unit ke- i dan waktu ke- t

Spesifikasi interaksi antar unit spasial dapat mencakup variabel dependen dengan spasial *lag* atau menyertakan aspek spasial dalam proses *error*nya, yang dikenal sebagai model spasial *lag* dan model spasial *error* (Elhorst, 2009).

1.4.9 Estimasi Parameter Model Spasial Data Panel

1. *Spatial Autoregressive Random Effect*

Spatial Autoregressive Random Effect (SAR-RE) merupakan model regresi data panel yang menggabungkan pengaruh spasial dan efek acak. Model ini memperhitungkan ketergantungan antar wilayah sehingga nilai variabel dependen pada suatu wilayah dipengaruhi oleh nilai variabel dependen di wilayah sekitarnya (*spatial lag*), serta memasukkan efek acak untuk menangkap variasi yang tidak teramati antar unit *cross-section* (Salima, 2018). Model SAR-RE ditunjukkan pada Persamaan (28).

$$y_{it} = \rho \sum_j^N W_{ij} y_{jt} + X_{it} \beta + u_i + \varepsilon_{it} \quad (28)$$

- y_{it} : nilai variabel dependen berukuran $NT \times 1$
 W_{ij} : matriks pembobot spasial W pada baris i dan kolom j
 ρ : koefisien parameter spasial *lag* dari variabel dependen
 X_{it} : matriks variabel independen berukuran $NT \times (K + 1)$
 β : vektor parameter regresi berukuran $((K + 1) \times 1)$
 u_i : vektor efek acak spesifik individu ke- i berukuran $N \times 1$
 ε_{it} : vektor *error* berukuran $NT \times 1$

2. *Spatial Error Model Random Effect*

Spatial Error Model Random Effect (SEM-RE) adalah model regresi data panel yang menangkap ketergantungan spasial pada komponen *error* serta memperhitungkan efek acak antar unit *cross-section*. Variasi *error* pada suatu wilayah dipengaruhi oleh variasi *error* di wilayah sekitarnya, bukan secara langsung melalui variabel dependen. Model SEM-RE ditunjukkan pada Persamaan (29) (Sari & Rosadi, 2013).

$$y_{it} = X_{it} \beta + u_i + \varepsilon_{it}; \quad \varepsilon_{it} = \lambda W_{ij} \varepsilon_{jt} + v_{it} \quad (29)$$

- y_{it} : nilai variabel dependen berukuran $NT \times 1$
 W_{ij} : matriks pembobot spasial W pada baris i dan kolom j
 λ : koefisien parameter spasial *error* dari variabel dependen
 X_{it} : matriks variabel independen $NT \times (K + 1)$
 β : vektor parameter regresi $(K + 1) \times 1$
 ε_{it} : vektor komponen *error* spasial $NT \times 1$
 u_i : vektor efek acak spesifik individu ke- i berukuran $NT \times 1$
 v_{it} : vektor *error* berukuran $NT \times 1$

Total *error* dalam model:

$$e_{it} = u_i + \varepsilon_{it}$$

Bentuk matriks dari persamaan (29) untuk *cross section* dalam waktu t .

$$y_t = X_t \beta + e_t \quad (30)$$

$$e_t = u + \varepsilon_t \quad (31)$$

$$\varepsilon_t = \lambda W \varepsilon_t + v_t \quad (32)$$

Persamaan (32) kemudian ditulis

$$\varepsilon_t - \lambda W \varepsilon_t = v_t$$

$$(I_N - \lambda W) \varepsilon_t = v_t$$

$$\varepsilon_t = (I_N - \lambda W)^{-1} v_t$$

Misalkan, $B = I_N - \lambda W$, diperoleh:

$$\varepsilon_t = B^{-1} v_t$$

Error term pada persamaan (31) dapat dituliskan dalam bentuk sebagai berikut:

$$e = (l_T \otimes I_N)u + (I_T \otimes B^{-1})v \quad (33)$$

dengan,

l_T : vektor unit elemen berdimensi $T \times 1$

I_T : matriks identitas berdimensi $T \times T$

I_N : matriks identitas berdimensi $N \times N$

Apabila u_i dianggap sebagai variabel acak, maka memiliki

$$E(u_i u_i^t) = \begin{cases} \sigma_u^2, & i = j \\ 0, & \text{yang lain} \end{cases}$$

Pada regresi panel linier *random effect*, $u \sim N(0, \sigma_u^2)$ dan $v_{it} \sim N(0, \sigma_v^2)$, sehingga matriks varians-kovarians dari persamaan (33)

$$\begin{aligned} \Omega &= E(ee^t) = E \left[\left((l_T \otimes I_N)u + (I_T \otimes B^{-1})v \right) \left((l_T \otimes I_N)u + (I_T \otimes B^{-1})v \right)^t \right] \\ &= E \left[\left((l_T \otimes I_N)u + (I_T \otimes B^{-1})v \right) (u^t (l_T \otimes I_N)^t + v^t (I_T \otimes B^{-1})^t) \right] \\ &= E(uu^t) (l_T l_T^t \otimes I_N I_N^t) + E(uv^t) (l_T l_T^t \otimes I_N B^{-1^t}) \\ &\quad + E(vu^t) (l_T l_T^t \otimes B^{-1} I_N^t) + E(vv^t) (I_T I_T^t \otimes B^{-1} B^{-1^t}) \\ \Omega &= \sigma_u^2 (l_T l_T^t \otimes I_N I_N^t) + \sigma_v^2 (I_T I_T^t B^{-1} B^{-1^t}) \end{aligned} \quad (34)$$

atau matriks varian kovarian juga dapat dinyatakan sebagai:

$$\begin{aligned} \Omega &= \frac{1}{T} l_T l_T^t \otimes (T \sigma_u^2 I_N + (B^t B)^{-1}) + \sigma_v^2 \left(\left(I_T - \frac{1}{T} l_T l_T^t \right) \otimes (B^t B)^{-1} \right) \\ &= \sigma_v^2 \left[\frac{1}{T} l_T l_T^t \otimes \left(T \frac{\sigma_u^2}{\sigma_v^2} I_N + (B^t B)^{-1} \right) + \left(\left(I_T - \frac{1}{T} l_T l_T^t \right) \otimes (B^t B)^{-1} \right) \right] \\ &= \sigma_v^2 \left[\frac{1}{T} l_T l_T^t \otimes (T \phi I_N + (B^t B)^{-1}) + \left(\left(I_T - \frac{1}{T} l_T l_T^t \right) \otimes (B^t B)^{-1} \right) \right] \\ \Omega &= \sigma_v^2 \Sigma \end{aligned}$$

dengan,

$$\Sigma = \frac{1}{T} l_T l_T^t \otimes (T \phi I_N + (B^t B)^{-1}) + \left(\left(I_T - \frac{1}{T} l_T l_T^t \right) \otimes (B^t B)^{-1} \right) \quad (35)$$

$$\Sigma^{-1} = \frac{1}{T} l_T l_T^t \otimes (T \phi I_N + (B^t B)^{-1})^{-1} + \left(I_T - \frac{1}{T} l_T l_T^t \right) \otimes (B^t B) \quad (36)$$

$$|\Sigma| = |T \phi I_N + (B^t B)^{-1}| \cdot |(B^t B)^{-1}|^{T-1} \quad (37)$$

Error total memiliki distribusi $e \sim N(0, \Sigma)$. $e = Y - X\beta$ dan $\Omega = \sigma_v^2 \Sigma$

Fungsi kepadatan peluang:

$$f(e | \sigma_v^2) = \frac{1}{(2\pi)^{NT/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} e^t (\sigma_v^2 \Sigma)^{-1} e \right)$$

Sehingga fungsi likelihood yang diperoleh:

$$\begin{aligned}
 L(\sigma_v^2, e) &= \prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma_v^2|\Sigma|}} \exp\left(-\frac{1}{2} \mathbf{e}^t (\sigma_v^2 \Sigma)^{-1} \mathbf{e}\right) \\
 &= \frac{1}{(2\pi\sigma_v^2)^{NT/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2\sigma_v^2} \mathbf{e}^t \Sigma^{-1} \mathbf{e}\right) \\
 L(\lambda, \beta, \sigma_v^2) &= \ln \frac{1}{(2\pi\sigma_v^2)^{NT/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2\sigma_v^2} \mathbf{e}^t \Sigma^{-1} \mathbf{e}\right) \\
 &= -\frac{NT}{2} \ln(2\pi\sigma_v^2) - \frac{1}{2} \ln|\Sigma| - \frac{1}{2\sigma_v^2} \mathbf{e}^t \Sigma^{-1} \mathbf{e} \\
 &= -\frac{NT}{2} \ln(2\pi\sigma_v^2) - \frac{1}{2} \ln|T\phi \mathbf{I}_N + (\mathbf{B}^t \mathbf{B})^{-1}| + (T-1) \sum_{i=1}^N \ln(1 - \lambda \mathbf{W}) - \\
 &\quad \frac{1}{2\sigma_v^2} \mathbf{e}^t \Sigma^{-1} \mathbf{e}
 \end{aligned} \tag{38}$$

$\ln|T\phi \mathbf{I}_N + (\mathbf{B}^t \mathbf{B})^{-1}|$ didefinisikan sebagai suatu fungsi akar karakteristik dari $\mathbf{W}$, sehingga diperoleh:

$$\ln|T\phi \mathbf{I}_N + (\mathbf{B}^t \mathbf{B})^{-1}| = \prod_{i=1}^N \ln \left| T\phi + \frac{1}{(1 - \lambda \mathbf{W})^2} \right| \tag{39}$$

Sehingga, dengan menggunakan persamaan (39), maka persamaan (38) menjadi:

$$\begin{aligned}
 L(\lambda, \beta, \sigma_v^2) &= -\frac{NT}{2} \ln(2\pi\sigma_v^2) - \frac{1}{2} \ln \left| T\phi + \frac{1}{(1 - \lambda \mathbf{W})^2} \right| + (T-1) \sum_{i=1}^N \ln(1 - \lambda \mathbf{W}) \\
 &\quad - \frac{1}{2\sigma_v^2} \mathbf{e}^t \Sigma^{-1} \mathbf{e} \\
 &= -\frac{NT}{2} \ln(2\pi\sigma_v^2) - \frac{1}{2} \ln|1 + T\phi(1 - \lambda \mathbf{W})^2| + \sum_{i=1}^N \ln(1 - \lambda \mathbf{W}) + \\
 &\quad T \sum_{i=1}^N \ln(1 - \lambda \mathbf{W}) - \sum_{i=1}^N \ln(1 - \lambda \mathbf{W}) - \frac{1}{2\sigma_v^2} \mathbf{e}^t \Sigma^{-1} \mathbf{e} \\
 &= -\frac{NT}{2} \ln(2\pi\sigma_v^2) - \frac{1}{2} \ln|1 + T\phi(1 - \lambda \mathbf{W})^2| + T \sum_{i=1}^N \ln(1 - \lambda \mathbf{W}) - \\
 &\quad \frac{1}{2\sigma_v^2} \mathbf{e}^t \Sigma^{-1} \mathbf{e}
 \end{aligned} \tag{40}$$

Estimasi Parameter β

Estimasi parameter β didapatkan dengan memaksimumkan fungsi *ln likelihood* pada persamaan (40) yaitu dengan cara menurunkan persamaan tersebut terhadap β dan menyamakannya dengan nol.

$$\frac{\partial l}{\partial \beta} = 0$$

$$\frac{\partial \left(-\frac{NT}{2} \ln(2\pi\sigma_v^2) - \frac{1}{2} \ln|1 + T\phi(1 - \lambda \mathbf{W})^2| + T \sum_{i=1}^N \ln(1 - \lambda \mathbf{W}) - \frac{1}{2\sigma_v^2} \mathbf{e}^t \Sigma^{-1} \mathbf{e} \right)}{\partial \beta} = 0$$

$$\frac{\partial \left(-\frac{1}{2\sigma_v^2} \mathbf{e}^t \Sigma^{-1} \mathbf{e} \right)}{\partial \beta} = 0$$

$$\frac{\partial \left(-\frac{1}{2\sigma_v^2} (\mathbf{y} - \mathbf{X}\beta)^t \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta) \right)}{\partial \beta} = 0$$

$$\frac{\partial \left(-\frac{1}{2\sigma_v^2} (\mathbf{y} \Sigma^{-1} \mathbf{y} - 2\mathbf{X}^t \beta \Sigma^{-1} \mathbf{y} + \mathbf{X}^t \beta \Sigma^{-1} \mathbf{X} \beta) \right)}{\partial \beta} = 0$$

$$-\frac{1}{2\sigma_v^2}(-2X^t\Sigma^{-1}y + 2\beta X^t\Sigma^{-1}X) = 0$$

$$-2X^t\Sigma^{-1}y + 2\beta X^t\Sigma^{-1}X = 0$$

$$X^t\Sigma^{-1}y - \beta X^t\Sigma^{-1}X = 0$$

$$X^t\Sigma^{-1}y = \beta X^t\Sigma^{-1}X$$

$$\hat{\beta} = (X^t\Sigma^{-1}X)^{-1}(X^t\Sigma^{-1}y) \quad (41)$$

1.4.10 Pengujian Model Regresi

1. Uji Signifikansi Parameter

Uji Wald merupakan metode yang sesuai untuk menguji signifikansi koefisien model dalam data panel spasial (Adiwinata & Annisa, 2024). Pengujian parameter λ dan β dilakukan sebagai berikut.

- a. Pengujian parameter λ digunakan hipotesis dan statistik uji yaitu:

$H_0: \lambda = 0$ (parameter tidak signifikan)

$H_1: \lambda \neq 0$ (parameter signifikan)

Statistik uji Wald (W) untuk parameter λ ditunjukkan pada persamaan (42):

$$W_\lambda = \left(\frac{\hat{\lambda}}{SE(\hat{\lambda})} \right)^2 \quad (42)$$

dengan $SE(\hat{\lambda})$ merupakan dugaan *error* baku untuk koefisien λ dan $\hat{\lambda}$ merupakan penduga koefisien parameter spasial *error*.

- b. Pengujian parameter β digunakan hipotesis dan statistik uji yaitu:

$H_0: \beta_k = 0$ (parameter tidak signifikan)

$H_1: \beta_k \neq 0$ (parameter signifikan)

Statistik uji Wald (W) untuk parameter λ ditunjukkan pada persamaan (43)

$$W_\beta = \left(\frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \right)^2 \quad (43)$$

dengan $SE(\hat{\beta}_k)$ merupakan dugaan *error* baku untuk koefisien β dan $\hat{\beta}_k$ merupakan penduga koefisien parameter pada variabel independen ke- k .

Kriteria pengambilan keputusan adalah tolak H_0 jika nilai $W \geq \chi^2_{(\alpha,1)}$ atau jika $p - \text{value} < \alpha$, artinya parameter signifikan.

2. Uji Asumsi

- a. Uji Normalitas

Salah satu uji yang dapat digunakan untuk normalitas *error* adalah uji *Kolmogorov-Smirnov* dengan hipotesis dan statistik uji yaitu:

$H_0: F(x) = F_0(x)$ (*error* berdistribusi normal)

$H_1: F(x) \neq F_0(x)$ (*error* tidak berdistribusi normal)

Statistik uji *Kolmogorov-Smirnov* dirumuskan dengan persamaan (44) (Sintia dkk., 2022).

$$D_{hitung} = \max|F_n(x) - F_0(x)| \quad (44)$$

Keterangan:

$F_n(x)$: probabilitas distribusi kumulatif dari data

$F_0(x)$: probabilitas distribusi kumulatif teoritis dari distribusi normal

Kriteria pengambilan keputusan adalah menolak H_0 jika $D \geq D_{(\alpha; NT)}$ atau $p - value < \alpha$, artinya *error* tidak berdistribusi normal.

b. Uji Multikolinearitas

Nilai *Variance Inflation Factor* (VIF) dalam model regresi digunakan untuk mendeteksi adanya multikolineritas (Musyarofah dkk., 2020). VIF dirumuskan pada persamaan (45).

$$VIF = \frac{1}{(1 - R_k^2)}; k = 1, 2, 3, \dots K \quad (45)$$

Dengan R_k^2 adalah nilai koefisien determinasi pada variabel ke- k . Nilai $VIF < 10$ menunjukkan tidak ada multikolineritas dan $VIF > 10$ menunjukkan adanya multikolineritas yang kuat.

c. Uji Homoskedastisitas

Asumsi homoskedastisitas, atau asumsi kesamaan varians, menyatakan bahwa seluruh nilai dari variabel independen memiliki varians *error* yang sama atau tetap. Varians yang tetap penting karena membantu menghasilkan estimasi parameter yang konsisten dan dapat dipercaya (Rizki dkk., 2022). Salah satu metode yang digunakan untuk mendeteksi adanya pelanggaran terhadap asumsi ini adalah uji *Breusch-Pagan*, dengan hipotesis

$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$ (terjadi homoskedastisitas)

H_1 : minimal ada satu $\sigma_i^2 \neq \sigma_j^2$ dimana $i \neq j$ (tidak terjadi homoskedastisitas)

Statistik uji yang digunakan pada uji *Breusch-Pagan* (Anselin, 1988):

$$BP = \left(\frac{1}{2} \right) \mathbf{f}^t \mathbf{Z} (\mathbf{Z}^t \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{f} \sim \chi^2_{(\alpha, K)} \quad (46)$$

dengan,

$$f_i = \left(\frac{e_i^2}{\sigma^2} - 1 \right); \sigma^2 = \frac{\mathbf{e}^t \mathbf{e}}{NT}$$

Keterangan:

e_i^2 : *error* untuk observasi ke- i

$\mathbf{Z}$: matriks variabel independen berukuran $NT \times (K + 1)$

K : jumlah variabel independen

NT : jumlah observasi

$\mathbf{e}$: vektor *error*

Daerah kritis tolak H_0 jika $BP \geq \chi^2_{(\alpha, K)}$ atau $p - value < \alpha$, artinya tidak terjadi homoskedastisitas.

1.4.11 Ukuran Kebaikan Model

Metode penilaian kebaikan model yang digunakan adalah koefisien determinasi (R^2). Perhitungan R^2 menggunakan persamaan (47).

$$R^2 = 1 - \frac{\mathbf{e}^t \mathbf{e}}{(\mathbf{y} - \bar{\mathbf{y}})^t (\mathbf{y} - \bar{\mathbf{y}})} \quad (47)$$

$\bar{y}$ adalah mean dari variabel dependen dan e adalah *error* pada masing-masing spasial data panel. R^2 dapat digunakan sebagai indikator kebaikan model dan memilih model yang terbaik berdasarkan nilai R^2 yang tertinggi (Ukra dkk., 2022).

Selain R^2 , untuk menentukan ukuran kebaikan model juga dapat menggunakan *Root Mean Square Error* (RMSE). Semakin kecil nilai RMSE, maka model yang dihasilkan semakin baik. Nilai ini diperoleh dengan mengakarkan nilai *Mean Square Error* (MSE). Rumus dari RMSE dituliskan pada persamaan (48).

$$RMSE = \sqrt{\frac{1}{NT} \sum_{i=1}^n e_i^2} \quad (48)$$

e_i adalah nilai *error* pada pengamatan ke- i dan NT adalah jumlah lokasi pengamatan.

1.4.12 Pendeteksian Pencilan

Pencilan merupakan nilai ekstrem observasi yang dapat memengaruhi model estimasi SEM-RE. Pencilan dapat diidentifikasi menggunakan nilai DFFITS dengan persamaan (49).

$$(DFFITS) = t_i \left(\frac{h_{ii}}{1 - h_{ii}} \right)^{\frac{1}{2}}, i = 1, 2, 3, \dots, n \quad (49)$$

dengan,

$$h_{ii} = \mathbf{X}(\mathbf{X}^t \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^t \Sigma^{-1}$$

Keterangan:

t_i : *studentized deleted error* ke- i

h_{ii} : *leverage* data ke- i

$\mathbf{X}$: matriks kovariat berukuran $(NT \times (K + 1))$

Σ^{-1} : invers dari matriks kovarians berukuran $(NT \times NT)$

Kriteria pengujian *outlier* yaitu, jika $|(DFFITS)| \geq 2\sqrt{(K + 1)/NT}$, K merupakan jumlah variabel independen dan NT merupakan jumlah observasi.

1.4.13 Regresi Robust

Regresi *robust* adalah metode analisis regresi yang digunakan untuk mengatasi masalah distribusi *error* yang tidak normal dan *outlier* (Ryan, 1997).

1. Fungsi Objektif

Fungsi objektif dalam regresi *robust* digunakan untuk membentuk fungsi pembobot yang mampu mereduksi pengaruh *outlier*. Salah satu pendekatan yang secara eksplisit memperhitungkan struktur *leverage* adalah fungsi pembobot *Schweppe*, yang didefinisikan sebagai: $w_i = \sqrt{1 - h_{ii}}$ dan $v(x_i) = w_i$. Dengan demikian, fungsi objektif *Schweppe* memperhitungkan variasi nilai h_{ii} antar observasi. Hal ini sangat penting pada data dengan struktur panel atau spasial, di mana nilai h_{ii} seringkali bervariasi secara signifikan. Berdasarkan pembobot *Schweppe* tersebut, diperoleh bentuk fungsi objektif yang ditunjukkan pada Persamaan (50).

$$w_{i,l} = w_i \frac{\psi\left(\frac{u_i}{v(x_i)}\right)}{(u_i)} = \frac{w_i}{u_i} \psi\left(\frac{u_i}{v(x_i)}\right) = \frac{\sqrt{1-h_{ii}}}{u_i} \psi\left(\frac{u_i}{\sqrt{1-h_{ii}}}\right) \quad (50)$$

Notasi $u_i = \frac{e_i}{\hat{\sigma}}$, dan $\hat{\sigma}_{GM} = 1,48 M_l$, sedangkan M_l adalah median dari $(NT - K)$ terbesar dari $|e_i|$ dan $\psi(x) = \max\{-L, \min(L, x)\}$ merupakan fungsi pengaruh *Huber* dengan $L = 2\sqrt{(K+1)/NT}$ (Utomo & Fitrianto, 2022).

2. Robust Generalized-M Estimator

Generalized M-Estimator (GM-Estimator) adalah pengembangan *M-Estimator* yang dirancang untuk mengatasi kekurangan *M-Estimator* dalam menangani *outlier* pada variabel x_i . GM-Estimator menggunakan fungsi pembobot w_i yang bergantung pada x_i untuk membatasi pengaruh *outlier*, sehingga meningkatkan akurasi estimasi parameter (Lin dkk., 2007). Secara umum GM-Estimator didefinisikan sebagai berikut (Utomo & Fitrianto, 2022):

$$\hat{\beta}_{GM} = \arg \min \sum_{i=1}^n w_i \rho\left(\frac{e_i}{v(x_i)}\right) \quad (51)$$

Simbol $v(x_i)$ merupakan pembobot untuk variabel x_i . Penduga $\hat{\beta}$ yang diperoleh merupakan skala *invariant*, sehingga digunakan nilai $\frac{e_i}{\hat{\sigma}}$ sebagai pengganti e_i . Nilai $\hat{\sigma}$ adalah faktor skala yang perlu diestimasi. Kemudian, persamaannya menjadi:

$$\hat{\beta}_{GM} = \sum_{i=1}^n w_i \rho\left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{v(x_i) \hat{\sigma}}\right) \quad (52)$$

Tahapan-tahapan estimasi parameter dengan GM-Estimator, yaitu (Said dkk., 2024):

- Menghitung nilai *error* $e_i = y_i - \hat{y}_i$
- Menghitung $\hat{\sigma}_{GM} = 1,48 M_l$
- Menghitung nilai $u_i = \frac{e_i}{\hat{\sigma}_{GM}}$
- Menghitung pembobot $w_{i,GM}$ (*Schweppe*) dengan rumus:

$$w_{i,GM} = \frac{\sqrt{1-h_{ii}}}{u_i} \psi\left(\frac{u_i}{\sqrt{1-h_{ii}}}\right), i = 1, 2, 3, \dots, NT$$

dengan $\psi(x_i) = \max\{-L, \min(L, x)\}$ merupakan fungsi pengaruh *Huber* dengan $L = 2\sqrt{(K+1)/NT}$.

- Mengestimasi parameter $\hat{\beta}^{l+1}$ dengan metode *Generalized Least Squared* (GLS) menggunakan bobot $w_{i,GM}$ sehingga diperoleh $e_i^{(1)}$.
- Mengulang tahap (b) sampai (f) sehingga memperoleh $\hat{\beta}_{RSEM-RE}$ konvergen.

1.4.14 Robust Spatial Error Model Random Effect

Parameter model RSEM-RE diestimasi menggunakan metode estimasi parameter regresi *robust* yaitu GM-Estimator. Langkah awal dalam melakukan estimasi parameter dengan metode GM-Estimator adalah dengan meminimumkan fungsi objektif pada persamaan (53).

$$\sum_{i=1}^n w_i \rho\left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{v(x_i) \hat{\sigma}}\right) \quad (53)$$

Persamaan (53) diturunkan terhadap β_j ($j = 0, 1, 2, \dots, k$) dan disamakan dengan nol, sehingga diperoleh persamaan:

$$\frac{\partial \sum_{i=1}^n w_i \rho \left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{v(x_i) \hat{\sigma}} \right)}{\partial \beta_j} = 0 \quad (54)$$

Misalkan $u_i = \frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{v(x_i) \hat{\sigma}}$, maka persamaan (54) dapat ditulis

$$\frac{\partial \sum_{i=1}^n w_i \rho(u_i)}{\partial \beta_j} = 0$$

Selanjutnya, dengan menggunakan aturan berantai maka diperoleh:

$$\frac{\partial \sum_{i=1}^n w_i \rho(u_i)}{\partial \beta_j} = \frac{\partial \sum_{i=1}^n w_i \rho(u_i)}{\partial u_i} \times \frac{\partial u_i}{\partial \beta_j} = 0$$

$$\frac{\partial \sum_{i=1}^n w_i \rho(u_i)}{\partial u_i} \times \frac{\partial \frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{v(x_i) \hat{\sigma}}}{\partial \beta_j} = 0$$

$$\sum_{i=1}^n w_i \rho'(u_i) \times -\frac{x_{ij}}{v(x_i) \hat{\sigma}} = 0$$

$$\sum_{i=1}^n w_i \rho' \left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{v(x_i) \hat{\sigma}} \right) \times -\frac{x_{ij}}{v(x_i) \hat{\sigma}} = 0$$

Masing-masing ruas dikali dengan $(v(x_i) \hat{\sigma})$ sehingga diperoleh:

$$\sum_{i=1}^n x_{ij} w_i \rho' \left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{v(x_i) \hat{\sigma}} \right) = 0$$

$\psi = \rho'$ adalah fungsi pengaruh yang merupakan turunan dari ρ sehingga dapat ditulis menjadi:

$$\sum_{i=1}^n x_{ij} w_i \psi \left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{v(x_i) \hat{\sigma}} \right) = 0 \quad (55)$$

Setelah itu, menentukan pembobot yang digunakan yaitu pembobot *Schweppe* dengan $w_i = \sqrt{1 - h_{ii}}$ dan $v(x_i) = w_i$, maka diperoleh:

$$w_{i,GM} = w_i \frac{\psi \left(\frac{u_i}{v(x_i)} \right)}{(u_i)} = \frac{w_i}{u_i} \psi \left(\frac{u_i}{v(x_i)} \right) = \frac{\sqrt{1 - h_{ii}}}{u_i} \psi \left(\frac{u_i}{\sqrt{1 - h_{ii}}} \right)$$

Notasi $u_i = \frac{e_i}{\hat{\sigma}}$, dan $\hat{\sigma} = 1,48 M_l$, sedangkan M_l adalah median dari $(NT - K)$ terbesar dari e_i dan $\psi(x) = \max\{-L, \min(L, x)\}$ merupakan fungsi pengaruh *Huber* dengan $L = 2\sqrt{(K+1)/NT}$ (Utomo & Fitrianto, 2022). Sehingga persamaan tersebut menjadi:

$$\sum_{i=1}^n x_{ij} w_i \left(y_i - \sum_{j=0}^k x_{ij} \beta_j \right) = 0; j = 0, 1, \dots, K \quad (56)$$

Maka persamaan (56) dalam bentuk matriks dapat ditulis:

$$\begin{aligned} X^t \omega (y - X\hat{\beta}) &= 0 \\ X^t \omega y - X^t \omega X \hat{\beta} &= 0 \\ X^t \omega y &= X^t \omega X \hat{\beta} \end{aligned} \quad (57)$$

dengan ω adalah matriks diagonal berukuran $NT \times NT$ dengan elemen diagonal $w_1, w_2, w_3, \dots, w_n$ dan NT adalah jumlah observasi. Setiap ω akan diperoleh estimator $\hat{\beta}$ dari persamaan (57) adalah

$$\begin{aligned} (X^t \omega X)^{-1} X^t \omega y &= (X^t \omega X)^{-1} X^t \omega X \hat{\beta} \\ (X^t \omega X)^{-1} X^t \omega y &= \hat{\beta} \\ \hat{\beta}_{GM} &= (X^t \omega X)^{-1} X^t \omega y \end{aligned} \quad (58)$$

Estimasi parameter model RSEM-RE menggunakan GLS yang diperoleh dengan meminimumkan jumlah kuadrat *error*. SEM-RE memiliki bentuk *error*:

$$e = y - X\beta$$

dengan asumsi $e \sim N(0, \Sigma)$

$$\begin{aligned} e^t \Sigma^{-1} e &= (y - X\beta)^t \Sigma^{-1} (y - X\beta) \\ &= y^t \Sigma^{-1} y - 2\beta^t X^t \Sigma^{-1} y + \beta^t X^t \Sigma^{-1} X \beta \end{aligned} \quad (59)$$

Kemudian, untuk memperoleh $\hat{\beta}$ dilakukan dengan menghitung turunan pertama dari persamaan (59) terhadap β dan menyamakan dengan nol, sehingga diperoleh:

$$\frac{\partial e^t \Sigma^{-1} e}{\partial \beta} = \frac{\partial (y^t \Sigma^{-1} y - 2\beta^t X^t \Sigma^{-1} y + \beta^t X^t \Sigma^{-1} X \beta)}{\partial \beta} = 0$$

$$-2X^t \Sigma^{-1} y + 2X^t \Sigma^{-1} X \beta = 0$$

$$2X^t \Sigma^{-1} X \beta = 2X^t \Sigma^{-1} y$$

$$X^t \Sigma^{-1} X \beta = X^t \Sigma^{-1} y$$

$$\hat{\beta}_{GLS} = (X^t \Sigma^{-1} X)^{-1} (X^t \Sigma^{-1} y) \quad (60)$$

Karena diketahui nilai estimasi y adalah $\hat{y} = X\hat{\beta}$ maka persamaan (58) disubstitusi ke persamaan (60) sehingga diperoleh Persamaan (61).

$$\begin{aligned} \hat{\beta}_{RSEM-REM} &= (X^t \Sigma^{-1} X)^{-1} X^t \Sigma^{-1} X \hat{\beta}_{GM} \\ &= (X^t \Sigma^{-1} X)^{-1} X^t \Sigma^{-1} X (X^t \omega X)^{-1} X^t \omega y \\ &= [(X^t \Sigma^{-1} X)^{-1} X^t \Sigma^{-1} X] (X^t \omega X)^{-1} X^t \omega y \\ &= (X^t \Sigma^{-1} \omega X)^{-1} X^t \Sigma^{-1} \omega y \end{aligned} \quad (61)$$

dengan ω adalah matriks diagonal berukuran $NT \times NT$ dengan elemen diagonalnya $w_1, w_2, w_3, \dots, w_{NT}$, N adalah banyaknya jumlah observasi dan T adalah jumlah *time series*.

Fungsi pembobot yang digunakan adalah *Schweppe* seperti pada persamaan (50). u_i adalah nilai *error* i yang telah distandarisasi terhadap perkiraan standar deviasi $u_i = \frac{e_i}{\hat{\sigma}_{GM}}$. Sebelum mendapatkan nilai u_i terlebih dahulu dilakukan perhitungan skala estimasi *robust* $\hat{\sigma}$ untuk GM-Estimator sebagai berikut.

$$\hat{\sigma}_{GM} = 1,48M_l$$

M_l adalah median dari $(NT - K)$ terbesar dari $|e_i|$. Estimasi parameter RSEM-RE diselesaikan dengan IRLS, ω dihitung menggunakan $\hat{\beta}_{RSEM-RE}$ dan skala estimasi $\hat{\sigma}_{GM}$. Untuk $\omega^{(l)}$ bobot yang diberikan akan diperoleh $\hat{\beta}_{RSEM-RE} = (X^t \Sigma^{-1} \omega X)^{-1} X^t \Sigma^{-1} \omega y$ sampai $\sum_{i=1}^n |e_i^l|$ konvergen yaitu selisih $\hat{\beta}_{RSEM-RE}^{l+1}$ dengan $\hat{\beta}_{RSEM-RE}^l$ kurang dari mendekati nol, dengan l adalah banyak iterasi.

1.4.15 Indeks Pembangunan Manusia

Indeks Pembangunan Manusia (IPM) merupakan ukuran penting untuk menilai keberhasilan pembangunan suatu negara atau wilayah, terutama dalam aspek peningkatan kualitas hidup masyarakatnya. IPM diperkenalkan oleh *United Nations Development Programme* (UNDP) pada tahun 1990 dan sejak saat itu dipublikasikan secara rutin dalam laporan tahunan *Human Development Report* (HDR). Dalam laporan tersebut, UNDP menyatakan bahwa “Manusia adalah kekayaan bangsa yang sesungguhnya” sebagai dasar pentingnya fokus pada pengembangan sumber daya manusia. Oleh karena itu, UNDP mengusulkan IPM atau *Human Development Index* (HDI), sebagai salah satu cara untuk mengukur kualitas modal manusia di suatu negara (Nizhamul dkk., 2023).

HDI berfungsi sebagai indikator yang digunakan untuk menilai mutu fisik dan non-fisik masyarakat. Mutu fisik biasanya diukur melalui angka harapan hidup, sementara mutu non-fisik diwakili oleh lamanya rata-rata penduduk bersekolah dan tingkat melek huruf. IPM menjadi indikator penting yang digunakan secara global, khususnya dalam kaitannya dengan pendidikan dan kualitas sumber daya manusia. Selain itu, IPM juga berfungsi sebagai alat untuk menggambarkan pencapaian pemerintah dalam bidang pembangunan manusia. UNDP mengintegrasikan tiga dimensi utama dalam pembangunan manusia, yaitu kemampuan untuk hidup panjang dan sehat, memperoleh pengetahuan, serta mencapai standar hidup yang layak. Dengan demikian, IPM tidak hanya mengukur aspek ekonomi, tetapi juga aspek kesehatan dan pendidikan yang berperan penting dalam menentukan kualitas hidup masyarakat di suatu negara atau wilayah (Nasruddin & Azizah, 2022).

Rentang nilai IPM digunakan untuk mengklasifikasikan tingkat pembangunan manusia suatu wilayah atau negara ke dalam berbagai kategori. Secara umum, klasifikasi IPM terbagi menjadi empat kategori utama (BPS, 2024b):

1. $IPM < 60$: Rendah, artinya kualitas Pendidikan, kesehatan, dan ekonomi masih sangat terbatas
2. $60 \leq IPM < 70$: Sedang, artinya sudah ada peningkatan, tapi belum merata
3. $70 \leq IPM < 80$: Tinggi, artinya capaian pembangunan manusia relatif baik
4. ≥ 80 : Sangat tinggi, artinya kondisi pembangunan manusia sudah sangat maju

IPM di Indonesia terus menunjukkan tren peningkatan dalam beberapa tahun terakhir. Pada tahun 2024, IPM Indonesia mencapai angka 75,02, meningkat sebesar 0,63 poin atau 0,85 persen dibandingkan tahun sebelumnya yang sebesar 74,39. Kenaikan ini mencerminkan adanya perbaikan di seluruh dimensi pembentuk IPM, yaitu kesehatan, pendidikan, dan standar hidup. Harapan hidup saat lahir naik menjadi 74,15 tahun, sedangkan harapan lama sekolah mencapai 13,21 tahun, dan rata-rata lama sekolah meningkat menjadi 8,85 tahun. Selain itu, standar hidup masyarakat juga membaik yang ditandai dengan naiknya rata-rata pengeluaran riil per kapita menjadi Rp12,34 juta per tahun. Secara spasial, DKI Jakarta tetap menjadi provinsi dengan IPM tertinggi yaitu 84,15, sementara Papua Pegunungan mencatat IPM terendah sebesar 54,43. Meskipun demikian, pertumbuhan IPM tertinggi pada tahun 2024 justru dicapai oleh Papua Pegunungan, yakni sebesar 1,83 persen. Hal ini menunjukkan adanya kemajuan yang cukup signifikan di wilayah yang sebelumnya tertinggal. Secara keseluruhan, peningkatan IPM menggambarkan keberhasilan pemerintah dalam mendorong pembangunan manusia, meskipun tantangan berupa ketimpangan antar wilayah masih perlu mendapatkan perhatian lebih lanjut.