

**PEMODELAN SPASIAL PADA DATA PANEL  
DENGAN *LEAST SQUARE DUMMY VARIABLE*  
UNTUK MENGIDENTIFIKASI FAKTOR YANG  
MEMENGARUHI PENGANGGURAN DI INDONESIA**

**SKRIPSI**



**SRI INDRIANI AMIL**

**H051181012**

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
NOVEMBER 2022**

**PEMODELAN SPASIAL PADA DATA PANEL  
DENGAN *LEAST SQUARE DUMMY VARIABLE*  
UNTUK MENGIDENTIFIKASI FAKTOR YANG  
MEMENGARUHI PENGANGGURAN DI INDONESIA**

**SKRIPSI**

**Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Sains pada  
Program Studi Statistika Departemen Statistika Fakultas Matematika dan  
Ilmu Pengetahuan Alam Universitas Hasanuddin**

**SRI INDRIANI AMIL**

**H051181012**

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN**

**MAKASSAR**

**NOVEMBER 2022**

**LEMBAR PERNYATAAN KEOTENTIKAN**

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

**Pemodelan Spasial pada Data Panel dengan *Least Square Dummy Variable* untuk Mengidentifikasi Faktor yang Memengaruhi Pengangguran di Indonesia**

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun

Makassar, 16 November 2022



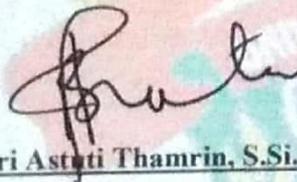
**Sri Idriani Amil**

**NIM H051181012**

**PEMODELAN SPASIAL PADA DATA PANEL DENGAN  
*LEAST SQUARE DUMMY VARIABLE* UNTUK  
MENGIDENTIFIKASI FAKTOR YANG MEMENGARUHI  
PENGANGGURAN DI INDONESIA**

Disetujui Oleh:

Pembimbing Utama



Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.  
NIP. 19740713 199903 2 001

Pembimbing Pertama



Siswanto, S.Si., M.Si.  
NIP. 19920107 201903 1 012

Ketua Program Studi



Dr. Nurtiti Sunusi, S.Si., M.Si.  
NIP. 19720117 199703 2 002

Pada 16 November 2022

**HALAMAN PENGESAHAN**

Skripsi ini diajukan oleh :

Nama : Sri Indriani Amil  
NIM : H051181012  
Program Studi : Statistika  
Judul Skripsi : *Pemodelan Spasial pada Data Panel dengan Least Square Dummy Variable* untuk Mengidentifikasi Faktor yang Memengaruhi Pengangguran di Indonesia

**Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.**

**DEWAN PENGUJI**

1. Ketua : Sri Astuti Thamrin, S.Si., M.Stat., Ph.D. (.....)
2. Sekretaris : Siswanto, S.Si., M.Si. (.....)
3. Anggota : Anisa, S.Si., M.Si. (.....)
4. Anggota : Sitti Sahrman, S.Si., M.Si. (.....)

Ditetapkan di : Makassar

Tanggal : 16 November 2022

## KATA PENGANTAR

### *Assalamu'alaikum Warahmatullahi Wabarakatuh*

Segala puji hanya milik Allah *Subhanallahu Wa Ta'ala* atas segala limpahan rahmat dan hidayah-Nya yang telah diberikan kepada penulis sampai saat ini. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa sallam*, kepada para keluarga, tabi'in, tabi'ut tabi'in, serta orang-orang sholeh yang haq hingga kadar Allah berlaku atas diri-diri mereka. *Alhamdulillahirobbil'aalamiin*, berkat rahmat dan kemudahan yang diberikan oleh Allah *Subhanallahu Wa Ta'ala*, penulis dapat menyelesaikan skripsi yang berjudul "***Pemodelan Spasial pada Data Panel dengan Least Square Dummy Variable untuk Mengidentifikasi Faktor yang Memengaruhi Pengangguran di Indonesia***" sebagai salah satu syarat memperoleh gelar sarjana pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulis tidak akan sampai pada titik ini, jikalau tanpa dukungan dan bantuan dari pihak yang selalu ada, peduli dan menyayangi penulis. Oleh karena itu, penulis haturkan rasa terima kasih yang setulus-tulusnya serta penghargaan yang setinggi-tingginya untuk orang tua penulis, Almarhum Ayahanda **Amil, S.H.** dan Ibunda tersayang, **Ros Arianty, S.ST., M.M.** yang telah memberikan dukungan penuh, pengorbanan, kesabaran hati, cinta dan kasih sayang, serta dengan ikhlas telah mengiri setiap langkah penulis dengan doa dan restunya. Kakakku **Nur Fitri Amil, S.H., Moh. Adhy Dharmawan Amil** dan adikku **Ana Masyita Putri Amil**, terima kasih telah menjadi kakak dan adik yang sangat baik, selalu ada dan selalu memberikan dorongan dukungan baik batin dan juga raga untuk penulis serta untuk keluarga besarku, terima kasih atas doa dan dukungannya selama ini.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada:

1. **Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.

3. **Ibu Dr. Nurtiti Sunusi S.Si., M.Si.**, selaku Ketua Departemen Statistika, segenap Dosen Pengajar dan Staf yang telah membekali ilmu dan kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
4. **Ibu Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.**, selaku Pembimbing Utama dan **Bapak Siswanto, S.Si., M.Si.**, selaku Pembimbing Pendamping yang dengan penuh kesabaran telah meluangkan waktu dan pemikirannya di tengah berbagai kesibukan dan prioritasnya untuk senantiasa memberikan arahan, dorongan, dan motivasi kepada penulis mulai dari awal hingga selesainya penulisan tugas akhir ini.
5. **Ibu Anisa, S.Si., M.Si.**, dan **Ibu Sitti Sahrinan, S.Si., M.Si.**, selaku Tim Penguji yang telah memberikan kritikan yang membangun dalam penyempurnaan penyusunan tugas akhir ini serta waktu yang telah diberikan kepada penulis.
6. **Ibu Sitti Sahrinan, S.Si., M.Si.**, selaku Penasehat Akademik penulis. Terima kasih atas segala bantuan, nasehat serta motivasi yang selalu diberikan kepada Penulis selama menjalani pendidikan di Departemen Statistika.
7. Bapak dan ibu dosen Fakultas MIPA yang telah mengajarkan ilmu pengetahuan sehingga dapat dijadikan dasar dalam penulisan skripsi ini.
8. **Badan Pusat Statistik Indonesia** yang telah membantu dalam penyediaan data untuk melakukan penelitian ini.
9. Keluargaku yang sangat saya sayangi, terima kasih telah memberikan banyak doa, dukungan, dan motivasi hingga penulis dapat menyelesaikan studi.
10. Sahabat-sahabatku, **Yustika, Hidayah, Ikki, Iffa, Uli**, dan **Intan**, terima kasih telah menjadi sahabat bagi penulis yang selalu memberikan motivasi, saran, doa, canda dan tawa hingga penulis dapat menyelesaikan skripsi ini.
11. *Murabbiyah* dan Teman-teman Ummu Fadhl, **Kak Indah, Marsya, Ana, Kiki, Ajeng, Risky**, dan **Itsnaini** terima kasih telah menjadi *murabbiyah* dan teman bagi penulis yang selalu memberikan doa, motivasi, saran dan mengajarkan berbagai hal yang sangat bermanfaat di dunia dan akhirat.

12. Teman-teman Statistika 2018, **Ica, Anugerah, Fitra, Hajra, Riska, Nurul, Nova, Restu, Fiska, Kholifah, Aisya, Ainun, Sonya, Ummi, Isra, Vivi, Juni, Ade**, dan semua teman Statistika 2018 yang tidak dapat penulis sebutkan satu-persatu, terima kasih telah memberikan motivasi dan berbagi suka dan duka selama perkuliahan.
13. Teman KKN Angkatan **106 Posko SULUT-SULTENG-GORONTALO** terimakasih telah membantu dan menjalin hubungan kerjasama yang baik selama menjalani proses KKN.
14. Kepada semua pihak yang tidak dapat penulis sebutkan satu-persatu, semoga segala dukungan dan partisipasi yang diberikan kepada penulis bernilai ibadah disisi Allah *Subhanallahu Wa Ta'ala*.

Penulis menyadari bahwa masih banyak kekurangan dalam skripsi ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini memberikan manfaat untuk pembaca.

*Wassalamu'alaikum Warahmatullahi Wabarakatuh*

Makassar, 16 November 2022



Sri Indriani Amil

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK  
KEPENTINGAN AKADEMIK**

---

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Sri Indriani Amil  
NIM : H051181012  
Program Studi : Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam  
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty- Free Right*)** atas tugas akhir saya yang berjudul:

**“Pemodelan Spasial pada Data Panel dengan *Least Square Dummy Variable* untuk Mengidentifikasi Faktor yang Memengaruhi Pengangguran di Indonesia”**

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada tanggal, 16 November 2022

Yang menyatakan



(Sri Indriani Amil)

## ABSTRAK

Analisis data panel adalah gabungan data *time series* dan *cross-section*. Metode estimasi data panel terdiri dari model *pooled Ordinary Least Square* (OLS) dan *Least Square Dummy Variable* (LSDV). Model *pooled OLS* menggabungkan data tanpa mempertimbangkan pengaruh perbedaan *time series* maupun *cross-section*. Sedangkan model LSDV mempertimbangkan pengaruh *cross-section*, sehingga model ini memiliki keunggulan dibandingkan model *pooled OLS*. Data panel memiliki unit *cross-section* berupa lokasi sehingga memerlukan analisis spasial untuk mengatasi keheterogenan ragam dari eror yang disebabkan oleh dependensi antar lokasi. Model spasial pada data panel dengan LSDV menggunakan *Spatial Autoregressive Model* (SAR) dengan variabel respon yang saling berkorelasi dan *Spatial Error Model* (SEM) dengan korelasi eror antar lokasi. Penelitian ini diaplikasikan pada data dengan variabel respon Tingkat Pengangguran Terbuka (TPT) dan variabel prediktor yaitu jumlah penduduk, rata-rata lama sekolah, laju PDRB, dan IPM di 34 Provinsi Indonesia dari tahun 2015-2020. Tujuannya untuk mengidentifikasi faktor yang memengaruhi TPT di Indonesia menggunakan pemodelan regresi spasial pada data panel dengan LSDV. Model SAR dengan LSDV memiliki nilai  $R^2$  (90.289%) yang lebih besar dibandingkan model SEM dengan LSDV (82.708%) dan model LSDV (87.864%). Selain itu nilai *root mean square error* pada model SAR dengan LSDV (0.58951) lebih kecil dari model SEM dengan LSDV (0.78669) dan model LSDV (0.65903). Berdasarkan model SAR dengan LSDV, faktor yang mempengaruhi TPT di Indonesia dari tahun 2015-2020 adalah laju PDRB dan IPM.

**Kata Kunci :** *Least Square Dummy Variable, Spatial Autoregressive Model dengan fixed effect, Spatial Error Model dengan fixed effect, Tingkat Pengangguran Terbuka*

## ABSTRACT

*Panel data analysis is a combination of time series and cross-section data. The panel data estimation method consists of the pooled Ordinary Least Square (OLS) and Least Square Dummy Variable (LSDV) models. The pooled OLS model combines data without considering the effect of differences in time series and cross-sections. Meanwhile, the LSDV model considers the effect of cross-section, so this model has an advantage over the pooled OLS model. Panel data has a cross-section unit in the form of a location, so it requires spatial analysis to overcome the heterogeneity of the variance of errors caused by dependencies between locations. The spatial model on panel data with LSDV uses SAR with correlated response variables and SEM with error correlation between locations. This research was applied to data with the response variable Open Unemployment Rate (TPT) and predictor variables, namely population, the average length of schooling, GRDP rate, and HDI in 34 Indonesian Provinces from 2015-2020. The aim is to identify the factors that influence TPT in Indonesia using spatial regression modeling on panel data with LSDV. The SAR model with LSDV has a greater  $R^2$  value (90.289%) than the SEM model with LSDV (82.708%) and the LSDV model (87.864%). In addition, the root means square error value in the SAR model with LSDV (0.58951) is smaller than the SEM model with LSDV (0.78669) and the LSDV model (0.65903). Based on the SAR model with LSDV, the factors that affect TPT in Indonesia from 2015-2020 are the GRDP and HDI rates.*

**Keywords :** *Least Square Dummy Variable, Open Unemployment Rate, Spatial Autoregressive model with fixed effect, Spatial Error Model with fixed effect*

## DAFTAR ISI

<b>HALAMAN SAMPUL</b> .....	i
<b>HALAMAN JUDUL</b> .....	ii
<b>LEMBAR PERNYATAAN KEOTENTIKAN</b> .....	iii
<b>LEMBAR PERSETUJUAN PEMBIMBING</b> .....	iv
<b>HALAMAN PENGESAHAN</b> .....	v
<b>KATA PENGANTAR</b> .....	vi
<b>ABSTRAK</b> .....	x
<b>ABSTRACT</b> .....	xi
<b>DAFTAR ISI</b> .....	xii
<b>DAFTAR TABEL</b> .....	xiv
<b>DAFTAR LAMPIRAN</b> .....	xv
<b>PENDAHULUAN</b> .....	1
1.1 Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Batasan Masalah .....	3
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	4
<b>TINJAUAN PUSTAKA</b> .....	5
2.1. Analisis Data Panel.....	5
2.1.1. <i>Pooled Ordinary Least Square</i> .....	5
2.1.2. <i>Least Square Dummy Variable</i> .....	6
2.1.3. Uji Chow .....	9
2.2. Analisis Spasial.....	9
2.2.1. Matriks Pembobot Spasial.....	10
2.2.2. Uji Autokorelasi Spasial.....	11
2.3. Analisis Spasial Data Panel .....	13
2.3.1. Model Panel dengan SAR .....	13
2.3.2. Model Panel dengan SEM.....	13
2.3.3. Uji <i>Lagrange Multiplier</i> .....	14
2.4. Koefisien Determinasi .....	15
2.5. <i>Root Mean Squared Error</i> .....	16

2.6. Tingkat Pengangguran Terbuka.....	16
2.7. Faktor-faktor yang Memengaruhi Pengangguran.....	17
<b>METODOLOGI PENELITIAN</b> .....	18
3.1 Data.....	18
3.2 Prosedur Penelitian .....	19
<b>HASIL DAN PEMBAHASAN</b> .....	23
4.1. Estimasi Parameter Model SAR dengan LSDV .....	23
4.2. Estimasi Parameter Model SEM dengan LSDV.....	27
4.3. Deskripsi Karakteristik Variabel Pengangguran dengan Peta Tematik... 32	
4.4. Koefisien Korelasi Variabel Prediktor Terhadap Variabel Respon.....	33
4.5. Analisis Data Panel.....	34
4.6. Uji Autokorelasi Spasial .....	37
4.7. Analisis Spasial Data Panel .....	39
4.8. Evaluasi Model .....	43
<b>PENUTUP</b> .....	45
5.1. Kesimpulan.....	45
5.2. Saran .....	45
<b>DAFTAR PUSTAKA</b> .....	46
<b>LAMPIRAN</b> .....	49

## DAFTAR TABEL

Tabel 3.1. Variabel yang digunakan dalam penelitian.....	18
Tabel 3.2. Struktur data panel yang digunakan.....	18
Tabel 4.1. Koefisien korelasi antar variabel respon dengan variabel prediktor ...	33
Tabel 4.2. Hasil estimasi model <i>pooled OLS</i> .....	34
Tabel 4.3. Hasil estimasi model LSDV .....	35
Tabel 4.4. Hasil uji Chow .....	36
Tabel 4.5. Perbandingan antar model regresi <i>pooled OLS</i> dan LSDV .....	37
Tabel 4.6. Hasil statistik uji Indeks Moran .....	39
Tabel 4.7. Hasil uji <i>lagrange multiplier</i> .....	40
Tabel 4.8. Hasil estimasi parameter model SAR dengan LSDV .....	41
Tabel 4.9. Hasil estimasi parameter model SAR (tanpa $X_1$ dan $X_2$ ) dengan LSDV .....	41
Tabel 4.10. Hasil estimasi parameter model SEM dengan LSDV .....	42
Tabel 4.11. Hasil estimasi parameter model SEM (tanpa $X_1$ dan $X_2$ ) dengan LSDV .....	43
Tabel 4.12. Evaluasi model terbaik.....	43

## DAFTAR LAMPIRAN

<b>Lampiran 1.</b> Data panel pada Tingkat Pengangguran Terbuka (TPT) dan faktor yang memengaruhi TPT .....	50
<b>Lampiran 2.</b> Intersep model LSDV .....	51
<b>Lampiran 3.</b> Data <i>longitude</i> dan <i>latitude</i> Indonesia .....	52
<b>Lampiran 4.</b> Matriks pembobot spasial .....	53
<b>Lampiran 5.</b> Intersep model SAR dengan LSDV .....	54
<b>Lampiran 6.</b> Intersep model SEM dengan LSDV .....	55

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pengangguran saat ini masih menjadi masalah sosial ekonomi yang perlu diselesaikan oleh pemerintah selain kemiskinan. Pengangguran juga merupakan fenomena makro ekonomi yang dapat memengaruhi manusia secara langsung. Pengangguran memiliki arti penduduk yang tidak bekerja tetapi sedang mencari pekerjaan. Tingkat Pengangguran Terbuka (TPT) merupakan indikator untuk melihat tingkat pengangguran (Amalia dan Sari, 2019). TPT hasil Sakernas Agustus 2020 sebesar 7.07%. Hal ini berarti dari 100 orang angkatan kerja, terdapat sekitar tujuh orang pengangguran. Pada Agustus 2020, TPT mengalami peningkatan yang cukup besar yaitu sebesar 1.84% poin dibandingkan dengan Agustus 2019. Oleh karena itu, perlu diketahui faktor-faktor yang memengaruhi TPT di Indonesia sehingga dapat diformulasikan kebijakan publik yang efektif untuk mengurangi pengangguran (Ilahi dkk., 2014).

Pengangguran dipengaruhi oleh beberapa faktor antara lain yaitu jumlah penduduk, pendidikan dalam hal ini adalah rata-rata lama sekolah, Produk Domestik Regional Bruto (PDRB), dan Indeks Pembangunan Manusia (IPM). Penelitian sebelumnya mengenai TPT oleh Arjun dkk (2019) menemukan bahwa faktor-faktor yang berpengaruh signifikan terhadap TPT di Kalimantan adalah rata-rata lama sekolah, tingkat partisipasi angkatan kerja, persentase penduduk miskin, laju pertumbuhan ekonomi, dan jumlah industri besar sedang. Hasil penelitian Agustina dkk (2022) tentang faktor TPT di Aceh menunjukkan bahwa laju pertumbuhan PDRB dan tingkat partisipasi angkatan kerja berpengaruh signifikan terhadap TPT. Menurut hasil penelitian Palindangan dan Bakar (2021) IPM memiliki pengaruh signifikan terhadap TPT di Kabupaten Mimika. Handayani (2019) menunjukkan bahwa jumlah penduduk, pendidikan, upah minimum dan PDRB secara bersama-sama berpengaruh signifikan terhadap tingkat pengangguran terbuka di Provinsi Jawa Tengah. Pengamatan terhadap perlakuan dalam bidang pendidikan maupun bidang lainnya, tidak cukup jika diamati hanya pada waktu bersamaan saja, perlu ditentukan pengamatan pada beberapa periode

waktu. Untuk itu, diperlukan data yang merupakan data gabungan antara data *cross section* dan *time series* yang disebut data panel (Astuti, 2010).

Data panel adalah gabungan antara data *time series* dan data *cross-section* (Pangestika, 2015). Metode pendekatan yang dapat dilakukan untuk mengestimasi model regresi data panel diantaranya yaitu *pooled Ordinary Least Square* (Soylu dkk., 2018), dan *Least Square Dummy Variable* (Appiah dkk., 2020). Menurut Sukendar dan Zainal dalam Astuti (2010) model *pooled Ordinary Least Square* (OLS) merupakan teknik yang paling sederhana dalam mengestimasi model regresi data panel. Data pada model *pooled OLS* dikombinasikan tanpa memerhatikan perbedaan antar lokasi dan antar waktu. Model regresi *Least Square Dummy Variable* (LSDV) dapat digunakan dalam menjelaskan heterogenitas lokasi melalui variabel *dummy* (Appiah dkk., 2020). Selain menggunakan data panel, analisis spasial diperlukan untuk mengatasi *error* heterogen yang dihasilkan akibat keterkaitan antar wilayah (autokorelasi spasial).

Analisis spasial merupakan analisis yang memasukkan pengaruh wilayah ke dalam model (Hikmah, 2017). Regresi spasial memiliki berbagai metode diantaranya yaitu *Spatial Autoregressive Model* (Rizqi, 2019) dan *Spatial Error Model* (Rahmawati dkk., 2015). *Spatial Autoregressive Model* (SAR) memiliki ketergantungan pada variabel respon dan *Spatial Error Model* (SEM) memiliki ketergantungan terhadap error antar lokasi (Hikmah, 2017). Untuk mengikutsertakan pengaruh wilayah ke dalam model, diperlukan matriks pembobot spasial. Matriks pembobot diperoleh dari informasi jarak atau ketetanggaan. Indonesia terdiri dari 34 provinsi yang berbentuk kepulauan sehingga matriks pembobot yang cocok digunakan adalah matriks invers jarak yang pada prinsipnya, bobot jarak antara suatu lokasi dengan lokasi di sekitarnya ditentukan oleh jarak antara kedua daerah tersebut. Menurut Anselin dalam Mahara dkk (2020) menyebutkan jika beberapa lokasi saling berdekatan ada kemungkinan bahwa lokasi tersebut memengaruhi satu sama lain dan sesuatu yang lebih dekat mempunyai pengaruh lebih besar dibandingkan yang agak jauh. Metode SAR dan SEM merupakan salah satu metode yang sering digunakan untuk masalah TPT.

Penelitian sebelumnya oleh Ira Rosianal Hikmah dan Yulial Hikmah (2020) tentang pemodelan spasial data panel untuk menentukan faktor-faktor yang memengaruhi kesehatan di Provinsi Papua, penelitian tersebut menghasilkan model panel SAR dengan pengaruh tetap terbukti lebih baik dari model panel biasa. Penelitian yang dilakukan oleh Reydita (2020) mengenai analisis spasial data panel pada indeks pembangunan manusia Provinsi Jawa Barat dan Provinsi Banten. Reydita (2020) menemukan bahwa model spasial data panel terbaik adalah model pengaruh tetap SAR dengan menggunakan matriks pembobot spasial *queen contiguity* tanpa transformasi.

Berdasarkan penjelasan diatas, peneliti tertarik mengangkat tema mengenai pemodelan spasial pada data panel dengan LSDV untuk mengidentifikasi faktor yang memengaruhi pengangguran di Indonesia. Analisis spasial pada data panel dengan LSDV dan matriks pembobot invers jarak yang dapat memperhitungkan efek spasial untuk menganalisis terdapat pengaruh wilayah terhadap TPT di 34 Provinsi Indonesia dari tahun 2015 sampai tahun 2020. Hal ini diharapkan dapat membantu pemerintah untuk mengatasi masalah pengangguran di Indonesia.

## **1.2. Rumusan Masalah**

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana memperoleh model regresi spasial pada data panel dengan LSDV yang terbaik dalam mengidentifikasi faktor tingkat pengangguran di 34 provinsi Indonesia tahun 2015-2020?
2. Bagaimana mengetahui faktor-faktor tingkat pengangguran yang memiliki pengaruh signifikan terhadap tingkat pengangguran di 34 Provinsi Indonesia berdasarkan model spasial pada data panel dengan LSDV yang terbaik?

## **1.3. Batasan Masalah**

Batasan masalah dalam penelitian ini adalah ruang lingkupnya terbatas pada beberapa hal yaitu:

1. Data yang digunakan adalah data tingkat pengangguran terbuka, jumlah penduduk, rata-rata lama sekolah, laju PDRB dan IPM dari tahun 2015 sampai tahun 2020.

2. Matriks pembobot spasial yang digunakan adalah invers jarak.
3. Model regresi spasial data panel yang digunakan adalah SAR dan SEM dengan LSDV.
4. Uji autokorelasi spasial menggunakan metode analisis Indeks Moran.

#### **1.4. Tujuan Penelitian**

Berdasarkan rumusan masalah, maka tujuan penelitian ini adalah sebagai berikut:

1. Memperoleh model regresi spasial pada data panel dengan LSDV yang terbaik dalam mengidentifikasi faktor tingkat pengangguran di 34 provinsi Indonesia dari tahun 2015-2020.
2. Mengetahui faktor-faktor tingkat pengangguran yang memiliki pengaruh signifikan terhadap tingkat pengangguran di 34 Provinsi Indonesia berdasarkan model regresi spasial pada data panel dengan LSDV yang terbaik.

#### **1.5. Manfaat Penelitian**

Manfaat yang diharapkan dari penelitian ini adalah mendapatkan model terbaik dari analisis spasial data panel. Kemudian dari model tersebut, dapat mengetahui faktor-faktor yang memengaruhi pengangguran di Indonesia. Sehingga dapat digunakan sebagai salah satu acuan atau pertimbangan untuk membantu pemerintah dalam penanganan masalah pengangguran di Indonesia.

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Analisis Data Panel

Menurut Wanner & Pevalin sebagaimana dikutip oleh Sembodo (2013) menyebutkan bahwa regresi panel merupakan sekumpulan teknik untuk memodelkan pengaruh variabel prediktor terhadap variabel respon pada data panel. Ada beberapa model regresi panel, salah satunya adalah model dengan *slope* konstan dan *intercept* bervariasi. Pangestika (2015), menyatakan model regresi panel yang hanya dipengaruhi oleh salah satu unit saja (unit *cross-sectional* atau unit waktu) disebut model komponen satu arah, sedangkan model regresi panel yang dipengaruhi oleh kedua unit (unit *cross-sectional* dan unit waktu) disebut model komponen dua arah. Secara umum terdapat dua pendekatan yang digunakan dalam menduga model dari data panel yaitu model tanpa pengaruh lokasi (*common effect*) dan model dengan pengaruh lokasi (*fixed effect*). Bentuk umum model regresi data panel adalah sebagai berikut:

$$Y_{it} = \alpha_{it} + \sum_{k=1}^p \beta_{ikt} X_{ikt} + \varepsilon_{it}$$

dengan  $i = 1, \dots, n$ ;  $t = 1, \dots, T$ ; dan  $k = 1, \dots, p$

$n$  : banyaknya unit lokasi

$T$  : banyaknya periode waktu

$p$  : banyaknya variabel prediktor

$Y_{it}$  : variabel respon unit lokasi ke- $i$  pada periode ke- $t$

$\alpha_{it}$  : intersep model regresi unit lokasi ke- $i$  pada periode ke- $t$

$\beta_k$  : *slope* dari variabel prediktor ke- $k$  dari unit lokasi ke- $i$  pada periode ke- $t$

$X_{ikt}$  : variabel prediktor ke- $k$  dari unit lokasi ke- $i$  pada periode ke- $t$

$\varepsilon_{it}$  : *error* untuk lokasi ke- $i$  pada periode ke- $t$

##### 2.1.1. Pooled Ordinary Least Square

Model *pooled Ordinary Least Square* (OLS) dikenal dengan estimasi *common effect* yaitu teknik yang paling sederhana dalam mengestimasi model

regresi data panel dengan cara mengombinasikan data *time series* dan *cross section*. Model ini hanya menggabungkan data tanpa memperhatikan perbedaan antar waktu dan antar lokasi. Pada pendekatan ini, digunakan metode OLS untuk mengestimasi model (Muslim, 2014).

Asumsi efek lokasi  $\alpha_i$  tidak ada, atau tidak ada heterogenitas lokasi, maka *pooled OLS* menghasilkan estimasi yang efisien dan konsisten (Appiah dkk., 2020). Persamaan model *pooled OLS* sama dengan regresi linier biasa yaitu sebagai berikut:

$$Y_{it} = \alpha + \sum_{k=1}^p \beta_k X_{ikt} + \varepsilon_{it} \quad (2.1)$$

dengan  $i = 1, \dots, n$ ;  $t = 1, \dots, T$ ; dan  $k = 1, \dots, p$

$n$  : banyaknya unit lokasi

$T$  : banyaknya periode waktu

$p$  : banyaknya variabel prediktor

$Y_{it}$  : variabel respon unit lokasi ke- $i$  pada periode ke- $t$

$\alpha$  : intersep model regresi

$\beta_k$  : *slope* dari variabel prediktor ke- $k$

$X_{ikt}$  : variabel prediktor ke- $k$  dari unit lokasi ke- $i$  pada periode ke- $t$

$\varepsilon_{it}$  : eror untuk lokasi ke- $i$  pada periode ke- $t$

### 2.1.2. *Least Square Dummy Variable*

Model *Least Square Dummy Variable* (LSDV) menggunakan variabel boneka atau *dummy* yang dikenal dengan sebutan model *Fixed Effect* (Muslim, 2014). Asumsi adanya heterogenitas lokasi, dapat membangun model LSDV dengan variabel respon dan variabel prediktor, serta *dummy* untuk memperhitungkan heterogenitas lokasi. Berikut model LSDV:

$$Y_{it} = \alpha_i + \sum_{k=1}^p \beta_k X_{ikt} + \varepsilon_{it}$$

Indeks  $i$  pada intersep ( $\alpha_i$ ) menunjukkan bahwa intersep dari masing-masing lokasi berbeda, namun intersep untuk unit *time series* tetap (konstan). Perbedaan

intersep antar lokasi dapat dijelaskan dengan menggunakan variabel *dummy*, sehingga persamaan dapat ditulis sebagai berikut (Hsiao, 2014):

$$Y_{it} = \alpha_i \sum_{j=2}^n D_{ij} + \sum_{k=1}^p \beta_k X_{ikt} + \varepsilon_{it} \quad (2.2)$$

dengan  $i = 1, \dots, n$ ;  $t = 1, \dots, T$ ; dan  $k = 1, \dots, p$

$n$  : banyaknya unit lokasi

$T$  : banyaknya periode waktu

$p$  : banyaknya variabel prediktor

$Y_{it}$  : variabel respon unit lokasi ke- $i$  pada periode ke- $t$

$\alpha_i$  : intersep model regresi untuk unit lokasi ke- $i$

$D_{ij}$  : variabel *dummy* dari lokasi ke- $j$ .  $D_{ij}$  bernilai 1 jika  $j = i$  dan bernilai 0 jika  $j \neq i$

$\beta_k$  : *slope* dari variabel prediktor ke- $k$

$X_{ikt}$  : variabel prediktor ke- $k$  dari unit lokasi ke- $i$  pada periode ke- $t$

$\varepsilon_{it}$  : eror untuk lokasi ke- $i$  pada periode ke- $t$

Estimasi parameter LSDV, pada Persamaan (2.2) dapat dituliskan dalam bentuk matriks sebagai berikut:

$$\begin{aligned} Y &= D\alpha + X\beta + \varepsilon \\ &= [D \quad X] \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \varepsilon \end{aligned}$$

Misalkan  $M = [D \quad X]$  dan  $\theta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$ , maka dapat di tulis menjadi:

$$Y = M\theta + \varepsilon$$

Selanjutnya mencari estimasi parameter  $\theta$ . Estimasi  $\theta$  menggunakan metode *least square* dengan cara meminimumkan fungsi total kuadrat eror.

$$S = \varepsilon' \varepsilon$$

$$S = (Y - M\theta)^2$$

$$S = (Y - M\theta)'(Y - M\theta)$$

$$\frac{dS}{d\theta} = \frac{d((Y - M\theta)'(Y - M\theta))}{d\theta}$$

$$\frac{dS}{d\theta} = \frac{d((Y' - \theta' M')(Y - M\theta))}{d\theta}$$

$$0 = \frac{d(Y'Y - Y'M\theta - \theta'M'Y + \theta'M'M\theta)}{d\theta}$$

$$0 = \frac{d(Y'Y - 2Y'M\theta + \theta'M'M\theta)}{d\theta}$$

$$0 = 0 - 2Y'M + 2\theta'M'M$$

$$2\theta'M'M = 2Y'M$$

$$2\theta'M'M = Y'M$$

$$2M'M\hat{\theta} = M'Y$$

Sehingga diperoleh:

$$\begin{bmatrix} D' \\ X' \end{bmatrix} \begin{bmatrix} D & X \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} D' \\ X' \end{bmatrix} Y$$

$$\begin{bmatrix} D'D & D'X \\ X'D & X'X \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} D'Y \\ X'Y \end{bmatrix}$$

$$D'D\hat{\alpha} + D'X\hat{\beta} = D'Y \quad (2.3)$$

$$X'D\hat{\alpha} + X'X\hat{\beta} = X'Y \quad (2.4)$$

Membentuk estimasi parameter  $\hat{\alpha}$  menggunakan Persamaan (2.3) yaitu sebagai berikut:

$$D'D\hat{\alpha} + D'X\hat{\beta} = D'Y$$

$$D'D\hat{\alpha} = D'Y - D'X\hat{\beta}$$

$$\hat{\alpha} = (D'D)^{-1}D'Y - (D'D)^{-1}D'X\hat{\beta} \quad \text{k. (2.5)}$$

Bentuk estimasi parameter  $\hat{\beta}$  diperoleh dengan mensubstitusikan Persamaan (2.5) ke dalam Persamaan (2.4).

$$X'D\hat{\alpha} + X'X\hat{\beta} = X'Y$$

$$X'D[(D'D)^{-1}D'Y - (D'D)^{-1}D'X\hat{\beta}] + X'X\hat{\beta} = X'Y$$

$$X'D(D'D)^{-1}D'Y - X'D(D'D)^{-1}D'X\hat{\beta} + X'X\hat{\beta} = X'Y$$

$$X'D(D'D)^{-1}D'Y + X'[I - D(D'D)^{-1}D']X\hat{\beta} = X'Y$$

Misalkan  $D(D'D)^{-1}D' = P$ , maka diperoleh:

$$X'PY + X'(I - P)X\hat{\beta} = X'Y$$

$$X'(I - P)X\hat{\beta} = X'Y - X'PY$$

$$X'(I - P)X\hat{\beta} = X'(I - P)Y$$

$$\hat{\beta} = [X'(I - P)X]^{-1}X'(I - P)Y$$

### 2.1.3. Uji Chow

Uji Chow merupakan uji signifikansi untuk menentukan menggunakan model *pooled* OLS atau model LSDV. Hipotesis dalam uji chow yaitu:

$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_{n-1} = 0$  (tidak terdapat pengaruh lokasi terhadap model atau model mengikuti model *pooled* OLS)

$H_1$  : minimal ada satu  $\alpha_i \neq 0$  (terdapat satu atau lebih pengaruh lokasi terhadap model atau model mengikuti model LSDV).

Statistik uji yang digunakan adalah:

$$F_0 = \frac{(RRSS - URSS)/(n - 1)}{URSS/(nT - n - p)}$$

dengan

$$RRSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$URSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$y_i$  : hasil observasi data ke- $i$

$\hat{y}_i$  : hasil prediksi data ke- $i$

$n$  : banyaknya unit lokasi

$T$  : banyaknya periode waktu

$p$  : banyaknya variabel prediktor

*Restricted residual sums of square* (RRSS) diperoleh dari hasil estimasi model *pooled* OLS dan *unrestricted residual sums of square* (URSS) diperoleh dari hasil estimasi model LSDV. Keputusan tolak  $H_0$  jika  $F_0 > F_{n-1, n(T-1)-p}$  atau jika  $p\text{-value} < \alpha$  (Baltagi, 2005).

## 2.2. Analisis Spasial

Data spasial berisi informasi atau letak geografis suatu wilayah. Lokasi pada data spasial harus diukur untuk menentukan efek spasial. Informasi lokasi dapat ditentukan dari dua sumber, yaitu hubungan ketetanggaan dan jarak. Hubungan ketetanggaan mencerminkan lokasi relatif dari unit spasial atau lokasi lain dalam ruang tertentu. Hubungan ketetanggaan unit spasial biasanya dibentuk berdasarkan peta (Thamrin dkk., 2020).

### 2.2.1. Matriks Pembobot Spasial

Pembobot spasial merupakan langkah penting dalam mengukur dan memodelkan ketergantungan spasial dalam data area (Bivand dkk., 2013). Matriks pembobotan spasial ( $\mathbf{W}$ ) pada dasarnya adalah matriks yang menggambarkan hubungan dalam suatu wilayah dan diperoleh dari informasi jarak atau ketetanggaan (Thamrin dkk., 2020). Hubungan ketetanggaan secara formal dinyatakan dalam matriks  $n \times n$  dengan elemen ( $W_{ij}$ ) mewakili ukuran hubungan antara lokasi  $i$  dan  $j$ . Spesifikasi matriks pembobot spasial dimulai dengan mengidentifikasi struktur ketetanggaan setiap sel (Dormann dkk., 2007). Misalkan  $\mathbf{W} = \{W_{ij}\}$  dengan adalah matriks ketetanggaan dengan bernilai 1 jika antara dua unit spasial saling bertetangga atau bersebelahan dan bernilai 0 jika antara dua unit spasial saling tidak bertetangga (Hikmah, 2017).

Beberapa pendekatan yang dapat dilakukan untuk menampilkan hubungan spasial antar lokasi, diantaranya adalah konsep persinggungan (*contiguity*) dan konsep jarak (Hikmah, 2017). Konsep persinggungan atau *contiguity* ada tiga jenis yakni *Rook Contiguity* (persinggungan sisi), *Bishop Contiguity* (persinggungan sudut) serta *Queen Contiguity* (persinggungan sisi dan sudut) (Ningtias, 2017). Sedangkan pada konsep jarak, elemen-elemen dari matriks pembobot spasial direpresentasikan dalam bentuk fungsi jarak. Pada prinsipnya, bobot jarak antara suatu lokasi dengan lokasi di sekitarnya ditentukan oleh jarak antara kedua daerah tersebut. Salah satu metode yang sering digunakan adalah metode invers jarak (*distance inverse*). Semakin pendek jarak antara lokasi maka bobot yang diberikan akan semakin besar (Hikmah, 2017). Secara matematis perhitungan matriks pembobot spasial invers jarak yaitu:

$$\mathbf{W} = [W_{ij}] = \begin{cases} 0, & i = j \\ \frac{1}{d_{ij}}, & i \neq j \\ \frac{1}{\sum_{j=1}^n \frac{1}{d_{ij}}}, & i \neq j \end{cases}$$

dengan

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

$d_{ij}$  : jarak lokasi  $i$  dan  $j$

$u_i$  : *longitude* lokasi ke- $i$

$u_j$  : longitude lokasi ke- $j$

$v_i$  : latitude lokasi ke- $i$

$v_j$  : latitude lokasi ke- $j$

$n$  : banyaknya unit lokasi

(Babak dan Deutsch, 2008).

### 2.2.2. Uji Autokorelasi Spasial

Autokorelasi spasial adalah taksiran dari korelasi antar nilai amatan yang berkaitan dengan lokasi spasial pada variabel yang sama. Autokorelasi spasial positif menunjukkan adanya kemiripan nilai dari lokasi-lokasi yang berdekatan dan cenderung berkelompok. Sedangkan autokorelasi spasial yang negatif menunjukkan bahwa lokasi-lokasi yang berdekatan mempunyai nilai yang berbeda dan cenderung menyebar.

Pengukuran autokorelasi spasial untuk data spasial dapat dihitung menggunakan metode *Moran's Index* (Indeks Moran), *Geary's C* dan *Tango's excess*. Indeks Moran merupakan metode yang paling banyak digunakan untuk menghitung autokorelasi spasial secara global. Metode ini dapat digunakan untuk mendeteksi permulaan dari keacakan spasial. Keacakan spasial ini dapat mengindikasikan adanya pola-pola yang mengelompok atau membentuk tren terhadap ruang. Perhitungan autokorelasi spasial dengan metode Indeks Moran dapat dilakukan dengan cara, yaitu:

Indeks Moran dengan matriks pembobot spasial terstandarisasi  $W^*$

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n W_{ij}^* (x_i - \bar{x})(x_j - \bar{x})}{S_0 \sum_{i=1}^n (x_i - \bar{x})^2}$$

dengan

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

$W_{ij}^*$  : elemen pada pembobot terstandarisasi antara daerah  $i$  dan  $j$

$I$  : Indeks Moran

$n$  : banyaknya lokasi kejadian

$x_i$  : nilai pada lokasi  $i$

$x_j$  : nilai pada lokasi  $j$

$\bar{x}$  : rata-rata dari jumlah variabel atau nilai

Nilai Indeks Moran tidak menjamin ketepatan pengukuran jika matriks pembobot yang digunakan adalah pembobot tak terstandarisasi. Untuk mengidentifikasi adanya autokorelasi spasial atau tidak, dilakukan uji signifikansi Indeks Moran (Wuryandari dkk., 2014). Uji hipotesis untuk Indeks Moran adalah sebagai berikut:

- i. Hipotesis  $H_0: I = 0$  (tidak terdapat autokorelasi spasial)
  - $H_1: I > 0$  (terdapat autokorelasi spasial positif) atau
  - $H_1: I < 0$  (terdapat autokorelasi spasial negatif)
- ii. Tingkat signifikansi ( $\alpha$ ) menunjukkan probabilitas atau peluang kesalahan yang ditetapkan peneliti dalam mengambil keputusan untuk menolak atau mendukung hipotesis nol, atau dapat diartikan juga sebagai tingkat kesalahan atau tingkat kekeliruan yang ditolerir oleh peneliti, yang diakibatkan oleh kemungkinan adanya kesalahan dalam pengambilan sampel (*sampling error*).
- iii. Statistik Uji:

$$Z(I) = \frac{I - E(I)}{\sqrt{Var(I)}}$$

dengan

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

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

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

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

$$S_2 = \sum_{i=1}^n (\sum_{j=1}^n W_{ij}^* + \sum_{j=1}^n W_{ji}^*)^2$$

$W_{ij}^*$  : elemen pada pembobot terstandarisasi antara daerah  $i$  dan  $j$

$I$  : Indeks Moran

$n$  : banyaknya lokasi kejadian

- iv. Kriteria Uji:

Tolak  $H_0$  pada taraf signifikansi  $\alpha$  jika  $Z(I) > Z_{1-\alpha}$  atau  $Z(I) < -Z_{1-\alpha}$  atau  $p\text{-value} < \alpha$  (Hikmah, 2017).

### 2.3. Analisis Spasial Data Panel

Analisis data panel spasial merupakan gabungan data *cross-section* dan *time series* dengan memperhitungkan pengaruh spasial. Model regresi linier pada data panel yang terdapat interaksi diantara unit-unit spasialnya. Model ini akan memiliki variabel spasial lag pada variabel respon yang disebut *Spatial Autoregressive Model* (SAR) atau variabel spasial proses pada eror yang biasanya disebut *Spatial Error Model* (SEM). Estimasi parameter-parameter ini menggunakan *Maximum Likelihood Estimation* (MLE) (Hikmah, 2017).

#### 2.3.1. Model Panel dengan SAR

Pada model ini, variabel respon bergantung pada pengamatan variabel respon pada unit-unit tetangga. Model lag spasial (autoregresif spasial) dinyatakan sebagai berikut (Hikmah, 2017):

$$Y_{it} = \rho \sum_{j=1}^n W_{ij} Y_{jt} + \sum_{k=1}^p \beta_k X_{ikt} + \alpha_i + \varepsilon_{it} \quad (2.6)$$

dengan

$\rho$  : koefisien autoregresif spasial

$W_{ij}$  : elemen pada matriks pembobot baris ke- $i$  dan kolom ke- $j$

$Y_{it}$  : variabel respon lokasi ke- $i$  pada waktu ke- $t$

$Y_{jt}$  : variabel respon lokasi tetangaan ke- $j$  pada waktu ke- $t$

$\beta_k$  : *slope* dari variabel prediktor ke- $k$

$X_{ikt}$  : variabel prediktor ke- $k$  dari unit lokasi ke- $i$  pada periode ke- $t$

$\alpha_i$  : intersep LSDV lokasi ke- $i$

$\varepsilon_{it}$  : eror lokasi pada amatan ke- $i$  pada periode ke- $t$

#### 2.3.2. Model Panel dengan SEM

Pada model ini, fokusnya terdapat pada bentuk eror-nya. Model eror spasial dinyatakan sebagai berikut:

$$Y_{it} = \sum_{k=1}^p \beta_k X_{ikt} + \alpha_i + \phi_{it} \quad (2.7)$$

dengan

$$\phi_{it} = \lambda \sum_{j=1}^N W_{ij} \phi_{jt} + \varepsilon_{it}$$

$\phi_{it}$  : bentuk eror dari autokorelasi spasial lokasi ke- $i$  dan periode ke- $t$

$\lambda$  : koefisien autokorelasi spasial

$W_{ij}$  : elemen pada matriks pembobot baris ke- $i$  dan kolom ke- $j$

$Y_{it}$  : variabel respon lokasi ke- $i$  pada waktu ke- $t$

$\beta_k$  : *slope* dari variabel prediktor ke- $k$

$X_{ikt}$  : variabel prediktor ke- $k$  dari unit lokasi ke- $i$  pada periode ke- $t$

$\alpha_i$  : intersep LSDV lokasi ke- $i$

$\varepsilon_{it}$  : eror lokasi pada amatan ke- $i$  pada periode ke- $t$

(Hikmah, 2017).

### 2.3.3. Uji Lagrange Multiplier

Uji *Lagrange Multiplier* (LM) digunakan untuk menguji efek interaksi spasial dalam data *cross-sectional* (Elhorst, 2014). Pengujian ini juga digunakan sebagai dasar untuk memilih model regresi spasial yang sesuai. LM eror signifikan maka model yang sesuai adalah SEM dan apabila LM lag signifikan maka model yang sesuai adalah SAR. Apabila LM eror maupun lag keduanya signifikan maka model yang sesuai adalah SARMA (*Spatial Autoregressive Moving Average*). Apabila LM eror maupun LM lag keduanya tidak signifikan maka model yang sesuai adalah OLS (Ningtias, 2017).

Hipotesis yang digunakan pada LM lag adalah:

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

$H_1 : \rho \neq 0$  (ada dependensi spasial lag)

Statistik uji yang digunakan pada LM lag adalah:

$$LM_{\rho} = \frac{\left[ \frac{\mathbf{e}'_{1 \times nT} (\mathbf{I}_{T \times T} \otimes \mathbf{W}_{n \times n}) \mathbf{Y}_{nT \times 1}}{\hat{\sigma}^2} \right]^2}{J}$$

dengan

$$J = \frac{1}{\hat{\sigma}^2} [(\mathbf{a}'_1 \mathbf{a}_2 \mathbf{a}_1) + TT_w \hat{\sigma}^2]$$

$$\mathbf{a}_1 = (\mathbf{I}_{T \times T} \otimes \mathbf{W}_{n \times n}) \mathbf{X}_{nT \times k} \hat{\boldsymbol{\beta}}_{k \times 1}$$

$$\mathbf{a}_2 = (\mathbf{I}_{nT \times nT} - \mathbf{X}_{nT \times k} (\mathbf{X}'_{k \times nT} \mathbf{X}_{nT \times k})^{-1} \mathbf{X}'_{k \times nT})$$

$$T_w = tr((\mathbf{W}'_{n \times n} + \mathbf{W}_{n \times n}) \mathbf{W}_{n \times n})$$

Pengambilan keputusan pada LM lag akan tolak  $H_0$  apabila nilai  $LM_{\rho} > \chi^2_{(\alpha, 1)}$ . Sedangkan uji LM Eror menggunakan uji hipotesis sebagai berikut ini:

Hipotesis LM eror:

$H_0 : \delta = 0$  (tidak ada dependensi spasial eror)

$H_1 : \delta \neq 0$  (ada dependensi spasial eror)

Statistik uji yang digunakan pada LM eror adalah:

$$LM_{\delta} = \frac{\left[ \frac{e'_{1 \times nT} (I_{T \times T} \otimes W_{n \times n}) e_{nT \times 1}}{\hat{\sigma}^2} \right]^2}{T \times T_w}$$

dengan

$$T_w = tr((W'_{n \times n} + W_{n \times n})W_{n \times n})$$

Pengambilan keputusan pada LM eror akan tolak  $H_0$  apabila nilai  $LM_{\delta} > \chi^2_{(\alpha,1)}$  (Elhorst, 2014).

#### 2.4. Koefisien Determinasi

Koefisien determinasi yang dinotasikan dengan  $R^2$ , merupakan suatu ukuran yang penting dalam regresi, karena dapat menginformasikan baik atau tidaknya model regresi yang terestimasi. Angka tersebut dapat mengukur seberapa dekat garis regresi yang terestimasi dengan data sesungguhnya. Koefisien determinasi dinotasikan dengan:

$$R^2 = 1 - \frac{SSE}{SST}$$

dengan

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2$$

$y_i$  : hasil observasi data ke- $i$

$\hat{y}_i$  : hasil prediksi data ke- $i$

$\bar{y}$  : hasil rata-rata observasi data

Nilai koefisien determinasi  $0 \leq R^2 \leq 1$ . Nilai ini mencerminkan seberapa besar variasi dari variabel respon dapat diterangkan oleh variabel prediktor. Bila nilai  $R^2 = 0$ , artinya variasi dari variabel respon tidak dapat diterangkan oleh variabel prediktor sama sekali. Sementara bila  $R^2 = 1$ , artinya variasi variabel respon secara keseluruhan dapat diterangkan oleh variabel prediktor (Pangestika, 2015).

### 2.5. *Root Mean Squared Error*

*Root Mean Squared Error* (RMSE) banyak digunakan dalam evaluasi model untuk melakukan pemilihan model terbaik. RMSE dapat dicari dengan menggunakan rumus berikut:

$$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}}$$

dengan

$y_i$  : hasil observasi data ke- $i$

$\hat{y}_i$  : hasil prediksi data ke- $i$

$n$  : banyaknya unit lokasi

(Chai dan Draxler, 2014).

### 2.6. **Tingkat Pengangguran Terbuka**

Pengangguran adalah angkatan kerja yang bersedia bekerja dari upah yang berlaku saat ini dan tidak dapat memperoleh pekerjaan (Soylu dkk., 2018). Pengangguran merupakan sebuah fenomena ekonomi, seseorang yang diklasifikasikan sebagai bagian dari angkatan kerja (16 tahun ke atas) secara aktif mencari pekerjaan tetapi tidak dapat menemukannya di pasar tenaga kerja. Pengangguran bisa bermacam-macam bentuknya. Bisa jadi karena perubahan struktur ekonomi. Kemajuan ekonomi karena perubahan teknologi dapat membuat lokasi menganggur. Pengangguran juga dapat disebabkan oleh perpindahan lokasi dari satu pekerjaan ke pekerjaan lain yang secara ekonomi didefinisikan sebagai pengangguran friksional (Appiah dkk., 2020).

Menurut Badan Pusat Statistika (BPS), terdapat 4 macam pengertian mengenai pengangguran terbuka. Yang pertama, pengangguran terbuka terjadi pada seseorang yang pada saat ini tidak mempunyai pekerjaan namun orang tersebut berusaha untuk mencari pekerjaan. Pengertian yang kedua, pengangguran terbuka terjadi kepada seseorang yang tidak mempunyai pekerjaan tetapi orang tersebut sedang berusaha mempersiapkan sebuah usaha. Definisi pengangguran terbuka yang ketiga terjadi kepada seseorang yang pada saat ini tidak memiliki sebuah pekerjaan namun orang tersebut tidak sedang mencari pekerjaan serta tidak berusaha mempersiapkan usaha dikarenakan orang tersebut yakin bahwa dia tidak bisa mendapatkan sebuah pekerjaan. Pengertian mengenai pengangguran

terbuka yang keempat yakni pengangguran yang terjadi kepada seseorang pada saat ini sudah memiliki pekerjaan akan tetapi orang tersebut belum memulai untuk bekerja. Rumus yang digunakan untuk bisa mengetahui besarnya persentase tingkat pengangguran kerja adalah sebagai berikut ini.

$$\text{Tingkat Pengangguran Terbuka} = \frac{\text{jumlah pengangguran}}{\text{jumlah angkatan kerja}} \times 100\%$$

Semakin pesatnya laju pertumbuhan penduduk mengakibatkan persaingan untuk mendapatkan pekerjaan semakin ketat. Ini menyebabkan tidak semua angkatan kerja dapat diserap oleh lapangan kerja. Sebagai akibatnya yakni terjadi pengangguran yang tidak terelakan lagi. Pengangguran terjadi karena jumlah angkatan kerja yang tidak sebanding dengan kesempatan kerja. Semakin tinggi laju pertumbuhan penduduk, maka menyebabkan angka pengangguran semakin meningkat (Ningtias, 2017).

## 2.7. Faktor-faktor yang Memengaruhi Pengangguran

Pengangguran dipengaruhi oleh beberapa faktor antara lain yaitu jumlah penduduk, pendidikan dalam hal ini adalah rata-rata lama sekolah, PDRB dan Indeks Pembangunan Manusia (IPM). Jumlah penduduk Indonesia yang besar memunculkan angkatan kerja baru tiap tahunnya yang berdampak pada tingkat pengangguran (Handayani, 2019). Pendidikan berpengaruh terhadap pekerjaan yang didapatkan. Semakin lama jangka waktu yang masyarakat habiskan untuk menempuh pendidikan maka semakin tinggi pula kemampuan dan kesempatan masyarakat tersebut untuk bekerja sehingga dapat menurunkan TPT (Ramiayu, 2015). Faktor PDRB yang meningkat diharapkan dapat menyerap tenaga kerja di suatu wilayah, karena dengan kenaikan PDRB dapat meningkatkan kapasitas produksi jasa (Muslim, 2014). Faktor IPM menurut Hukum Okun menyatakan peningkatan IPM mendorong peningkatan pertumbuhan ekonomi. Pertumbuhan ekonomi diharapkan dapat meningkatkan peluang dan permintaan tenaga kerja yang dapat menyerap masyarakat dan membantu mengurangi angka pengangguran (Palindangan dan Bakar, 2021).