SKRIPSI

ESTIMASI PARAMETER MODEL *MIXED GEOGRAPHICALLY WEIGHTED REGRESSION* MENGGUNAKAN METODE *GENERALIZED M-ESTIMATOR*

Disusun dan diajukan oleh

FAKHRIYYAH DJ. JUNUS

H051171301

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2021

ESTIMASI PARAMETER MODEL *MIXED GEOGRAPHICALLY WEIGHTED REGRESSION* MENGGUNAKAN METODE *GENERALIZED M-ESTIMATOR*

SKRIPSI

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

FAKHRIYYAH DJ. JUNUS

H051171301

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2021

LEMBAR PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama	: Fakhriyyah Dj. Junus
NIM	: H051171301
Program Studi	: Statistika
Jenjang	: Sarjana (S1)

Menyatakan dengan ini bahwa karya tulis saya yang berjudul

ESTIMASI PARAMETER MODEL *MIXED GEOGRAPHICALLY WEIGHTED REGRESSION* MENGGUNAKAN METODE *GENERALIZED M-ESTIMATOR*

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

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 03 November 2021

METERAI TEMPEL
10000
SEPULUH RIBU RUPIAH
5709EAJX485197159

FAKHRIYYAH DJ. JUNUS

NIM. H051171301

ESTIMASI PARAMETER MODEL *MIXED GEOGRAPHICALLY WEIGHTED REGRESSION* MENGGUNAKAN METODE *GENERALIZED M-ESTIMATOR*

Disetujui Oleh:

Pembimbing Utama,

Andi Kresna Jaya, S.Si., M.Si.
NIP. 19731228 200003 1001

Pembimbing Pertama,

Dr. La Podje Talangko, M.Si.
NIP. 19551219 187011 001

Ketua Departemen Statistika

Dr. Nurtiti Sunusi, S.Si., M.Si.
NIP. 19720117 199703 2002

Pada Tanggal: 03 November 2021

LEMBAR PENGESAHAN

ESTIMASI PARAMETER MODEL *MIXED GEOGRAPHICALLY WEIGHTED REGRESSION* MENGGUNAKAN METODE *GENERALIZED M-ESTIMATOR*

Disusun dan diajukan oleh

FAKHRIYYAH DJ. JUNUS

H051171301

Telah dipertahankan dihadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

pada tanggal 03 November 2021

dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama,	**Pembimbing Pertama,**
Andi Kresna Jaya, S.Si., M.Si.	**Dr. La Podje Talangko, M.Si.**
NIP. 19731228 200003 1001	**NIP. 19551219 187011 001**

Ketua Departemen Statistika

Dr. Nurtiti Sunusi, S.Si., M.Si.

NIP. 19720117 199703 2002

KATA PENGANTAR

Assalamu'alaikum Warohmatullahi Wabarokatuh.

Alhamdulillahi robbil'alamin, Puji syukur kepada **Allah** ***Subhanahu Wa Ta'ala*** atas segala limpahan rahmat, nikmat, dan hidayah-Nya yang diberikan kepada penulis sehingga dapat menyelesaikan penulisan skripsi dengan judul **"Estimasi Parameter Model *Mixed Geographically Weighted Regression* Menggunakan Metode *Generalized M-estimator*"** sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Salam dan sholawat *InsyaAllah* senantiasa tercurah kepada **Nabi Muhammad, Rasulullah** ***Shallallahu'alaihi Wasallam***, yang telah memberikan petunjuk cinta dan kebenaran dalam kehidupan.

Dalam penyelesaian skripsi ini, penulis telah melewati perjuangan panjang dan pengorbanan yang tidak sedikit dengan segala keterbatasan kemampuan dan pengetahuan. Namun berkat rahmat dan izin-Nya serta dukungan dari berbagai pihak yang turut membantu baik moril maupun material sehingga akhirnya tugas akhir ini dapat terselesaikan. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang setinggi-tingginya dan penghargaan yang tak terhingga kepada Ayahanda **H. Djarham Jamin Junus, S.Pd.** dan Ibunda tercinta **Dra. Hi. Martin Umar, S.H.** yang telah menjadi inspirasi, membesarkan dan mendidik penulis dengan penuh kesabaran dan dengan limpahan cinta, kasih sayang, dan doa kepada penulis yang tak pernah habis, untuk kakak tersayang **Nur Azizah Dj. Junus, S.Pd.** dan adik tersayang **Afnizar Dj. Junus** serta keluarga besar penulis yang selalu mendoakan, memberikan dukungan dan motivasi, serta menjadi penyemangat untuk segera menyelesaikan masa studi penulis.

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

1. **Ibu Prof. Dr. Dwia Aries Tina Pulubuhu, MA,** 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 yang telah meluangkan waktunya untuk memberikan masukan dan motivasi. Segenap dosen pengajar dan staf **Departemen Statistika** yang telah membekali ilmu dan kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
4. **Bapak Andi Kresna Jaya, S.Si., M.Si.** selaku Pembimbing Utama penulis yang telah ikhlas meluangkan waktu dan pemikirannya untuk memberikan arahan, pengetahuan, motivasi dan bimbingan ditengah kesibukan beliau serta menjadi tempat berkeluh kesah untuk penulis.
5. **Bapak Dr. La Podje Talangko, M.Si. Rohimahullah** selaku Pembimbing pertama penulis yang telah ikhlas meluangkan waktunya ditengah kesibukan untuk memberikan arahan bagi penulis.
6. **Ibu Dr. Dr. Georgina Maria Tinungki, M.Si.** selaku Dosen Penguji sekaligus Penasehat Akademik penulis dan **Bapak Dr. Nirwan Ilyas, M.Si.** selaku Dosen Penguji, terima kasih telah ikhlas meluangkan waktunya ditengah kesibukan untuk memberikan arahan berupa saran dan kritikan yang membangun dalam penyempurnaan penyusunan tugas akhir ini.
7. Spesial untuk sahabat tercinta penulis, **Nur Aprilia Dzulhijjah, Riska Rasyid, Miftahul Jannah, Nurul Wahyuni, Sakinah Oktoni, Munadiah Apriliani, Nurul Annisa, Nurhidayatullah, Fitri,** dan **Risnawati Azali** yang telah menjadi sahabat terbaik sejak awal perkuliahan dan senantiasa mendengarkan curhatan, memberikan dorongan, semangat, nasehat dan motivasi dalam setiap keadaan sehingga penulis bisa mendapatkan lebih banyak pelajaran hidup dan *alhamdulillah* tetap bahagia walau hidup dalam perantauan.
8. Sahabat terbaik sejak di bangku SMA, **Ahdiyaty Rahmi Suaib, Syifa Allifa Utrujjah,** dan **Balqis Miftahussa'adah** yang sampai saat ini masih setia mendengarkan keluh kesah penulis dan senantiasa membantu berbagai hal yang menjadi hambatan penulis dalam menjalani proses pendidikan dibangku kuliah.

9. Teman-teman **Statistika 2017**, terima kasih atas kebersamaan, suka, dan duka selama menjalani pendidikan di Departemen Statistika. Penulis senang bisa mengenal kalian semua, terkhusus **Eva Riyantie, Annisa Haura Salsa Fatih Yusuf, Putri Henri, Nurkamalia, Siti Ihza Arsella Kasim** dan **Aqilah Salsabila Rahman,** terima kasih atas setiap cerita cita dan duka, motivasi, nasehat pembelajaran hidup sehingga memberi kesan perkuliahan yang berwarna.
10. Teman-teman **DISKRIT 2017**, terima kasih untuk cerita sekaligus kenangan selama proses yang telah dilalui. Banyak pengalaman dan pelajaran berharga yang telah penulis dapatkan dari teman-teman selama berporoses.
11. Kakak-kakak, teman-teman, dan adik-adik **Keluarga Mahasiswa FMIPA Unhas** terkhusus anggota keluarga **Himatika FMIPA Unhas** dan **Himastat FMIPA Unhas**, terima kasih atas ilmu yang mungkin tidak bisa didapatkan penulis di proses perkuliahan dan telah menjadi keluarga selama penulis kuliah di Universitas Hasanuudin.
12. Kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih setinggi-tingginya untuk segala dukungan dan partisipasi yang diberikan kepada penulis semoga bernilai ibadah di sisi Allah *Subhanahu Wa Ta'ala.*

Penulis berharap skripsi ini dapat memberikan tambahan pengetahuan baru bagi para pembelajar statistika. Penulis menyadari bahwa dalam penulisan tugas akhir ini masih banyak terdapat kekurangan. Oleh karena itu, dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga dapat bermanfaat bagi pihak-pihak yang berkepentingan. *Aamiin Yaa Rabbal Alamin.*

Makassar, 03 November 2021

Fakhriyyah Dj. Junus

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama	: Fakhriyyah Dj. Junus
NIM	: H051171301
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:

"Estimasi Parameter Model *Mixed Geographically Weighted Regression* Menggunakan Metode *Generalized M-estimator*"

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 03 November 2021

Yang menyatakan

Fakhriyyah Dj. Junus

ABSTRAK

Mixed Geographically Weighted Regression (MGWR) merupakan suatu model gabungan regresi global dan model GWR, sehingga parameter model yang dihasilkan bersifat lokal dan sebagian lainnya bersifat global. Dalam menganalisis data dengan model MGWR, terkadang ditemukan adanya pencilan yang dapat berpengaruh besar terhadap koefisien regresi sehingga membuat estimasi parameter menjadi bias. Oleh karena itu, untuk mengatasi hal tersebut diperlukan sebuah metode yang lebih *robust* (kekar) terhadap keberadaan pencilan dalam mengestimasi parameter model MGWR. Pada penelitian ini, estimasi parameter dilakukan menggunakan metode *GM-estimator* dengan fungsi pembobot *Tukey Bisquare*. Selanjutnya, model MGWR diterapkan pada data persentase penduduk miskin di Provinsi Sulawesi Selatan tahun 2019 dengan faktor-faktor yang mempengaruhinya diantaranya indeks pembangunan manusia (x_1), persentase penduduk (x_2), tingkat pengangguran terbuka (x_3), persentase pengeluaran per kapita untuk makanan (x_4), persentase rumah tangga dengan sumber air minum kemasan, air minum isi ulang, dan leding (x_5) serta persentase rumah tangga dengan luas lantai rumah < 50 m^2 (x_6). Hasil yang diperoleh adalah variabel yang berpengaruh signifikan secara lokal terhadap persentase penduduk miskin di Provinsi Sulawesi Selatan tahun 2019 adalah indeks pembangunan manusia (x_1) dan persentase rumah tangga dengan luas lantai rumah < 50 m^2 (x_6). Adapun variabel yang berpengaruh signifikan secara global terhadap persentase penduduk miskin di Provinsi Sulawesi Selatan tahun 2019 adalah tingkat pengangguran terbuka (x_3) dan persentase pengeluaran per kapita untuk makanan (x_4).

Kata Kunci: *GM-estimator*, MGWR, Pencilan, Persentase Penduduk miskin, *Tukey Bisquare*.

ABSTRACT

The Mixed Geographically Weighted Regression (MGWR) model is a combination of a global regression and a GWR model, with some model parameters being local and others being global. When using the MGWR model to analyze data, it is sometimes discovered that there are outliers that have a large impact on the regression coefficient, causing the parameter estimates to be biased. To solve this, a more robust method for estimating the parameters of the MGWR model against outliers is required. The GM-estimator method with the Tukey Bisquare weighting function was used to estimate parameters in this research. Furthermore, the MGWR model is used to analyze data on the percentage of poverty population of South Sulawesi Province in 2019, considering factors such as the human development index (x_1), population percentage (x_2), open unemployment rate (x_3), percentage of food expenditure per capita (x_4), percentage of households with bottled drinking water sources, refilled drinking water, and plumbing (x_5), as well as the percentage of households with a floor area of more than 50 m^2 (x_6). The human development index (x_1) and the percentage of households with a floor space of less than 50 m^2 (x_6) are variables that have a significant local effect on the percentage of poverty population in South Sulawesi Province in 2019. The open unemployment rate (x_3) and the percentage of food expenditure per capita (x_4) are two variables that have a significant global influence on the percentage of poverty population in South Sulawesi Province in 2019.

Keyword: GM-estimator, MGWR, Outliers, Percentage of Poverty Population, Tukey Bisquare

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan metode analisis statistika yang digunakan untuk mempelajari hubungan antara variabel respon dan variabel prediktor. Model regresi yang melibatkan satu variabel respon dan lebih dari satu variabel prediktor disebut regresi linear berganda. Model regresi linear berganda sering disebut dengan model regresi global (Widarjono, 2007). Pada model regresi global, diasumsikan bahwa nilai taksiran parameternya akan bernilai sama untuk setiap lokasi pengamatan atau berlaku secara global (Wahyunik, 2014). Asumsi tersebut akan menghasilkan kesalahan apabila terdapat perbedaan kondisi antar lokasi pengamatan yang dapat dipengaruhi oleh faktor geografis, sosial budaya atau faktor lainnya. Perbedaan kondisi antar lokasi pengamatan dapat menimbulkan efek heterogenitas spasial, sehingga data antar pengamatan sangat sulit dianalisis menggunakan regresi global. Oleh karena itu, diperlukan sebuah model regresi yang melibatkan pengaruh heterogenitas spasial ke dalam model. Salah satu metode regresi yang dapat menangani data yang memiliki heterogenitas spasial adalah *Geographically Weighted Regression* (GWR) (Fotheringham, dkk., 2002).

Model GWR merupakan pengembangan dari model regresi global dengan menambahkan faktor letak geografis sehingga setiap lokasi pengamatan memiliki nilai parameter yang berbeda-beda. Model ini menghasilkan penaksir parameter yang bersifat lokal untuk setiap lokasi pengamatan. Namun pada saat pengujian parameter variabel prediktor, terkadang diperoleh beberapa variabel yang tidak signifikan atau tidak mempunyai pengaruh lokasi sehingga variabel tersebut berpengaruh secara global terhadap model (Widayaka, 2016). Oleh karena itu, model GWR kemudian dikembangkan menjadi model *Mixed Geographically Weighted Regression* (MGWR).

Model MGWR adalah suatu metode pemodelan yang menggabungkan model regresi global dengan model GWR. Penaksiran parameter model MGWR terdiri dari penaksiran parameter yang besifat global menggunakan metode *Ordinary Least Square*

(OLS) dan parameter yang bersifat lokal menggunakan metode *Weighted Least Square* (WLS) (Fotheringham, dkk., 2002). Beberapa penelitian yang telah mengkaji tentang model MGWR, diantaranya dilakukan oleh Lumaela, dkk (2013) yang memodelkan *Chemical Oxigen Demand* (COD) sungai di Surabaya menggunakan MGWR yang menghasilkan variabel prediktor global dan prediktor lokal yang signifikan. Wahyunik (2014) mengkaji tentang penggunaan MGWR dengan fungsi pembobot *fixed gaussian kernel* pada data indeks kesehatan di Provinsi Sulawesi Selatan. Widayaka (2016) mengkaji pendekatan MGWR yang diaplikasikan untuk menyelidiki variabel-variabel yang berpengaruh terhadap penentuan nilai PDRB di Provinsi Jawa Tengah dengan memperhatikan letak daerah untuk mengestimasi parameternya. Semua penelitian tersebut menggunakan metode WLS dan OLS dalam menaksir parameter lokal dan global model MGWR.

Data yang diperoleh berdasarkan tiap lokasi pengamatan seringkali memuat pencilan. Pencilan merupakan pengamatan yang jauh dari pusat data yang mungkin dapat berpengaruh besar terhadap koefisien regresi sehingga membuat estimasi parameter menggunakan WLS dan OLS menjadi bias. Oleh karena itu, diperlukan sebuah metode yang lebih *robust* (kekar) terhadap keberadaan pencilan dalam mengestimasi parameter model MGWR. Imayati (2015) mengkaji tentang model MGWR pada data yang mengandung pencilan menggunakan metode *robust M-estimator*. Huber (1964) mengatakan bahwa *M-estimator* tidak resisten terhadap pencilan pada variabel prediktor, sehingga Mallows (1975) memperkenalkan sebuah metode yang merupakan pengembangan dari *M-estimator* dengan tujuan untuk membatasi pengaruh pencilan pada variabel prediktor dengan menambahkan sebuah fungsi pembobot baru yang disebut *Generalized M-estimator* atau *GM-estimator*.

Beberapa penelitian terkait *GM-estimator* diantaranya, Song dkk, (1998) menggunakan *GM-estimator* untuk memodelkan pola perubahan oksidasi amonia menjadi asam nitrat. Chave dan Thompson (2003) menggunakan *GM-estimator* untuk menyelesaikan masalah pencilan pada variabel prediktor (*leverage point*). Kedua penelitian tersebut menggunakan *GM-estimator* yang diaplikasikan pada regresi global yang mana tidak mempertimbangkan efek spasial pada suatu pengamatan. Kristanto

(2016) mengkaji model GWR menggunakan metode *GM-estimator* pada data angka putus sekolah tingkat SMA di Jawa Timur. Namun, terkadang pada model GWR terdapat beberapa variabel yang tidak signifikan atau tidak mempunyai pengaruh lokasi sehingga digunakan model MGWR yang dapat menghasilkan variabel yang bersifat global dan lokal menggunakan metode *GM-estimator*.

Data spasial dapat ditunjukkan pada data kemiskinan yang diamati pada setiap kabupaten/kota di Provinsi Sulawesi Selatan tahun 2019. Berdasarkan hasil Susenas Maret 2019, jumlah penduduk miskin di Provinsi Sulawesi Selatan berjumlah 767,80 ribu jiwa atau 8.69% dari total penduduk. Jumlah penduduk miskin di Sulawesi Selatan terus mengalami fluktuasi setiap tahunnya dan sebagian besar berada di pedesaan. Pada Maret 2019 penduduk miskin dipedesaan mencapai 597,69 ribu jiwa sementara didaerah perkotaan hanya 170,10 ribu jiwa sehingga memungkinkan adanya pencilan pada data. Data kemiskinan pada suatu daerah dan faktor-faktor yang mempengaruhinya akan berbeda untuk setiap daerah, tergantung pada kondisi lokasi pengamatan yang diamati. Namun, tidak dapat dipungkiri bahwa kemungkinan terdapat beberapa faktor yang berpengaruh secara global pada suatu lokasi pengamatan. Oleh karena itu, dalam penelitian ini penulis akan mengkaji tentang estimasi model MGWR menggunakan *GM-estimator* yang diaplikasikan pada data persentase penduduk miskin di Provinsi Sulawesi Selatan tahun 2019. Berdasarkan uraian tersebut, penulis mengajukan bahan skripsi dengan judul "**Estimasi Parameter Model *Mixed Geographically Weighted Regression* Menggunakan Metode *Generalized M-estimator***".

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka diperoleh rumusan masalah sebagai berikut:

1. Bagaimana estimasi parameter model MGWR menggunakan metode *GM-estimator*?
2. Bagaimana penerapan model MGWR menggunakan metode *GM-estimator* pada data persentase penduduk miskin di Provinsi Sulawesi Selatan tahun 2019?

1.3 Batasan Masalah

Untuk mendekati sasaran yang diharapkan, maka perlu diadakan pembatasan permasalahan sebagai berikut:

1. Metode estimasi parameter model MGWR menggunakan metode *GM-estimator* dengan fungsi pembobot *Tukey Bisquare.*
2. Penelitian ini diaplikasikan pada data persentase penduduk miskin, indeks pembangunan manusia, persentase penduduk, tingkat pengangguran terbuka, persentase pengeluaran per kapita untuk makanan, persentase rumah tangga dengan sumber air minum kemasan, air minum isi ulang, dan leding serta persentase rumah tangga dengan luas lantai rumah < 50 m^2 di Provinsi Sulawesi Selatan tahun 2019 yang diperinci berdasarkan 24 kabupaten/kota.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, tujuan dari penelitian ini adalah sebagai berikut:

1. Memperoleh estimasi parameter model MGWR menggunakan metode *GM-estimator*.
2. Memperoleh hasil penerapan model MGWR menggunakan metode *GM-estimator* pada data persentase penduduk miskin di Provinsi Sulawesi Selatan tahun 2019.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Menambah wawasan dan pengetahuan tentang estimasi parameter model MGWR menggunakan metode *GM-estimator*.
2. Memberikan informasi kepada pemerintah untuk menetapkan kebijakan dalam rangka mengatasi kemiskinan di Provinsi Sulawesi Selatan.

BAB II

TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi merupakan suatu metode analisis statistika yang mempelajari hubungan antara variabel respon dan variabel prediktor. Dalam analisis statistika, hubungan yang sebenarnya tidak dapat diketahui secara pasti, tetapi model hubungannya dapat diestimasi berdasarkan data pengamatan. Model regresi yang melibatkan satu variabel respon dan satu variabel prediktor disebut regresi linear sederhana. Menurut Pambudi (2012) model regresi linear sederhana dapat dinyatakan dalam persamaan berikut:

$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i, \quad i = 1,2,\dots,n \tag{2.1}$$

Model regresi linear sederhana pada persamaan (2.1) dapat diperluas menjadi model regresi linear berganda dengan melakukan penambahan pada variabel prediktor. Jika terdapat k variabel prediktor ($k \geq 2$) maka model regresi linear berganda dapat dinyatakan dalam bentuk persamaan berikut (Pambudi, 2012):

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i, \quad i = 1,2,\dots,n \tag{2.2}$$

dengan

y_i : variabel respon pada pengamatan ke-i

β_0 : nilai konstanta/intersep

β_j : koefisien regresi, $j = 1,2,\dots,k$

x_{ij} : variabel prediktor ke-j pada pengamatan ke-i

ε_i : residual yang diasumsikan berdistribusi normal dengan *mean* nol dan *variansi* σ^2

Parameter $\beta_0, \beta_1, \dots, \beta_k$ merupakan parameter yang tidak diketahui, sehingga perlu diestimasi. Estimasi parameter yang biasa digunakan adalah metode kuadrat terkecil atau *Ordinary Least Square* (OLS) dengan meminimumkan jumlah kuadrat dari residual dan disamakan dengan nol. Berdasarkan persamaan (2.2) dapat ditulis:

$$S(\beta_j) = \sum_{i=1}^{n} {\varepsilon_i}^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2 \tag{2.3}$$

Persamaan (2.3) dapat ditulis dalam bentuk matriks sebagai berikut:

$$S = \sum_{i=1}^{n} \boldsymbol{\varepsilon_i}^2 = \boldsymbol{\varepsilon}^t \boldsymbol{\varepsilon} = (\boldsymbol{Y} - \boldsymbol{X\beta})^t(\boldsymbol{Y} - \boldsymbol{X\beta})$$
$$= \boldsymbol{Y}^t\boldsymbol{Y} - \boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{Y} - \boldsymbol{Y}^t\boldsymbol{X\beta} + \boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{X\beta}$$
$$= \boldsymbol{Y}^t\boldsymbol{Y} - 2\boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{Y} + \boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{X\beta} \tag{2.4}$$

Untuk memperoleh penduga parameter $\boldsymbol{\hat{\beta}}$, maka persamaan (2.4) diturunkan terhadap $\boldsymbol{\beta}^t$ dan hasilnya disamakan dengan nol maka diperoleh:

$$\frac{\partial S}{\partial \boldsymbol{\beta}^t} = \frac{\partial(\boldsymbol{Y}^t\boldsymbol{Y} - 2\boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{Y} + \boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{X\beta})}{\partial \boldsymbol{\beta}^t} = 0$$
$$-2\boldsymbol{X}^t\boldsymbol{Y} + 2\boldsymbol{X}^t\boldsymbol{X\beta} = 0$$
$$\boldsymbol{X}^t\boldsymbol{X\beta} = \boldsymbol{X}^t\boldsymbol{Y}$$
$$(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{X\beta} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{Y}$$
$$\boldsymbol{\hat{\beta}} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{Y} \tag{2.5}$$

2.2 Pengujian Heterogenitas Spasial

Perbedaan karakteristik antara satu titik lokasi pengamatan dengan titik lokasi pengamatan lainnya menyebabkan terjadinya heterogenitas spasial. Pengujian heterogenitas spasial dapat dilakukan dengan menggunakan statistik uji *Breusch-Pagan* dengan hipotesisnya sebagai berikut (Anselin, 1988):

H_0 : $\sigma^2(u_1, v_1) = \cdots = \sigma^2(u_n, v_n) = \sigma^2$ (tidak terjadi heterogenitas spasial)

H_1 : minimal ada satu $\sigma^2(u_i, v_i) \neq \sigma^2, i = 1,2, \ldots, n$ (terjadi heterogenitas spasial)

Adapun statistik uji *Breusch-Pagan* (BP) adalah sebagai berikut:

$$BP = \frac{1}{2}\boldsymbol{f}^t\boldsymbol{Z}(\boldsymbol{Z}^t\boldsymbol{Z})^{-1}\boldsymbol{Z}^t\boldsymbol{f} \tag{2.6}$$

dengan elemen vektor $\boldsymbol{f}$ adalah $f_i = \left(\frac{\varepsilon_i{}^2}{\sigma^2} - 1\right)$, ε_i merupakan residual pengamatan ke-i, σ^2 merupakan ragam residual dan $\boldsymbol{Z}$ matriks berukuran $n \times (k+1)$ berisi vektor variabel prediktor yang telah terstandarisasi ($\boldsymbol{z}$) untuk setiap lokasi pengamatan dengan k adalah banyaknya variabel prediktor. Nilai statistik uji BP mengikuti distribusi χ^2. Kriteria keputusan dalam uji ini adalah tolak H_0 jika nilai BP $> \chi^2_{(\alpha,k)}$ sehingga dapat disimpulkan terjadi heterogenitas spasial (Anselin, 2009 dalam Ramadhan, 2012).

2.3 Fungsi pembobot

Fungsi dari pembobot adalah untuk memberikan hasil pendugaan parameter yang berbeda untuk setiap lokasi pengamatan. Pada analisis spasial, pendugaan parameter di lokasi pengamatan ke-i akan lebih dipengaruhi oleh titik-titik yang dekat dengan lokasi pengamatan tersebut. Oleh karena itu, pemilihan pembobot spasial yang digunakan dalam menduga parameter menjadi sangat penting. Matriks pembobot dapat diperoleh berdasarkan informasi jarak dari ketetanggaan atau dengan kata lain jarak antara satu lokasi pengamatan dengan lokasi pengamatan yang lain. Apabila lokasi ke-j terletak pada koordinat (u_j, v_j) maka akan diperoleh jarak *euclidean* (d_{ij}) antara lokasi pengamatan ke-i dan lokasi pengamatan ke-j dengan menggunakan persamaan sebagai berikut (Amaliasari, 2015):

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \tag{2.7}$$

dengan

d_{ij} : jarak *euclidean* antara lokasi pengamatan ke-i dengan lokasi pengamatan ke-j

u_i : koordinat *longitude* lokasi pengamatan ke-i

u_j : koordinat *longitude* lokasi pengamatan ke-j

v_i : koordinat *latitude* lokasi pengamatan ke-i

v_j : koordinat *latitude* lokasi pengamatan ke-j

Pembobotan pada model GWR dapat menggunakan fungsi kernel yang mana pada penelitian ini menggunakan fungsi pembobot *adaptive gaussian kernel* yang dapat dituliskan sebagai berikut:

$$W_{ij}(u_i, v_i) = exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h_i}\right)^2\right) \tag{2.8}$$

dengan

d_{ij} : jarak *euclidean* antara lokasi pengamatan ke-i dengan lokasi pengamatan ke-j

h_i : nilai *bandwidth* pada lokasi pengamatan ke- i

2.4 Penentuan *Bandwidth*

Bandwidth merupakan radius (h) suatu lingkaran sehingga sebuah titik lokasi pengamatan yang berada dalam radius lingkaran tersebut masih dianggap berpengaruh dalam membentuk parameter di titik lokasi pengamatan ke- i (Fotheringham, dkk., 2002). Nilai *bandwidth* yang sangat kecil akan mengakibatkan penaksiran parameter di lokasi pengamatan ke-i semakin bergantung pada titik lokasi pengamatan yang memiliki jarak terdekat dengan lokasi pengamatan ke-i, sehingga varians yang dihasilkan akan semakin besar. Sebaliknya, jika nilai *bandwidth* sangat besar maka akan mengakibatkan bias yang semakin besar (Dwinata, 2012). Oleh karena itu, diperlukan pemilihan *bandwidth* optimum menggunakan metode *Cross Validation* (CV). *Bandwidth* optimum adalah *bandwidth* yang menghasilkan CV minimum. Secara sistematis nilai CV dapat dituliskan sebagai berikut:

$$CV(h) = \sum_{i=1}^{n} (y_i - \hat{y}_{\neq i}(h))^2 \tag{2.9}$$

dengan

n : banyaknya pengamatan

y_i : pengamatan ke-i

$\hat{y}_{\neq i}(h)$: penduga y_i dengan pengamatan lokasi ke-i dihilangkan dari proses pendugaan.

Proses untuk mendapatkan *bandwidth* yang menghasilkan CV minimum dapat dilakukan menggunakan teknik *Golden Section Search* (Fotheringham, dkk., 2002). Proses ini dilakukan dengan cara mengevaluasi fungsi dengan tiga nilai yang berbeda, misalnya a, b dan c dengan $a < b < c$. Nilai a merupakan batas bawah nilai *bandwidth* yang diperoleh dari minimum d_{ij} sedangkan c merupakan batas atas nilai *bandwidth* yang diperoleh dari maksimum d_{ij} dan $b = (a + c)/2$. Nilai fungsi yang dihasilkan pada tiga titik tersebut adalah $f(a), f(b),$ dan $f(c)$ yang disebut sebagai triplet. Fungsi tersebut dievaluasi kembali pada suatu nilai baru d yang nilainya dapat ditentukan diantara a dan b atau diantara b dan c sehingga diperoleh nilai fungsi baru, yaitu $f(d)$.

Kemudian salah satu nilai a atau c dibuang untuk membentuk triplet baru. Aturan yang digunakan pada teknik *Golden Section Search* adalah:

Jika $f(b) < f(d)$: triplet baru yang digunakan adalah $a < b < d$

Jika $f(b) > f(d)$: triplet baru yang digunakan adalah $b < d < c$

Proses tersebut berulang hingga diperoleh dua nilai $f(d)$ yang dihasilkan mendekati sama atau selisihnya lebih kecil daripada suatu nilai yang ditentukan, misalnya 1×10^{-6} (Caraka dan Yasin, 2017).

2.5 *Geographically Weighted Regression*

Geographically Weighted Regression (GWR) merupakan pengembangan dari model regresi linear berganda yang menghasilkan penduga parameter yang bersifat lokal untuk setiap lokasi pengamatan. Model GWR dikembangkan untuk mengatasi pengaruh heterogenitas spasial yang disebabkan oleh perbedaan kondisi antar tiap lokasi pengamatan. Parameter yang dihasilkan dari model GWR akan berbeda-beda untuk setiap lokasi pengamatan. Menurut Fotheringham, dkk, (2002) Model GWR dapat ditulis sebagai berikut:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{k} \beta_j(u_i, v_i)x_{ij} + \varepsilon_i, \qquad i = 1,2, \dots, n \tag{2.10}$$

dengan

y_i : variabel respon pada lokasi pengamatan ke-i

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

$\beta_0(u_i, v_i)$: nilai intersep model GWR

$\beta_j(u_i, v_i)$: koefisien regresi ke-j pada lokasi pengamatan ke-i

x_{ij} : variabel prediktor ke-j pada lokasi pengamatan ke-i

ε_i : residual yang diasumsikan berdistribusi normal dengan *mean* nol dan *varians* σ^2

Pendugaan parameter model GWR dapat dilakukan menggunakan metode *Weighted Least Square* (WLS) dengan memberikan pembobot yang berbeda pada setiap lokasi pengamatan (Fotheringham, dkk., 2002). Misalkan setiap titik lokasi (u_i, v_i) diberikan pembobot yaitu $W_i(u_i, v_i)$ dengan $i = 1,2, \dots, n$ maka parameter pada lokasi pengamatan (u_i, v_i) diestimasi dengan menambahkan unsur pembobot

$W_i(u_i, v_i)$ kemudian meminimumkan jumlah kuadrat residual dari persamaan (2.10) sehingga diperoleh (Azizah, 2013):

$$\sum_{i=1}^{n} W_i(u_i, v_i)\varepsilon_i^2 = \sum_{i=1}^{n} W_i(u_i, v_i)\left(y_i - \beta_0(u_i, v_i) - \sum_{j=1}^{k} \beta_j(u_i, v_i)x_{ij}\right)^2 \quad (2.11)$$

Persamaan (2.11) dimisalkan koordinat $(u_i, v_i) = l$ dan dapat dinyatakan dalam bentuk matriks sebagai berikut:

$$\begin{aligned}
\boldsymbol{\varepsilon}^t \boldsymbol{W_l} \boldsymbol{\varepsilon} &= (\boldsymbol{Y}^t - \boldsymbol{\beta_l}^t \boldsymbol{X}^t)\boldsymbol{W_l}(\boldsymbol{Y} - \boldsymbol{X\beta_l}) \\
&= \boldsymbol{Y}^t\boldsymbol{W_l}\boldsymbol{Y} - \boldsymbol{Y}^t\boldsymbol{W_l}\boldsymbol{X\beta_l} - \boldsymbol{\beta_l}^t\boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{Y} + \boldsymbol{\beta_l}^t\boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{X\beta_l} \\
&= \boldsymbol{Y}^t\boldsymbol{W_l}\boldsymbol{Y} - (\boldsymbol{Y}^t\boldsymbol{W_l}\boldsymbol{X\beta_l})^t - \boldsymbol{\beta_l}^t\boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{Y} + \boldsymbol{\beta_l}^t\boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{X\beta_l} \\
&= \boldsymbol{Y}^t\boldsymbol{W_l}\boldsymbol{Y} - \boldsymbol{\beta_l}^t\boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{Y} - \boldsymbol{\beta_l}^t\boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{Y} + \boldsymbol{\beta_l}^t\boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{X\beta_l} \\
&= \boldsymbol{Y}^t\boldsymbol{W_l}\boldsymbol{Y} - \boldsymbol{2\beta_l}^t\boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{Y} + \boldsymbol{\beta_l}^t\boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{X\beta_l} \qquad (2.12)
\end{aligned}$$

dengan

$$\boldsymbol{\beta_l} = \begin{pmatrix} \beta_0(u_i, v_i) \\ \beta_1(u_i, v_i) \\ \vdots \\ \beta_p(u_i, v_i) \end{pmatrix} \text{ dan } \boldsymbol{W_l} = diag\big(W_1(u_i, v_i), W_2(u_i, v_i), \dots, W_n(u_i, v_i)\big)$$

Untuk mendapatkan pendugaan parameter $\widehat{\boldsymbol{\beta}}_{\boldsymbol{l}}$, maka persamaan (2.12) diturunkan terhadap $\boldsymbol{\beta_l}^t$ dan disamakan dengan nol, sehingga diperoleh (Azizah, 2013):

$$\begin{aligned}
\frac{\partial \boldsymbol{\varepsilon}^t \boldsymbol{W_l} \boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta_l}^t} = 0 - 2\boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{Y} + \boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{X\beta_l} + \boldsymbol{W_l}\big(\boldsymbol{X}^t\boldsymbol{\beta_l}^t\boldsymbol{X}\big)^t &= 0 \\
-2\boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{Y} + \boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{X\beta_l} + \boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{X\beta_l} &= 0 \\
-2\boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{Y} + 2\boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{X\beta_l} &= 0 \\
\boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{Y} &= \boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{X\beta_l} \\
\widehat{\boldsymbol{\beta}}_{\boldsymbol{l}} &= (\boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{W_l}\boldsymbol{Y}
\end{aligned}$$

Sehingga diperoleh estimasi parameter model GWR pada lokasi ke pengamatan-i:

$$\widehat{\boldsymbol{\beta}}(u_i, v_i) = (\boldsymbol{X}^t\boldsymbol{W}(u_i, v_i)\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^t\boldsymbol{W}(u_i, v_i)\boldsymbol{Y} \qquad (2.13)$$

2.6 Uji Variabilitas Model GWR

Uji variabilitas model GWR dilakukan untuk menentukan sifat lokal dan global dari suatu variabel dengan menggunakan pengujian pengaruh lokasi secara parsial (Leung, dkk., 2000). Uji parsial digunakan untuk mengetahui adanya perbedaan pengaruh yang signifikan dari variabel prediktor x_j antara satu lokasi dengan lokasi pengamatan lainnya. Pengujian ini dapat dilakukan dengan hipotesis sebagai berikut:

H_0 : $\beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots = \beta_k(u_i, v_i)$ untuk $i = 1,2, \ldots, n$

H_1 : minimal ada satu $\beta_1(u_i, v_i) \neq \beta_k(u_n, v_n)$ untuk $i = 1,2, \ldots, n$

Statistik uji:

$$F = \frac{{V_j}^2 / tr\left(\frac{1}{k}\boldsymbol{B_j}^t \left[I - \frac{1}{k}\boldsymbol{J}\right]\boldsymbol{B_j}\right)}{\boldsymbol{Y}^t(\boldsymbol{I}-\boldsymbol{S})^t(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{Y}/b_1} \tag{2.14}$$

dengan

$$V_j^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_j(u_i, v_i) - \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_j(u_i, v_i)\right)^2 = \frac{1}{n}\boldsymbol{\beta_j}^{\boldsymbol{t}}\left[\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right]\boldsymbol{\beta_j}$$

$$\boldsymbol{B}_j = \begin{bmatrix} {\varepsilon_k}^t(\boldsymbol{X}^t\boldsymbol{W}(u_1, v_1)\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{W}(u_1, v_1) \\ {\varepsilon_k}^t(\boldsymbol{X}^t\boldsymbol{W}(u_2, v_2)\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{W}(u_2, v_2) \\ \vdots \\ {\varepsilon_k}^t(\boldsymbol{X}^t\boldsymbol{W}(u_n, v_n)\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{W}(u_n, v_n) \end{bmatrix}$$

$$\boldsymbol{\beta_j}(u_i, v_i) = \begin{bmatrix} \boldsymbol{\beta}_0(u_i, v_i) \\ \boldsymbol{\beta}_1(u_i, v_i) \\ \vdots \\ \boldsymbol{\beta}_k(u_i, v_i) \end{bmatrix}$$

$$b_i = tr([(\boldsymbol{I}-\boldsymbol{S})^t(\boldsymbol{I}-\boldsymbol{S})]^t), i = 1,2$$

$$c_i = tr\left(\frac{1}{n}\boldsymbol{B}_j^t\left[\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right]\boldsymbol{B}_k\right)^t, i = 1,2$$

$$\boldsymbol{S}_l = \begin{bmatrix} \boldsymbol{x}_{l1}^t[\boldsymbol{X_l}^t\boldsymbol{W}(u_i, v_i)\boldsymbol{X_l}]^{-1}\boldsymbol{X_l}^t\boldsymbol{W}(u_i, v_i) \\ \boldsymbol{x}_{l2}^t[\boldsymbol{X_l}^t\boldsymbol{W}(u_i, v_i)\boldsymbol{X_l}]^{-1}\boldsymbol{X_l}^t\boldsymbol{W}(u_i, v_i) \\ \vdots \\ \boldsymbol{x}_{ln}^t[\boldsymbol{X_l}^t\boldsymbol{W}(u_i, v_i)\boldsymbol{X_l}]^{-1}\boldsymbol{X_l}^t\boldsymbol{W}(u_i, v_i) \end{bmatrix}$$

$$\boldsymbol{S} = \boldsymbol{S}_l + (\boldsymbol{I}-\boldsymbol{S}_l)\boldsymbol{X}_g\left[\boldsymbol{X_g}^t(\boldsymbol{I}-\boldsymbol{S}_l)^t(\boldsymbol{I}-\boldsymbol{S}_l)\right]^{-1}\boldsymbol{X_g}^t(\boldsymbol{I}-\boldsymbol{S}_l)^t(\boldsymbol{I}-\boldsymbol{S}_l)$$

Matriks $\boldsymbol{J}$ merupakan matriks berukuran $n \times n$ dengan semua elemenya adalah 1 dan ε_k adalah vektor kolom berukuran $(k+1)$ yang bernilai satu untuk elemen ke-j

dan nol untuk lainnya. Nilai statistik uji F mengikuti distribusi F dengan derajat bebas $df_1 = \frac{{c_1}^2}{c_2}$ dan $df_2 = \frac{{b_1}^2}{b_2}$. Pada taraf signifikansi sebesar α maka H_0 akan ditolak jika nilai statistik $F \geq F_{\alpha,df_1,df_2}$ yang artinya variabel prediktor bersifat lokal. Sebaliknya apabila nilai statistik uji $F < F_{\alpha,df_1,df_2}$, artinya variabel prediktor bersifat global (Suritman, 2020).

2.7 *Mixed Geographically Weighted Regression*

Mixed Geographically Weighted Regression (MGWR) merupakan model kombinasi atau gabungan antara regresi global dengan model GWR yang mempertimbangkan situasi dimana beberapa variabel prediktor yang mempengaruhi respon bersifat global dan variabel prediktor lainnya bersifat lokal. Model MGWR dengan q variabel prediktor diantaranya bersifat lokal dan $(k-q)$ variabel prediktor bersifat global dengan mengasumsikan bahwa nilai intersep model bersifat lokal. Model MGWR dapat dituliskan sebagai berikut (Khairunnisa, 2015):

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{q} \beta_j\,(u_i, v_i)x_{ij} + \sum_{j=q+1}^{k} \beta_j\, x_{ij} + \varepsilon_i, i = 1,2 \dots, n \quad (2.15)$$

dengan

y_i : variabel respon pada lokasi pengamatan ke-i

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

$\beta_0(u_i, v_i)$: nilai intersep pada lokasi pengamatan ke-i

$\beta_j(u_i, v_i)$: koefisien regresi variabel prediktor ke-k pada lokasi pengamatan ke-i

β_j : koefisien regresi variabel prediktor ke-k

x_{ij} : variabel prediktor ke-j pada lokasi pengamatan ke-i

ε_i : residual pengamatan ke-i yang diasumsikan berdistribusi normal dengan *mean* nol dan *varian*s konstan σ^2

Estimasi parameter model MGWR dapat menggunakan pendekatan WLS secara lokal dengan membentuk matriks pembobot untuk setiap lokasi pengamatan. Model MGWR persamaan (2.15) dapat dinyatakan dalam bentuk matriks sebagai berikut (Khairunnisa, 2015):

$$\boldsymbol{Y} = \boldsymbol{X_l}\boldsymbol{\beta_l}(u_i, v_i) + \boldsymbol{X_g}\boldsymbol{\beta_g} + \boldsymbol{\varepsilon} \quad (2.16)$$

Persamaan (2.16) dapat dituliskan dalam bentuk GWR sebagai berikut:

$$\widetilde{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{X_g}\boldsymbol{\beta_g} = \boldsymbol{X_l}\boldsymbol{\beta_l}(u_i, v_i) + \boldsymbol{\varepsilon} \tag{2.17}$$

Estimasi parameter model MGWR pada lokasi (u_i, v_i) seperti halnya model GWR, maka diperoleh persamaan rumus berikut:

$$\begin{aligned}
Q &= \boldsymbol{\varepsilon}^t \boldsymbol{W}(u_i, v_i)\boldsymbol{\varepsilon} \\
&= \left(\widetilde{\boldsymbol{Y}} - \boldsymbol{X_l}\boldsymbol{\beta_l}(u_i, v_i)\right)^t \boldsymbol{W}(u_i, v_i)\left(\widetilde{\boldsymbol{Y}} - \boldsymbol{X_l}\boldsymbol{\beta_l}(u_i, v_i)\right) \\
&= \left(\widetilde{\boldsymbol{Y}} - \boldsymbol{X_l}\boldsymbol{\beta_l}(u_i, v_i)\right)^t \boldsymbol{W}(u_i, v_i)\left(\widetilde{\boldsymbol{Y}} - \boldsymbol{X_l}\boldsymbol{\beta_l}(u_i, v_i)\right) \\
&= \widetilde{\boldsymbol{Y}}^t \boldsymbol{W}(u_i, v_i)\widetilde{\boldsymbol{Y}} - \widetilde{\boldsymbol{Y}}^t \boldsymbol{W}(u_i, v_i)\boldsymbol{X_l}\boldsymbol{\beta_l}(u_i, v_i) - \boldsymbol{\beta_l^t}(u_i, v_i)\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_i, v_i)\widetilde{\boldsymbol{Y}} + \\
&\quad \boldsymbol{\beta_l^t}(u_i, v_i)\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_i, v_i)\boldsymbol{X_l}\boldsymbol{\beta_l}(u_i, v_i) \\
&= \widetilde{\boldsymbol{Y}}^t \boldsymbol{W}(u_i, v_i)\widetilde{\boldsymbol{Y}} - 2\boldsymbol{\beta_l^t}(u_i, v_i)\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_i, v_i)\widetilde{\boldsymbol{Y}} + \boldsymbol{\beta_l^t}(u_i, v_i)\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_i, v_i)\boldsymbol{X_l}\boldsymbol{\beta_l}(u_i, v_i)
\end{aligned}$$

Untuk memperoleh penduga parameter $\boldsymbol{\beta_l}(u_i, v_i)$, maka persamaan diatas diturunkan terhadap $\boldsymbol{\beta_l^t}(u_i, v_i)$ dan hasilnya disamakan dengan nol maka diperoleh:

$$\begin{aligned}
\frac{\partial Q}{\partial \boldsymbol{\beta_l^t}(u_i, v_i)} &= 0 \\
-2\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_i, v_i)\widetilde{\boldsymbol{Y}} + \boldsymbol{2}\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_i, v_i)\boldsymbol{X_l}\boldsymbol{\beta_l}(u_i, v_i) &= 0 \\
\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_i, v_i)\boldsymbol{X_l}\boldsymbol{\beta_l}(u_i, v_i) &= 0 \\
\boldsymbol{\beta_l}(u_i, v_i) &= \left(\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_i, v_i)\boldsymbol{X_l}\right)^{-1}\boldsymbol{X_l}^{t}\boldsymbol{W}(u_i, v_i)\widetilde{\boldsymbol{Y}}
\end{aligned}$$

Sehingga estimasi parameter model GWR pada lokasi pengamatan ke-i adalah sebagai berikut (Fotheringham, dkk., 2002):

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{l}}(u_i, v_i) = \left(\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_i, v_i)\boldsymbol{X_l}\right)^{-1}\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_i, v_i)\widetilde{\boldsymbol{Y}} \tag{2.18}$$

dengan

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{l}}(u_i, v_i) = \left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{0}}(u_i, v_i), \widehat{\boldsymbol{\beta}}_{\boldsymbol{1}}(u_i, v_i), \dots \widehat{\boldsymbol{\beta}}_{\boldsymbol{q}}(u_i, v_i)\right)^t$$

Misalkan $x_{li}{}^t = (1, x_{1i}, x_{2i}, \dots, x_{qi})$ adalah elemen baris ke-i dari matriks $\boldsymbol{X_l}$, maka nilai dugaan untuk $\tilde{y}$ pada lokasi pengamatan ke-i dapat diperoleh sebagai berikut:

$$\begin{aligned}
\hat{\tilde{y}}_i &= x_{li}{}^t\widehat{\boldsymbol{\beta}}_{\boldsymbol{l}}(u_i, v_i) \\
&= x_{li}{}^t\left((\boldsymbol{X_l^t}\boldsymbol{W}(u_i, v_i)\boldsymbol{X_l})^{-1}\boldsymbol{X_l^t}\boldsymbol{W}(u_i, v_i)\widetilde{\boldsymbol{Y}}\right)
\end{aligned} \tag{2.19}$$

Sehingga untuk keseluruhan pengamatan dapat ditulis:

$$\hat{\widetilde{\boldsymbol{Y}}} = \left(\hat{\widetilde{\boldsymbol{Y}}}_{\boldsymbol{1}}, \hat{\widetilde{\boldsymbol{Y}}}_{\boldsymbol{2}}, \dots, \hat{\widetilde{\boldsymbol{Y}}}_{\boldsymbol{n}}\right) = \boldsymbol{S_l}\widetilde{\boldsymbol{Y}}$$

dengan,

$$\boldsymbol{S}_l = \begin{bmatrix} x_{l1}^t[\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_1,v_1)\boldsymbol{X_l}]^{-1}\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_1,v_1) \\ x_{l2}^t[\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_2,v_2)\boldsymbol{X_l}]^{-1}\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_2,v_2) \\ \vdots \\ x_{ln}^t[\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_n,v_n)\boldsymbol{X_l}]^{-1}\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_n,v_n) \end{bmatrix}$$

Selanjutnya mensubtitusikan elemen dari $\widehat{\boldsymbol{\beta}}_{\boldsymbol{l}}(u_i,v_i)$ kedalam model MGWR sebagai berikut:

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{X_l}\boldsymbol{\beta_l}(u_i,v_i) + \boldsymbol{X_g}\boldsymbol{\beta_g} + \boldsymbol{\varepsilon} \\ \boldsymbol{Y} &= \boldsymbol{S}_l\widetilde{\boldsymbol{Y}} + \boldsymbol{X_g}\boldsymbol{\beta_g} + \boldsymbol{\varepsilon} \\ \boldsymbol{Y} &= \boldsymbol{S}_l\left(\boldsymbol{Y} - \boldsymbol{X_g}\boldsymbol{\beta_g}\right) + \boldsymbol{X_g}\boldsymbol{\beta_g} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= (\boldsymbol{I} - \boldsymbol{S}_l)\boldsymbol{Y} - (\boldsymbol{I} - \boldsymbol{S}_l)\boldsymbol{X_g}\boldsymbol{\beta_g} \end{aligned}$$

dengan menggunakan metode kuadrat terkecil (OLS) maka diperoleh estimasi untuk parameter global yaitu:

$$\boldsymbol{\varepsilon}^t\boldsymbol{\varepsilon} = \left((\boldsymbol{I} - \boldsymbol{S}_l)\boldsymbol{Y} - (\boldsymbol{I} - \boldsymbol{S}_l)\boldsymbol{X_g}\boldsymbol{\beta_g}\right)^t\left((\boldsymbol{I} - \boldsymbol{S}_l)\boldsymbol{Y} - (\boldsymbol{I} - \boldsymbol{S}_l)\boldsymbol{X_g}\boldsymbol{\beta_g}\right)$$

$$\boldsymbol{\varepsilon}^t\boldsymbol{\varepsilon} = \boldsymbol{Y}^t(\boldsymbol{I} - \boldsymbol{S}_l)^t(\boldsymbol{I} - \boldsymbol{S}_l)\boldsymbol{Y} - 2\boldsymbol{X_g}^t\boldsymbol{\beta_g}^t(\boldsymbol{I} - \boldsymbol{S}_l)^t(\boldsymbol{I} - \boldsymbol{S}_l)\boldsymbol{Y} + \boldsymbol{X_g}^t\boldsymbol{\beta_g}^t(\boldsymbol{I} - \boldsymbol{S}_l)^t(\boldsymbol{I} - \boldsymbol{S}_l)\boldsymbol{X_g}\boldsymbol{\beta_g}$$

Jika persamaan $\boldsymbol{\varepsilon}^t\boldsymbol{\varepsilon}$ diturunkan terhadap $\boldsymbol{\beta_g}^t$ dan hasilnya disamakan dengan nol maka diperoleh estimasi parameter model global yaitu (Fotheringham, dkk., 2002):

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{g}} = \left(\boldsymbol{X_g}^{\boldsymbol{t}}(\boldsymbol{I} - \boldsymbol{S}_l)^{\boldsymbol{t}}(\boldsymbol{I} - \boldsymbol{S}_l)\boldsymbol{X_g}\right)^{-1}\boldsymbol{X_g}^{\boldsymbol{t}}(\boldsymbol{I} - \boldsymbol{S}_l)^{\boldsymbol{t}}(\boldsymbol{I} - \boldsymbol{S}_l)\boldsymbol{Y} \tag{2.20}$$

Dengan mensubstitusikan nilai $\widehat{\boldsymbol{\beta}}_{\boldsymbol{g}}$ pada persamaan (2.18) maka diperoleh estimasi parameter lokal model MGWR pada lokasi pengamatan ke-i adalah sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{l}}(u_i,v_i) = \left(\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_i,v_i)\boldsymbol{X_l}\right)^{-1}\boldsymbol{X_l}^{\boldsymbol{t}}\boldsymbol{W}(u_i,v_i)\left(\boldsymbol{Y} - \boldsymbol{X_g}\widehat{\boldsymbol{\beta}}_{\boldsymbol{g}}\right) \tag{2.21}$$

2.8 Pengujian Hipotesis Model MGWR

Pengujian hipotesis model MGWR terdiri dari pengujian serentak dan pengujian parsial parameter variabel global dan lokal model MGWR.

2.8.1 Pengujian Serentak Parameter Model MGWR

Menurut Purhadi dan Yasin (2012), uji ini digunakan untuk menguji secara serentak signifikansi dari parameter variabel model MGWR. Pengujian serentak parameter variabel model MGWR terdiri dari dua diantaranya pengujian pada parameter variabel global dan dilanjutkan dengan pengujian pada parameter variabel lokal. Berikut hipotesis yang digunakan pada pengujian serentak parameter variabel global model MGWR:

H_0 : $\beta_{q+1} = \beta_{q+2} = \cdots = \beta_k = 0$

H_1 : minimal ada satu $\beta_j \neq 0$ untuk suatu $j = q+1, q+2, \ldots, k$

Statistik uji:

$$F_1 = \frac{\boldsymbol{Y}^t[(\boldsymbol{I}-\boldsymbol{S}_l)^t(\boldsymbol{I}-\boldsymbol{S}_l)-(\boldsymbol{I}-\boldsymbol{S})^t(\boldsymbol{I}-\boldsymbol{S})]\boldsymbol{Y}/r_1}{\boldsymbol{Y}^t(\boldsymbol{I}-\boldsymbol{S})^t(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{Y}/u_1} \sim F_{(\alpha,p)} \quad (2.22)$$

dengan $r_i = tr\left([(\boldsymbol{I}-\boldsymbol{S}_l)^t(\boldsymbol{I}-\boldsymbol{S}_l)-(\boldsymbol{I}-\boldsymbol{S})^t(\boldsymbol{I}-\boldsymbol{S})]^i\right), i = 1,2$ dan $u_t = tr([(\boldsymbol{I}-\boldsymbol{S}_l)^t(\boldsymbol{I}-\boldsymbol{S}_l)]^t), t = 1,2$. Nilai statistik uji F_1 mengikuti distribusi F dengan derajat bebas df_1 dan df_2. Jika diberikan taraf signifikansi sebesar α maka H_0 ditolak apabila nilai statistik uji $F_1 \geq F_{\alpha,df_1,df_2}$ dengan $df_1 = \frac{{r_1}^2}{r_2}$ dan $df_2 = \frac{{u_1}^2}{u_2}$.

Selanjutnya dilakukan pengujian serentak parameter variabel lokal model MGWR dengan hipotesis sebagai berikut:

H_0 : $\beta_1(u_i,v_i) = \beta_2(u_i,v_i) = \cdots = \beta_q(u_i,v_i) = 0$ untuk setiap $i = 1,2,\ldots,n$

H_1 : minimal ada satu $\beta_j(u_i,v_i) \neq 0$ untuk setiap $j = 1, 2, \ldots, q$

Statistik uji:

$$F_2 = \frac{\boldsymbol{Y}^t\left[\left(\boldsymbol{I}-\boldsymbol{S_g}\right)^t\left(\boldsymbol{I}-\boldsymbol{S_g}\right)-(\boldsymbol{I}-\boldsymbol{S})^t(\boldsymbol{I}-\boldsymbol{S})\right]\boldsymbol{Y}/c_1}{\boldsymbol{Y}^t(\boldsymbol{I}-\boldsymbol{S})^t(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{Y}/u_1} \sim F_{(\alpha,p)} \quad (2.23)$$

dengan $c_i = tr\left(\left[\left(\boldsymbol{I}-\boldsymbol{S}_g\right)^t\left(\boldsymbol{I}-\boldsymbol{S}_g\right)-(\boldsymbol{I}-\boldsymbol{S})^t(\boldsymbol{I}-\boldsymbol{S})\right]^i\right), i = 1,2$ dan adalah $\boldsymbol{S_g} = \boldsymbol{X_g}\left({\boldsymbol{X_g}}^t\boldsymbol{X_g}\right)^{-1}{\boldsymbol{X_g}}^t$. Nilai statistik uji F_2 mengikuti distribusi F dengan derajat bebas df_1 dan df_2. Jika diberikan taraf signifikansi sebesar α maka H_0 ditolak apabila nilai statistik uji $F_2 \geq F_{\alpha,df_1,df_2}$ dengan $df_1 = \frac{{c_1}^2}{c_2}$ dan $df_2 = \frac{{u_1}^2}{u_2}$ (Suritman, 2020).

2.8.2 Pengujian Parsial Parameter Model MGWR

Menurut Purhadi dan Yasin (2012), uji ini digunakan untuk mengetahui variabel global dan lokal yang berpengaruh signifikan secara parsial terhadap variabel respon pada model MGWR. Pengujian ini terdiri dari pengujian parsial parameter variabel global dan parameter variabel lokal. Untuk pengujian parsial parameter variabel global digunakan hipotesis yaitu:

H_0 : $\beta_j = 0$ untuk setiap $j = q+1, q+2, \dots, k$

H_1 : minimal ada satu $\beta_j \neq 0$ untuk suatu $j = q+1, q+2, \dots, k$

Statistik uji:

$$t_{g\_hitung} = \frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{g_{jj}}} \sim t_{(\alpha/_{2}, df)} \tag{2.24}$$

dengan g_{jj} adalah elemen diagonal ke-j dari hasil perkalian matriks $\boldsymbol{GG^t}$. Matriks $\boldsymbol{G} = \left[\boldsymbol{X_g^t}(\boldsymbol{I}-\boldsymbol{S}_l)^t(\boldsymbol{I}-\boldsymbol{S}_l)\boldsymbol{X}_g\right]^{-1}\boldsymbol{X_g^t}(\boldsymbol{I}-\boldsymbol{S}_l)^t(\boldsymbol{I}-\boldsymbol{S}_l)$ dan $\hat{\sigma}^2 = \frac{\boldsymbol{Y}^t(\boldsymbol{I}-\boldsymbol{S})^t(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{Y}}{tr\left((\boldsymbol{I}-\boldsymbol{S})^t(\boldsymbol{I}-\boldsymbol{S})\right)}$. Pada taraf signifikansi α, maka dapat diambil keputusan tolak H_0 jika $\left|t_{g\_hitung}\right| > t_{\alpha/_{2}, df}$ dengan $df = \frac{{u_1}^2}{u_2}$.

Uji hipotesis selanjutnya ditunjukkan untuk mengetahui variabel lokal yang berpengaruh signifikan terhadap variabel respon pada model MGWR. Untuk menguji signifikasi suatu variabel lokal digunakan hipotesis yaitu:

H_0 : $\beta_j(u_i, v_i) = 0$ untuk setiap $j = 1, 2, \dots, q$ dan $i = 1,2, \dots, n$

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

Statistik uji:

$$t_{l\_hitung} = \frac{\hat{\beta}_j(u_i, v_i)}{\hat{\sigma}\sqrt{m_{jj}}} \sim t_{(\alpha/_{2}, df)} \tag{2.25}$$

dengan m_{jj} adalah elemen diagonal ke-j dari hasil perkalian matriks $\boldsymbol{MM^t}$. Matriks $\boldsymbol{M} = [\boldsymbol{X}_l^t\boldsymbol{W}(u_i, v_i)\boldsymbol{X}_l]^{-1}\boldsymbol{X}_l^t\boldsymbol{W}(u_i, v_i)\left(\boldsymbol{I} - \boldsymbol{X}_g\boldsymbol{G}\right)$. Pada taraf signifikansi sebesar α, maka dapat diambil keputusan tolak H_0 jika $\left|t_{l\_hitung}\right| > t_{\alpha/_{2}, df}$ dengan $df = \frac{{u_1}^2}{u_2}$ (Suritman, 2020).

2.9 Pencilan (*Outlier*)

Istilah pencilan atau *outlier* merujuk pada suatu pengamatan yang tidak mengikuti pola umum pada persamaan regresi yang dihasilkan atau pola data secara keseluruhan. Secara tidak langsung pencilan merupakan hasil observasi yang menyimpang dari pola yang terbentuk oleh sebagian besar data (Ghozali, 2011). Data pencilan dapat dikenali dengan pemeriksaan secara visual dari data mentah menggunakan metode grafis seperti diagram pencar/*scatter plot* dan *boxplot* variabel prediktor dan variabel respon. Namun, ketika terdapat lebih dari dua variabel prediktor, beberapa pencilan akan sulit dideteksi dengan pemeriksaan visual sehingga diperlukan suatu metode yang dapat digunakan dalam mengidentifikasi pencilan yang berpengaruh terhadap koefisien regresi antara lain:

2.9.1 Nilai *Leverage*

Metode yang digunakan dalam mengidentifikasi pencilan pada variabel prediktor adalah nilai pengaruh (*leverage value*). Nilai *leverage* berasal dari nilai diagonal matriks *hat* ($\boldsymbol{H}$) yang berukuran $n \times n$.

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^t \tag{2.26}$$

dengan $\boldsymbol{X}$ merupakan matriks $n \times (k+1)$ dengan n merupakan banyaknya data pengamatan dan k merupakan banyaknya variabel prediktor. Diagonal dari $\boldsymbol{H}$ berisi nilai-nilai *leverage* (h_{ii}). *Leverage* untuk kasus ke- i merupakan nilai dari baris ke-i dan kolom ke-i dari $\boldsymbol{H}$.

$$h_{ii} = \boldsymbol{x}_i{}^t(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{x}_i \tag{2.27}$$

dengan $\boldsymbol{x}_i{}^t$ adalah baris ke-i dari $\boldsymbol{X}$. Jika nilai h_{ii} lebih besar dari $\frac{2(k+1)}{n}$ maka pengamatan ke-i dikatakan pencilan terhadap x.

2.9.2 Metode DFFITS

Metode DFFITS (*Difference in fit Standardized*) digunakan untuk mendeteksi pencilan yang berpengaruh pada model regresi. Nilai DFFITS dapat didefinisikan sebagai berikut (Montgomery dan Peck, 1982):

$$DFFITS_i = t_i\left(\frac{h_{ii}}{1-h_{ii}}\right)^{\frac{1}{2}} \tag{2.28}$$

dengan

$$t_i = \varepsilon_i\sqrt{\frac{n-k-1}{JKG(1-h_{ii})-\varepsilon_i{}^2}}$$

ε_i adalah residual pengamatan ke-i dan JKG merupakan jumlah kuadrat residual. Suatu pengamatan ke-i dikatakan pencilan apabila pengamatan tersebut memiliki nilai $|DFFITS_i| > \sqrt{\frac{k+1}{n}}$ dengan k banyaknya parameter dalam model dan n banyaknya pengamatan (Widodo dan Dewayanti, 2016).

2.10 Fungsi Objektif

Fungsi yang digunakan untuk mencari fungsi pembobot pada regresi *robust* adalah fungsi objektif. Fungsi objektif yang diturunkan terhadap u_i akan menghasilkan suatu fungsi yang disebut fungsi pengaruh $\psi(u_i)$. Fungsi pengaruh yang dibagi dengan u_i akan menghasilkan fungsi pembobot yang digunakan dalam perhitungan regresi *robust*. Fungsi pembobot yang digunakan adalah *Tukey Bisquare*. Diberikan suatu fungsi objektif sebagai berikut:

$$\rho(u_i) = \begin{cases} \frac{c^2}{6}\left\{1 - \left[1 - \left(\frac{u_i}{c}\right)^2\right]^3\right\}, |u_i| \leq c \\ \frac{c^2}{6} \qquad\qquad\qquad\qquad |u_i| > c \end{cases} \tag{2.29}$$

dengan fungsi pengaruh yaitu:

$$\rho'(u_i) = \frac{\partial(\rho(u_i))}{\partial u_i} = \psi(u_i) = \begin{cases} \varepsilon_i \left[1 - \left(\frac{u_i}{c}\right)^2\right]^3, |u_i| \leq c \\ 0, \qquad\qquad |u_i| > c \end{cases} \tag{2.30}$$

Sehingga diperoleh fungsi pembobot

$$\omega_i = \omega(u_i) = \frac{\psi(\varepsilon_i)}{u_i} = \begin{cases} \left[1 - \left(\frac{u_i}{c}\right)^2\right]^3, |u_i| \leq c \\ 0, \qquad\qquad |u_i| > c \end{cases} \tag{2.31}$$

dengan $c = 4.685$ adalah konstanta *tunning* yang menghasilkan efisiensi tinggi dengan residual berdistribusi normal dan dapat memberikan perlindungan terhadap pencilan.

2.11 *GM-estimator*

Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi residual tidak normal atau terdapat beberapa pencilan yang berpengaruh pada hasil analisis regresi. Metode ini merupakan alat penting untuk menganalisis data yang dipengaruhi oleh pencilan sehingga dihasilkan model yang *robust* (kekar) terhadap pencilan (Draper dan Smith, 1998). Salah satu metode *robust* yang dapat digunakan adalah *GM-estimator*. *GM-estimator* merupakan pengembangan dari *M-estimator*, ketika *M-estimator* kurang resisten terhadap pencilan pada variabel x_i. Ide dasar yang melatarbelakangi *GM-estimator* adalah untuk membatasi pengaruh pencilan pada variabel x_i dengan menggunakan fungsi pembobot (η), dengan fungsi pembobot (η) hanya bergantung pada x_i. Secara umum *GM-estimator* didefinisikan sebagai berikut:

$$GM = \sum_{i=1}^{n} \eta(x_i)\rho(u_i) \tag{2.32}$$

dengan $\eta(x_i)$ merupakan fungsi pembobot untuk variabel x_i dan $\rho(u_i)$ adalah fungsi objektif dengan $u_i = \frac{\varepsilon_i}{\hat{\sigma}}$. Nilai $\hat{\sigma}$ merupakan skala *robust* yang perlu diestimasi. Estimasi $\hat{\sigma}$ yang digunakan adalah:

$$\hat{\sigma} = \frac{MAD}{0.6745} = \frac{med|\varepsilon_i - med(\varepsilon_i)|}{0.6745} \tag{2.33}$$

Pemilihan konstanta 0.6745 membuat $\hat{\sigma}$ suatu estimator yang mendekati tak bias dari σ jika n besar dan residual berdistribusi normal. Pembuktian jika $\hat{\sigma} = \frac{MAD}{s}$, maka $s = 0.6745$ diperoleh dari perhitungan berikut:

$$\begin{aligned}
P\left(|Z| \leq \frac{MAD}{\hat{\sigma}}\right) &= \frac{1}{2} \\
P\left(-\frac{MAD}{\hat{\sigma}} \leq Z \leq \frac{MAD}{\hat{\sigma}}\right) &= \frac{1}{2} \\
P\left(Z \leq \frac{MAD}{\hat{\sigma}}\right) &= \frac{1}{4} \\
P\left(Z > \frac{MAD}{\hat{\sigma}}\right) &= \frac{3}{4} \\
\frac{MAD}{\hat{\sigma}} &= 0.6745 \\
\frac{MAD}{0.6745} &= \frac{MAD}{0.6745} \\
s &= 0.6745
\end{aligned}$$

dengan demikian persamaan (2.32) menjadi,

$$GM = \sum_{i=1}^{n} \eta(x_i)\rho\left(\frac{y_i - \sum_{j=0}^{k} \beta_j x_{ij}}{\hat{\sigma}}\right) \tag{2.34}$$

Misalkan $\rho(u_i)$ mempunyai turunan $\frac{\partial\rho(u_i)}{\partial(u_i)} = \psi(u_i, \beta)$, sehingga untuk meminimumkan persamaan (2.34) akan digunakan turunan parsial pertama dari $\rho(u_i)$ terhadap β_j dengan $j = 0,1, \dots k$ kemudian disamakan dengan nol sehingga menghasilkan persamaan berikut:

$$\frac{\partial\rho(u_i)}{\partial(u_i)} = \frac{\partial\rho(u_i)}{\partial(u_i)} . \frac{\partial(u_i)}{\partial\beta_j} = 0$$

$$= \sum_{i=1}^{n} x_{ij}\eta(x_i)\psi\left(\frac{y_i - \sum_{j=0}^{k}\beta_j x_{ij}}{\hat{\sigma}}\right) = 0 \tag{2.35}$$

dengan ψ merupakan fungsi pengaruh yang digunakan untuk memperoleh fungsi pembobot *robust*. Secara umum fungsi pembobot *GM-estimator* $\eta(x_i)$ terdefinisi pada nilai *leverage* dengan $\eta(x_i) = \sqrt{1 - h_{ii}}$. Nilai *leverage* (h_{ii}) berbatas pada $0 \leq h_{ii} \leq 1$ yang diperoleh dari matriks diagonal pada *hat* matriks simetris $(n \times n)$ pada persamaan (2.27). Berdasarkan persamaan (2.35) Draper dan Smith (1998) memberikan solusi dengan mendefinisikan fungsi pembobot (ω_i) sebagai berikut:

$$\omega_i = \omega(u_i) = \frac{\psi\left(\frac{y_i - \sum_{j=0}^{k}\beta_j x_{ij}}{\hat{\sigma}}\right)}{\frac{y_i - \sum_{j=0}^{k}\beta_j x_{ij}}{\hat{\sigma}}} \tag{2.36}$$

Kemudian kedua ruas pada persamaan (2.36) dikalikan $\frac{y_i - \sum_{j=0}^{k}\beta_j x_{ij}}{\hat{\sigma}}$ sehingga diperoleh:

$$\omega_i\left(\frac{y_i - \sum_{j=0}^{k}\beta_j x_{ij}}{\hat{\sigma}}\right) = \psi\left(\frac{y_i - \sum_{j=0}^{k}\beta_j x_{ij}}{\hat{\sigma}}\right) \tag{2.37}$$

sehingga persamaan (2.35) dapat ditulis:

$$\sum_{i=1}^{n} x_{ij}\eta(x_i)\omega_i\left(\frac{y_i - \sum_{j=0}^{k}\beta_j x_{ij}}{\hat{\sigma}}\right) = 0, \quad j = 0,1,2 \dots, k \tag{2.38}$$

Persamaan (2.38) dapat diselesaikan menggunakan metode iterasi yang dinamakan *Iteratively Reweighted Least Square* (IRLS). Prosedur estimasi menggunakan metode IRLS dapat dilakukan dengan beberapa langkah sebagai berikut:

a. Melakukan estimasi parameter regresi menggunakan *least square*, sehingga diperoleh $\hat{y}_i{}^0$ dan diperoleh residual $\varepsilon_i{}^0$ dimana $\varepsilon_i{}^0 = y_i - \hat{y}_i{}^0, i = 1,2, \dots, n$ yang diperlakukan sebagai nilai awal.
b. Menentukan estimator skala *robust* $\hat{\sigma}^0$ berdasarkan nilai residual yang diperoleh pada langkah (a) dan fungsi pembobot awal $\omega_i{}^0 = \frac{\psi(u_i{}^{(0)})}{u_i{}^{(0)}}$, dimana $\omega_i{}^0$ dihitung menggunakan fungsi *Tukey Bisquare.*
c. Mencari estimasi parameter pada iterasi pertama $r = 1$ menggunakan nilai residual $\varepsilon_i{}^{(0)}$ dan $\omega_i{}^{(0)}$.

$$\widehat{\boldsymbol{\beta}}^{(\boldsymbol{r})} = (\boldsymbol{X}^t\boldsymbol{\eta}\boldsymbol{\omega}^{(\boldsymbol{r}-\mathbf{1})}\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{\eta}\boldsymbol{\omega}^{(\boldsymbol{r}-\mathbf{1})}\boldsymbol{Y} \tag{2.39}$$

d. Menghitung $\sum_{i=1}^{n}\left|y_i - \hat{y}_i{}^{(1)}\right|$ atau $\sum_{i=1}^{n}\left|\varepsilon_i{}^{(1)}\right|$.
e. Mengulangi langkah (b) sampai (d) hingga mendapatkan nilai $\sum_{i=1}^{n}\left|\varepsilon_i{}^{(r)}\right|$ yang konvergen ke suatu nilai yaitu:

$$\left|\widehat{\boldsymbol{\beta}}^{(\boldsymbol{r})} - \widehat{\boldsymbol{\beta}}^{(\boldsymbol{r}-1)}\right| < \varepsilon$$

dengan ε adalah nilai bilangan positif kecil yaitu 0.001.

2.12 Kemiskinan

Kemiskinan merupakan suatu kondisi ketidakmampuan secara ekonomi untuk memenuhi standar hidup rata-rata masyarakat di suatu daerah. Kondisi ketidakmampuan ini ditandai dengan rendahnya kemampuan pendapatan untuk memenuhi kebutuhan pokok baik berupa pangan, sandang, maupun papan. Provinsi Sulawesi Selatan merupakan salah satu wilayah di Indonesia yang masih menghadapi permasalahan kemiskinan meskipun menjadi salah satu provinsi yang mempunyai tingkat pertumbuhan ekonomi cukup baik tetapi angka kemiskinan di Sulawesi Selatan masih terbilang cukup tinggi (Azwar dan Subekan, 2016). Berdasarkan data resmi yang dirilis Badan Pusat Statistik (BPS) hingga akhir Desember 2020, penduduk miskin di Provinsi Sulawesi Selatan Maret 2019 berjumlah 767.800 jiwa mengalami penurunan sebesar 24.830 jiwa dibandingkan kondisi maret 2018 yang berjumlah 792.603 jiwa. Persentase penduduk miskin Maret 2019 sebesar 8,69 persen juga mengalami penurunan 0,37 poin persen dibandingkan Maret 2018 yang besarnya 9,06 persen.

Untuk mengukur kemiskinan, BPS menggunakan konsep kemampuan memenuhi kebutuhan dasar (*basic needs approach*). Pada pendekatan ini, kemiskinan dipandang sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran. BPS menggunakan metode *Head Count Index* untuk mengetahui persentase penduduk miskin pada suatu wilayah, yang dapat ditunjukkan pada persamaan berikut:

$$P_0 = \frac{1}{n}\sum_{i=1}^{q}\left[\frac{z - y_i}{z}\right]^0 \qquad (2.40)$$

dengan

P_0 : persentase penduduk miskin

z : garis kemiskinan

q : banyaknya penduduk yang berada dibawah garis kemiskinan

n : jumlah penduduk

y_i : rata-rata pengeluaran per kapita sebulan penduduk yang berada dibawah garis kemiskinan ($y_i < z$)

Penduduk yang memiliki rata-rata pengeluaran per kapita perbulan dibawah garis kemiskinan dikategorikan sebagai penduduk miskin. Garis Kemiskinan (GK) merupakan penjumlahan dari Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non Makanan (GKNM). GKM merupakan nilai pengeluaran kebutuhan minimum makanan yang disetarakan dengan 2100 kkal per kapita perhari sedangkan GKNM adalah kebutuhan minimum untuk perumahan, sandang, pendidikan dan kesehatan. Adapun faktor-faktor penyebab kemiskinan yang digunakan pada penelitian ini terdiri dari:

1. Indeks Pembangunan Manusia

   Indeks Pembangunan Manusia (IPM) merupakan indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup masyarakat, sehingga dapat menentukan peringkat atau level pembangunan suatu wilayah atau negara. Terdapat tiga komponen dasar penyusun indeks pembangunan manusia, yaitu dimensi kesehatan, pendidikan, dan standar hidup layak. Adapun rumus yang digunakan untuk menghitung IPM adalah sebagai berikut (BPS, 2020):

$$IPM = \sqrt[3]{I_{kesehatan} + I_{pendidikan} + I_{pengeluaran}}$$

dengan

$I_{kesehatan}$: indeks untuk dimensi kesehatan

$I_{pendidikan}$: indeks untuk dimensi pendidikan

$I_{pengeluaran}$: indeks untuk dimensi pengeluaran

2. Persentase Penduduk

   Kemiskinan adalah ketidakmampuan untuk memenuhi kebutuhan dasar seperti makan, tempat tinggal, dan kesehatan. Dalam proses pemenuhan kebutuhan hidup erat kaitannya dengan tingkat pertumbuhan penduduk karena jika jumlah penduduk tinggi disuatu wilayah diindikasikan akan mempengaruhi kesediaan lahan tempat tinggal, kebutuhan hidup, dan melimpahnya tenaga kerja. Kelebihan tenaga kerja akan mengakibatkan upah menjadi turun. Upah tersebut hanya dapat digunakan untuk membiayai taraf hidup minimum sehingga perekonomian akan mengalami keterhambatan. Untuk menghitung persentase penduduk di suatu wilayah maka dapat dinyatakan dengan rumus sebagai berikut (BPS, 2020):

$$Persentase\ penduduk = \frac{Jumlah\ penduduk\ di\ suatu\ wilayah}{Jumlah\ total\ penduduk} \times 100\%$$

3. Tingkat Pengangguran Terbuka

   Faktor lain yang dapat mempengaruhi kemiskinan adalah pengangguran. Menurut BPS (2020) pengangguran didefinisikan sebagai (1) penduduk yang aktif mencari pekerjaan, (2) penduduk yang sedang mempersiapkan usaha/pekerjaan baru, (3) penduduk yang tidak mencari pekerjaan karena merasa tidak mungkin mendapat pekerjaan. (4) kelompok penduduk yang tidak aktif mencari pekerjaan dengan alasan sudah mempunyai pekerjaan tetapi belum mulai bekerja. Adapun alat ukur yang digunakan untuk melihat angka pengangguran adalah Tingkat Pengangguran Terbuka (TPT). Tingkat pengangguran terbuka adalah persentase jumlah pengangguran terhadap jumlah angkatan kerja. Angkatan kerja merupakan penduduk usia kerja (15 tahun ke atas) yang bekerja atau punya pekerjaan namun sementara tidak bekerja, dan pengangguran. Untuk menghitung tingkat pengangguran terbuka maka dapat dinyatakan dengan rumus berikut (BPS, 2020):

$$TPT = \frac{Jumlah\ pengangguran}{Jumlah\ angkatan\ kerja} \times 100\%$$

4. Persentase Pengeluaran per Kapita Untuk Makanan

   Badan Pusat Statistik mendefinisikan bahwa pengeluaran per kapita adalah pengeluaran rumah tangga dibagi dengan banyaknya anggota rumah tangga. Persentase pengeluaran per kapita untuk makanan adalah pengeluaran per kapita untuk makanan dibagi dengan total pengeluaran per kapita (makanan + non makanan). Selain itu, BPS memberikan konsep bahwa pengeluaran untuk konsumsi makanan dihitung selama seminggu terakhir yang selanjutnya dikonversikan ke dalam pengeluaran rata-rata perbulan. Persentase pengeluaran per kapita untuk makanan dapat dinyatakan dengan rumus sebagai berikut (BPS, 2020):

$$PPM = \frac{Pengeluaran\ per\ kapita\ untuk\ makanan}{Total\ pengeluaran\ per\ kapita\ (makanan + non\ makanan)} \times 100\%$$

5. Persentase Rumah Tangga dengan Sumber Air Minum Kemasan, Air Minum Isi Ulang, dan Leding.

   Berdasarkan peraturan pemerintah Nomor 122 tahun 2015 tentang sistem penyediaan air minum mendefinisikan kebutuhan pokok air minum sehari-hari adalah air untuk memenuhi kebutuhan hidup sehari-hari yang digunakan untuk keperluan minum, masak, mandi, cuci, peturasan, dan ibadah (Badan Pusat Statistik, 2020). Air minum kemasan, air minum isi ulang, dan leding termasuk dalam sumber air minum layak. Persentase rumah tangga yang memiliki akses terhadap sumber air minum layak dapat dinyatakan dengan rumus sebagai berikut (BPS, 2020):

$$PRTSAM = \frac{a}{b} \times 100\%$$

   a : jumlah rumah tangga yang memiliki akses terhadap sumber air minum layak

   b : jumlah rumah tangga

6. Persentase Rumah Tangga dengan Luas Lantai Rumah < 50 m^2

   Luas lantai adalah luas lantai yang ditempati dan digunakan untuk keperluan sehari-hari (sebatas atap). Bagian-bagian yang digunakan bukan untuk keperluan sehari-hari tidak dimasukkan dalam perhitungan luas lantai seperti lumbung padi, kandang ternak, lantai jemur (lamporan semen), dan ruangan khusus untuk usaha (misalnya warung). Adanya data persentase rumah tangga dengan luas lantai rumah < 50 m2 untuk melihat kesejahteraan rumah tangga dari sisi perumahan. Rumah tangga dengan luas laintai rumah < 50 m2 dapat dinyatakan dengan rumus sebagai berikut (BPS, 2020):

$$PRTLR = \frac{c}{d} \times 100\%$$

   c : jumlah seluruh luas lantai < 50 m^2 yang dihuni rumah tangga

   d : jumlah rumah tangga