PEMODELAN *ROBUST GEOGRAPHICALLY WEIGHTED REGRESSION* MENGGUNAKAN METODE ESTIMASI MM DAN *LEAST ABSOLUTE DEVIATION*

ROBUST GEOGRAPHICALLY WEIGHTED REGRESSION MODELING USING THE METHODS OF MM ESTIMATION AND LEAST ABSOLUTE DEVIATION

AQILAH SALSABILA RAHMAN

PROGRAM STUDI STATISTIKA
SEKOLAH PASCASARJANA
UNIVERSITAS HASANUDDIN
MAKASSAR
2024

PEMODELAN *ROBUST GEOGRAPHICALLY WEIGHTED REGRESSION* MENGGUNAKAN METODE ESTIMASI MM DAN *LEAST ABSOLUTE DEVIATION*

Tesis

sebagai salah satu syarat untuk mencapai gelar Magister

Program Studi Statistika

Disusun dan diajukan oleh :

AQILAH SALSABILA RAHMAN

H062221016

kepada

PROGRAM MAGISTER STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2024

UNIVERSITAS HASANUDDIN

TESIS

PEMODELAN *ROBUST GEOGRAPHICALLY WEIGHTED REGRESSION* MENGGUNAKAN METODE ESTIMASI MM DAN *LEAST ABSOLUTE DEVIATION*

AQILAH SALSABILA RAHMAN
H062221016

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Magister Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin
pada tanggal 24 Januari 2024
dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama

Prof. Dr. Dr. Georgina Maria Tinungki, M.Si.
NIP. 19620926 198702 2 001

Pembimbing Pendamping

Dr. Erna Tri Herdiani, S.Si., M.Si.
NIP. 19750429 200003 2 001

Ketua Program Studi
Magister Statistika

Dr. Erna Tri Herdiani, S.Si., M.Si.
NIP. 19750429 200003 2 001

Dekan Fakultas MIPA
Universitas Hasanuddin

Dr. Eng. Amiruddin, M.Si.
NIP. 19720515 199702 1 002

iii

TESIS

UNIVERSITAS HASANUDDIN

PERNYATAAN KEASLIAN TESIS DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, tesis berjudul "**Pemodelan *Robust Geographically Weighted Regression* Menggunakan Metode Estimasi MM dan *Least Absolute Deviation***" adalah benar karya saya dengan arahan dari komisi pembimbing Prof. Dr. Dr. Georgina Maria Tinungki, M.Si. sebagai Pembimbing Utama dan Dr. Erna Tri Herdiani, S.Si, M.Si. sebagai Pembimbing Pendamping. Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka tesis ini. Sebagian dari isi tesis ini telah dipublikasikan di Jurnal **(International Journal of Research Publication, Volume 131, Halaman 334-341, dan DOI: 10.47119/IJRP1001311820235415)** sebagai artikel dengan judul "**Robust Geographically Weighted Regression Model on Poverty Data in South Sulawesi in 2019**".

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Makassar, 24 Januari 2024
Yang menyatakan,

METERAI TEMPEL
D8DAKX569660050

Aqilah Salsabila Rahman
NIM H062221016

PERNYATAAN KEASLIAN TESIS

DAN PELIMPAHAN HAK CIPTA

UCAPAN TERIMA KASIH

Segala Puji Syukur kepada **Allah Subhanahu wa Ta'ala**, atas berkat, bimbingan dan kasih karunia-Nya yang dilimpahkan kepada penulis, serta shalawat dan salam semoga selalu tercurahkan kepada Nabi yang paling dimuliakan, pemimpin orang-orang bertakwa, **Muhammad shallallahu ‘alaihi wa sallam** yang dinanti-nantikan syafaatnya di akhirat kelak. Limpahan doa kepada keluarga serta sahabat Rasulullah. *Alhamdulillah*, berkat rahmat dan karunia serta mukzizat-Nya sehingga penulis dapat menyelesaikan tesis ini sebagai salah satu syarat akademik untuk meraih gelar Magister Sains pada Program Studi Magister Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulisan tugas akhir ini tentunya tidak lepas dari bantuan berbagai pihak baik moril maupun materil. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang tak terhingga untuk Ayah tercinta **Ir. Abdul Rahman Suyuti, M.M.** dan Ibu tercinta **Andi Suryani, S.E.** yang tak kenal lelah mendoakan, memberikan dukungan, dan selalu melimpahkan cinta dan kasih sayangnya kepada penulis sehingga menjadi motivasi terbesar bagi penulis dalam menyelesaikan tugas akhir ini. Serta adik penulis **Achmad Fauzi Rahman**, terima kasih telah memberikan semangat, motivasi, dukungan, dan doa yang diberikan kepada penulis.

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.
2. Bapak **Dr. Eng. Amiruddin,** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. Ibu **Dr. Anna Islamiyati, S.Si., M.Si.** selaku Ketua Departemen Statistika beserta seluruh jajarannya, serta segenap dosen pengajar dan staff Departemen Statistika.
4. Ibu **Dr. Erna Tri Herdiani, S.Si, M.Si** selaku Ketua Program Magister Statistika sekaligus pembimbing pendamping yang telah bersedia meluangkan waktu dan pikirannya untuk membimbing dan memberi arahan [illegible]an penuh kesabaran serta selalu memotivasi penulis serta menjadi [illegible]at berkeluh kesah untuk penulis selama menjadi mahasiswa hingga akhir [illegible]lisan tugas akhir ini.

UNIVERSITAS HASANUDDIN

5. Ibu **Prof. Dr. Dr. Georgina Maria Tinungki, M.Si.** selaku Pembimbing Utama penulis yang telah ikhlas meluangkan waktu dan pemikirannya untuk memberikan arahan, pengetahuan, motivasi dan bimbingan di tengah kesibukan beliau serta menjadi tempat berkeluh kesah untuk penulis.
6. Ibu **Prof. Dr. Nurtiti Sunusi, S.Si., M.Si, Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.**, dan **Dr. Nirwan, M.Si.** selaku Tim Penguji. Terima kasih telah meluangkan waktu dan telah memberikan kritik dan saran yang membangun dalam penyempurnaan penyusunan tugas akhir ini.
7. Sahabat terbaik penulis, **Muhammad Fathur Rahman**, yang senantiasa mendengarkan curahan hati, keluh, dan kesah, serta banyak memberikan bantuan, motivasi, semangat, dan turut mendoakan dalam penyelesaian tugas akhir ini.
8. Sahabat spesial penulis selama menjalani kuliah perkuliahan, **Zakiah Fitri**, **Wulan Abdullah**, **Nurhidayatullah**, **Anisa Haura**, **Zulfa Putri Asmawi**, yang bersedia mendengarkan curahan hati, memberikan dorongan, semangat, dan motivasi dalam setiap keadaan sehingga penulis bisa mendapatkan lebih banyak pelajaran hidup.
9. Sahabat penghuni Ruang Diskusi, **Khairunnisa Abdullah**, **Serly Malinda**, **Mahrani**, **Nur Ikhwana**, yang menjadi motivasi, menemani, dan saling memberi dukungan dan bantuan dalam penyelesaian tugas akhir ini. Terkhusus kak **Muhammad Fadil**, penulis berterima kasih atas bantuan yang luar biasa dalam penyusunan tugas akhir ini.
10. Teman-teman Mahasiswa Program Magister Statistika angkatan I sampai angkatan VIII, terkhusus kak **Samsul Arifin** dan **Samsir Aditya Ania**, serta seluruh pihak yang sedikit banyaknya telah berpartisipasi dalam penulisan tugas akhir ini yang tidak dapat penulis sebutkan satu persatu.

Semoga Allah Subhanahu wa Ta'ala memberikan balasan yang berlipat ganda kepada semuanya yang telah membantu dalam penulisan tugas akhir ini. Penulis berharap semoga tulisan ini dapat memberikan manfaat bagi semua pihak yang memerlukannya. Dengan penuh kerendahan hati, penulis meminta maaf atas segala kekurangan dalam penulisan tugas akhir ini.

Makassar, 24 Januari 2024

Penulis

ABSTRAK

AQILAH SALSABILA RAHMAN. **Pemodelan *Robust Geographically Weighted Regression* Menggunakan Metode Estimasi MM dan *Least Absolute Deviation*** (dibimbing oleh Georgina Maria Tinungki dan Erna Tri Herdiani).

Keberhasilan pembangunan suatu daerah dapat dilihat dari pertumbuhan ekonominya. Laju pertumbuhan ekonomi dihitung berdasarkan Produk Domestik Regional Bruto (PDRB). Secara spasial, struktur perekonomian Indonesia mengalami fluktuasi setiap tahunnya. Kasus fluktuatif tersebut dapat ditekan jika faktor-faktor yang mempengaruhi PDRB dapat diketahui. Hal itu dapat diperoleh dari hasil estimasi atau kesimpulan yang diperoleh menggunakan analisis regresi. Namun, tidak jarang kesimpulan yang diperoleh keliru. Kesalahan penarikan kesimpulan bukanlah hal sepele, yang jika diabaikan, dapat menyebabkan hasil analisis menjadi meragukan dan tidak valid. Dalam banyak kasus, data yang diamati tidak normal karena keberadaan pencilan dan nilai ekstrim. Hal ini mengandalkan asumsi normalitas dan keragaman spasial. Model *Robust Geographically Weighted Regression* (RGWR) menawarkan solusi pada kasus heterogenitas spasial dengan keberadaan pencilan. Terdapat banyak metode robust dikembangkan oleh para ahli. Salah satunya adalah metode Estimasi MM dan *Least Absolute Deviation*. Metode Estimasi MM dilakukan dengan meminimumkan suatu fungsi galat. Sedangkan, metode LAD dilakukan dengan meminimumkan jumlah mutlak galat. Model RGWR memungkinkan setiap lokasi pengamatan memiliki model yang berbeda-beda dan hanya berlaku pada lokasi pengamatan tersebut. Pemodelan diterapkan pada angka PDRB di Indonesia tahun 2021, sehingga diperoleh bahwa metode LAD lebih baik daripada metode Estimasi MM untuk memodelkan faktor-faktor yang mempengaruhi angka PDRB di Indonesia tahun 2021.

Kata Kunci: *Robust*; *Geographically Weighted Regression*; Estimasi MM; *Least Absolute Deviation*; Produk Domestik Regional Bruto.

ABSTRACT

AQILAH SALSABILA RAHMAN. **Robust Geographically Weighted Regression Modeling Using The Methods Of MM Estimation and Least Absolute Deviation** (dibimbing oleh Georgina Maria Tinungki dan Erna Tri Herdiani).

The rate of economic growth in a region indicates how well its development is going. The economic growth rate is determined based on Gross Regional Domestic Product (GRDP). Spatially, the structure of the Indonesian economy fluctuates every year. This fluctuating case can be suppressed if the factors that influence GRDP can be discovered. This can be obtained from estimation results or conclusions using regression analysis. However, it is not uncommon for the conclusions obtained to be wrong. The existence of errors in decision making are not trivial and, if neglected, might result in dubious and invalid analysis results. In many instances, the observed data is not normal because of outliers and extreme values. It is based on the normality and spatial diversity assumptions. It relies on the assumptions of normality and spatial diversity. The Robust Geographically Weighted Regression (RGWR) model proposes a solution to the case of spatial heterogeneity in the presence of outliers. Experts develop numerous robust methods. One of these is the MM and Least Absolute Deviation estimate method. The MM-Estimation method is done by minimizing an error function. Whereas the LAD method is done by minimizing the absolute amount of error. The RGWR model allows for a distinct model for each observation location, which solely applies to that observation location. After modeling the GRDP figures in Indonesia in 2021, it was discovered that the LAD method outperforms the MM-Estimation method in modeling the factors affecting the GRDP figures in Indonesia in 2021.

Kata Kunci: Robust; Geographically Weighted Regression; MM-Estimation; Least Absolute Deviation; Gross Regional Domestic Product.

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1. Latar Belakang

Keberhasilan pembangunan suatu daerah dapat dilihat dari tingkat pertumbuhan ekonominya. Pertumbuhan ekonomi yang tinggi dan berkelanjutan merupakan kondisi utama bagi kelangsungan pembangunan ekonomi. Pertumbuhan ekonomi merupakan aspek yang sangat penting bagi kemajuan suatu wilayah dimana kondisi perekonomian wilayah setiap tahun diharapkan terus membaik dan tidak mengalami perlambatan. Definisi laju pertumbuhan ekonomi merupakan pertumbuhan produksi barang dan jasa di suatu wilayah perekonomian dalam selang waktu tertentu. Tujuan laju pertumbuhan ekonomi diantaranya untuk mengukur kemajuan ekonomi sebagai hasil pembangunan nasional. Selain itu, pertumbuhan ekonomi juga digunakan sebagai dasar pembuatan perkiraan penerimaan wilayah untuk perencanaan pembangunan nasional atau sektoral dan regional, serta sebagai dasar pembuatan perkiraan bisnis, khususnya persamaan penjualan (Badan Pusat Statistik, 2015).

Laju pertumbuhan ekonomi dihitung berdasarkan Produk Domestik Regional Bruto (PDRB). Perekonomian di suatu wilayah dikatakan tumbuh dan berkembang jika barang dan jasa yang diproduksi pada periode tertentu lebih besar dibandingkan dengan periode sebelumnya, yang kemudian diturunkan menjadi nilai tambah. PDRB dibedakan menjadi dua yaitu PDRB menurut harga konstan dan menurut harga berlaku. PDRB menurut harga konstan didasarkan pada tahun 2010. Tahun 2010 dipilih sebagai tahun dasar PDRB atas harga konstan dikarenakan perekonomian Indonesia pada tahun 2010 relatif stabil, selain itu terjadi perubahan struktur ekonomi selama beberapa tahun terakhir terutama dalam bidang teknologi dan informasi yang berpengaruh terhadap munculnya produk-produk baru. PDRB menurut harga berlaku menggambarkan nilai tambah barang dan jasa yang dihitung menggunakan harga-harga tahun berjalan. PDRB atas dasar harga berlaku digunakan untuk melihat struktur perekonomian, sedangkan PDRB Atas Dasar Harga Konstan (ADHK) bertujuan engukur pertumbuhan ekonomi (Departemen Statistik Ekonomi dan Bank Indonesia, 2014).

ecara spasial, struktur perekonomian Indonesia pada triwulan III 2023 sumbangkan oleh kelompok provinsi di Pulau Jawa dengan peranan terhadap PDB sebesar 57.12 persen, diikuti Pulau Sumatra sebesar 22.16 persen, Pulau Kalimantan sebesar 8.08 persen, Pulau Sulawesi sebesar 7.25 persen, Pulau Bali dan Nusa Tenggara sebesar 2.80 persen, serta Pulau Maluku dan Papua sebesar 2.59 persen. Laju pertumbuhan ekonomi triwulan III 2023 dibanding triwulan III 2022 mengalami pertumbuhan yang melambat di hampir seluruh wilayah Indonesia. Pertumbuhan tertinggi dicatat oleh kelompok provinsi di Pulau Maluku dan Papua sebesar 9.25 persen, diikuti Pulau Sulawesi sebesar 6.44 persen, Pulau Kalimantan dan Pulau Jawa sebesar 4.83 persen, Pulau Sumatera sebesar 4.50 persen dan pertumbuhan terendah di Pulau Bali dan Nusa Tenggara yang tumbuh sebesar 3.43 persen (Badan Pusat Statistik, 2023).

Kasus fluktuatif laju pertumbuhan PDRB dapat ditekan jika faktor-faktor yang mempengaruhi PDRB dapat diketahui. Hubungan angka PDRB dengan faktor-faktor penyebabnya dapat diketahui dengan menggunakan metode analisis regresi. Regresi linier berganda memodelkan hubungan antara variabel terikat dan variabel bebas (Sinaga dkk., 2022).

Beberapa penelitian telah dilakukan untuk mengkaji faktor-faktor yang mempengaruhi PDRB. Wulandari (2019) mengatakan bahwa tenaga kerja, pendapatan daerah, upah minimum kabupaten/kota, dan indeks pembangunan manusia mempengaruhi PDRB di Pulau Jawa. Yuliana, dkk (2022) mengatakan bahwa realisasi investasi dalam negeri dan jumlah proyek mempengaruhi PDRB di Kalimantan Timur. Tripena (2022) mengatakan bahwa sektor industri dan sektor perdagangan secara signifikan mempengaruhi PDRB di Kawasan Barlingmascakeb. Sari (2023) memastikan bahwa investasi, tenaga kerja, dan indeks pembangunan manusia mempengaruhi PDRB di Pulau Jawa.

Regresi linier berganda merupakan metode yang memodelkan hubungan antara variabel respon dan variabel prediktor (Sinaga dkk., 2022). Metode estimasi yang digunakan untuk menganalisis regresi adalah metode kuadrat terkecil (Montgomery dan Peck, 1992). Dalam menggunakan suatu metode, terdapat asumsi yang harus dipenuhi agar metode tersebut dapat digunakan. Menurut Gujarati (2007) metode kuadrat terkecil mempunyai asumsi-asumsi tertentu. Salah satu asumsi yang harus dipenuhi ialah kehomogenan ragam. Menurut Erda (2018) kehomogenan ragam sulit didapat karena adanya perbedaan karakteristik pada ayah yang mengakibatkan terjadinya keragaman spasial. Keragaman nerupakan suatu keadaan saat pengukuran hubungan antar variabel beda antara satu lokasi pengamatan dengan lokasi pengamatan yang lainnya (Fotheringham dkk, 2002). Sehingga, diperlukan pendekatan analisis yang memperhatikan keadaan geografis yang dalam hal ini disebut analisis spasial.

*S*alah satu metode analisis spasial yang dapat digunakan untuk melakukan analisis dengan pemberian pembobot berdasarkan jarak masing-masing lokasi pengamatan secara geografis serta asumsi memiliki keragaman spasial, yaitu dengan menggunakan model *Geographically Weighted Regression* (GWR) (Fotheringham dkk., 2002). Hasil dari analisis ini adalah model persamaan yang nilai-nilai parameternya berlaku hanya pada masing-masing lokasi pengamatan dan berbeda dengan lokasi pengamatan lainnya.

Terdapat beberapa penelitian yang menggunakan model *Geographically Weighted Regression*, di antaranya Sari (2013) memodelkan *Geographically Weighted Regression* pada penderita diare di Jawa Tengah dengan fungsi pembobot *kernel bisquare*. Nadya, dkk (2015) memodelkan *Geographically Weighted Regression* pada kasus pneumonia pada balita di provinsi Jawa Barat. Fadli (2018) memodelkan *Geographically Weighted Regression* dengan fungsi pembobot *tricube* terhadap Angka Kematian Ibu (AKI) di Kabupaten Kutai Kartanegara pada Tahun 2015. Zhao, dkk (2020) menganalisis sebaran spasial PM$_{25}$ menggunakan model *Geographically Weighted Regression* dengan analisis *geodetector* dan analisis komponen utama. Permai, dkk (2021) menganalisis desentralisasi fiskal yang mempengaruhi kinerja ekonomi menggunakan *Geographically Weighted Regression*.

Seluruh penelitian yang baru saja disebutkan menggunakan metode estimasi parameter *Weighted Least Squares* (WLS). Menurut Wulandari (2019) metode tersebut tidak kekar terhadap keberadaan pencilan. Sehingga, apabila terdapat pencilan pada data, maka akan mengakibatkan terciptanya koefisien estimator yang bias dan mengakibatkan kekeliruan dalam menyimpulkan hubungan regresi. Menurut Lainun, dkk (2018) alternatif yang dapat dilakukan pada analisis regresi untuk mengatasi adanya pencilan pada data, yaitu menggunakan metode regresi kekar (*robust*). Regresi *robust* (kekar) adalah metode regresi yang digunakan ketika data pengamatan mengandung pencilan yang mempengaruhi model. Metode ini penting untuk menganalisis data yang mengandung pencilan agar model yang dihasilkan kekar terhadap keberadaan (Lainun dkk, 2018). Oleh karena itu, metode analisis *Robust* *hically Weighted Regression* (RGWR) sesuai untuk menganalisis data yang mengandung pencilan dan mencakup beberapa lokasi pengamatan dengan asumsi memiliki keragaman spasial.

Pada regresi *robust* terdapat beberapa metode estimasi yang dapat digunakan, yakni estimasi *Least Median of Squares* (LMS), estimasi *Least Trimmed Squares* (LTS), estimasi S, estimasi M, dan estimasi MM (Chen, 2002). Sedangkan, menurut Wang dan Scott (1994) metode *Least Absolute Deviation* (LAD) adalah metode yang paling sederhana dan membuat fungsi kernel menjadi lebih *robust*.

Almetwally dan Almongy (2018) membandingkan antara metode Estimasi M, Estimasi S, dan Estimasi MM menggunakan aplikasi dan simulasi dan diperoleh bahwa Estimasi MM merupakan metode terbaik berdasarkan hasil simulasi Monte Carlo. Wang, dkk (2019) menggunakan model *Geographically Weighted Regression* untuk memodelkan pengaruh urbanisasi terhadap emisi CO_2 dan diperoleh bahwa model GWR dapat dengan baik menunjukkan pengaruh urbanisasi terhadap emisi CO_2. Putra, dkk (2019) menggunakan model *Robust Geographically Weighted Regression* dengan menggunakan metode Estimasi S dan diperoleh bahwa metode Estimasi S dapat dengan efektif menghasilkan model yang lebih representatif pada data yang mengandung pencilan berdasarkan kriteria *Root Mean Squares of Error* (RMSE) dan *Mean Absolute Deviation* (MAD). Isnaeni, dkk (2019) menggunakan model *Robust Geographically Weighted Regression* menggunakan metode estimasi M dan diperoleh bahwa metode Estimasi M lebih baik berdasarkan kriteria MAD. Wulandari (2019) menggunakan model *Robust Geographically Weighted Regression* dengan menggunakan metode estimasi M dan metode *Least Absolute Deviation*, sehingga diperoleh bahwa metode LAD menghasilkan estimasi yang lebih baik menggunakan kriteria *Mean Absolute Percentages of Error* (MAPE).

Berdasarkan uraian di atas, maka peneliti tertarik untuk memodelkan faktor-faktor yang mempengaruhi Produk Domestik Regional Bruto (PDRB) atas dasar harga konstan 2010 di Indonesia Tahun 2021 menggunakan *Robust Geographically Weighted Regression* (RGWR) dengan metode estimasi MM dan *Least Absolute Deviation* (LAD).

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah yang akan dikaji adalah sebagai berikut.

1. Bagaimana pemodelan *Robust Geographically Weighted Regression* (RGWR) dengan metode Estimasi MM dan metode *Least Absolute Deviation* (LAD)?
2. Bagaimana hasil penerapan model *Robust Geographically Weighted Regression* (RGWR) dengan metode Estimasi MM dan metode *Least Absolute Deviation* (LAD) pada Produk Domestik Regional Bruto (PDRB) di Indonesia Tahun 2021?

1.3. Batasan Penelitian

Agar penelitian ini sesuai dengan rumusan masalah dan tujuan masalah, maka perlu adanya batasan masalah. Dalam penelitian ini digunakan model *Robust Geographically Weighted Regression* dengan metode Estimasi MM dan *Least Absolute Deviation* (LAD). Sehingga, penelitian ini dibatasi pada data yang memiliki keragaman spasial dan mengandung pencilan.

1.4. Tujuan Masalah

Berdasarkan rumusan masalah, maka tujuan dari penelitian ini antara lain sebagai berikut.

1. Mendapatkan model *Robust Geographically Weighted Regression* (RGWR) dengan metode Estimasi MM dan metode *Least Absolute Deviation* (LAD).
2. Mendapatkan hasil penerapan model *Robust Geographically Weighted Regression* (RGWR) dengan metode Estimasi MM dan dan metode *Least Absolute Deviation* (LAD) pada Produk Domestik Regional Bruto (PDRB) di Indonesia Tahun 2021.

1.5. Manfaat Penelitian

Adapun manfaat yang diharapkan dapat diambil dari penelitian ini, yaitu dapat menambah wawasan dan pengetahuan tentang pemodelan *Robust* [illegible]*hically Weighted Regression* (RGWR) dengan menggunakan metode [illegible] MM dan *Least Absolute Deviation* serta memperoleh informasi terkait [illegible]erapan model *Robust Geographically Weighted Regression* (RGWR) [illegible]etode Estimasi MM dan dan *Least Absolute Deviation* pada Produk

Domestik Regional Bruto (PDRB) di Indonesia Tahun 2021 agar dapat dijadikan sebagai salah satu acuan untuk menetapkan kebijakan dalam rangka meningkatkan angka Produk Domestik Regional Bruto (PDRB) di Indonesia baik bagi pemerintah ataupun instansi terkait.

BAB II
TINJAUAN PUSTAKA

2.1. Produk Domestik Regional Bruto (PDRB)

Laju pertumbuhan ekonomi dihitung berdasarkan Produk Domestik Regional Bruto (PDRB). Perekonomian di suatu wilayah dikatakan tumbuh dan berkembang jika barang dan jasa yang diproduksi pada periode tertentu lebih besar dibandingkan dengan periode sebelumnya, yang kemudian diturunkan menjadi nilai tambah. PDRB dibedakan menjadi dua yaitu PDRB menurut harga konstan dan menurut harga berlaku. PDRB menurut harga konstan didasarkan pada tahun 2010. Tahun 2010 dipilih sebagai tahun dasar PDRB atas harga konstan dikarenakan perekonomian Indonesia pada tahun 2010 relatif stabil, selain itu terjadi perubahan struktur ekonomi selama beberapa tahun terakhir terutama dalam bidang teknologi dan informasi yang berpengaruh terhadap munculnya produk-produk baru. PDRB menurut harga berlaku menggambarkan nilai tambah barang dan jasa yang dihitung menggunakan harga-harga tahun berjalan. PDRB atas dasar harga berlaku digunakan untuk melihat struktur perekonomian, sedangkan PDRB Atas Dasar Harga Konstan (ADHK) bertujuan untuk mengukur pertumbuhan ekonomi (Departemen Statistik Ekonomi dan Moneter Bank Indonesia, 2014).

Secara spasial, struktur perekonomian Indonesia pada triwulan III 2023 masih disumbangkan oleh kelompok provinsi di Pulau Jawa dengan peranan terhadap PDB sebesar 57.12 persen, diikuti Pulau Sumatra sebesar 22.16 persen, Pulau Kalimantan sebesar 8.08 persen, Pulau Sulawesi sebesar 7.25 persen, Pulau Bali dan Nusa Tenggara sebesar 2.80 persen, serta Pulau Maluku dan Papua sebesar 2.59 persen. Laju pertumbuhan ekonomi triwulan III 2023 dibanding triwulan III 2022 mengalami pertumbuhan yang melambat di hampir seluruh wilayah Indonesia. Pertumbuhan tertinggi dicatat oleh kelompok provinsi di Pulau Maluku dan Papua sebesar 9.25 persen, diikuti Pulau Sulawesi sebesar 6.44 persen, Pulau Kalimantan dan Pulau Jawa sebesar 4.83 persen, Pulau Sumatera sebesar 4.50 persen dan pertumbuhan terendah di Pulau Bali dan Nusa a yang tumbuh sebesar 3.43 persen (Badan Pusat Statistik, 2023).

roduk Domestik Regional Bruto (PDRB) merupakan salah satu indikator ntuk mengetahui kondisi ekonomi di suatu daerah dalam suatu periode (Bank Indonesia, 2015). Penerapan data pertumbuhan PDRB sesuai dengan penerapan regresi spasial yang sangat mungkin dipengaruhi perbedaan karakteristik antar wilayah dilihat dari kondisi geografis, sosial budaya, potensi wilayah, sumber daya manusia, dan termasuk posisinya terhadap provinsi lain di sekitarnya. Penelitian mengenai PDRB sebelumnya dilakukan Fatulloh (2013) dan Handayani et al. (2017) menggunakan GWR dan *Geographically Temporal Weighted Regression* (GTWR) untuk mengestimasi parameter PDRB. Wulandari (2019) mengatakan bahwa tenaga kerja, pendapatan daerah, upah minimum kabupaten/kota, dan indeks pembangunan manusia mempengaruhi PDRB di Pulau Jawa. Yuliana, dkk (2022) mengatakan bahwa realisasi investasi dalam negeri dan jumlah proyek mempengaruhi PDRB di Kalimantan Timur. Tripena (2022) mengatakan bahwa sektor industri dan sektor perdagangan secara signifikan mempengaruhi PDRB di Kawasan Barlingmascakeb. Sari (2023) memastikan bahwa investasi, tenaga kerja, dan indeks pembangunan manusia mempengaruhi PDRB di Pulau Jawa. Widyaningsih dan Fitrianingrum (2022) melakukan pemodelan spasial pada data PDRB di Pulau Jawa sebelum dan ketika pandemi.

Widyaningsih dan Fitrianingrum (2022) mengatakan bahwa Pulau Jawa yang menjadi pusat perekonomian nasional memiliki rata-rata PDRB yang lebih tinggi dibandingkan dengan pulau-pulau di Indonesia lainnya. Tetapi karena terdampak pandemi berkepanjangan, aktivitas ekonomi menjadi terbatas dan mengakibatkan perekonomian nasional melemah. Karena pulau Jawa adalah pulau dengan penduduk terpadat di Indonesia dan pusat perekonomian nasional, menjadikan berkurangnya aktivitas perekonomian di pulau Jawa memiliki dampak paling besar bagi perekonomian nasional.

2.2. Regresi Linier Berganda

Regresi linier berganda merupakan metode yang memodelkan hubungan antara variabel respon dan variabel prediktor (Sinaga, 2022). Model regresi linier untuk p variabel prediktor secara umum ditulis sebagai berikut (Fotheringham, 2002).

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + e_i \qquad (2.\ 1)$$

$$i = 1,2,\dots,n;\ k = 1,2,\dots p$$

ariabel respon pengamatan ke-i

x_{ki} : variabel prediktor pengamatan ke- i

β_0 : intersep dari model

β_k : koefisien regresi

e_i : galat pengamatan ke-i

Dari persamaan (2.1) di atas jika diuraikan menjadi persamaan-persamaan berikut.

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + e_1$$
$$y_2 = \beta_0 + \beta_1 x_1 + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + e_2$$
$$\vdots$$
$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + e_n$$

Dari persamaan-persamaan di atas didadpatkan bentuk matriks seperti berikut:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

yang dapat ditulis menjadi:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (2.\ 2)$$

Metode estimasi yang digunakan untuk menganalisis regresi adalah metode kuadrat terkecil (*Ordinary Least Square*/OLS) (Montgomery dan Peck, 1992), yaitu dengan cara meminimumkan jumlah kuadrat galat berdasarkan persamaan (2.2).

$$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$
$$= \mathbf{Y}^{\mathrm{T}}\mathbf{Y} - \mathbf{2}\mathbf{Y}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}$$

Sehingga, diperoleh bentuk estimasi parameter sebagai berikut.

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \qquad (2.\ 3)$$

Berdasarkan persamaan (2.3), maka $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$.

Dalam menggunakan suatu metode, terdapat asumsi yang harus dipenuhi agar metode tersebut dapat digunakan. Menurut Gujarati (2007) metode kuadrat terkecil mempunyai asumsi-asumsi tertentu. Salah satu asumsi yang harus dipenuhi ialah kehomogenan ragam. Menurut Erda (2018) kehomogenan ragam sulit didapat karena adanya perbedaan karakteristik pada suatu wilayah yang ...batkan terjadi keragaman spasial. Keragaman spasial merupakan suatu ... saat pengukuran hubungan antar variabel berbeda-beda antara satu ...engamatan dengan lokasi pengamatan yang lainnya (Fotheringham, 2002). Sehingga, diperlukan pendekatan analisis yang memperhatikan keadaan geografis yang dalam hal ini disebut analisis spasial.

2.3. Analisis Spasial

Analisis spasial adalah analisis yang digunakan untuk mendapatkan informasi pengamatan yang dipengaruhi efek ruang atau lokasi (Nurhuda, 2018). Data spasial adalah data yang berkaitan dengan lokasi, berdasarkan geografi yang terdiri dari lintang-bujur dan wilayah. Analisis data spasial tidak dapat dilakukan secara global, artinya setiap lokasi mempunyai karakteristik sendiri. Sebagian besar pendekatan analisisnya merupakan eksplorasi data yang disajikan dalam bentuk peta tematik (Pfeiffer, 2008). Pada model regresi spasial terdapat dua efek spasial yaitu, ketergantungan spasial (*spatial dependence*) dan keragaman spasial (*spatial heterogenity*) (Anselin, 1988).

Adanya ketergantungan spasial menunjukkan bahwa pengamatan pada lokasi yang satu dipengaruhi oleh pengamatan di lokasi yang lain. Untuk mengetahui hal tersebut, perlu dilakukan identifikasi kebenaran efek spasial pada data yang digunakan. Salah satu pengujian yang dapat digunakan untuk mengidentifikasi adanya ketergantungan spasial, yaitu dengan pengujian *Moran's I*. Pengujian *Moran's I* merupakan pengujian yang dilakukan untuk melihat apakah pengamatan di suatu lokasi berpengaruh terhadap pengamatan di lokasi lain yang letaknya saling berdekatan.

Sedangkan, keragaman spasial merujuk pada adanya keberagaman dalam hubungan secara kewilayahan (Daulay dan Simamora, 2023).

2.4. Keragaman Spasial

Keragaman spasial disebabkan karena adanya perbedaan karakteristik antar titik lokasi pengamatan. Pendeteksian adanya keragaman spasial pada data dapat dilakukan dengan Uji B*reusch-Pagan* (BP) dengan prosedur pengujian hipotesis sebagai berikut (Anselin, 1988).

i. Hipotesis pengujian

$H_0 : {\sigma_1}^2 = {\sigma_2}^2 = \cdots = {\sigma_n}^2 = \sigma^2$ (Tidak terdapat keragaman spasial)

H_1 : ada ${\sigma_i}^2 \neq \sigma^2$ (Terdapat keragaman spasial)

$1,2, \dots, n$.

ii. Statistik uji

$$BP = \left(\frac{1}{2}\right)\mathbf{f}^T\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{f} \sim \chi^2_{(\alpha,k)} \qquad (2.\ 4)$$

dengan elemen vektor $\mathbf{f}$, $f_i = \left(\frac{e_i^2}{\sigma^2} - 1\right)$

dan

e_i : galat pengamatan ke-*i*

σ^2 : ragam galat e_i

$\mathbf{Z}$: matriks berukuran $n \times (k+1)$ berisi vektor dari yang telah terstandardisasi ($\mathbf{z}$) untuk setiap lokasi dan *k* merupakan banyaknya variabel prediktor.

iii. Kriteria Keputusan

Tolak H_0 dan terima H_1 jika $BP > \chi^2_{(\alpha,k)}$ sehingga dapat disimpulkan terdapat keragaman spasial.

2.5. *Geographically Weighted Regression*

Model *Geographically Weighted Regression* (GWR) adalah salah satu metode yang digunakan untuk mengestimasi data yang memiliki keragaman spasial dengan estimasi menggunakan metode kuadrat terkecil terboboti (Fotheringham, 2002). GWR menggunakan pembobot berdasarkan jarak satu lokasi pengamatan dengan lokasi pengamatan lainnya. Model GWR dapat ditulis sebagai berikut (Huang, 2010):

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \qquad (2.\ 5)$$

$$i = 1,2,\ldots,n; k = 1,2,\ldots p$$

dengan

y_i : variabel respon lokasi ke-*i*

x_{ik} : variabel prediktor *k* pada lokasi ke-*i*

$\beta_0(u_i, v_i)$: intersep model pada lokasi ke-*i*

$\beta_k(u_i, v_i)$: koefisien regresi untuk setiap lokasi ke-*i*

(u_i, v_i) : koordinat (lintang, bujur) pada lokasi ke-*i*

ε_i : galat pada lokasi ke-*i* yang diasumsikan identik, bebas, dan berdistribusi normal dengan *mean* nol dan variansi konstan.

ari persamaan (2.5) di atas jika diuraikan menjadi persamaan-
an berikut.

$$y_1 = \beta_0(u_1, v_1) + \beta_1(u_1, v_1)x_{11} + \beta_2(u_1, v_1)x_{12} + \cdots + \beta_k(u_1, v_1)x_{1k} + \varepsilon_1$$
$$y_2 = \beta_0(u_2, v_2) + \beta_1(u_2, v_2)x_{21} + \beta_2(u_2, v_2)x_{22} + \cdots + \beta_k(u_2, v_2)x_{2k} + \varepsilon_2$$
$$\vdots$$
$$y_n = \beta_0(u_n, v_n) + \beta_1(u_n, v_n)x_{n1} + \beta_2(u_n, v_n)x_{n2} + \cdots + \beta_k(u_n, v_n)x_{nk} + \varepsilon_n$$

Dari persamaan-persamaan di atas didapatkan bentuk matriks seperti berikut.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \left[\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \odot \begin{bmatrix} \beta_0(u_1, v_1) & \beta_1(u_1, v_1) & \cdots & \beta_k(u_1, v_1) \\ \beta_0(u_2, v_2) & \beta_1(u_2, v_2) & \cdots & \beta_k(u_2, v_2) \\ \vdots & \vdots & \ddots & \vdots \\ \beta_0(u_n, v_n) & \beta_1(u_n, v_n) & \cdots & \beta_k(u_n, v_n) \end{bmatrix}\right] \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

yang dapat ditulis dalam matriks sebagai berikut (Düzgün dan Kemeç, 2008):

$$\underset{(n\times 1)}{\mathbf{Y}} = \left(\underset{(n\times(k+1))}{\mathbf{X}} \odot \underset{(n\times(k+1))}{\boldsymbol{\beta}(u_i, v_i)}\right) \underset{((k+1)\times 1)}{\mathbf{1}} + \underset{(n\times 1)}{\boldsymbol{\varepsilon}}$$

dengan $\odot$ merupakan operator perkalian Hadamard, yaitu perkalian elemen-elemen yang bersesuaian dari dua matriks yang ukurannya sama (Schott, 1997).

Metode estimasi parameter pada model GWR adalah dengan metode kuadrat terkecil terboboti (*Weighted Least Square*/WLS), yaitu dengan memberikan pembobot yang berbeda untuk setiap lokasi pengamatan data tersebut dikumpulkan. Misalkan, pembobot untuk setiap lokasi pengamatan ke-i adalah $W_i(u_i, v_i)$ dengan $i = 1,2, \ldots, n$, maka parameter lokasi $(u_i,\ v_i)$ diduga dengan menambahkan unsur pembobot dan kemudian meminimumkan jumlah kuadrat galat berikut ini.

$$\sum_i^n W_i(u_i, v_i)\varepsilon_i^2 = \left\{\sum_i^n W_i(u_i, v_i)(y_i - \beta_0(u_i, v_i) - \beta_1(u_i, v_i)x_{i1} - \beta_2(u_i, v_i)x_{i2} - \cdots - \beta_k(u_i, v_i)x_{ik})^2\right\} \qquad (2.\ 6)$$

dengan $\mathbf{W}(u_i, v_i) = diag[w_1(u_i, v_i), w_2(u_i, v_i), \ldots, w_n(u_i, v_i)]$.

Penyelesaian persamaan (2.6) dalam bentuk matriks dinyatakan sebagai berikut.

$$)\boldsymbol{\varepsilon} = \mathbf{Y}^T\mathbf{W}(u_i, v_i)\mathbf{Y} - 2\boldsymbol{\beta}^T(u_i, v_i)\mathbf{X}^T\mathbf{W}(u_i, v_i)\mathbf{Y} + \boldsymbol{\beta}^T(u_i, v_i)\mathbf{X}^T\mathbf{W}(u_i, v_i)\mathbf{X}\boldsymbol{\beta}(u_i, v_i)$$

ehingga, diperoleh bentuk estimasi parameter dari model GWR pada -i adalah sebagai berikut (Düzgün dan Kemeç, 2008).

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

$$\widehat{\boldsymbol{\beta}}(u_i, v_i) = (\mathbf{X}^T\mathbf{W}_i\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}_i\mathbf{Y} \qquad (2.\ 7)$$

Berdasarkan persamaan (2.7) nilai estimasi y untuk lokasi ke-i dapat diperoleh dengan

$$\hat{y}_i = \mathbf{x}_i\widehat{\boldsymbol{\beta}}_i \qquad (2.\ 8)$$

$$\hat{y}_i = \mathbf{x}_i\widehat{\boldsymbol{\beta}}(u_i, v_i) = \mathbf{x}_i\big((\mathbf{X}^T\mathbf{W}_i\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}_i\mathbf{Y}\big) = \mathbf{x}_i\big((\mathbf{X}^T\mathbf{W}(u_i, v_i)\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}(u_i, v_i)\mathbf{Y}\big)$$

Selanjutnya, berdasarkan persamaan (2.8) galat ε_i dapat dinyatakan sebagai berikut.

$$\varepsilon_i = \mathrm{y}_i - \hat{y}_i = \mathrm{y}_i - \mathbf{x}_i\widehat{\boldsymbol{\beta}}(u_i, v_i) \qquad (2.\ 9)$$

dengan $\mathbf{X} = [\mathbf{x}_1; \mathbf{x}_2; \dots; \mathbf{x}_n]^{\boldsymbol{T}}$ (Wheeler, 2014).

2.6. Pembobot Spasial

Matriks pembobot spasial ($\boldsymbol{W}$) dapat diperoleh berdasarkan informasi jarak dari ketetanggaan (*neighborhood*) atau jarak antara satu lokasi pengamatan dengan lokasi pengamatan yang lain. Elemen matriks pembobot GWR, yaitu W_{ij} ditentukan berdasarkan kedekatan titik regresi i dengan titik pengamatan j. Titik pengamatan yang lebih dekat ke titik regresi diberi bobot lebih besar daripada titik pengamatan yang lebih jauh. Berdasarkan Fotheringham (2002) salah satu fungsi Kernel yang dapat dijadikan sebagai pembobot spasial dalam analisis spasial, yaitu fungsi *Adaptive Kernel Bisquare* seperti yang dituliskan pada persamaan berikut:

$$W_{ij}(u_i, v_i) = \begin{cases} \left(1 - \left(\dfrac{d_{ij}}{h_i}\right)^2\right)^2 & ;\ d_{ij} \le h_i \\ 0 & ;\ d_{ij} > h_i \end{cases} \qquad (2.\ 10)$$

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2} \qquad (2.\ 11)$$

dengan $i, j = 1,2, \dots, n$ dan d_{ij} merupakan jarak *euclidean* antar lokasi pengamatan dan h_i adalah radius dari titik pusat lokasi ke-i atau disebut juga *adaptive bandwith*.

Salah satu metode yang digunakan untuk menentukan *bandwidth* optimum adalah metode validasi silang (*cross validation* / CV) dan secara matematis dapat dituliskan sebagai berikut (Fotheringham, 2002).

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

dengan $\hat{y}_{\neq i}(h)$ adalah nilai estimasi y_i pada pengamatan di lokasi (u_i, v_i) dihilangkan dari proses estimasi. Untuk mendapatkan nilai h yang optimal, maka diperoleh dari h yang menghasilkan nilai CV yang minimum.

2.7. Pencilan

Pencilan adalah keganjilan yang menunjukkan sebuah data yang tidak mencirikan hal yang sama dengan data lainnya (Draper dan Smith, 1998). Fotheringham (2002) menyebutkan bahwa pengaruh pencilan merupakan masalah utama dalam regresi dan masalah untuk GWR. Jika pencilan dimasukkan ke dalam estimasi model, maka akan memberikan hasil yang kurang akurat. Sehingga, beberapa cara untuk mengidentifikasi pencilan akan menjadi alat yang berguna dalam GWR. Johnson dalam Lainun, dkk (2018) mendefinisikan pencilan sebagai pengamatan dalam kumpulan data yang tampaknya tidak konsisten dengan data lainnya.

Salah satu metode yang dapat digunakan untuk mengidentifikasi keberadaan pencilan, yaitu dengan metode *Boxplot* menggunakan nilai kuartil dan jangkauan untuk mendeteksi adanya pencilan. Kuartil 1, 2, dan 3 akan membagi data yang telah diurutkan sebelumnya menjadi empat bagian. Rentang Antar Kuartil atau *Interquartile Range* (IQR) didefinisikan sebagai selisih kuartil 1 terhadap kuartil 3, atau $IQR = Q_3 - Q_1$. Data dikatakan pencilan saat nilainya kurang dari $1{,}5 \times IQR$ terhadap Q_1 dan nilai yang lebih dari $1{,}5 \times IQR$ terhadap Q_3.

2.8. Regresi *Robust*

Regresi *robust* adalah metode regresi yang digunakan ketika data pengamatan mengandung pencilan (Lainun dkk, 2018). Metode ini penting untuk menganalisis data yang terkontaminasi oleh pencilan agar model yang dihasilkan kekar terhadap keberadaan pencilan. Regresi *robust* tetap menggunakan seluruh data, tetapi dengan memberikan bobot yang kecil untuk data pencilan (Soemartini, 2007). Andersen (2008) menyebutkan dua kriteria kekekaran suatu penduga yang patut diperhitungkan, yaitu *breakdown point* dan efesiensi estimator. *Breakdown point* adalah proporsi atau persentase data yang terkontaminasi yang masih dapat ditangani suatu estimator sebelum menghasilkan estimasi yang keliru (Andersen, 2008). Sedangkan, estimasi memiliki efisiensi yang cukup baik pada berbagai jika ragamnya mendekati ragam minimum untuk setiap sebaran ery & Peck, 1982). Pada regresi *robust* terdapat beberapa metode yang dapat digunakan, yakni estimasi *Least Median of Squares* (LMS), estimasi *Least Trimmed Squares* (LTS), estimasi S, estimasi M, dan estimasi MM (Chen, 2002). Sedangkan, menurut Wang dan Scott (1994) metode *Least Absolute Deviation* (LAD) adalah metode yang paling sederhana dan membuat fungsi kernel menjadi lebih *robust*.

2.9. Estimasi S

Estimasi S pertama kali diperkenalkan oleh Rousseeuw P. J. (1984), dan dinamakan estimasi S karena estimasi ini berdasarkan pada skala sisaan dari estimasi M. Regresi *robust* memiliki beberapa metode dalam mengestimasi, salah satunya, yaitu metode Estimasi S. Estimasi S merupakan estimasi yang dapat mencapai *breakdown point* sampai 50%. Sehingga, metode estimasi ini dapat mengatasi setengah dari pencilan.

Estimasi S didefinisikan sebagai $\hat{\beta}_S = min_\beta \hat{\sigma}_s(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ dengan menentukan nilai estimator skala *robust* ($\hat{\sigma}_s$) yang memenuhi:

$$\min \sum_{i=1}^{n} \rho(\varepsilon_i) \quad \text{dan} \quad \min \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ik}\beta}{\hat{\sigma}_S}\right) \qquad (2.\ 12)$$

dengan

$$\hat{\sigma}_s = \begin{cases} \dfrac{\text{median}|\varepsilon_i - \text{median}(\varepsilon_i)|}{0{,}6745}, & iterasi = 1 \\ \sqrt{\dfrac{1}{nK}\displaystyle\sum_{i=1}^{n} w_i \varepsilon_i^2}\,, & iterasi > 1 \end{cases} \qquad (2.\ 13)$$

dan $K = 0{,}199$ (Susanti, 2014).

Penyelesaian dari persamaan (2.12) adalah dengan cara mencari turunannya terhadap β sehingga diperoleh sebagai berikut:

$$\sum_{i=1}^{n} \rho'\left(\frac{y_i - \sum_{j=0}^{k} x_{ik}\beta}{\hat{\sigma}_s}\right) = 0$$

$$\sum_{i=1}^{n} x_i \psi\left(\frac{y_i - \sum_{j=0}^{k} x_{ik}\beta}{\hat{\sigma}_s}\right) = 0 \qquad (2.\ 14)$$

dengan $k = 1,2, \dots, p$ dan ψ disebut fungsi *influence* yang merupakan turunan dari

), turunan dari fungsi ρ adalah

$$\psi(\varepsilon_i^*) = \rho'(\varepsilon_i^*) = \begin{cases} \varepsilon_i^* \left(1 - \left(\dfrac{\varepsilon_i^*}{c}\right)^2\right)^2, & |\varepsilon_i^*| \le c \\ 0, & |\varepsilon_i^*| > c \end{cases}$$

dengan w_i merupakan fungsi pembobot *Tukey's bisquare*.

$$w_i = \begin{cases} \begin{cases} \left[1 - \left[\frac{\varepsilon_i^*}{c}\right]^2\right]^2, & |\varepsilon_i^*| \leq c \\ 0, & |\varepsilon_i^*| \leq c \end{cases}, & iterasi = 1 \\ \frac{\rho(\varepsilon_i^*)}{(\varepsilon_i^*)^2}, & iterasi > 1 \end{cases}$$

dengan $\varepsilon_i^* = \frac{\varepsilon_i}{\sigma}$ dan $c = 1{,}547$. Dengan demikian dari persamaan (2.14) akan dihitung dengan *Iteratively Reweighted Least Square* (IRLS) sehingga bisa mencapai konvergen.

2.10. Estimasi M

Menurut Chen (2002) regresi *robust* memiliki beberapa metode dalam mengestimasi, salah satunya adalah metode M (*Maximum Likelihood Type*). Metode ini merupakan metode yang baik dalam perhitungan maupun secara teoritis. Metode ini diperkenalkan oleh Huber pada tahun 1973. Dimana dalam metode ini menganalisis data dengan mengasumsikan bahwa sebagian besar pencilan yang terdeteksi berada pada variabel prediktor.

Estimasi M dicari berdasarkan konsep dari estimasi *maksimum likelihood* yang memaksimumkan fungsi *likelihood* (Maronna dkk, 2006). Misalkan adalah suatu fungsi untuk ε_i^* dan σ dengan $\varepsilon_i^* = \frac{y_i - \sum_{j=0}^{k} x_{ik}\beta}{\hat{\sigma}_M}$. Menurut Montgomery dan Peck (2006), pada prinsipnya estimasi M merupakan estimasi yang meminimumkan suatu fungsi galat ρ sebagai berikut:

$$\hat{\beta}_M = \min \sum_{i=1}^{n} \rho\left(\frac{\varepsilon_i}{\hat{\sigma}_M}\right) = \min \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ik}\beta}{\hat{\sigma}_M}\right) \qquad (2.\ 15)$$

Untuk memperoleh persamaan (2.15), dengan menyelesaikan persamaan:

$$\min \sum_{i=1}^{n} \rho(\varepsilon_i^*) = \min \sum_{i=1}^{n} \rho\left(\frac{\varepsilon_i}{\hat{\sigma}_M}\right) = \min \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ik}\beta}{\hat{\sigma}_M}\right) \qquad (2.\ 16)$$

dengan dipilih estimator untuk σ adalah

$$\hat{\sigma}_M = \frac{\text{MAD}}{0{,}6745} = \frac{\text{median}|\varepsilon_i - \text{median}(\varepsilon_i)|}{0{,}6745}$$

Pemilihan konstanta $0{,}6745$ membuat $\hat{\sigma}$ suatu penduga yang mendekati tak bias ...a n besar dan galat berdistribusi normal (Montgomery dan Peck, 2006). ... adalah fungsi yang memberikan kontribusi pada masing-masing galat pada fungsi objektif. Untuk meminimumkan persamaan (2.11), dicari turunan parsial pertama dari $\hat{\beta}_M$ terhadap β, sehingga diperoleh persamaan:

$$\sum_{i=1}^{n} x_i \psi \left(\frac{y_i - \sum_{j=0}^{k} x_{ik}\beta}{\hat{\sigma}_M} \right) = 0 \qquad (2.\ 17)$$

dengan $\psi = \rho'$ dan x_{ik} adalah pengamatan ke-i pada variabel bebas ke-k.
Draper dan Smith (1998) memberikan penyelesaian persamaan 2.17, yaitu dengan mendifinisikan suatu fungsi pembobot:

$$\omega(\varepsilon_i) = \frac{\psi\left(\frac{y_i - \sum_{j=0}^{k} x_{ik}\beta}{\hat{\sigma}_M} \right)}{\left(\frac{y_i - \sum_{j=0}^{k} x_{ik}\beta}{\hat{\sigma}_M} \right)} \qquad (2.\ 18)$$

Karena $\varepsilon_i^* = \frac{\varepsilon_i}{\sigma}$ sebagai pengganti ε_i, maka persamaan (2.18) menjadi

$$w_i(\varepsilon_i^*) = \frac{\psi(\varepsilon_i^*)}{\varepsilon_i^*} = \begin{cases} \varepsilon_i^* \left(1 - \left(\frac{\varepsilon_i^*}{c} \right)^2 \right)^2 & , |\varepsilon_i^*| \le c \\ 0 & , |\varepsilon_i^*| > c \end{cases}$$

$$w_i(\varepsilon_i^*) = \begin{cases} \left| 1 - \left(\frac{\varepsilon_i^*}{c} \right)^2 \right|^2 & , |\varepsilon_i^*| \le c \\ 0 & , |\varepsilon_i^*| > c \end{cases}$$

Untuk fungsi pembobot *Tukey's Bisquare*, konstanta yang digunakan adalah $c = 4{,}685$.

2.11. Estimasi MM

Estimasi MM menggabungkan estimasi *high breakdown point* dan efisiensi statistik yang dikenalkan oleh Yohai tahun 1987. Langkah pertama dalam estimasi ini adalah mencari estimator dengan Estimasi S, kemudian menetapkan parameter-parameter regresi menggunakan Estimasi M. Estimasi S menjamin nilai *breakdown point* tinggi dan Estimasi M membuat estimator mempunyai efisiensi tinggi. Pada umumnya digunakan fungsi *Tukey Bisquare* baik pada Estimasi S maupun Estimasi M. Bentuk dari metode Estimasi MM (Yohai, 1987).

$$\min \sum_{i=1}^{n} \rho \left(\frac{\varepsilon_i}{\hat{\sigma}_S} \right) \qquad (2.\ 19)$$

stimasi MM bertujuan untuk mendapatkan nilai estimasi dengan nilai *vn* tinggi dan lebih efisien.

2.12. *Least Absolute Deviation*

Least Absolute Deviation (LAD) dikembangkan pertama kali oleh Roger Joseph Boscovich pada tahun 1957. Apabila d_{ij} adalah jarak (u_i, v_i) antar lokasi pengamatan ke-i, maka koefisien regresi pada persamaan (2.5) diestimasi menggunakan metode LAD dapat dilakukan dengan meminimumkan jumlah mutlak galat yang dinyatakan sebagai berikut (Zhang dan Mei, 2011).

$$\min \sum_{i=1}^{n} |\varepsilon_i|\, W_i \tag{2. 20}$$

dengan $i = 1,2, \dots, n$ dan

$$W_{ij} = \begin{cases} \left(1 - \left(\dfrac{d_{ij}}{h_i}\right)^2\right)^2 & ;\ d_{ij} \le h_i \\ 0 & ;\ d_{ij} > h_i \end{cases}$$

Apabila dilakukan dekomposisi pada galat ε_i, maka

$$\varepsilon_i = \varepsilon_i^+ - \varepsilon_i^-$$

$$|\varepsilon_i| = \varepsilon_i^+ + \varepsilon_i^-$$

dengan ε_i^+ adalah galat yang nilainya lebih dari atau sama dengan nol dan ε_i^- adalah galat yang nilainya kurang dari nol. Sehingga, ε_i^+ dan ε_i^- dapat dinyatakan sebagai $\varepsilon_i^+ = \varepsilon_i I(\varepsilon_i \ge 0)$ dan $\varepsilon_i^- = \varepsilon_i I(\varepsilon_i < 0)$. Adapun $I(A)$ merupakan fungsi indikator I dari himpunan A dengan $I(A) = \begin{cases} 1 & , \varepsilon_i \epsilon A \\ 0 & , lainnya \end{cases}$.

Solusi untuk menghasilkan parameter tidak dapat dilakukan dengan proses diferensiasi seperti pada metode WLS. Menurut Wagner (1959) penyelesaian solusi dari regresi dengan LAD dapat diselesaikan dengan menggunakan algoritma simpleks.

2.13. Algoritma Simpleks

Algoritma simpleks adalah algoritma yang dikembangkan oleh Barrodale dan Robert pada tahun 1974. Algoritma simpleks memberikan solusi permasalahan optimasi linier yang melibatkan beberapa variabel keputusan dengan bantuan komputasi (Davino, 1998). Optimasi linier adalah suatu cara/teknik aplikasi matematika untuk menyelesaikan persoalan pengalokasian ... umber terbatas di antara beberapa aktivitas yang bertujuan untuk ... mumkan keuntungan atau meminimumkan biaya yang dibatasi oleh ... batasan tertentu, atau dikenal juga dengan teknik optimalisasi. Adapun istilah-istilah yang terdapat dalam algoritma simpleks, yaitu (Rafflesia dan Widodo, 2014):

1. Fungsi tujuan merupakan fungsi yang menggambarkan tujuan atau sasaran dalam permasalahan program linier yang berkaitan dengan pemanfaatan sumber daya secara optimal untuk memperoleh keuntungan maksimum atau untuk penggunaan biaya minimum.
2. Fungsi kendala/pembatas merupakan bentuk rumusan terhadap kendala yang dihadapi dalam mencapai tujuan.
3. Iterasi adalah tahapan perhitungan dengan nilai dalam perhitungan itu tergantung dari nilai tabel sebelumnya.
4. Variabel non basis adalah variabel yang nilainya diatur menjadi nol pada sembarang iterasi. Dalam terminologi umum, jumlah variabel non basis selalu sama dengan derajat bebas dalam sistem persamaan.
5. Variabel basis pada solusi awal, variabel basis merupakan variabel *slack* (jika fungsi kendala merupakan pertidaksamaan ≤) atau variabel *artificial* (jika fungsi kendala menggunakan pertidaksamaan ≥ atau =). Secara umum, jumlah variabel basis selalu sama dengan jumlah fungsi kendala.
6. Variabel *slack* adalah variabel yang ditambahkan ke fungsi kendala untuk mengkonversikan pertidaksamaan ≤ menjadi persamaan (=). Pada solusi awal, variabel *slack* akan berfungsi sebagai variabel basis.
7. Variabel *surplus* adalah variabel yang dikurangkan dari model matematik kendala untuk mengkonversikan pertidaksamaan ≥ menjadi persamaan (=). Pada solusi awal, variabel *surplus* tidak dapat berfungsi sebagai variabel basis. Dalam kasus regresi menggunakan metode LAD, variabel *surplus* adalah deviasi bawah.
8. Variabel *artificial* adalah variabel yang ditambahkan ke model matematik kendala dengan bentuk ≥ atau = untuk difungsikan sebagai variabel basis awal. Dalam kasus regresi menggunakan metode LAD variabel *artificial* adalah deviasi atas.
9. Kolom pivot (kolom kerja) adalah kolom yang memuat variabel masuk. Koefisien pada kolom ini akn menjadi pembagi nilai kanan untuk menentukan baris pivot (baris kerja).

pivot (baris kerja) adalah salah satu baris dari antara variabel basis yang uat variabel keluar.

11. Elemen pivot (elemen kerja) adalah elemen yang terletak pada perpotongan kolom dan baris pivot. Elemen pivot akan menjadi dasar perhitungan untuk tabel simpleks berikutnya.
12. Variabel masuk adalah variabel yang terpilih untuk menjadi variabel basis pada iterasi berikutnya. Variabel masuk dipilih satu dari antara variabel non basis pada setiap iterasi.
13. Variabel keluar adalah variabel yang keluar dari variabel basis pada iterasi berikutnya dan digantikan oleh variabel masuk. Variabel keluar dipilih satu dari antara variabel basis pada setiap iterasi.
14. Solusi atau nilai kanan merupakan nilai sumber daya pembatas yang masih tersedia. Pada solusi awal, nilai kanan atau solusi sama dengan jumlah sumber daya pembatas awal yang ada, karena aktivitas belum dilaksanakan.

Algoritma simpleks memerlukan sebuah tabel simpleks atau yang biasa dikenal dengan tabulasi simpleks seperti Tabel 2.1.

Tabel 2. 1. Tabel Simpleks

c_j			0	0	...	0	d_{11}^+	...	d_{1n}^+	d_{11}^-	...	d_{1n}^-
c_b	v_b	w_b					x_1		x_n			
d_{11}^+	x_1	b_1	a_{ij}									
d_{11}^+	x_2	b_2										
⋮	⋮	⋮										
d_{11}^+	x_n	b_n										
z_j												
$c_j - z_j$												

Sumber: Khairunisa dalam Puteri, 2019.

dengan d_i^+ merupakan pembobot deviasi atas dan d_i^- deviasi bawah. Pengisian Tabel 2.1 diuraikan sebagai berikut.

1. Baris c_j diisi dengan koefisien variabel yang menjadi non basis.
2. Baris c_b diisi dengan koefisien variabel yang menjadi basis.
3. Baris v_b diisi dengan variabel yang menjadi basis (variabel yang menyusun matriks identitas). Dalam hal ini diisi dengan variabel *artificial*, yaitu deviasi atas.
4. Baris w_b diisi dengan nilai ruas kanan dari kendala.
5. Baris z_j diisi dengan rumus $z_j = \sum_{j=1}^{n} d_i a_{ij}$.

dapun proses algoritma simpleks sebagai berikut.

gubah terlebih dahulu masalah optimasi linier ke bentuk standar, yang n hal ini fungsi tujuan dan kendala-kendala diubah ke dalam bentuk amaan. Seperti yang sudah dijelaskan di atas, dengan menambahkan variable *slack*, variable *surplus*, dan variable *artificial* terhadap kendala yang berbentuk pertidaksamaan.

2. Menentukan kolom pivot (variabel masuk), yaitu untuk masalah maksimum memilih $c_j - z_j$ yang terbesar, sedangkan untuk masalah minimum memilih $c_j - z_j$ yang terkecil.
3. Menentukan baris pivot (variabel keluar), yaitu dari nilai rasio antara nilai ruas kiri (b_i) dengan koefisien kolom pivot (a_{ij}), pilih yang terkecil (untuk masalah minimum atau maksimum). Rasio =$\frac{b_i}{aij}$, dengan a rasio > 0.
4. Menentukan pivot dari perpotongan antara kolom pivot dan baris pivot yang dinamakan elemen pivot atau elemen penentu iterasi algoritma simpleks dan akan diubah nilainya menjadi 1.
5. Melakukan operasi baris dasar (OBD) berdasarkan pivot untuk baris lainnya, termasuk baris $c_j - z_j$ dengan nilai elemen-elemen yang termasuk di dalam kolom pivot dijadikan nol (selain elemen yang dijadikan pivot).

Proses iterasi untuk masalah maksimum berhenti jika semua nilai pada baris $c_j - z_j \leq 0$, berarti solusi sudah optimal. Apabila masih ada $c_j - z_j > 0$ (positif), maka iterasi algoritma simpleks masih berlanjut. Untuk masalah minimum berhenti jika semua nilai psada baris $c_j - z_j \geq 0$. Apabila masih ada $c_j - z_j < 0$ (negatif), maka iterasi algoritma simpleks masih berlanjut (Khairunisa dalam Puteri, 2019).

2.14. *Robust Geographically Weighted Regression*

Model *Geographically Weighted Regression* yang dihasilkan menggunakan metode estimasi MM dan LAD kemudian disebut sebagai model *Robust Geographically Weighted Regression* (RGWR). Model yang digunakan pada RGWR sama dengan model yang digunakan pada model GWR. Begitu pula pada fungsi matriks pembobot yang digunakan, yang membedakan adalah pada kriteria yang digunakan pada pemilihan *bandwidth* yang optimum (Zhang dan Mei, 2011). Kriteria pemilihan *bandwidth* optimum pada RGWR dapat dilakukan dengan prosedur kriteria *Absolute Cross Validation* (ACV). Menurut Wang dan Scott (1994), nilai ACV tidak terpengaruh oleh keberadaan pencilan dan akibatnya skor ACV lebih kekar daripada nilai CV. Kriteria ACV menggunakan nilai mutlak dari [illegible]riabel respon dan nilai estimasi $\hat{y}_{\neq i}(h)$ yang dirumuskan sebagai berikut.

$$ACV(h) = \sum_{i=1}^{n} |y_i - \hat{y}_{\neq i}(h)|$$

Nilai optimum dari *bandwith* (h) dapat dipilih dengan cara yang sama dengan kriteria pada CV, yaitu dengan memilih nilai h yang menghasilkan ACV terkecil.

2.15. Pemilihan Model Terbaik

Hartono dalam Hermanto dan Rizqika (2019) menyebutkan di antara kriteria yang dapat digunakan dalam pemilihan model terbaik, yaitu dengan kriteria *Mean Squares of Error* dan Koefisien Determinasi ($\mathbf{R}^2$).

2.15.1. *Mean Squares of Error* (MSE)

Model regresi yang dipilih dari metode ini adalah model regresi yang menghasilkan nilai MSE paling kecil.

$$MSE = \frac{JKG}{df_{galat}} \tag{2. 21}$$

dengan

JKG : Jumlah Kuadrat Galat

df_{galat} : derajat kebebasan galat

2.15.2. Koefisien Determinasi (R^2)

Pengujian Kesesuaian (*Goodness of Fit*) dilakukan dengan menghitung koefisien determinasi (R^2) model. Menurut Gujarati (1993), besaran koefisien determinasi (R^2) merupakan besaran yang paling lazim digunakan untuk mengukur kecocokan model (*goodness of fit*) garis regresi. Nilai R^2 GWR didapatkan dengan persamaan matematis sebagai berikut (Fotheringham, 2002).

$$R^2 = \frac{JKR}{JKT} \times 100\% = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \times 100\% \tag{2. 22}$$

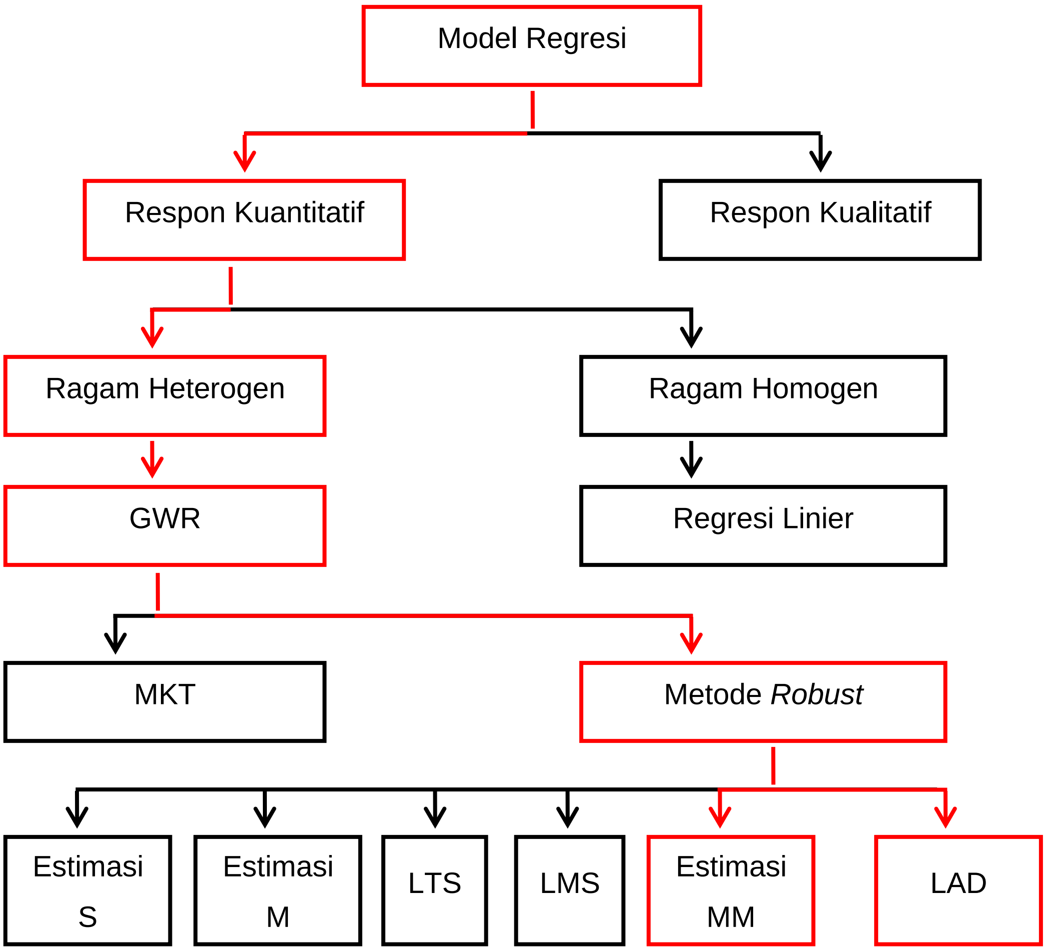
dengan

JKR : Jumlah Kuadrat Regresi

JKT : Jumlah Kuadrat Total

2.16 Kerangka Konsep

Kerangka konsep pada penelitian memiliki kaitan atau hubungan antara konsep satu dengan konsep yang lainnya dari masalah yang ingin diteliti. Kerangka konsep diperoleh dari konsep ilmu atau teori yang dipakai sebagai landasan penelitian. Adapun gambaran kerangka konsep dalam penelitian ini seperti pada Gambar 2.1 berikut.



Gambar 2. 1 Kerangka Konsep