PERBANDINGAN METODE *ROBUST* *LEAST MEDIAN OF SQUARE* (LMS) DAN PENDUGA S UNTUK MENANGANI *OUTLIER* PADA REGRESI LINIER BERGANDA

Skripsi

disusun sebagai salah satu syarat

untuk memperoleh gelar Sarjana Sains

Program Studi Matematika

oleh

Laeli Sidik Febrianto

4111412053

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI SEMARANG

2016

PERNYATAAN KEASLIAN TULISAN

Saya menyatakan bahwa skripsi ini bebas plagiat, kecuali yang secara tertulis dirujuk dalam skripsi ini dan disebutkan dalam daftar pustaka. Apabila dikemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya bersedia menerima sanksi sesuai ketentuan perundang-undangan.

Semarang, Juni 2016

Laeli Sidik Febrianto
4111412053

PENGESAHAN

Skripsi yang berjudul

Perbandingan Metode *Robust Least Median of Square* (LMS) dan Penduga S untuk Menangani *Outlier* pada Regresi Linier Berganda.

disusun oleh

Laeli Sidik Febrianto

4111412053

telah dipertahankan di hadapan sidang Panitia Ujian Skripsi FMIPA Universitas Negeri Semarang pada tanggal 2 Juni 2016

Panitia:

Ketua

Prof. Dr. Zaenuri, S.E., M.Si., Akt.
NIP. 196412231988031001

Sekretaris

Drs. Arief Agoestanto, M.Si.
NIP. 196807221993031005

Ketua Penguji

Drs. Sugiman, M.Si.
NIP. 196401111989011001

Anggota Penguji /
Pembimbing I

Dr. Nur Karomah Dwidayati, M.Si.
NIP. 196605041990022001

Anggota Penguji /
Pembimbing II

Putriaji Hendikawati, S.Si., M.Pd., M.Sc.
NIP. 198208182006042001

MOTTO DAN PERSEMBAHAN

MOTTO

- Man Jadda Wajada (Umar bin Abd. Aziz)
- Dan (ingatlah juga), tatkala Tuhanmu memaklumkan; "Sesungguhnya jika kamu bersyukur, pasti Kami akan menambah (nikmat) kepadamu, dan jika kamu mengingkari (nikmat-Ku), maka sesungguhnya azab-Ku sangat pedih". (QS. 'Ibrahim [14] : 7)
- Orang-orang yang sukses telah belajar membuat diri mereka melakukan hal yang harus dikerjakan ketika hal itu memang harus dikerjakan, entah mereka menyukainya ataupun tidak. (Aldus Huxley)
- Jika dirimu sendiri tidak mampu menjadi motivasi untukmu meraih sukses, jadikan orang tuamu sebagai motivasi untukmu meraih sukses.

PERSEMBAHAN

- Untuk kedua orang tua tercinta, Ibu Sukiyem dan Bapak Suratno
- Untuk Adikku tersayang, Dwi Aditya
- Untuk keluarga besar tercinta
- Untuk Universitas Negeri Semarang

KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah memberikan nikmat dan karunia-Nya serta kemudahan sehingga penulis dapat menyelesaikan skripsi yang berjudul "Perbandingan Metode *Robust Least Median of Square* (LMS) dan Penduga S untuk Menangani *Outlier* pada Regresi Linier Berganda".

Penyusunan skripsi ini dapat diselesaikan berkat kerjasama, bantuan, dan dorongan dari berbagai pihak. Oleh karena itu penulis mengucapkan terima kasih kepada:

1. Prof. Dr. Fathur Rokhman, M.Hum., Rektor Universitas Negeri Semarang.
2. Prof. Dr. Zaenuri, S.E., M.Si., Akt., Dekan FMIPA Universitas Negeri Semarang.
3. Drs. Arief Agoestanto, M.Si., Ketua Jurusan Matematika FMIPA Universitas Negeri Semarang.
4. Drs. Mashuri, M.Si., Ketua Prodi Matematika FMIPA Universitas Negeri Semarang.
5. Dr. Nur Karomah Dwidayati, M.Si., selaku Dosen Pembimbing I yang telah memberikan bimbingan, pengarahan, nasehat, dan saran selama penyusunan skripsi ini.
6. Putriaji Hendikawati, S.Si., M.Pd., M.Sc., selaku Dosen Pembimbing II yang telah memberikan bimbingan, pengarahan, nasehat, dan saran selama penyusunan skripsi ini.

7. Drs. Sugiman, M.Si., selaku Dosen Penguji yang telah memberikan penilaian dan saran dalam perbaikan skripsi ini.
8. Drs. Mashuri, M.Si., selaku Dosen Wali saya yang telah memberikan bimbingan dan arahan.
9. Dosen-dosen Matematika Universitas Negeri Semarang yang telah membekali penulis dengan berbagai ilmu selama mengikuti perkuliahan sampai akhir penulisan skripsi ini.
10. Ibu dan Bapak tercinta, Ibu Sukiyem dan Bapak Suratno yang senantiasa memberikan dukungan dan doa yang tiada putusnya.
11. Adik tersayang, Dwi Aditya yang selalu memberikan semangat dan doa.
12. Bidikmisi Universitas Negeri Semarang yang telah memberikan dukungan secara materiil maupun non-materiil.
13. Teman-teman HIMATIKA dan MSC yang telah memberikan banyak pengalaman organisasi.
14. Sahabat dan teman-teman Jurusan Matematika FMIPA Unnes.
15. Semua pihak yang tidak dapat disebutkan satu per satu yang telah memberikan bantuan.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kekurangan. Oleh karena itu, penulis mengharapkan saran dan kritik yang membangun dari pembaca.

Semarang, Juni 2016

Penulis

ABSTRAK

Febrianto, Laeli Sidik. 2016. *Perbandingan Metode Robust Least Median of Square (LMS) dan Penduga S untuk Menangani Outlier pada Regresi Linier Berganda*. Skripsi, Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang. Pembimbing Utama: Dr. Nur Karomah Dwidayati, M.Si. dan Pembimbing Pendamping: Putriaji Hendikawati, S.Si., M.Pd., M.Sc.

Kata Kunci: *Outlier*, Metode *Robust*, LMS, Penduga S.

Analisis regresi linier berganda digunakan untuk mengukur pengaruh lebih dari satu variabel bebas (X) terhadap variabel terikat (Y). Estimasi parameter analisis regresi umumnya diselesaikan dengan *Ordinary Least Square* (OLS). Pada kenyataannya banyak ditemukan kasus bahwa data mengandung *outlier* yang menyebabkan estimasi koefisien garis regresi dengan OLS menjadi tidak tepat, sehingga diperlukan metode regresi *robust*. *Least Median of Square* (LMS) dan Penduga S merupakan metode-metode dalam regresi *robust*. Permasalahan yang dikaji dalam penelitian ini adalah menentukan metode terbaik dalam mengatasi permasalah *outlier*.

Penelitian ini menggunakan simulasi dengan data rekap Anggaran Pendapatan dan Belanja Daerah (APBD) kabupaten/kota di Pulau Jawa tahun 2010 dengan variabel bebas meliputi Pendapatan Asli Daerah (X_1), Dana Bagi Hasil (X_2), Dana Alokasi Umum (X_3), Luas Wilayah (X_4), dan variabel terikat yaitu Belanja Modal (Y). Analisis dimulai dengan uji asumsi normalitas, linieritas, keberartian simultan, keberartian parsial, multikolinearitas, heteroskedastisitas, dan autokorelasi. Model regresi yang dapat diterima yaitu regresi data transformasi logaritma dari data APBD dengan variabel bebas meliputi Pendapatan Asli Daerah ($logX_1$) dan Dana Bagi Hasil ($logX2$), serta variabel terikat yaitu Belanja Modal ($logY$). Pendeteksian *outlier* menggunakan metode *boxplot* dan *Cook's Distance* menunjukan bahwa terdapat *outlier*, sehingga dilakukan pendugaan parameter regresi *robust* dengan metode LMS dan Penduga S. Metode LMS menghasilkan nilai AIC sebesar 25,54423 dan SIC sebesar 27,76414, sedangkan dengan metode Penduga S menghasilkan nilai AIC sebesar 40,22523 dan SIC sebesar 43,72099.

Penentuan metode terbaik dengan membandingkan nilai AIC dan SIC. Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan bahwa LMS merupakan metode regresi *robust* terbaik dibandingkan metode Penduga S, karena metode LMS memiliki nilai AIC dan SIC yang lebih kecil dibandingkan dengan metode Penduga S.

DAFTAR ISI

Halaman

DAFTAR TABEL

DAFTAR GAMBAR

Gambar Halaman

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1 Latar Belakang

Statistika memegang peranan penting dalam memecahkan masalah yang terjadi pada berbagai macam bidang. Seperti bidang ekonomi, kependudukan, kesehatan, dan kemiliteran. Adanya permasalahan-permasalahan yang terjadi pada bidang-bidang tersebut, maka statistikawan berusaha memberikan solusi berupa suatu hasil analisis yang berkualitas yang pada akhirnya dapat digunakan untuk pengambilan keputusan.

Analisis regresi memiliki beberapa kegunaan (Draper dan Smith, 1992), diantaranya untuk tujuan deskripsi dari fenomena data atau kasus yang sedang diteliti, untuk tujuan kontrol, dan sebagai prediksi. Regresi mampu mendeskripsikan fenomena data melalui terbentuknya suatu model hubungan yang bersifat numerik. Regresi juga dapat digunakan untuk melakukan pengendalian (kontrol) terhadap suatu kasus atau hal-hal yang sedang diamati melalui penggunaan model regresi yang diperoleh. Selain itu, model regresi juga dapat dimanfaatkan untuk melakukan prediksi variabel terikat. Analisis Regresi Linier Berganda digunakan untuk mengukur pengaruh antara lebih dari satu variabel prediktor (variabel bebas) terhadap variabel terikat.

Ordinary Least Square (OLS) (Draper dan Smith, 1992) merupakan salah satu metode yang sering digunakan untuk mendapatkan nilai-nilai penduga parameter dalam pemodelan regresi. Penggunaan OLS memerlukan beberapa asumsi klasik yang harus dipenuhi oleh komponen residual atau galat dalam model yang dihasilkan. Beberapa asumsi itu antara lain: (1) Residual mengikuti distribusi normal, (2) varians dari residual adalah konstan dan homoskedastisitas, (3) tidak ada autokorelasi, (4) tidak ada multikolinearitas di antara variabel bebas.

Jika asumsi-asumsi klasik dalam metode OLS terpenuh maka penduga parameter yang diperoleh bersifat *Best Linear Unbiased Estimasi* (BLUE). Pada berbagai kasus tidak jarang ditemui hal-hal yang menyebabkan tidak terpenuhinya asumsi klasik. Salah satu penyebabnya adalah adanya pencilan (*outlier*) dalam data amatan.

Data *outlier* (Makkulau *et al*., 2010) adalah data pengamatan yang berada jauh (ekstrim) dari pengamatan-pengamatan lainnya. *Outlier* mungkin ada karena adanya data terkontaminasi, yaitu adanya kesalahan pada saat melakukan pengambilan sampel pada populasi. *Outlier* yang disebabkan oleh data terkontaminasi dapat dihapuskan dari data penelitian atau jika memungkinkan dapat dilakukan sampling ulang. Jika setelah dilakukan beberapa sampling ulang namun data *outlier* tetap muncul maka data tersebut tidak dapat dihapuskan dari data penelitian, karena analisis data yang dihasilkan akan tidak mencerminkan populasi yang diteliti.

Menurut Sembiring (Paludi, 2009:57), *outlier* adalah pengamatan yang jauh dari pusat data yang mungkin berpengaruh besar terhadap koefisien regresi. Keberadaan data *outlier* akan mengganggu dalam proses analisis data. Kaitannya dalam analisis regresi, *outlier* dapat menyebabkan hal-hal berikut.

1. Residual yang besar dari model yang terbentuk atau $E(e) \neq 0$
2. Varians pada data tersebut menjadi lebih besar
3. Taksiran interval memiliki rentang yang lebar.

Pendeteksian *outlier* merupakan tahapan yang perlu dilakukan terutama jika estimasi modelnya dengan OLS, yang dikenal cukup peka terhadap *outlier*. Pendeteksian *outlier* dapat dilakukan dengan beberapa metode diantaranya dengan metode *Boxplot* dan metode *Cook's Distance.* Metode *Boxplot* merupakan metode yang mempergunakan nilai kuartil dan jangkauan untuk mendeteksi *outlier*, sehingga pada metode ini dapat mengetahui adanya *outlier* untuk masing-masing variabel. Sedangkan menggunakan metode *Cook's Distance* dapat mengetahui adanya *outlier* secara simultan pada variabel bebas.

Saat ada asumsi yang tidak terpenuhi, maka penggunaan metode OLS akan memberikan kesimpulan yang bersifat kurang baik atau nilai penduga parameternya bersifat bias sehingga berakibat interpretasi hasil yang diperoleh menjadi tidak valid (Nurcahyadi, 2010). Oleh karena itu, saat asumsi klasik tidak terpenuhi maka metode OLS perlu dihindari. Untuk mengatasinya diperlukan metode lain supaya analisis data dengan adanya data *outlier* tetap tahan (*robust*) terhadap asumsi yang diterapkan pada analisis datanya. Metode tersebut dikenal dengan metode *robust*.

Regresi *robust* diperkenalkan oleh Andrews (1972), yaitu metode regresi yang digunakan ketika distribusi dari residual tidak normal atau adanya beberapa *outlier* yang berpengaruh pada model (Wijayanti, 2015). Metode ini merupakan alat penting untuk menganalisa data yang dipengaruhi oleh *outlier* sehingga dihasilkan model yang *robust* atau *resistance* terhadap *outlier*. Menurut Chen (2014) regresi *robust* terdiri dari 5 metode penduga, yaitu estimasi *robust* M, estimasi *robust least median of square* (LMS), estimasi *robust least trimmed square* (LTS), estimasi *robust* S dan estimasi *robust* MM.

Penggunaan metode LMS data *outlier* yang ada tidak dibuang begitu saja, tetapi diproses dan dieliminasi melalui sebuah iterasi. Metode ini mempunyai keuntungan untuk mengurangi pengaruh dari residual terhadap keakuratan koefisien regresi. Menggunakan median dari kuadrat residual, penduga yang dihasilkan akan lebih kekar dalam menghadapi *outlier*.

Pada penelitian Oktarinanda (2014) mengenai perbandingan efisiensi metode LTS dan metode LMS dalam estimasi parameter regresi *robust.* Perbandingan keakuratan model menggunakan koefisien determinasi dan RMSE diperolehkan kesimpulan bahwa data yang digunakan dalam penelitian lebih sesuai menggunakan penduga LMS dalam menduga parameter regresi. Selain itu, parameter duga yang dihasilkan LMS relatif lebih efisien daripada LTS karena ragam parameter duga dari metode LMS lebih kecil daripada LTS. Dengan kata lain metode LMS lebih efisien.

Penduga S bertujuan untuk memperoleh penduga dengan nilai simpangan baku terkecil. Pendugaan parameter dengan Penduga S dapat menghasilkan penduga yang bersifat *robust* terhadap *outlier* berpengaruh. Hasil penelitian Permana (2014) menunjukan metode Penduga S merupakan metode yang lebih baik dibandingkan metode *Least Trimmed Square* (LTS) untuk menangani *outlier* pada regresi karena memiliki nilai *Mean Square Error* (MSE) lebih kecil.

Penelitian ini difokuskan pada metode estimasi parameter dengan menggunakan metode *Robust* LMS dan Penduga S. Dari kedua model regresi *robust* yang dihasilkan dari kedua metode tersebut diperoleh metode terbaik berdasarkan nilai AIC dan SIC terkecil. Pada penelitian ini menggunakan bantuan *software* SPSS16 dan SAS 9.1.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka permasalahan yang dikaji dalam penelitian ini adalah:

1. Bagaimana hasil estimasi regresi linier berganda metode *robust* LMS dan Penduga S pada data Anggaran Pendapatan dan Belanja Daerah (APBD) kabupaten/kota di Pulau Jawa tahun 2010?
2. Manakah metode regresi *robust* terbaik di antara metode LMS dan metode Penduga S?

1.3 Batasan Masalah

Batasan masalah yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Identifikasi *outlier* hanya menggunakan metode *boxplot* dan *Cook's distance*.
2. Penelitian hanya menggunakan metode LMS dan metode Penduga S?
3. Paket program yang mendukung penelitian adalah SPSS16 dan SAS 9.1.

1.4 Tujuan Penelitian

Berdasarkan permasalahan yang dikaji, penelitian ini mempunyai tujuan:

1. Memperoleh hasil estimasi regresi linier berganda metode *robust* LMS dan Penduga S pada data Anggaran Pendapatan dan Belanja Daerah (APBD) kabupaten/kota di Pulau Jawa tahun 2010.
2. Memperoleh metode regresi *robust* terbaik di antara metode LMS dan metode Penduga S.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dalam penelitian ini diantaranya :

1.5.1 Bagi Mahasiswa

Bagi mahasiswa manfaat dari penelitian ini adalah agar dapat:

1. Memperoleh pengetahuan tentang data *outlier*.
2. Memperoleh pengetahuan mengenai prosedur untuk memperoleh hasil estimasi regresi linier berganda metode *robust* LMS dan Penduga S pada data Anggaran Pendapatan dan Belanja Daerah (APBD) kabupaten/kota di Pulau Jawa tahun 2010.
3. Menggunakan model regresi *robust* terbaik di antara metode LMS dan metode Penduga S untuk keperluan peramalan.

1.5.2 Bagi Pembaca

Bagi pembaca manfaat dari penelitian ini adalah agar dapat:

1. Menambah atau memperkaya khasanah kepustakaan Jurusan Matematika.
2. Menambah topik kajian tentang metode LMS dan metode Penduga S.
3. Meramalkan data yang mengandung *outlier* menggunakan model regresi *robust* tanpa membuang data *outlier* yang ada.

1.6 Sistematika Penulisan

Secara garis besar skripsi ini dibagi menjadi tiga bagian yaitu bagian awal skripsi, bagian isi skripsi, dan bagian akhir skripsi. Berikut ini dijelaskan masing-masing bagian skripsi.

1.6.1 Bagian awal skripsi

Bagian awal skripsi meliputi halaman judul, abstrak, halaman pengesahan, halaman motto dan persembahan, kata pengantar, daftar isi, daftar gambar, daftar tabel, dan daftar lampiran.

1.6.2 Bagian isi skripsi

Bagian isi terdiri dari lima bab. Adapun lima bab tersebut sebagai berikut.

Bab 1 Pendahuluan

Pada bab Pendahuluan dikemukakan tentang alasan pemilihan judul, permasalahan, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan skripsi.

Bab 2 Landasan Teori

Pada bab Landasan Teori dikemukakan konsep-konsep yang dijadikan landasan teori seperti regresi linier berganda, uji asumsi regresi linier berganda, *outlier*, deteksi *outlier*, regresi *robust*, dan kriteria pemilihan motode terbaik. Selain itu, penelitian terdahulu dan kerangka berpikir juga dikemukakan pada bab ini.

Bab 3 Metode Penelitian

Pada bab Metode Penelitian berisi penentuan masalah fokus penelitian, klasifikasi penelitian berdasarkan tujuan dan pendekatan, pengumpulan data, penyelesaian masalah, dan penarikan kesimpulan.

Bab 4 Pembahasan

Pada bab Pembahasan berisi hasil penelitian dan pembahasan sebagai jawaban atas permasalahan.

Bab 5 Penutup

Pada bab Penutup dikemukakan kesimpulan dari pembahasan dan saran yang berkaitan dengan kesimpulan.

1.6.3 Bagian akhir skripsi

Bagian akhir skripsi meliputi daftar pustaka dan lampiran-lampiran yang mendukung.

BAB 2

LANDASAN TEORI

2.1. Tinjauan Pustaka

2.1.1 Regresi Linier Berganda

Istilah regresi pertama kali dalam konsep statistik digunakan oleh Sir Francis Galton dimana yang bersangkutan melakukan kajian yang menunjukkan bahwa tinggi badan anak-anak yang dilahirkan dari para orang tua yang tinggi cenderung bergerak (*regress*) kearah ketinggian rata-rata populasi secara keseluruhan (http:// www.jonathansarwono.info/regresi/regresi.pdf diakses 11-12-2015). Galton memperkenalkan kata regresi (*regression*) sebagai nama proses umum untuk memprediksi satu variabel, yaitu tinggi badan anak dengan menggunakan variabel lain, yaitu tinggi badan orang tua. Pada perkembangan berikutnya hukum Galton mengenai regresi ini ditegaskan lagi oleh Karl Pearson dengan menggunakan data lebih dari seribu. Pada perkembangan berikutnya, para ahli statistik menambahkan isitilah regresi berganda (*multiple regression*) untuk menggambarkan proses dimana beberapa variabel digunakan untuk memprediksi satu variabel lainnya.

Regresi dalam pengertian moderen menurut Gujarati (http:// www.jonathansarwono.info/regresi/regresi.pdf diakses 11-12-2015) ialah kajian terhadap ketergantungan satu variabel, yaitu variabel bebas terhadap satu atau lebih variabel lainnya atau yang disebut sebagai variabel-variabel eksplanatori dengan tujuan untuk membuat estimasi dan / atau memprediksi rata-rata populasi atau nilai rata-rata variabel tergantung dalam kaitannya dengan nilai-nilai yang sudah diketahui dari variabel eksplanatorinya. Meski analisis regresi berkaitan dengan ketergantungan atau dependensi satu variabel terhadap variabel-variabel lainnya hal tersebut tidak harus menyiratkan sebab-akibat (*causation*). Untuk mendukung pendapatnya ini, Gujarati mengutip pendapat Kendal dan Stuart yang diambil dari buku yang berjudul "*The Advanced Statistics*" terbit pada tahun 1961 yang mengatakan bahwa, "suatu hubungan statistik betapapun kuat dan sugestifnya tidak akan pernah dapat menetapkan hubungan sebab akibat (*causal connection*); sedang gagasan mengenai sebab akibat harus datang dari luar statistik, yaitu dapat berasal dari teori atau lainnya".

Menurut Draper dan Smith (1992) analisis regresi linier berganda digunakan untuk mengukur pengaruh antara lebih dari satu variabel prediktor (variabel bebas) terhadap variabel terikat. Bentuk umum model regresi linier berganda dengan p variabel bebas seperti pada persamaan (2.1) berikut.

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + e_i \tag{2.1}$$

dengan:

Y_i adalah variabel terikat untuk pengamatan ke-i, untuk i = 1, 2, ..., n.

$\beta_0, \beta_1, \beta_2, \dots, \beta_p$ adalah parameter

$X_1, X_2, \dots, X_p$ adalah variabel bebas

e_i adalah residual untuk pengamatan ke-i yang diasumsikan berdistribusi normal yang saling bebas dan identik dengan rata-rata 0 (nol) dan variansi σ^2.

Bentuk notasi matriks persamaan (2.1) dapat ditulis menjadi persamaan (2.2) berikut (Draper dan Smith, 1992).

$$Y = X\beta + e \tag{2.2}$$

dengan:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, X = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & Y_n & X_{np-1} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

Seringkali persamaan (2.1) ditaksir oleh model dengan persamaan (2.3) sebagai berikut (Draper dan Smith, 1992).

$$\hat{Y}_i = b_0 + b_1X_1 + b_2X_2 + \cdots + b_pX_p \tag{2.3}$$

2.1.2 Residual

Menurut Gujarati (2004) residual dalam regresi linear sederhana merupakan selisih dari nilai prediksi dengan nilai yang sebenarnya atau $e_i = Y_i - \hat{Y}_i$. Namun penggunaan jarak $e_i = Y_i - \hat{Y}_i$ tidaklah memuaskan. Dengan meminimumkannya diperoleh hasil yang wajar seperti berikut:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \tag{2.4}$$

Menurut Sungkawa (2009) jika nilai pengamatan terletak dalam garis regresi maka nilai residualnya sama dengan nol. Jadi jika total jarak atau nilai mutlak dari residual sama dengan nol ($\sum_{i=1}^{n} |e_i| = 0$) berarti semua nilai pengamatan berada pada garis regresi. Semakin besar nilai residualnya maka garis regresi semakin kurang tepat digunakan untuk memprediksi. Yang diharapkan adalah total residu kecil sehingga garis regresi cukup baik untuk digunakan.

Menurut Sungkawa (2009) nilai residual akan semakin besar jika terdapat data *outlier* dan dapat menurunkan nilai koefisien regresi. Untuk menunjukkan apakah model regresi tersebut layak atau tidak maka beberapa persyaratan harus dipenuhi, diantaranya anggapan nilai residu menyebar normal. Jika ini dipenuhi maka jelas total residualnya sama dengan nol ($\sum_{i=1}^{n}|e_i| = 0$). Jadi apabila nilainya jauh dari nol maka perlu dilakukan pengecekan (normalitas serta adanya *outlier*).

2.1.3 *Ordinary Least Square* (OLS)

Menurut Sembiring (2003) salah satu penduga model untuk bentuk regresi linier adalah dengan metode *Ordinary Least Square* (OLS). Konsep dari metode ini adalah meminimumkan jumlah kuadrat residual (selisih antara data sebenarnya dengan data dugaan) dari model regresi yang terbentuk. OLS pertama kali diperkenalkan oleh Carl Freidrich Gauss, seorang ahli matematika dari Jerman. Metode ini merupakan metode yang paling banyak digunakan dalam pembentukan model regresi atau mengestimasi parameter regresi dibandingkan dengan metode-metode yang lain.

Untuk mengestimasi koefisien garis regresi $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ pada p data suatu penelitian adalah (Sembiring, 2003):

$$J = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - b_0 - b_1X_1 - b_2X_2 - \cdots - b_pX_p\right)^2 \qquad (2.5)$$

dan itu harus bernilai minimum. Pada persamaan (2.5) nilai X dan Y berasal dari pengamatan. Jika J berubah diturunkan terhadap $b_0, b_1, b_2, \dots, b_p$, kemudian menyamakannya dengan nol, sehingga diperoleh (Sembiring, 2003):

$$\frac{\partial J}{\partial b_o} = -2\sum_{i=1}^{n}\left(Y_i - b_0 - b_1X_1 - b_2X_2 - \cdots - b_pX_p\right) = 0 \tag{2.6}$$

$$\frac{\partial J}{\partial b_1} = -2\sum_{i=1}^{n}\left(Y_i - b_0 - b_1X_1 - b_2X_2 - \cdots - b_pX_p\right)X_1 = 0 \tag{2.7}$$

$$\vdots$$

$$\frac{\partial J}{\partial b_p} = -2\sum_{i=1}^{n}\left(Y_i - b_0 - b_1X_1 - b_2X_2 - \cdots - b_pX_p\right)X_p = 0 \tag{2.8}$$

Persamaan (2.6), (2.7), dan (2.8) dapat disederhanakan menjadi (Sembiring, 2003):

$$nb_0 + b_1\sum_{i=1}^{n} X_1 + b_2\sum_{i=1}^{n} X_1X_2 + \cdots + b_p\sum_{i=1}^{n} X_1X_p = \sum_{i=1}^{n} Y_i \tag{2.9}$$

$$b_0\sum_{i=1}^{n} X_1 + b_1\sum_{i=1}^{n} {X_1}^2 + b_2\sum_{i=1}^{n} X_1X_2 + \dots + b_p\sum_{i=1}^{n} X_1X_p = \sum_{i=1}^{n} X_2Y_i \tag{2.10}$$

$$\vdots$$

$$b_0\sum_{i=1}^{n} X_p + b_1\sum_{i=1}^{n} X_1X_p + b_2\sum_{i=1}^{n} {X_2}^2X_p + \dots + b_p\sum_{i=1}^{n} {X_p}^2 = \sum_{i=1}^{n} X_pY_i \tag{2.11}$$

Apabila dinyatakan dalam bentuk matriks, persamaan normal (2.9), (2.10), (2.11) menjadi (Sembiring, 2003):

$$X^TXb = X^TY \tag{2.12}$$

Dengan demikian b sebagai penduga β dapat diperoleh melalui rumus (Sembiring, 2003):

$$b = (X^TX)^{-1}X^TY \tag{2.13}$$

2.1.4 Uji Asumsi

Model regresi yang diperoleh dari OLS merupakan model regresi yang menghasilkan estimator linier tidak bias yang terbaik (*Best Linear Unbias Estimator*/BLUE). Kondisi ini akan terjadi jika beberapa asumsi yang disebut dengan asumsi klasik dipenuhi. Menurut Gujarati (Suliyanto, 2008) dalam bukunya yang berjudul *Basic Econometrics*, mengemukakan ada beberapa asumsi klasik diantaranya:

1. Nilai residual berdistribusi normal.
2. Model regresi adalah linier, yaitu linier dalam parameter.
3. Variabel bebas berpengaruh secara signifikan terhadap variabel terikat.
4. Tidak terdapat multikolinearitas yang sempurna.
5. Homoskedastisitas atau varian dari residual adalah konstan.
6. Tidak terdapat autokorelasi antara nilai residual.

2.1.4.1 Uji Normalitas

Menurut Suliyanto (2008) uji normalitas dimaksudkan untuk mengetahui apakah residual yang telah distandardisasi berdistribusi normal atau tidak. Nilai residual dikatakan berdistribusi normal jika nilai residual tersebut sebagaian besar mendekati nilai rata-ratanya sehingga bila residual tersebut berdistribusi normal maka jika digambarkan dalam bentuk kurva, kurva tersebut akan berbentuk lonceng (*ell-shaped curve*) yang kedua sisinya melebar sampai tidak terhingga. Melihat pengertian uji normalitas tersebut maka uji normalitas disini tidak dilakukan per variabel (*univariate*) tetapi hanya terhadap nilai residual terstandarisasinya saja (*multivariate*). Tidak terpenuhinya normalitas pada umumnya disebabkan karena distribusi data yang dianalisis tidak normal, karena tedapat *outlier* dalam data yang diambil. Nilai *outlier* ini dapat terjadi karena adanya kesalahan dalam pengambilan sampel, bahkan karena kesalahan dalam melakukan input data atau memang karena karakteristik data tersebut memang aneh.

Untuk mendeteksi apakah nilai residual terstandardisasi berdistribusi normal atau tidak, dapat digunakan uji Kolmogorov-Smirnov (Suliyanto, 2008). Uji ini dilakukan dengan menggunakan langkah-langkah:

1. Membuat persamaan regresi.
2. Mencari nilai prediksinya $(\hat{Y})$.
3. Mencari nilai residualnya $(Y - \hat{Y})$.
4. Mengurutkan nilai residual terstandardisasi dari yang terkecil sampai yang terbesar.
5. Mencari nilai Z_r relatif kumulatif.
6. Mencari nilai Z_t teoritis berdasarkan Tabel Z.
7. Menghitung selisih nilai Z_r dengan Z_t dan diberi simbol K.
8. Mencari nilai K mutlak terbesar dan beri nama dengan K_{hitung}.
9. Bandingkan nilai K_{hitung} dengan Tabel Kolmogorov-Smirnov (K_{tabel}).
10. Menarik kesimpulan dengan kriteria jika $K_{hitung} < K_{tabel}$ maka residual terstandardisasi berdistribusi normal, atau jika sig > 0,05 (pada perhitungan menggunakan SPSS16) maka residual berdistribusi normal.

Konsekuensi jika asumsi normalitas tidak terpenuhi adalah nilai prediksi yang diperoleh akan bias dan tidak konsisten. Untuk mengatasi jika asumsi normalitas tidak terpenuhi dapat digunakan bebarapa metode berikut (Suliyanto, 2008):

1. Menambah jumlah data.
2. Melakukan transformasi data menjadi log atau LN atau bentuk lainnya.
3. Menghilangkan data yang dianggap sebagai penyebab data tidak normal.

2.1.4.2 Uji Linieritas

Menurut Suliyanto (2008) pengujian linieritas perlu dilakukan untuk mengetahui model yang dibuktikan merupakan model linier atau tidak. Uji linieritas dilakukan agar diperoleh informasi apakah model empiris sebaiknya linier, kuadrat, atau kubik. Apabila salah dalam menentukan model regresi maka nilai prediksi yang dihasilkan akan menyimpang jauh sehingga nilai prediksinya akan menjadi bias.

Menurut Suliyanto (2008) uji Lagrange Multipler (LM-Test) merupakan salah satu metode yang digunakan untuk mengukur linieritas yang dikembangkan oleh Engle pada tahun 1982. Prinsip metode ini adalah membandingkan antara nilai X^2_{hitung} dengan nilai X^2_{tabel} dengan df=(n,α). Langkah-langkahnya adalah sebagai berikut:

1. Membuat persamaan regresinya.
2. Mencari nilai prediksinya $(\hat{Y})$.
3. Mencari nilai residualnya $(Y - \hat{Y})$.
4. Menguadratkan semua nilai variabel bebas.
5. Meregresikan kuadrat variabel bebas terhadap nilai residualnya.
6. Mencari nilai koefisien determinasinya (R^2).
7. Menghitung nilai $X^2_{hitung} = (n\ X\ R^2)$ seperti pada persamaan (2.14)

$$X^2_{hitung} = (n\ X\ R^2) \quad (2.14)$$

   dimana n adalah jumlah pengamatan.

8. Menarik kesimpulan uji linieritas, dengan kriteria jika$X^2_{hitung} < X^2_{tabel}$ dengan df=(n,α) maka model dinyatakan linier. Demikian juga sebaliknya.

2.1.4.3 Uji Keberartian Simultan

Menurut Suliyanto (2008) pengujian keberartian simultan digunakan untuk menguji ketepatan model. Uji signifikansi simultan sering disebut uji F, digunakan untuk menguji apakah variabel bebas yang digunakan dalam model secara simultan (bersama-sama) mampu menjelaskan perubahan nilai variabel tak bebas atau tidak.

Menurut Suliyanto (2008) uji signifikansi simultan menggunakan nilai F_{hitung} yang dapat diperoleh dengan perhitungan rumus seperti pada persamaan (2.15). Nilai F_{hitung} juga dapat diperoleh menggunakan SPSS16. Kriteria pengujian keberartian simultan yaitu jika $F_{hitung} \geq F_{tabel}$ atau nilai sig $(\hat{\alpha}) <$ 0,05 maka variabel bebas yang digunakan dalam model secara simultan (bersama-sama) mampu menjelaskan perubahan nilai variabel tak bebas. Nilai F_{tabel} diperoleh dari Tabel distribusi *F* dengan df: α,(k-1),(n-k).

$$F = \frac{R^2/(k-1)}{1-R^2/(n-k)} \qquad (2.15)$$

2.1.4.4 Uji Keberartian Parsial

Pengujian keberartian parsial perlu dilakukan untuk mengetahui keberartian masing-masing variabel bebas terhadap variabel tak bebas. Uji keberartian parsial menggunakan nilai t_{hitung}, dengan kriteria pengujian jika $t_{hitung} > t_{tabel}$ atau nilai sig $(\hat{\alpha}) < 0{,}05$ maka variabel bebas memiliki pengaruh yang berarti terhadap variabel tak bebas. Nilai t_{tabel} diperoleh dengan df: α,(n-k). Nilai t_{hitung} dapat diperoleh menggunakan rumus dengan persamaan (2.16) atau menggunakan SPSS16.

$$t_{hitung} = \frac{b_j}{Sb_j} \tag{2.16}$$

Keterangan:

b_j = koefisien regresi

Sb_j = kesalahan baku koefisien regresi

2.1.4.5 Uji Multikolinearitas

Menurut Suliyanto (2008) pengertian kolinearitas sering dibedakan dengan multikolinearitas. Kolinearitas berarti terjadi korelasi linier yang mendekati sempurna antara kedua variabel bebas. Sedangkan multikolinearitas berarti terjadi korelasi linier yang mendekati sempurna antara lebih dari dua variabel bebas. Multikolinearitas bisa terjadi saat adanya kesalahan spesifikasi model (*spesification model*). Hal ini dapat terjadi karena seorang peneliti memasukan variabel bebas yang seharusnya dikeluarkan dari model empiris. Dapat juga terjadi karena seorang peneliti mengeluarkan variabel bebas yang seharusnya dimasukkan dalam model empiris. Selain itu, adanya model yang berlebihan (an overdetermined model) juga dapat menyebabkan multikolinearitas. Hal ini terjadi ketika model empiris (jumlah variabel bebas) yang digunakan melebihi jumlah data (observasi).

Untuk mendeteksi adanya masalah multikolinearitas, metode yang paling sering digunakan yaitu dengan menggunakan nilai VIF (*Variance Inflation Factor*). Untuk mengetahui ada tidaknya multikolinearitas antar variabel, salah satu caranya dengan melihat nilai *Variance Inflation Factor* (VIF) dari masing-masing variabel bebas terhadap variabel terikatnya. Menurut Gujarati (Suliyanto, 2008), jika nilai VIF tidak lebih dari 10 maka model dikatakan tidak mengandung multikolinearitas.

Beberapa akibat yang timbul jika hasil estimasi model empiris mengalami masalah multikolinearitas diantaranya (Suliyanto, 2008):

1. Penaksir OLS tidak bisa ditentukan (*indeterminate*) meskipun hasil estimasi yang dihasilkan masih BLUE (*Best Linear Unbiased Estimator*).
2. Interval kepercayaan cenderung meningkat lebih besar sehingga mendorong untuk menerima hipotesis nol (antara lain koefisien populasi adalah nol).
3. Nilai t-statistik koefisien dari satu atau beberapa variabel bebas secara statistik tidak signifikan sehingga dapat menyebabkan dikeluarkannya suatu variabel bebas dalam model regresi, padahal variabel bebas tersebut memiliki peran yang sangat penting dalam menjelaskan variabel terikat.
4. Penaksir-penaksir OLS dan kesalahan bakunya cenderung tidak stabil dan sangat sensitif bila terjadi perubahan data, meskipun perubahan itu sangat kecil.
5. Jika multikolinearitas sangat tinggi maka mungkin R^2 bisa tinggi namun sangat sedikit taksiran koefisien regresi yang signifikan secara statistik.

Menurut Suliyanto (2008) beberapa cara untuk mengatasi multikolinear diantaranya memperbesar ukuran sampel, menghilangkan salah satu atau lebih variabel bebas, menggabungkan data *time series* dan data *cross section*, atau melakukan transformasi data.

2.1.4.6 Uji Heteroskedastisitas

Menurut Suliyanto (2008) heteroskedastisitas berarti ada varians variabel dalam model yang tidak sama (konstan). Sebaliknya jika varian variabel dalam model memiliki nilai yang sama (konstan) disebut sebagai homoskedastisitas.

Untuk menguji adanya masalah heteroskedastisitas dapat dilakukan dengan menggunakan metode Glejser. Uji Glejser (Suliyanto, 2008) dilakukan dengan meregresikan semua variabel bebas terhadap nilai mutlak residualnya. Jika terdapat pengaruh variabel bebas yang signifikan terhadap nilai mutlak residualnya (sig < 0,05) maka dalam model terdapat masalah heteroskedastisitas.

Menurut Gujarati (Suliyanto, 2008) ada beberapa konsekuensi sebagai akibat dari adanya masalah heteroskedastisitas dalam model persamaan regresi diantaranya:

1. Walaupun penaksir OLS masih linier dan masih tak bias, tetapi akan mempunyai varian yang tidak minimum lagi serta tidak efisien dalam sampel kecil. Lebih lanjut penaksir OLS juga tidak efisien dalam sampel besar.
2. Formulasi untuk menaksir varian dari estimasi OLS secara umum adalah bias, dimana bila menaksir secara apriori, seorang peneliti tidak dapat mengatakan bahwa bias tersebut akan positif atau negatif. Akibatnya interval kepercayaan dan uji hipotesis yang didasarkan pada uji t dan nilai distribusi F tidak dapat dipercaya.
3. Prediksi yang didasarkan pada koefisien parameter variabel bebas dari data asli akan mempunyai varian yang tinggi sehingga prediksi tidak efisien.

Menurut Suliyanto (2008) perbaikan model apabila terjadi masalah heteroskedastisitas diantaranya melakukan transformasi model regresi dengan membagi model regresi dengan salah satu variabel independen yang digunakan dalam model regresi tersebut atau melakukan transformasi logaritma dan LN.

2.1.4.7 Uji Autokorelasi

Menurut Suliyanto (2008) uji autokorelasi bertujuan untuk mengetahui apakah ada korelasi antara anggota serangkaian data observasi yang diuraikan menurut waktu (*time series*) atau ruang (*cross section*).

Menurut Gujarati (Suliyanto, 2008) ada beberapa cara untuk mendeteksi adanya masalah autokorelasi salah satunya yaitu Uji Durbin Watson (Uji DW). Uji DW pertama kali diperkenalkan oleh J. Durbin dan G. S. Watson tahun 1951. Rumus yang digunakan untuk Uji DW adalah

$$DW = \frac{\sum(e - e_{t-1})^2}{\sum e} \tag{2.17}$$

Keterangan:

DW = Nilai Durbin-Watson *Test*

e = Nilai residual

e_{t-1} = Nilai residual satu baris/periode sebelumnya

dengan kriteria pengujian tertera pada Tabel 2.1.

Tabel 2.1 Kriteria Pengujian Autokorelasi dengan Durbin-Watson

DW	Kesimpulan
< dL	Ada autokorelasi positif
dL s.d. dU	Ragu-ragu
dU s.d. 4-dU	Tidak ada autokorelasi
4-dU s.d. 4-dL	Ragu-ragu
>4-dL	Ada autokorelasi negatif

Menurut Gujarati (Suliyanto, 2008) menyebutkan beberapa konsekuensi dari munculnya masalah autokorelasi dalam analisis regresi bahwa penaksir OLS *unbiased* dalam penyampelan berulang dan konsisten, tetapi sebagaimana dalam kasus heteroskedastisitas, penaksir OLS tidak lagi efisien (mempunyai varian minimum), baik dalam sampel kecil maupun sampel besar.

Menurut Suliyanto (2008) untuk memperbaiki autokorelasi dapat dilakukan dengan cara diantaranya dengan membuat persamaan perbedaan yang digeneralisasikan atau dengan metode perbedaan pertama.

2.1.5 Pencilan (*outlier*)

Menurut Rousseeuw *et al* (1987) pencilan (*outlier*) adalah data yang tidak mengikuti pola umum pada model regresi yang dihasilkan, atau tidak mengikuti pola data secara keseluruhan. Dalam suatu himpunan data biasanya terdapat 10% amatan yang merupakan *outlier* (Hampel *et al*., 1986). Jumlah maksimum *outlier* dalam data yang diperbolehkan adalah 50%.

Menurut Paludi (2009: 57) *Outlier* merupakan suatu keganjilan dan menandakan suatu titik yang sama sekali tidak tipikal dari data lainnya. Apabila dalam pengamatan terdapat data *outlier*, maka alternatif langkah yang diambil adalah menghilangkan atau membuang data *outlier* tersebut secara langsung terlebih dahulu sebelum dilakukan analisis lebih lanjut. Data *outlier* tersebut dapat dibuang secara langsung jika data tersebut diperoleh dari kesalahan teknis peneliti, seperti kesalahan mencatat amatan atau ketika menyiapkan peralatan.

Menurut Paludi (2009: 57) keberadaan data *outlier* akan mengganggu dalam proses analisis data dan harus dihindari dari beberapa hal. Dalam kaitannya dalam analisis regresi, *outlier* dapat menyebabkan hal-hal berikut:

1. Residual yang besar dari model yang terbentuk $E(e_i) \neq 0$
2. Varians pada data tersebut menjadi lebih besar
3. Taksiran interval memeliki rentang yang lebar

Menurut Draper & Smith (Paludi, 2009) adanya *outlier* berpengaruh akan memberikan nilai penduga parameternya bersifat bias sehingga berakibat interpretasi hasil yang diperoleh menjadi tidak valid. Namun menghindari *outlier* berpengaruh (menghapus *outlier* berpengaruh) dalam melakukan analisis bukanlah hal yang tepat untuk dilakukan. Adakalanya *outlier* memberikan informasi yang tidak bisa diberikan oleh titik data lainnya, misalnya *outlier* timbul karena kombinasi keadaan yang tidak biasa yang mungkin saja sangat penting dan perlu diselidiki lebih jauh.

2.1.6 Deteksi *Outlier*

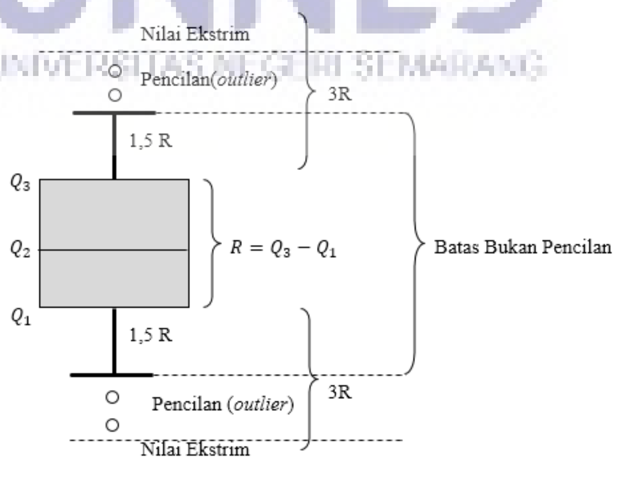
Menurut Nurcahyadi (2010: 17) ketika peneliti mendeteksi *outlier*, perlakuan pertamanya adalah melihat kemungkinan bahwa *outlier* merupakan data yang terkontaminasi. Data *outlier* dapat dikenali dengan pemeriksaan visual dari data mentahnya (*raw*) atau dari diagram pencar dari variabel dependen. Jika terdapat lebih dari dua variabel independen, beberapa *outlier* mungkin akan sangat sulit dideteksi dengan pemeriksaan visual. Oleh karena itu, dibutuhkan alat bantu pada pemeriksaan visual yang dapat membantu dalam pendeteksian outlier.

Data *outlier* harus dilihat terhadap posisi dan sebaran data yang lainnya sehingga akan dievaluasi apakah data *outlier* tersebut perlu dihilangkan atau tidak. Berdasarkan penelitian Ardiyanti (2011) ada berbagai macam metode pendeteksian data *outlier* yang berpengaruh dalam koefisien regresi diantaranya adalah metode *boxplot*, *Leverage Value*, *Standardized Residual*, *Breakdown Point*, dan *Cook's Distance*. Dalam penelitian ini hanya dibahas metode *boxplot* dan *Cook's Distance* dalam pendeteksian *outlier*.

2.1.7 Metode *Bloxplot*

Menurut Paludi (2009:58) metode *boxplot* merupakan metode yang mempergunakan nilai kuartil dan jangkauan untuk mendeteksi *Outlier*. Kuartil 1, 2, dan 3 akan membagi sebuah urutan data menjadi empat bagian. Jangkauan (IQR, *Interquartile Range*) didefinisikan sebagai selisih kuartil 1 terhadap kuartil 3, atau $IQR = Q_3 - Q_1$.

Data *Outlier* dapat ditentukan yaitu nilai yang kurang dari 1,5*IQR terhadap kuartil 1 dan nilai yang lebih dari dari 1,5*IQR terhadap kuartil 3 (Paludi, 2009: 58).



Gambar 2.1. Skema Identifikasi *Outlier* Menggunakan *Boxplot*

2.1.8 Metode *Cook's Distance*

Cook's Distance diperkenalkan oleh Cook (Yaffe, 2002:44). *Cook's Distance* merupakan salah satu ukuran untuk mendeteksi adanya *outlier* dalam data.

$$D_i = \frac{e_i^2 h_{ii}}{pMSE(1-h_{ii})^2}, \tag{2.18}$$

dengan e_i adalah residual ke$-i$, MSE adalah rata-rata jumlah kuadrat residual, h_{ii} merupakan nilai *leverage* untuk kasus ke$-i$, dan p banyaknya variabel independen ditambah konstan. Nilai *leverage* merupakan elemen-elemen diagonal dari matriks H.

$$H = X_i({X_i}^T X_i)^{-1} {X^T}_i \tag{2.19}$$

suatu data yang mempunyai nilai $D_i > \frac{4}{n}$ disebut *outlier*.

2.1.9 Regresi *Robust*

Regresi *robust* merupakan alat yang penting untuk menganalisis data yang terdeteksi sebagai data *outlier*. Regresi *robust* digunakan untuk mendeteksi *outlier* dan memberikan hasil yang resisten terhadap adanya data *outlier*. Sedangkan menurut Aunuddin (1988), regresi *robust* ini ditujukan untuk mengatatasi adanya data ekstrim serta meniadakan pengaruhnya terhadap hasil pengamatan tanpa terlebih dulu mengadakan identifikasi.

Metode ini merupakan metode yang mempunyai sifat (Aunuddin,1988):

1. Sama baiknya dengan OLS ketika semua asumsi terpenuhi dan tidak terdapat titik data yang berpengaruh.
2. Dapat menghasilkan model regresi yang lebih baik daripada OLS ketika asumsi tidak dipenuhi dan terdapat titik data yang berpengaruh.

3. Perhitungannya cukup sederhana dan mudah dimengerti, tetapi dilakukan secara iteratif sampai diperoleh dugaan terbaik yang mempunyai standar residual parameter yang paling kecil.

Pada regresi *robust* terdapat beberapapa estimasi, yaitu :

2.1.9.1 *M-Estimation*

Salah satu regresi *robust* yang penting dan paling luas digunakan adalah *M-Estimation*. Pada prinsipnya *M-Estimation* merupakan estimasi yang meminimumkan suatu fungsi residual ρ dan residualnya (Pradewi, 2012: 3).

$$\beta_{min} = \sum_{i=1}^{n} \rho(e_i) = \sum_{i=1}^{n} \rho\left(y_i - \sum_{j=0}^{k} x_{ij}\rho_j\right) \qquad (2.20)$$

Berdasarkan penelitian Pradewi (2012) pada estimasi parameter regresi *robust* M metode iterasi diperlukan, karena residualnya tidak dapat dihitung sampai diperoleh model yang cocok dan parameter regresi juga tidak dapat dihitung tanpa mengetahui nilai *Iteratively reweighted least squares* (IRLS) adalah metode yang banyak digunakan.

2.1.9.2 *Least Median of Squares (LMS)*

Menurut Rosseeuw dan Leroy (1987) prinsip dasar metode regresi *robust* penduga *Least Median of Squares* (LMS) adalah mencocokkan sebagian besar data setelah *outlier* teridentifikasi sebagai titik yang tidak berhubungan dengan data. Jika pada OLS hal yang perlu dilakukan adalah meminimumkan kuadrat residual $(\sum_{i=1}^{n} e_i^2)$, maka pada LMS hal yang perlu dilakukan adalah meminimumkan median kuadrat residual yaitu:

$$M_j = min\{med\ e_i^2\} = min\{M_1, M_2, \ldots, M_s\} \qquad (2.21)$$

dengan e_i^2 adalah kuadrat residual hasil taksiran dengan OLS.

Menurut Rosseeuw dan Leroy (1987) untuk mendapatkan nilai M_1, dicari himpunan bagian dari matriks X sejumlah h_i pengamatan, yaitu:

$$h_i = h_1 = \left[\frac{n}{2}\right] + \left[\frac{P+1}{2}\right] \quad (2.22)$$

di mana n banyaknya data, dan p banyaknya parameter ditambah satu.

Menurut Rosseeuw dan Leroy (1987) pada proses perhitungan, nilai h_i harus selalu dalam bentuk bilangan bulat oleh karena itu, jika nilai hi bukan dalam bentuk bilangan bulat maka dilakukan pembulatan ke atas. Selanjutnya untuk mencari M_2, ditentukan himpunan bagian data dari matriks X sejumlah h_2 pengamatan, yaitu :

$$h_i = h_2 = \left[\frac{n}{2}\right] + \left[\frac{P+1}{2}\right] \quad (2.23)$$

di mana $n = h_1$ dan $p = 3$.

Demikian seterusnya, sampai iterasi berahir pada iterasi ke-s yaitu saat $h_s = h_{s+1}$. Jadi akan diperoleh nilai M_j seperti pada persamaan (2.21).

Menurut Rosseeuw dan Leroy (1987) karena LMS merupakan penduga pada regresi *robust*, maka sama hal nya dengan penduga lain pada regresi *robust*, prinsip dasar dari LMS adalah dengan memberikan bobot w_{ii} pada data sehingga data *outlier* tidak mempengaruhi model parameter taksiran. Bobot w_{ii} ditentukan berdasarkan taksiran *robust standard deviation* yang diperoleh berdasarkan hasil perhitungan M_j dan $\hat{\sigma}$.

Berdasarkan Rousseeuw (Parmikanti, *et al.*, 2013: 625-626), bobot w_{ii} dirumuskan dengan ketentuan sebagai berikut:

$$w_{ii} = \begin{cases} 1, & \text{jika } \left|\frac{e_i}{\hat{\sigma}}\right| \leq 2{,}5 \\ 0, & \text{lainnya} \end{cases} \quad (2.24)$$

dengan

$$\hat{\sigma} = 1{,}4826\left[1 + \frac{5}{n-p}\right]\sqrt{M_j}. \tag{2.25}$$

Setelah bobot w_{ii} dihitung, dapat dibentuk matriks W sebagai berikut:

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix} \tag{2.26}$$

dengan entri matriks $w_{ij} = 0$, dimana $i \neq j$.

Setelah terbentuk matriks W, maka penaksir parameter regresi LMS dapat dihitung dengan menggunakan rumus (Parmikanti, *et al.*, 2013: 625-626):

$$\hat{\beta}_{LMS} = (X^T W X)^{-1}(X^T W Y) \tag{2.27}$$

Adapun algoritma pendugaan parameter regresi *robust* dengan metode LMS secara teoritis sebagai berikut:

1. Mendapatkan nilai M_1, dicari himpunan bagian data dari matriks X sejumlah h_i pengamatan, yaitu $h_i = h_1 = \left[\frac{n}{2}\right] + \left[\frac{P+1}{2}\right]$ dengan n banyaknya data dan p banyaknya parameter ditambah satu.
2. Melakukan langkah 1 sampai iterasi berahir pada iterasi ke-s yaitu saat $h_s = h_{s+1}$.
3. Membentuk matriks M_j seperti pada persamaan (2.21).
4. Menghitung bobot w_{ii} seperti pada persamaan (2.24).
5. Membentuk matriks W seperti pada persamaan (2.26).
6. Menghitung penduga parameter seperti pada persamaan (2.27).

2.1.9.3 Least Trimmed Squares (LTS)

LTS diusulkan oleh Rousseuw (1998) sebagai alternatif *robust* untuk mengatasi kelemahan *ordinary least squares (OLS)*, yaitu dengan menggunakan sebanyak $h (h \leq n)$ kuadrat residual yang diturunkan nilainya.

$$\min_{b} \sum_{i=1}^{h} e_i^2 \tag{2.28}$$

dengan

$$h = \frac{3n + p + 1}{4} \tag{2.29}$$

keterangan:

e_i^2 = kuadrat residual yang diurutkan dari terkecil ke terbesar

$e_1^2 < e_2^2 < \cdots < e_i^2 < \cdots < e_h^2 < \cdots < e_n^2$

n = banyaknya sampel

p = parameter regresi

Jumlah *h* menunjukkan sejumlah subset data dengan kuadrat fungsi objektif terkecil. Nilai *h* pada persamaan (2.29) akan membangun *breakdown point* yang besar sebanding dengan 50%. Untuk mendapatkan nilai residual pada LTS, digunakan algoritma LTS menurut Rousseeauw dan Van Driessen (1999) dalam Willems dan Aels (2005) adalah gabungan FAST-LTS dan C-*step*, yaitu dengan mengestimasi parameter b_0, kemudian menentukan n residual dengan menggunakan rumus $e_i^2 = (y_i - x_i b)^2$ yang bersesuaian dengan b. Setelah itu menghitung $\sum_{i=1}^{h_0} e_i^2$, dengan $h_0 = \frac{(3n+p+1)}{4}$ pengamatan dengan nilai e_i^2 terkecil.

Tahapan-tahapan tersebut dilakukan sampai diperoleh nilai residual terkecil dan konvergen.

2.1.9.4 Penduga S (S-Estimation)

Estimasi-S pertama kali diperkenalkan oleh Rousseeuw dan Yohai (Susanti, *et al.*, 2013: 256-257), Estimasi-S didefinisikan sebagai $\hat{\beta} = min_{\beta}\ \hat{\sigma}_s(e_1, e_2, \dots, e_n)$ dengan menentukan nilai estimator skala *robust* ($\hat{\sigma}_s$) yang minimum dan memenuhi

$$min \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta}{\hat{\sigma}_s}\right) \quad (2.30)$$

dengan

$$\hat{\sigma}_s = \sqrt{\frac{1}{nK}\sum_{i=1}^{n} w_i\, {e_i}^2} \quad (2.31)$$

dengan

$K = 0{,}199$, $w_i = w_{\sigma}(u_i) = \frac{\rho(u_i)}{{u_i}^2}$, dan dipilih estimasi awal

$$\hat{\sigma}_s = \frac{median|e_i - median(e_i)|}{0{,}6745} \quad (2.32)$$

Penyelesaian persamaan (2.30) dengan cara mencari turunannya terhadap $\hat{\beta}$ sehingga diperoleh (Susanti, *et al.*, 2013: 256-257):

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

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta}{\hat{\sigma}_s}\right) = 0 \quad j = 0,1, \dots, k \quad (2.33)$$

ψ disebut fungsi pengaruh yang merupakan turunan dari ρ ($\rho' = \psi$), turunan dari fungsi ρ adalah

$$\psi(u_i) = \rho'(u_i) = \begin{cases} u_i\left(1 - \left(\frac{u_i}{c}\right)^2\right)^2 & , |u_i| \le c \\ 0 & , |u_i| > c \end{cases} \quad (2.34)$$

dengan w_i merupakan fungsi pembobot IRLS

$$w_i(u_i) = \frac{\psi(u_i)}{c} = \begin{cases} \dfrac{u_i\left(1-\left(\frac{u_i}{c}\right)^2\right)^2}{u_i} & , |u_i| \leq c \\ 0 & , |u_i| > c \end{cases} \tag{2.35}$$

$$w_i(u_i) = \begin{cases} \left(1-\left(\frac{u_i}{c}\right)^2\right)^2 & , |u_i| \leq c \\ 0 & , |u_i| > c \end{cases} \tag{2.36}$$

dengan $u_i = \frac{e_i}{\hat{\sigma}_s}$ dan $c = 1{,}547$. Persamaan (2.33) dapat diselesaikan dengan IRLS sehingga mencapai konvergen.

Adapun algoritma pendugaan parameter regresi *robust* dengan metode Penduga S secara teoritis sebagai berikut:

1. Menghitung residual awal yang diperoleh dari OLS.
2. Menghitung standar deviasi residual $\hat{\sigma}_s$untuk mendapat nilai u_i.
3. Menghitung nilai pembobot w_i.
4. Menghitung OLS terbobot untuk mendapatkan penduga kuadrat terkecil terbobot dengan rumus seperti pada persamaan (2.37)

$$\hat{\beta}_s = (X^T W X)^{-1}(X^T W Y) \tag{2.37}$$

5. Menjadikan residual langkah (4) sebagai residual awal langkah (3) sehingga diperoleh nilai $\hat{\sigma}_s$ dan pembobot w_i yang baru.
6. Melakukan pengulangan iterasi IRLS (langkah 1 sampai 5) sampai didapatkan kekonvergenan sehingga diperoleh $\hat{\beta}_0{}^s, \hat{\beta}_1{}^s, \dots, \hat{\beta}_p{}^s$ yang merupakan estimasi-S.

2.1.9.5 MM-Estimation

MM-*Estimation* adalah metode yang pertama kali diperkenalkan oleh Yohai (Irfagutami, 2014: 45-47) yaitu dengan yang menggabungkan estimasi *high breakdown point* dan efisiensi statistik. Langkah pertama dalam estimasi ini adalah mencari estimator S dengan menjamin nilai *breakdown point*, kemudian menetapkan parameter-parameter regresi menggunakan estimasi M. Pada umumnya digunakan fungsi Tukey Bisquare β baik pada estimasi S maupun estimasi M.

Bentuk dari metode MM-*Estimation* adalah (Irfagutami, 2014: 45-47):

$$\tilde{\beta}_{MM} = \arg\min \sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}}\right) = \arg min \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}}\right) \qquad (2.38)$$

MM-*Estimation* juga menggunakan *Iteratively Reweighted Least Squares* (IRLS) untuk mencari estimasi parameter regresi.

Adapun langkah-langkah dalam proses *MM-Estimation* adalah:

1. Menghitung estimator awal koefisien $\hat{\beta}_j^{(1)}$dan residual $e_i^{(1)}$ dari regresi *robust* dengan *high breakdown point* (estimasi S) dengan bobot huber / bisquare (dilihat sebagai bentuk estimasi M).
2. Residual $e_i^{(1)}$ pada langkah pertama digunakan untuk menghitung skala estimasi $\hat{\sigma}_s^{(1)}$ dan dihitung pula pembobot awal $w_i^{(1)}$.
3. Residual $e_i^{(1)}$ dengan skala estimasi $\hat{\sigma}_s$ pada langkah kedua digunakan dalam iterasi awal sebagai penaksir WLS untuk menghitung koefisien regresi $\sum_{i=1}^{n} w_i^{(1)} \left(\frac{e_i^{(1)}}{\hat{\sigma}_s^{(1)}}\right) x_i = 0,\ w_i^{(1)}$ .merupakan pembobot Huber/bisquare.

4. Menghitung bobot baru $w_i^{(2)}$ dengan skala estimasi dari iterasi awal WLS.
5. Mengulang langkah (2),(3),(4) (dengan skala estimasi tetap konstan) sampai mendapatkan $\sum_{i=1}^{n}\left|e_i^{(m)}\right|$ konvergen (selisih $\hat{\beta}_j^{(m+1)}$ dan $\hat{\beta}_j^{(m)}$ mendekati 0, dengan banyak m iterasi).

2.1.10 Ukuran Pemilihan Model Terbaik

Metode pemilihan model terbaik dalam analisis regresi biasanya dengan membandingkan nilai MAE dan MSE yang dirumuskan sebagai berikut:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i| \tag{2.39}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \tag{2.40}$$

Tujuan optimalisasi statistik seringkali dilakukan untuk memilih suatu model agar nilai MSE minimal, tetapi ukuran ini mempunyai dua kelemahan. Pertama ukuran ini menunjukkan pencocokkan (*fitting*) suatu model terhadap data historis. Pencocokan seperti ini tidak selalu mengimplikasikan peramalan yang baik. Suatu model yang terlalu cocok (*over fitting*) dengan deret data berarti sama dengan memasukkan unsur random sebagai bagian proses bangkitan, adalah sama buruknya dengan dengan tidak berhasil mengenai pola non acak dalam data. Kekurangan kedua dalam MSE sebagai ukuran ketepatan model adalah berhubungan dengan kenyataan bahwa metode berbeda akan menggunakan prosedur yang berbeda pula dalam fase pencocokan. Selain itu MAE dan MSE tidak memudahkan perbandingan antar deret berskala yang berbeda dan untuk selang waktu yang berlainan, karena MAE dan MSE merupakan ukuran absolut yang sangat tergantung pada skala dari data deret waktu. Lagi pula, interpretasi nilai

MSE tidak bersifat intuitif, karena ukuran ini menyangkut pengkuadratan sederetan nilai.

Menurut Fathurahman (2009:37-39) kelemahan dari metode R^2, diantaranya adalah: (1) metode R2 hanya digunakan untuk peramalan *in sample* yaitu apakah prediksi model bisa sedekat mungkin dengan data yang ada, (2) tidak ada jaminan bahwa dengan metode R^2 mampu meramalkan nilai di masa mendatang (*out of sample*) dengan baik, (3) metode R^2 harus digunakan dengan syarat variabel tidak bebas (respon) harus sama, (4) nilai R^2 tidak pernah menurun, jika terus ditambahkan variabel prediktor di dalam model walaupun variabel prediktor tersebut kurang atau tidak relevan.

Menurut Widarjono (Fathurahman, 2009:37-39) beberapa metode lain yang dapat digunakan untuk mendapatkan model regresi terbaik, diantaranya adalah dengan metode *Akaike's Information Criterion* (AIC) dan *Schwarz Information Criterion* (SIC). Kedua metode tersebut mempunyai kelebihan dibanding menggunakan metode koefisien determinasi (R^2) yang banyak digunakan selama ini. Untuk menghitung nilai AIC dan SIC digunakan rumus sebagai berikut:

$$AIC = 2{,}718^{\frac{2k}{n}} \frac{\sum_{i=1}^{n} {e_i}^2}{n} \quad (2.41)$$

$$SIC = n^{\frac{k}{n}} \frac{\sum_{i=1}^{n} {e_i}^2}{n} \quad (2.42)$$

dengan

k = jumlah parameter yang diestimasi dalam model regresi

n = jumlah observasi

e = residual

Menurut Fathurahman (2009:37-39) kelebihan AIC dan SIC adalah terutama pada pemilihan model regresi terbaik untuk tujuan peramalan (*forecasting*), yaitu dapat menjelaskan kecocokan model dengan data yang ada (*insample forecasting*) dan nilai yang terjadi di masa mendatang (*out of sample forecasting*).

2.1.11 Anggaran Pendapatan dan Belaja Daerah

Menurut Iskandar (2012:12) anggaran sebagai proses alokasi sumberdaya yang terbatas untuk memenuhi kebutuhan yang tidak terbatas, dan anggaran merupakan rencana kerja dalam satuan mata uang untuk suatu periode tertentu. Pasal 1 ayat 8 Undang-undang Nomor 17 Tahun 2003 tentang keuangan negara menjelaskan pengertian Anggaran Pendapatan dan Belanja Daerah (APBD) adalah rencana keuangan tahunan pemerintah daerah yang disetujui oleh DPRD. APBD ditetapkan dalam bentuk Peraturan Daerah (Perda) mempunyai fungsi pemerintahan untuk mencapai tujuan bernegara. APBD terdiri atas anggaran pendapatan, anggaran belanja, dan pembiayaan.

2.1.11.1 Belanja Modal

Menurut Nordiawan dan Hertianti (Iskandar, 2012:13-14) Belanja Modal adalah pengeluaran yang manfaatnya cenderung melebihi satu tahun anggaran dan akan menambah jumlah aset atau kekayaan organisasi sektor publik, yang selanjutnya akan menambah anggaran oprasional untuk biaya pmeliharaannya. Pengertian belanja modal menurut Permendagri Nomor 13 Tahun 2006 sebagaimana terakhir telah diubah menjadi Permendagri Nomor 59 Tahun 2007 tentang Pedoman Pengelolaan Keuangan Daerah adalah pengeluaran yang dilakukan dalam rangka pengadaan aset tetap berwujud yang mempunyai nilai manfaat lebih dari 12 (dua belas) bulan untuk digunakan dalam kegiatan pemerintah.

2.1.11.2 Pendapatan Asli Daerah

Menurut Sholikhah (2014) Pendapatan Asli Daerah dihasilkan dari penggalian potensi kekayaan-kekayaan daerah yang berupa pajak daerah, retribusi, laba BUMD, dan lain-lain Pendapatan Asli Daerah yang dipisahkan.Pendapatan Asli Daerah merupakan komponen penerimaan daerah yang sangat penting untuk membantu laju pembangunan di daerah. Adanya Pendapatan Asli Daerah dijadikan sebagai modal untuk membiayai belanja di daerah. Pendapatan Asli Daerah ini juga menjadi tolok ukur dari keberhasilan daerah dalam mewujudkan daerah mandiri yang sesuai dengan konsep desentralisasi. Setiap penyusunan APBD, alokasi Belanja Modal harus disesuaikan dengan kebutuhan daerah dengan mempertimbangkan PAD yang diterima.

2.1.11.3 Dana Bagi Hasil

Menurut Sholikhah (2014) Dana Bagi Hasil (DBH) adalah dana yang bersumber dari pendapatan APBN yang ditransfer dari pemerintah pusat untuk mendukung penerapan desentralisasi. Indikator DBH adalah DBH Pajak dan DBH Bukan Pajak. DBH merupakan sumber pendapatan daerah yang cukup potensial dan merupakan salah satu modal dasar Pemerintah Daerah dalam mendapatkan dana pembangunan dan memenuhi belanja daerah yang bukan berasal dari PAD selain DAU dan DAK. Secara teoritis Pemerintah Daerah akan mampu menetapkan Belanja Modal yang semakin besar jika anggaran DBH semakin besar pula, begitupun Sebaliknya semakin kecil Belanja Modal yang akan ditetapkan jika anggaran DBH semakin kecil.

2.1.11.4 Dana Alokasi Umum

Menurut Sholikhah (2014) Dana Alokasi Umum (DAU) menjadi salah satu dana perimbangan yang juga mempunyai sumbangsih yang cukup besar dalam mendukung pembangunan daerah. Berkaitan dengan perimbangan keuangan antara pemerintah pusat dan daerah, hal tersebut merupakan konsekuensi adanya penyerahan kewenangan pemerintah pusat kepada Pemerintah Daerah. Dengan demikian, terjadi transfer yang cukup signifikan didalam APBN dari pemerintah pusat ke Pemerintah Daerah, dan Pemerintah Daerah secara leluasa dapat menggunakan dana ini apakah untuk memberi pelayanan yang lebih baik kepada masyarakat atau untuk keperluan lain yang tidak penting.

Menurut Sholikhah (2014) DAU mempunyai korelasi yang positif dan signifikan terhadap Belanja Modal sehingga dapat diasumsikan bahwa DAU mempunyai peranan yang sangat besar dalam pengalokasian Belanja Modal. Semakin besar transfer DAU dari Pemerintah Pusat maka alokasi Belanja Modal akan naik juga. Ketika alokasi Belanja Modal suatu daerah cukup besar, hal tersebut membuat rawan akan adanya perilaku korup dari berbagai pihak. Alokasi Belanja Modal begitu mudah di markup karena sering ditujukan untuk pengadaan barang maupun bangunan yang nominalnya jarang bisa ditaksir oleh masyarakat awam.

2.1.11.5 Luas Wilayah

Anggaran belanja modal didasarkan pada kebutuhan daerah akan sarana dan prasarana, baik untuk kelancaran pelaksanaan tugas pemerintahan maupun untuk fasilitas publik. Daerah dengan wilayah yang lebih luas membutuhkan sarana dan prasarana yang lebih banyak sebagai syarat untuk pelayanan kepada publik bila dibandingkan dengan daerah dengan wilayah yang tidak begitu luas (Kusnandar dan Dodik, 2009).

Berdasarkan penelitian yang dilakukan oleh Kusnandar dan Dodik (2009) , luas wilayah daerah memang mempunyai pengaruh yang positif terhadap anggaran Belanja Modal namun jika dianalisis, daerah yang mempunyai wilayah yang cukup luas hal itu justru akan memakan biaya pembangunan yang cukup besar. Untuk melaksanakan pembangunan tersebut, maka pemerintah harus menyediakan anggaran yang cukup besar jika ingin daerah tersebut benar-benar maju dan sejahtera. Untuk mewujudkan itu semua maka pemerintah harus cerdas dalam mengalokasikan penerimaan dan pengeluaran yang akan dibawa oleh pemerintah untuk mewujudkan daerah yang sejahtera.

Kaitan antara Luas Wilayah Daerah daerah dengan alokasi Belanja Modal yang kemudian dihubungkan dengan adanya hubungan keagenan hal ini dapat terlihat ketika suatu daerah ingin melakukan pemekaran wilayah dimana disitu terjadi konflik antara daerah dan pusat. Daerah mengalami kecemburuan sosial pada pusat karena alokasi dan distribusi pendapatan yang dikembalikan dari pemerintah pusat ke daerah dari hasil eksplorasi sumber-sumber daya di daerah dirasa kurang adil.

2.2. Penelitian Terdahulu

Berdasarkan penelitian Oktarinanda (2014) mengenai perbandingan efisiensi metode LTS dan metode LMS dalam estimasi parameter regresi *robust*. Perbandingan keakuratan model menggunakan koefisien determinasi dan RMSE diperolehkan kesimpulan bahwa data yang digunakan dalam penelitian lebih sesuai menggunakan penduga LMS dalam menduga parameter regresi. Selain itu, parameter duga yang dihasilkan LMS relatif lebih efisien daripada LTS karena ragam parameter duga dari metode LMS lebih kecil daripada LTS. Dengan kata lain metode LMS lebih efisien.

Berdasarkan penelitian Permana (2014) mengenai perbandingan metode LTS dan penduga-S sebagai metode pendugaan parameter memberikan kesimpulan bahwa keberadaan *outlier* berpengaruh dapat mempengaruhi nilai koefisien regresi yang dihasilkan metode LTS dan penduga S. Perubahan nilai koefisien regresi terjadi pada nilai intersep dan pada nilai koefisien yang lain. Berdasarkan kriteria *Mean Square Error* (MSE) terkecil diperoleh penduga S lebih baik digunakan untuk menduga parameter regresi linier berganda pada data yang mengandung *outlier*.

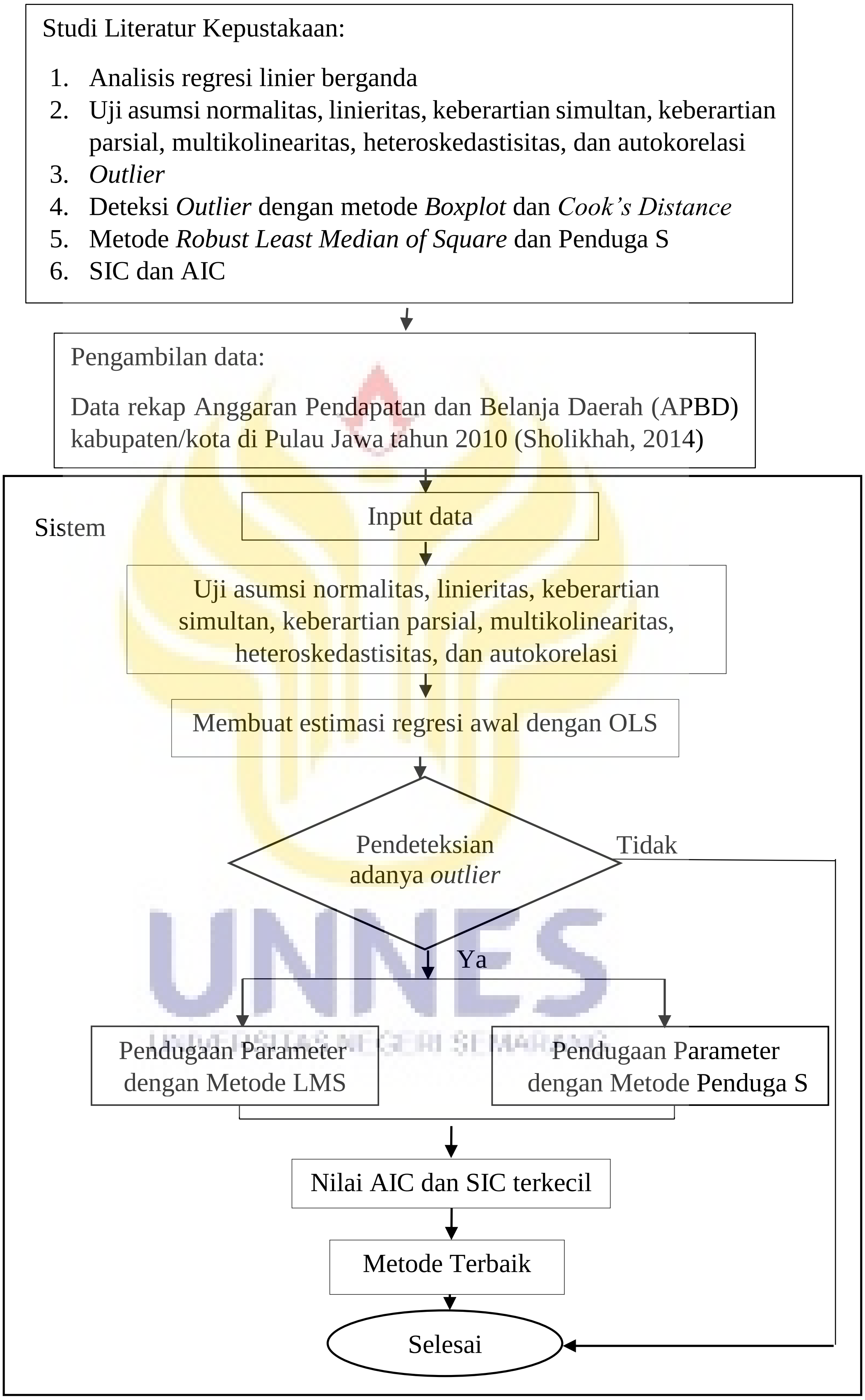
2.3. Kerangka Berpikir

Estimasi koefisien regresi pada umumnya digunakan metode estimasi OLS. Namun metode ini sangat sensitif terhadap kehadiran *outlier*. Hasil estimasi koefisien regresi dengan metode OLS menjadi tidak tepat jika terdapat *outlier* dalam data.

Pendeteksian *outlier* merupakan tahapan yang perlu dilakukan jika estimasi model regresi dengan OLS, yang dikenal cukup peka terhadap *outlier*. Pendeteksian *outlier* dapat dilakukan diantaranya dengan metode *boxplot* dan *Cook's Distance.* Jika dalam tahap pendeteksian *outlier* tidak terdapat *outlier* maka estimasi model dengan OLS diterima. Apabila terdapat *outlier* maka diperlukan suatu metode yang bersifat *robust* terhadap keberadaan *outlier*.

Regresi *robust* merupakan salah satu cara untuk mengatasi kelemahan OLS terhadap *outlier* pada data. Regresi *robust* menghasilkan estimasi model yang resisten terhadap pengaruh *outlier*. Pada regresi *robust* terdapat beberapa metode estimasi, diantaranya estimasi-M, estimasi LMS, estimasi LTS, Penduga S, dan Penduga MM. Penelitian ini difokuskan pada metode LMS dan Penduga S. Dari kedua metode tersebut dibandingkan sehingga diperoleh metode terbaik.

Penelitian bermula dengan pengumpulan data, yakni data rekap Anggaran Pendapatan dan Belanja Daerah (APBD) kabupaten/kota di Pulau Jawa tahun 2010. Dari data tersebut diuji asumsi normalitas, linieritas, keberartian simultan, keberartian parsial, multikolinearitas, heteroskedastisitas, dan autokorelasi. Lalu, dibuat estimasi regresi awal dengan OLS, dan mendeteksi *outlier* menggunakan metode *boxplot* dan *Cook's Distance*. Apabila dalam pendeteksian *outlier* diperoleh *outlier* dalam data maka dilakukan pendugaan parameter regresi *robust* dengan metode LMS dan penduga S. Untuk memperoleh metode terbaik dengan membandingkan nilai AIC dan SIC dari hasil kedua metode tersebut.



Gambar 2.2 Diagram Alir Kerangka Berpikir

BAB 5

PENUTUP

5.1. Simpulan

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan sebagai berikut.

1. Hasil estimasi regresi linier berganda data Anggaran Pendapatan dan Belanja Daerah (APBD) kabupaten/kota di Pulau Jawa tahun 2010 dengan metode *robust* LMS sebagai berikut.

   $$log\hat{Y}_{LMS} = 4{,}946 + 0{,}072 logX_1 + 0{,}520 logX_2$$

   dengan X_1 adalah Pendapatan Asli Daerah (PAD), X_2 adalah Dana Bagi Hasil (DBH), dan Y adalah Belanja Modal (BM).

   Hasil estimasi regresi linier berganda data Anggaran Pendapatan dan Belanja Daerah (APBD) kabupaten/kota di Pulau Jawa tahun 2010 dengan metode *robust* Penduga S sebagai berikut.

   $$log\hat{Y}_{S} = 4{,}717 + 0{,}119 logX_1 + 0{,}492 logX_2$$

   dengan X_1 adalah Pendapatan Asli Daerah (PAD), X_2 adalah Dana Bagi Hasil (DBH), dan Y adalah Belanja Modal (BM).

2. LMS merupakan metode regresi *robust* terbaik dibandingkan metode Penduga S.

5.2. Saran

1. Peneliti sebaiknya tidak membuang *outlier* dalam data observasi, karena apabila data *outlier* dibuang maka akan mengubah karakteristik atau makna penting suatu data sehingga interpretasi hasil analisis jauh dari fakta.
2. Penelitian lanjutan sebaiknya mencoba metode-metode estimasi regresi *robust* yang lain sebagai alternatif untuk memperoleh model regresi yang lebih baik.
3. Untuk mempermudah dalam melakukan analisis regresi *robust,* peneliti sebaiknya menggunakan program SAS9.1, karena lebih efektif dalam mengestimasi regresi dengan metode *robust.*

DAFTAR PUSTAKA

Aunuddin. 1988. *Analisis Data*. Bogor : PAU-Institut Pertanian Bandung.

Chen, C. 2014. Robust Regression and Outlier Detection with the Robustreg Procedure. *Proceedings International*. America: SAS Institute Inc. Tersedia di http://www2.sas.com/proceedings/sugi27/p265-27.pdf [diakses 5-4-2015].

Draper, N. R., & Smith, H. 1992. *Analisis Regresi Terapan*. Jakarta: Gramedia Pustaka Utama.

Fathurahman, M. 2009. Pemilihan Model Regresi Terbaik Menggunakan Metode Akaike's Information Criterion dan Schwarz Information Criterion. *Jurnal Informatika Mulawarman*, 4(3): 37-39. Tersedia di https:// informatikamulawarman.files.wordpress.com [diakses 1-12-2015].

Gujarati, D. N. 2004. *Basic Econometrics* (4^{th} ed). New York: The McGraw-Hill Companies.

http://www.jonathansarwono.info/regresi/regresi.pdf [diakses 1-12-2015].

Irfagutami, *et al*. 2014. Perbandingan Regresi Robust Penduga MM dengan Metode Random Sample Consensus dalam Menangani Pencilan. *E-Jurnal Matematika*, 3(2): 45-52. Tersedia di http://download.portalgaruda.org [diakses 3-4-2015].

Iskandar, M. A. 2012. *Pengaruh Belanja Modal, Dana Perimbangan, dan Kemandirian Fiskal Terhadap Pertumbuhan Ekonomi Daerah (Studi Empiris pada Pemerintah Kabupaten/Kota di Pulau Jawa Periode 2006-2010)*. Skripsi. Salemba: FE Universitas Indonesia.

Makkulau, *et al*. 2010. Pendeteksian Outlier dan Penentuan Faktor-Faktor yang Mempengaruhi Produksi Gula dan Tetes Tebu dengan Metode Likelihood Displacement Statistic-Lagrange. *Jurnal Teknik Industri*, 12(2): 95-100. Tersedia di http://ced.petra.ac.id/index.php [diakses 1-4-2015].

Meianto, *et al*. ____. Pengaruh Dana Alokasi Umum, Dana Alokasi Khusus, Pendapatan Asli Daerah, dan Luas Wilayah terhadap Belanja Modal pada Kabupaten/Kota di Sumatera Selatan. *Jurnal Akuntansi Palembang*. Tersedia di http://eprints.mdp.ac.id/1472/1/Jurnal%20Edy%20Meianto%202011210030.pdf [diakses 8-2-2016]

Nurcahyadi, H. 2010. *Analisis Regresi pada Data Outlier dengan Menggunakan Least Trimmed Square (LTS) dan MM-Estimasi*. Skripsi. Jakarta: Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah. Tersedia di http://repository.uinjkt.ac.id [diakses 10-5-2015].

Oktarinanda, A. 2014. Perbandingan Efisiensi Metode Least Trimmed Square (Lts) dan Metode Least Median Square (LMS) dalam Estimasi Parameter Regresi Robust. *Jurnal Statistik*, *2*(3): 177-180. Tersedia di http://statistik.studentjournal.ub.ac.id/ [diakses 1-4-2015].

Paludi, S. 2009. Identifikasi dan Pengaruh Keberadaan Data Pencilan (Outlier). *Majalah Panorama Nasional*, Januari-Juni. Hlm. 56-62. Tersedia di http://stein.ac.id/e-journal/pn_6/PN_6.pd [diakses 2-4-2015].

Permana, A.T. 2014. Perbandingan Metode Least Trimmed Square (LTS) dan Penduga-S sebagai Metode Pendugaan Parameter Regresi Robust. *Jurnal FMIPA Universitas Brawijaya*, 2(2): 125-128. Tersedia di http://statistik.studentjournal.ub.ac.id/ [diakses 1-4-2015].

Parmikanti, K., E. Rusyaman, & E. Suryamah. 2013. Model Regresi Kandungan Batubara Menggunakan Metode Least Median Of Squares. *Prosiding Seminar Nasional Sains dan Teknologi Nuklir*. Bandung: Badan Tenaga Nuklir Nasional. Tersedia di http://digilib.batan.go.id [diakses 1-4-2015].

Pradewi, E.D, & Sudarno. 2012. Kajian Estimasi-M IRLS Menggunakan Fungsi Pembobot Huber dan Bisquare Tukey pada Data Ketahanan Pangan di Jawa Tengah. *Jurnal Media Statistika*, 5(1): 1-10. Tersedia di http://ejournal.undip.ac.id [diakses 2-4-2015].

Rousseeuw, P.J., & Leroy, A.M. 1987. *Robust Regression and Outlier Detection*. Canada: John Wiley & Sons, Inc.

Rousseeuw, P.J. 1984. Least Median of Squares Regression. *Journal of the American Statistical Association*, 79(388): 871-880. Tersedia di http://www.cse.yorku.ca [diakses 3-4-2015].

Sembiring, R. K. 2003. *Analisis Regresi* (2th ed.). Bandung: Institut Teknologi Bandung.

Solikhah, I. 2014. *Analisis Belanja Modal pada Pemerintah Kabupaten/Kota di Jawa.* Skripsi. Semarang: FE Universitas Negeri Semarang.

Suliyanto. 2008. *Teknik Proyeksi Bisnis*. Yogyakarta: C.V. Andi Offset.

Sungkawa, Iwa. 2009. Penditeksian Pencilan (Outlier) dan Residual pada Regresi Linier. *Jurnal Informatika Pertanian*, 18(2): 95-105. Tersedia di http:// www.litbang.pertanian.go.id/warta-ip/pdf-file/2.iwa_ipvol18-2-2009.pdf [diakses 1-4-2015].

Susanti, Y., *et al*. 2013. Optimasi Model Regresi Robust untuk Memprediksi Produksi Kedelai di Indonesia. *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika*. Yogyakarta: FMIPA Universitas Negeri Yogyakarta. Tersedia di http://eprints.uny.ac.id/10850/1/S%20-%2031.pdf [diakses 5-4-2015].

Wandira, A. G. 2013. *Pengaruh Pendapatan Asli Daerah (PAD), Dana Alokasi Umum (DAU), Dana Alokasi Khusus (DAK), dan Dana Bagi Hasil (DBH) terhadap Pengalokasian Belanja Modal*. Skripsi. Semarang. FE Universitas Negeri Semarang. Tersedia di http://lib.unnes.ac.id/17630/1/7211409047.pdf [diakses 8-2-2016].

Wijayanti, L.U. 2015. *Analisis Perbandingan Regresi Robust Estimasi-M Huber dan Estimasi-S dalam Mengatasi Outlier*. Skripsi. Yogyakarta: FMIPA Universitas Negeri Yogyakarta. Tersedia di http://eprints.uny.ac.id/17923/ [diakses 4-4-2015].

Yaffe, R. A. 2002. *Robust Regression Modelling With STATA Lecture Notes.* Avenue: Social Science and Mapping Group Academic Computing Service. Tersedia di http://faculty.ksu.edu.sa/72563/Documents/Robust%20Regression.pdf [diakses 4-4-2015].