ANALISIS REGRESI *RIDGE ROBUST-MM* UNTUK MENGATASI DATA MULTIKOLINEARITAS DAN TIDAK NORMAL

(Skripsi)

Oleh

WAHYU HIDAYAT TULLAH

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2018

ABSTRACT

RIDGE ROBUST-MM **REGRESSION ANALYSIS TO HANDLE MULTICOLLINEARITY AND NON-NORMALITY**

By

WAHYU HIDAYAT TULLAH

Ridge robust-MM regression is a method that can solve multicollinearity and non-normality data in regression model. The purpose of this study is to know the performance of *ridge robust-MM* regression to handle multicollinearity and non-normality data and compare its etimates with OLS. The results show that *ridge robust-MM* regression gives better estimator in handling multicollinearity and non-normality than OLS based on the value of regression coefficient, MSE and AMSE .

Keywords: *ridge robust-MM* regression, multicollinearity, non-normality, MSE, AMSE

ABSTRAK

ANALISIS REGRESI *RIDGE ROBUST-MM* UNTUK MENGATASI DATA MULTIKOLINEARITAS DAN TIDAK NORMAL

Oleh

WAHYU HIDAYAT TULLAH

Regresi *ridge robust-MM* merupakan sebuah metode yang dapat digunakan untuk mengatasi data multikolinearitas dan tidak normal dalam model analisis regresi. Tujuan dari penelitian ini adalah untuk mengetahui kinerja metode regresi *ridge robust-MM* dalam menangani data yang mengandung multikolinearitas dan tidak normal serta membandingkan nilai dugaan nya dengan menggunakan MKT. Hasil dari penelitian ini menunjukkan bahwa regresi *ridge robust-MM* lebih baik dalam menangani data multikolinearitas dan tidak normal dibandingkan dengan MKT berdasarkan nilai dari koefisien regresi, MSE dan AMSE.

Kata Kunci: regresi *ridge robust-MM*, multikolinearitas, data tidak normal, MSE, AMSE

ANALISIS REGRESI *RIDGE ROBUST-MM* UNTUK MENGATASI DATA MULTIKOLINEARITAS DAN TIDAK NORMAL

Oleh

WAHYU HIDAYAT TULLAH

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA SAINS

pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2018

Judul Skripsi : **ANALISIS REGRESI *RIDGE ROBUST-MM* UNTUK MENGATASI DATA MULTIKOLINEARITAS DAN TIDAK NORMAL**

Nama Mahasiswa : **Wahyu Hidayat Tullah**

Nomor Pokok Mahasiswa : 1417031125

Program Studi : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Drs. Eri Setiawan, M.Si.	**Ir. Netti Herawati, M.Sc., Ph.D.**
NIP. 19581101 198803 1 002	NIP. 19650125 199003 2 001

2. Ketua Jurusan Matematika

Prof. Dra. Wamiliana, M.A., Ph.D.
NIP. 19631108 198902 2 001

MENGESAHKAN

1. Tim Penguji

Ketua : **Drs. Eri Setiawan, M.Si.** ....

Sekretaris : **Ir. Netti Herawati, M.Sc., Ph.D.** ....

Penguji
Bukan Pembimbing: **Prof. Drs. Mustofa Usman, M.A., Ph.D.** ....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Prof. Warsito, S.Si., D.E.A., Ph.D.
NIP. 19710212 199512 1 001

Tanggal Lulus Ujian Skripsi : **16 Mei 2018**

PERNYATAAN SKRIPSI MAHASISWA

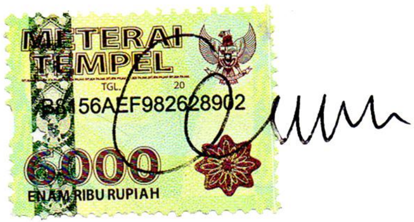
Yang bertanda tangan di bawah ini :

Nama	: **Wahyu Hidayat Tullah**
Nomor Pokok Mahasiswa	: **1417031125**
Jurusan	: **Matematika**
Judul Skripsi	: **Analisis Regresi *Ridge Robust-MM* untuk Mengatasi Data Multikolinearitas dan Tidak Normal**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 16 Mei 2018

Yang Menyatakan



Wahyu Hidayat Tullah

RIWAYAT HIDUP

Penulis bernama lengkap Wahyu Hidayat Tullah, lahir di Yukum Jaya pada 3 April 1997. Penulis merupakan anak tunggal, pasangan bapak Paimin dan ibu Iswahyuni.

Penulis menempuh pendidikan dasar di SD Negeri 4 Yukum Jaya dari tahun 2003-2009. Kemudian melanjutkan pendidikan di SMP Negeri 1 Terbanggi Besar dan lulus pada tahun 2011. Kemudian menempuh pendidikan di SMA Negeri 1 Terbanggi Besar dan lulus pada tahun 2014.

Pada tahun 2014, penulis diterima sebagai mahasiswi di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN). Pada tahun 2015-2017 penulis menjadi anggota organisasi HIMATIKA (Himpunan Mahasiswa Matematika). Pada tahun 2017 penulis melakukan Kerja Praktik di Kantor Badan Pusat Statistik Provinsi Lampung dan sebagai salah satu bentuk pengabdian kepada masyarakat penulis melaksanakan Kuliah Kerja Nyata di Desa Kelaten, Kecamatan Penengahan, Kabupaten Lampung Selatan.

KATA INSPIRASI

“Sesungguhnya sesudah kesulitan itu ada kemudahan. Maka apabila kamu telah selesai (dari suatu urusan), kerjakanlah dengan sungguh-sungguh (urusan yang lainnya). Dan hanya kepada Tuhanmu lah hendaknya kamu berharap.”

(QS. Al-Insyirah : 6-8)

“Maka nikmat Tuhan kamu yang manakah yang kamu dustakan?.”

(Q.S Ar-Rahman)

“Barang siapa yang ingin do’anya terkabul dan terlepas dari kesulitannya, maka hendaklah ia mengatasi (meringankan) kesulitan/kesusahan orang lain.”

(HR. Ahmad)

“Barang siapa yang menempuh jalan untuk menuntut ilmu, niscaya Allah subhanahu wata’ala akan memudahkan baginya jalan menuju surga.”

(H.R. Muslim)

PERSEMBAHAN

Karyaku yang sederhana ini kupersembahkan kepada:

Bapak dan Ibu

Terima kasih kepada Bapak dan Ibu yang selalu mendo'akan kesuksesanku, memberi semangat, nasihat, dukungan serta kasih sayang yang tiada henti.

Siti Ulfa Nabila

Terima kasih kepada Siti Ulfa Nabila yang selalu menemani, memberikan do'a, semangat dan motivasi yang tiada henti.

Sahabat-sahabatku Agus, Rahmad, Sadha, Dracjat, Wayan, Alvin, Nandra dan Darmawansyah

Terima kasih kepada para sahabatku yang selalu memberikan semangat, do'a, dan motivasi, serta kenangan indah selama ini.

Almamater dan Negeriku

SANWACANA

Puji syukur kehadirat Allah SWT yang telah memberikan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul "Analisis Regresi *Ridge Robust-MM* untuk Mengatasi Data Multikolinearitas dan Tidak Normal" dengan baik dan tepat pada waktunya.

Penulis menyadari bahwa skripsi ini dapat terselesaikan dengan baik karena dukungan, bimbingan, saran, serta do'a dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis menyampaikan terima kasih kepada:

1. Bapak Drs. Eri Setiawan, M.Si., selaku dosen pembimbing satu yang telah memberikan bimbingan, pengarahan, dan saran kepada penulis dalam mengerjakan skripsi.
2. Ibu Ir. Netti Herawati, M.Sc., Ph.D., selaku pembimbing dua yang telah memberikan saran serta pembelajaran yang sangat bermanfaat dalam menyelesaikan skripsi.
3. Bapak Prof. Drs. Mustofa Usman, M.A., Ph.D., selaku pembahas dan penguji skripsi yang telah memberikan evaluasi, arahan, dan saran demi perbaikan skripsi.
4. Bapak Subian Saidi, S.Si., M.Si., selaku dosen pembimbing akademik.
5. Ibu Prof. Dra. Wamiliana, M.A., Ph.D., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Seluruh dosen Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

7. Bapak Prof. Warsito, S.Si., D.E.A., Ph.D. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
8. Bapak dan Ibu tercinta yang selalu mendo'akan kesuksesan dunia dan akhirat.
9. Siti Ulfa Nabila yang telah mendo'akan, memberi saran serta keceriaan, dan memperlihatkan sudut pandang lain dari sebuah kehidupan.
10. Sahabat-sahabat tersayang, Agus, Rahmad, Sadha, Dracjat, Wayan, Nandra, Darma yang telah mendo'akan, memberi dukungan dan kenangan indah kepada penulis.
11. Zhofar, Febi, Yani, Ratna dan teman-teman satu bimbingan lainnya, terima kasih atas semangat dan saran selama penyelesaian skripsi.
12. HIMATIKA yang telah memberikan pengalaman berharga.
13. Teman-teman mahasiswa Jurusan Matematika angkatan 2014.

Penulis menyadari bahwa dalam penulisan skripsi ini masih jauh dari sempurna, sehingga informasi tambahan, saran, dan kritik untuk pengembangan lebih lanjut sangat diharapkan. Semoga skripsi ini dapat bermanfaat bagi kita semua.

Bandar Lampung, 16 Mei 2018

Penulis

Wahyu Hidayat Tullah

DAFTAR ISI

Halaman

IV. HASIL DAN PEMBAHASAN

DAFTAR TABEL

Tabel Halaman

DAFTAR GAMBAR

Gambar Halaman

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis regresi merupakan salah satu metode statistika yang dapat digunakan untuk menyelidiki atau mengetahui hubungan serta membangun hubungan antara dua variabel atau lebih. Variabel tersebut terdiri dari variabel yang dijelaskan disebut dengan variabel terikat (dilambangkan dengan Y) dan variabel penjelas yang disebut variabel bebas (dilambangkan dengan X). Analisis regresi sering diterapkan dalam beberapa bidang keilmuan, diantaranya mencakup bidang ekonomi, ilmu-ilmu sosial, ilmu biologi, bidang pendidikan, serta dalam bidang teknik.

Metode penduga terbaik dalam analisis regresi adalah Metode Kuadrat Terkecil (MKT). Sebuah model regresi dikatakan baik atau cocok apabila memenuhi asumsi-asumsi klasik, yaitu galat menyebar normal dengan rata-rata nol, ragam dari galat bersifat homogen, galat tidak mengalami autokolerasi, dan tidak terjadi multikolinearitas antarvariabel bebas. Pendugaan parameter dengan menggunakan metode kuadrat terkecil akan menghasilkan penduga yang tak bias, akan tetapi metode ini peka terhadap adanya penyimpangan asumsi. Apabila terdapat salah satu asumsi regresi yang tidak terpenuhi maka penduga dengan metode kuadrat terkecil tidak lagi efisien untuk digunakan. Salah satu penyimpangan asumsi yang tak jarang ditemukan adalah munculnya data yang menyimpang dari sekumpulan data lainnya yang disebut dengan data pencilan sehingga asumsi normalitas tidak terpenuhi. Munculnya pencilan dapat berpengaruh terhadap model regresi yang dihasilkan. Sehingga dibutuhkan suatu metode untuk mengatasi data yang tidak memenuhi asumsi normalitas yaitu dengan menggunakan regresi *robust.* Terdapat beberapa jenis metode *robust.* Salah satu jenis metode *robust* adalah penduga MM (*method of moment*). Penduga MM merupakan penggabungan antara metode penduga S yang memiliki *break-down* yang tinggi dan metode penduga M yang memiliki efisiensi tinggi yang termasuk jenis penduga *maximum likelihood.*

Selain itu dalam analisis regresi, asumsi yang memungkinkan tidak terpenuhi apabila variabel bebas yang digunakan lebih dari satu adalah terjadinya multikolinearitas. Adanya multikolinearitas dapat menyebabkan kesalahan tanda (positif atau negatif) dari nilai dugaan koefisien kuadrat terkecil. Salah satu metode analisis yang dapat digunakan untuk mengatasi masalah multikolinearitas adalah dengan menggunakan regresi *ridge*.

Apabila dalam model regresi terdapat penyimpangan asumsi kenormalan galat yang disebabkan oleh adanya data pencilan dan terdapat multikolinearitas secara simultan, dibutuhkan suatu metode yang dapat mengatasi kedua penyimpangan asumsi tersebut yaitu digunakan metode regresi *ridge robust*. Regresi *ridge robust* merupakan penggabungan metode regresi *ridge* dan metode regresi *robust* untuk mendapatkan nilai dugaan parameter yang stabil dan resisten terhadap pencilan dan menangani masalah multikolinearitas. Metode *ridge robust* ini diawali dengan mencari pembobot dan nilai dugaan parameter pada regresi *robust-MM* kemudian nilai dugaan parameter yang diperoleh digunakan untuk mencari nilai dugaan parameter dari metode regresi *ridge robust.*

Oleh sebab itu, dalam penelitian ini akan dikaji tentang seberapa baik analisis regresi *ridge robust* dengan penduga MM dibandingkan dengan metode kuadrat terkecil dalam mengatasi masalah multikolinearitas dan data tidak menyebar normal.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Membandingkan nilai dugaan parameter regresi *ridge robust-MM* dan metode kuadrat terkecil untuk data yang tidak menyebar normal dan terdapat multikolinearitas.
2. Melihat seberapa baik kinerja metode regresi *ridge robust-MM* dalam mengatasi kasus multikolinieritas dan data tidak menyebar normal berdasarkan nilai MSE dan AMSE.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah menambah pengetahuan serta diharapkan dapat menjadi masukkan bagi para peneliti, mahasiswa, dan pembaca tentang metode regresi *ridge robust-MM* (*Method of Moment*) untuk menganalisis data yang tidak menyebar normal dan terdapat multikolinearitas.

II. TINJAUAN PUSTAKA

2.1 Analisis Regresi

Menurut Usman dan Warsono (2001) analisis regresi merupakan salah satu metode statistik yang dapat digunakan untuk menyelidiki, meramalkan atau membangun model hubungan antara dua variabel atau lebih. Dalam analisis regresi dibedakan menjadi dua jenis variabel yaitu variabel bebas atau variabel prediktor dan juga variabel terikat atau variabel respon. Variabel terikat adalah variabel yang akan diestimasi nilainya dan biasa diplot pada sumbu tegak (sumbu-Y). Sedangkan variabel bebas adalah variabel yang diasumsikan memberikan pengaruh terhadap variasi variabel terikat dan biasanya diplot pada sumbu datar (sumbu-X).

Misalkan diasumsikan model hubungan antara variabel X dan Y adalah linier dan ingin menentukan garis dugaan terbaiknya, maka harus menyadari bahwa garis dugaan dari masalah yang sebenarnya diharapkan mampu memprediksi dengan tepat setiap individu Y oleh setiap individu X. Aspek yang sangat penting dari analisis regresi adalah pengumpulan data karena kesimpulan dari analisis sangat tergantung pada data yang dikumpulkan. Pengumpulan data yang baik akan memberikan banyak manfaat, termasuk penyederhanaan analisis dan membangun model yang secara umum dapat dipergunakan dan dipertanggungjawabkan

2.2 Analisis Regresi Linear Berganda

Menurut Usman dan Warsono (2001) analisis regresi linier berganda merupakan analisis hubungan secara linier antara dua atau lebih variabel bebas dengan satu variabel terikat. Persamaan umum garis regresi untuk regresi linear berganda adalah :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \qquad (2.1)$$

dengan,

i = banyaknya pengamatan dengan $i = 1,2,3, \ldots, n$

j = banyaknya variabel bebas dengan $j = 1,2,3, \ldots, k$

Y_I = variabel terikat pengamatan ke-i

X_{ki} = variabel bebas pengamatan ke-i

β_0 = konstanta (parameter)

β_k = koefisien regresi atau *slope* (parameter) ke-k

ε_i = sisaan (galat) pengamatan ke-i

Dalam regresi linear berganda yang akan diduga adalah β_0 dan β_j artinya $(\beta_1, \beta_2, \ldots, \beta_k)$. Persamaan linear untuk pendugaan garis regresi linear ditulis dalam bentuk :

$$\hat{Y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \ldots + b_k x_{ki} \qquad (2.2)$$

dengan

$\hat{Y}_i$ = nilai dugaan variabel terikat pengamatan ke-i

x_{ki} = nilai variabel bebas pengamatan ke-i

b_0 = titik potong garis regresi pada sumbu-y atau nilai dugaan Ŷ bila x=0

b_k = gradien garis regresi (perubahan nilai dugaan Ŷ per satuan perubahan nilai x) ke-k

Model regresi linear berganda dapat juga ditulis dalam bentuk matriks yaitu :

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad (2.3)$$

dengan

Y = vektor pengamatan berukuran n x 1

X = matriks variabel bebas berukuran n x k

β = vektor parameter yang akan ditaksir berukuran k x 1

$\boldsymbol{\varepsilon}$ = vektor galat berukuran n x 1

2.3 Asumsi Analisis Regresi Linear

Menurut Drapper dan Smith (1992), agar mampu memiliki kesimpulan yang benar tentang parameter β_0 dan β_k, pemenuhan asumsi-asumsi model regresi harus terpenuhi. Asumsi-asumsi tersebut adalah :

1. Nilai ε_i adalah bebas satu dengan yang lainnya atau korelasi $(\varepsilon_i , \varepsilon_j)=0$. Untuk asumsi pertama yang menyatakan *independent*, artinya ε_i merupakan variabel acak dengan nilai tengah nol dan σ^2 yang tidak diketahui. Jadi, $E(\varepsilon_i)=0$,

$V(\varepsilon_i)=\sigma^2$. ε_i dan ε_j tidak berkolerasi, $i\neq j$, sehingga $cov(\varepsilon_i, \varepsilon_j)=0$. Jadi, ε_i merupakan variabel acak normal, dengan nilai tengah nol dan ragam σ^2_ε dengan kata lain ε_i~N $(0, \sigma^2_\varepsilon)$.

2. Nilai tengah dari Y adalah fungsi linear dari X, yaitu jika dihubungkan titik-titik dari nilai tengah yang berbeda, maka akan diperoleh garis lurus $\mu_{(y|x)} = \beta_0 + \beta_k X_{ki}$.
3. Ragam galat homogen (homokedastik) yaitu galat memiliki nilai ragam yang sama antara galat ke-i dan galat ke-j. Secara matematis ditulis $Var(\varepsilon_i)= \sigma^2$.
4. Ragam galat menyebar normal dengan rata-rata nol dan suatu ragam tertentu. Asumsi keempat menyatakan untuk sembarang kombinasi tetap dari variabel bebas X, variabel tak bebas Y berdistribusi normal atau yang biasa disebut asumsi kenormalan. Dengan kata lain ε_i~N $(0, \sigma^2_\varepsilon)$.

2.4 Normalitas

Menurut Drapper dan Smith (1992) uji normalitas berguna pada tahap awal dalam metode pemilihan analisis data. Tujuan uji normalitas adalah untuk mengetahui apakah variabel pengganggu atau residual memiliki distribusi normal. Pengujian ini diperlukan karena untuk melakukan uji t dan uji F mengasumsikan bahwa nilai residual mengikuti distribusi normal.

Untuk pengujian normalitas, pada pengujian ini dilakukan dengan uji normalitas Kolmogorov Smirnov. Uji Kolmogorov Smirnov meggunakan hipotesis :

H_0 : Data residual berdistribusi normal

H_1 : Data residual tidak berdistribusi normal

Konsep dasar dari uji normalitas Kolmogorov Smirnov adalah dengan membandingkan distribusi data (yang akan diuji normalitasnya) dengan distribusi normal baku. Distribusi normal baku adalah data yang telah ditransformasi ke dalam bentuk *Z-score* dan diasumsikan normal. Jika nilai signifikansi berada dibawah nilai α yang ditentukan (dalam hal ini $\alpha = 5\%$) berarti nilai data residual yang diuji memiliki perbedaan yang signifikan dengan data normal baku, artinya data residual tersebut tidak berdistribusi normal (H_0 ditolak). Selanjutnya, jika nilai signifikansi berada diatas nilai α yang ditentukan berarti tidak terdapat perbedaan yang signifikan antara data residual yang diuji dengan data normal baku, artinya data residual tersebut berdistribusi normal (H_0 diterima).

2.5 Multikolinearitas

Istilah multikolinearitas pertama kali ditemukan oleh Ragnar Frisch pada tahun 1934 yang artinya terdapat hubungan linier diantara beberapa atau semua variabel bebas dalam model regresi. Masalah multikolinearitas hanya ditemukan dalam regresi linier berganda. Model yang baik adalah model yang bebas dari multikolinearitas. Suatu model yang bebas dari multikolinearitas adalah model yang memiliki nilai *variance inflation factor* (VIF) < 10 apabila *variance inflation factor* (VIF) > 10 mengindikasi terdapat multikolinearitas (Myers, 1990).

Menurut Gujarati (1995) salah satu cara untuk menguji gejala multikolinearitas dalam model regresi adalah dengan melihat nilai TOL (*tolerance*) dan VIF (*Variance Inflation Factor*) dari masing-masing variabel bebas terhadap variabel terikatnya. Uji untuk mengetahui gejala multikolinearitas dengan melihat nilai VIF dan TOL tersebut dilakukan dengan menggunakan langkah-lngkah:

1. Menghitung VIF dari $X_{1.}$
2. Meregresikan variabel bebas selain X_1 terhadap $X_{1.}$
3. Menghitung koefisien determinasi dari regresi variabel bebas selain X_1 terhadap X_1 dan diperoleh Rj^2.
4. Menghitung nilai TOL dengan rumus TOL = $(1 – Rj^2)$.
5. Menghitung nilai VIF dengan rumus $VIF = \frac{1}{TOL}$.

Masalah multikolinearitas bisa timbul karena berbagai sebab. Pertama, karena sifat-sifat yang terkandung dalam kebanyakan variabel ekonomi berubah bersama-sama sepanjang waktu. Besaran-besaran ekonomi dipengaruhi oleh faktor-faktor yang sama. Oleh karena itu, sekali faktor-faktor yang mempengaruhi itu menjadi operatif, maka seluruh variabel akan cenderung berubah dalam satu arah. Dalam data *time series*, pertumbuhan dan faktor-faktor kecenderungan merupakan penyebab utama adanya multikolinearitas. Kedua, penggunaan nilai lag (*lagget values*) dari variabel-variabel bebas tertentu dalam model regresi. Mengingat sifat yang sangat mendasar dari data, multikolinearitas diperkirakan terdapat pada sebagian besar hubungan-hubungan ekonomi. Oleh karena itu, perhatian sesungguhnya bukan lagi terletak pada ada atau tidaknya multikolinearitas, tetapi lebih pada akibat-akibat yang ditimbulkan oleh adanya multikolinearitas dalam sampel (Sumodiningrat, 1998).

2.6 Konsekuensi Multikolinearitas

Menurut Gujarati (1995) jika asumsi pada model regresi linear klasik terpenuhi, maka penaksir kuadrat terkecil / *Ordinary Least Square* (OLS) dari koefisien regresi linear adalah linear, tak bias dan mempunyai varian minimum dalam arti penaksir tersebut adalah penaksir tak bias kolinear terbaik / *Best Linear Unbiased Estimator* (BLUE), meskipun multikolinearitas sangat tinggi, penaksir kuadrat terkecil biasa masih tetap memenuhi syarat BLUE, tetapi penaksir tersebut tidak stabil.

Multikolinearitas berpengaruh terhadap estimasi kuadrat terkecil dari koefisien regresi. Akan diperlihatkan bagaimana $\hat{\beta}$, variansi $\left(\hat{\beta}_j\right)$ dan kovariansi $\left(\hat{\beta}_j, \hat{\beta}_h\right)$ dengan j dan $h = 1,2,\ldots,k$. Misalkan ada dua variabel bebas (X_1, X_2) dan Y variabel terikat sehingga diperoleh model

$$Y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon \tag{2.4}$$

Persamaan normal dengan metode kuadrat terkecil adalah

$$(\boldsymbol{X^T X})\boldsymbol{\hat{\beta}} = \boldsymbol{X^T Y} \tag{2.5}$$

$$\boldsymbol{X^T X} = \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}, \qquad \boldsymbol{X^T Y} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}$$

diperoleh

$$[\boldsymbol{X^T X}]^{-\mathbf{1}} = \begin{bmatrix} \dfrac{1}{1 - r_{12}^2} & \dfrac{-r_{12}}{1 - r_{12}^2} \\ \dfrac{-r_{12}}{1 - r_{12}^2} & \dfrac{1}{1 - r_{12}^2} \end{bmatrix}$$

Elemen diagonal utama dari matriks $[\boldsymbol{X^T X}]^{-\mathbf{1}}$ merupakan *Varians Inflasion Factor* (VIF) yaitu :

$$C_{jj} = \frac{1}{1-R_j^2} \ ; j = 1,2, \dots, k \tag{2.6}$$

dengan

R_j^2 = koefisien determinansi dari regresi X_j

C_{jj} = *varians inflasion factor* (VIF)

r_{12} = korelasi antara $X_1 \ dan \ X_2$

r_{X_jY} = korelasi antara X_j dan Y

$$(\boldsymbol{X^T X})\boldsymbol{\hat{\beta}} = \boldsymbol{X^T Y}$$
$$\begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}$$

sehingga,

$$\hat{\beta}_1 = \frac{r_{1y} - r_{12}r_{2y}}{(1 - r_{12}^2)}$$
$$\hat{\beta}_2 = \frac{r_{2y} - r_{12}r_{1y}}{(1 - r_{12}^2)}$$

Jika ada multikolinieritas antara X_1 dan X_2 yang sangat erat dan $r_{12} \to 1$. Variansi dan kovariansi koefisien regresi menjadi sangat besar karena $V(\hat{\beta}_j) = C_{jj}r^2 \to \infty$ seperti $|r_{12}| \to 1$, galat $Cov(\hat{\beta}_1, \hat{\beta}_2) = C_{12}\delta^2 \to \pm\infty$, variansi yang besar untuk $\hat{\beta}_j$ menyatakan bahwa koefisien regresi adalah perkiraan yang sangat lemah. Pengaruh multikolinieritas adalah untuk memperkenalkan sebuah ketergantungan linier yang dekat dalam kolom matriks. Selanjutnya jika kita mengasumsikan $X_1^T Y \to X_2^T Y$, seperti $|r_{12}| \rightarrow 1$, perkiraan koefisien regresi menjadi sama besarnya, tetapi berlawanan tanda, yaitu $\hat{\beta}_1 = -\hat{\beta}_2$.

Masalah yang sama terjadi bila masalah multikolinieritas disajikan dan ada lebih dari dua variabel bebas. Umumnya elemen diagonal matriks $\boldsymbol{C} = [\boldsymbol{X^T X}]^{-\mathbf{1}}$ dapat ditulis sebagai berikut :

$$C_{jj} = \frac{1}{1 - R_j^2}$$

R_j^2 dihasilkan dari meregresikan X_1 pada variabel bebas lainnya. Konsekuensinya kita biasa menyebut

$$VIF(\hat{\beta}_j) = \frac{1}{1 - R_j^2} \tag{2.7}$$

Varians inflasion factor (VIF) untuk $\hat{\beta}_j$ ini adalah ukuran penting dalam perkiraan multikolinieritas.

Menurut Sumodiningrat (1998) dalam hal terdapat multikolinearitas sempurna, penaksir dengan kuadrat terkecil bisa menjadi tak tentu dan variansi serta standar deviasinya menjadi tak terhingga. Sedangkan jika multikolinearitas tinggi, tetapi tidak sempurna maka konsekuensinya adalah sebagai berikut :

a. Meskipun penaksir melalui kuadrat terkecil biasa didapatkan, standar deviasinya cenderung besar jika derajat kolinearitas antarvariabel bertambah.

b. Karena standar deviasi besar, internal kepercayaan bagi parameter populasi yang relevan akan menjadi besar.

c. Taksiran-taksiran parameter kuadrat terkecil biasa dan standar deviasi akan menjadi sangat sensitif terhadap perubahan.

d. Jika multikolinearitas tinggi, mungkin bisa tinggi namun tidak satu pun (sangat sedikit) taksiran koefisien regresi yang signifikan secara statistik.

2.7 Metode Kuadrat Terkecil (MKT)

Menurut Montgomery (2006) metode kuadrat terkecil merupakan suatu metode yang digunakan untuk mengestimasi $\beta_0, \beta_1, \beta_{2,...,} \beta_k$ dengan cara meminimumkan jumlah kuadrat galat. Parameter $\beta_0, \beta_1, \beta_{2,...,} \beta_k$ tidak diketahui dan perlu dicari nilai estimasinya.

Dari persamaan umum regresi linear berganda dapat ditulis :

$$Q(\beta_j) = \sum_{i=1}^{n} ei^2 = \sum_{i=1}^{n} (\text{yi} - \beta_0 - \beta_1 x_{i1} - \beta x_{i2} - \cdots - \beta_k x_{ik})^2 \qquad (2.8)$$

Pada notasi matriks jumlah kuadrat galat e_i^2 dapat ditulis sebagai berikut

$$\mathbf{e_i^T}\,\mathbf{e_i} = [\,e_1 \quad e_2 \quad \ldots \quad e_i\,]\begin{pmatrix} e1 \\ e2 \\ ei \end{pmatrix} = e_1^2 + e_2^2 + \ldots + e_i^2 = e_i^2$$

Berdasarkan persamaan umum regresi linear berganda dengan matriks diperoleh

$$\boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{X\beta}$$

Oleh karena itu, perkalian matriks galat menjadi :

$$\mathbf{e_i^T\,e_i = (Y - X\beta)^T\,(Y - X\beta)}$$

$$\mathbf{e_i^T\, e_i = Y^TY - Y^T\beta - X^T\beta^T + X^T\beta^TX}$$

(karena $\mathbf{X^T\beta^TY = Y^TX\beta}$)

$$\mathbf{e_i^T\, e_i = Y^TY - 2Y^T\beta^TY + X^T\beta^TX\beta}$$

Untuk mencari nilai-nilai β yaitu dengan meminimumkan jumlah kuadrat galat, kemudian dicari turunan dari $Q(\beta_j)$ secara parsial terhadap β_j , j = 1,2, . . . , k dan disamakan dengan nol, sehingga diperoleh persamaan normal :

$$\frac{\partial Q}{\partial \beta o} = -2\ \sum_{i=1}^{n}(\mathrm{yi} - \beta_0 - \beta_1 x_{i1} - \beta x_{i2} - \ldots - \beta_k x_{ik}) = 0$$

$$\frac{\partial Q}{\partial \beta 1} = -2\sum_{i=1}^{n}\ (\mathrm{yi} - \beta_0 - \beta_1 x_{i1} - \beta_2\, x_{i2} - \ldots - \beta_k\, x_{ik} x_{i1}) = 0$$

$$\frac{\partial Q}{\partial \beta 2} = -2\sum_{i=1}^{n}\ (\mathrm{yi} - \beta_0 - \beta_1 x_{i1} - \beta_2\, x_{i2} - \ldots - \beta_k\, x_{ik} x_{i2}) = 0$$

$$\frac{\partial Q}{\partial \beta k} = -2\sum_{i=1}^{n}\ (\mathrm{yi} - \beta_0 - \beta_1 x_{i1} - \beta_2\, x_{i2} - \ldots - \beta_k\, x_{ik} x_{ik}) = 0$$

Setelah disusun kembali dan mengganti semua parameter dengan estimatornya, maka sistem persamaan diatas dapat ditulis :

$$nb_0 + b_1\sum_{i=1}^{n} x_{i1} + b_2\sum_{i=1}^{n} x_{i2} + \ldots + b_k\sum_{i=1}^{n} x_{ik} = \sum_{i=1}^{n} Yi$$

$$b_0\sum_{i=1}^{n} x_{i1} + b_1\sum_{i=1}^{n} x_{i2}^2 + b_2\sum_{i=1}^{n} x_{i2}x_{i1} \ldots + b_k\sum_{i=1}^{n} x_{ik}\, x_{i1} = \sum_{i=1}^{n} x_{i1}\ y_i$$

$$b_0\sum_{i=1}^{n} x_{i2} + b_1\sum_{i=1}^{n} x_{i2}x_{i1} + b_2\sum_{i=1}^{n} x_{i2}^2 \ldots + b_k\sum_{i=1}^{n} x_{ik}\, x_{i2} = \sum_{i=1}^{n} x_{i2}\ y_i$$

$$b_0\sum_{i=1}^{n} x_{ik} + b_1\sum_{i=1}^{n} x_{i1}\, x_{ik} + b_2\sum_{i=1}^{n} x_{i2}\, x_{ik} \ldots + b_k\sum_{i=1}^{n} Xik^2 = \sum_{i=1}^{n} x_{ik}\, x_{ik}$$

persamaan ini merupakan persamaan normal. Jika ditulis dalam bentuk matriks maka bentuknya menjadi

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i2} & \dots & \sum_{i=1}^{n} X_{ik} \\ \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i1}^2 & \sum_{i=1}^{n} X_{i2}X_{i1} & \dots & \sum_{i=1}^{n} X_{ik}X_{i1} \\ \sum_{i=1}^{n} X_{i2} & \sum_{i=1}^{n} X_{i1}X_{i2} & \sum_{i=1}^{n} X_{i2}^2 & \dots & \sum_{i=1}^{n} X_{ik}X_{i2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} X_{ik} & \sum_{i=1}^{n} X_{i1}X_{ik} & \sum_{i=1}^{n} X_{i2}X_{ik} & \dots & \sum_{i=1}^{n} X_{ik}^2 \end{bmatrix} \begin{bmatrix} b_o \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & \dots & X_{n1} \\ X_{12} & X_{22} & \dots & X_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{ik} & X_{2k} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}$$

Atau secara lengkap jika ditulis kedalam bentuk matriks menjadi

$$[\boldsymbol{X^T X}]\widehat{\boldsymbol{\beta}} = [\boldsymbol{X^T Y}]$$

Pada persamaan diatas kedua ruasnya dikalikan invers dari matriks $(\boldsymbol{X^T X})^{-\mathbf{1}}$**,** sehingga diperoleh :

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

$$\boldsymbol{I}\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T Y}$$

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

sehingga diperoleh estimator untuk MKT adalah

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T Y} \tag{2.9}$$

Sifat-sifat penduga metode kuadrat terkecil adalah sebagai berikut :

1. $\hat{\beta}$ linear

   $\hat{\beta}$ linear jika $\hat{\beta}$ merupakan fungsi linear dari β

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= (\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T Y} \\ &= (\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T}(\boldsymbol{X\beta} + \boldsymbol{\varepsilon}) \\ &= (\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T X\beta} + (\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T \varepsilon} \\ &= \boldsymbol{I\beta} + (\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T \varepsilon} \end{aligned}$$

2. $\hat{\beta}$ tak bias

   $\hat{\beta}$ adalah penduga tak bias jika $E(\hat{\beta}) = \beta$

$$\begin{aligned} \boldsymbol{E}(\widehat{\boldsymbol{\beta}}) &= \boldsymbol{E}((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}) \\ &= \boldsymbol{E}((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{E}((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}) \\ &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{E}(\boldsymbol{\varepsilon}) \\ &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta} \end{aligned}$$

   Sehingga $\hat{\beta}$ merupakan penduga tak bias dari β

3. $\hat{\beta}$ memiliki variansi minimum

$$\begin{aligned} \boldsymbol{var}(\widehat{\boldsymbol{\beta}}) &= \boldsymbol{E}[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\boldsymbol{T}}] \\ &= \boldsymbol{E}[(\,(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} - \boldsymbol{\beta})((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} - \boldsymbol{\beta})^{\mathrm{T}}] \\ &= \boldsymbol{E}[\,(\,(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) - \boldsymbol{\beta})\,((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) - \boldsymbol{\beta})^{\mathrm{T}}] \\ &= \boldsymbol{E}[((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}) - \boldsymbol{\beta})((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} + \\ &\quad (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}) - \boldsymbol{\beta})^{\mathrm{T}}] \\ &= \mathbf{E}\,[(\boldsymbol{\beta} + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon} - \boldsymbol{\beta})\,(\boldsymbol{\beta} + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon} - \boldsymbol{\beta})^{\mathrm{T}}] \\ &= \mathbf{E}\,[((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon})\,((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon})^{\mathrm{T}}] \\ &= \mathbf{E}\,[((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon})\,((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon})^{\mathrm{T}}] \\ &= \mathbf{E}\,[((\boldsymbol{X}^T\boldsymbol{X})^{-1}\,\boldsymbol{X}^T\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\,\boldsymbol{X}\,(\boldsymbol{X}^T\boldsymbol{X})^{-1}] \\ &= \mathbf{E}\,[((\boldsymbol{X}^T\boldsymbol{X})^{-1}\,\boldsymbol{X}^T\,\boldsymbol{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)\,\boldsymbol{X}\,(\boldsymbol{X}^T\boldsymbol{X})^{-1}] \\ &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\,\boldsymbol{X}^T\boldsymbol{X}\,(\boldsymbol{X}^T\boldsymbol{X})^{-1}\,\boldsymbol{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] \\ &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{\sigma}^2 \end{aligned}$$

$Var(\hat{\beta}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\,\sigma^2$ merupakan varians terkecil dari semua penaksir linear tak bias, hal ini dijamin dengan teorema Gauss-Markov. Untuk menunjukkan bahwa $Var(\hat{\beta})$ adalah varians yang paling minimum. Maka akan diasumsikan penduga lain yang linear dan tak bias, kemudian dibuktikan bahwa variansinya lebih besar dari $Var(\hat{\beta})$.

Misalkan $\hat{\beta}^*$ adalah penduga yang linier dan tak bias bagi β. Asumsikan bahwa :

$$\widehat{\boldsymbol{\beta}}^* = ((\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T} + \boldsymbol{Z})\boldsymbol{y}$$

dimana Z adalah matriks konstanta $(k \times n)$ yang diketahui

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^* &= ((\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T} + \boldsymbol{Z})\boldsymbol{y} \\
&= ((\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T} + \boldsymbol{Z})(\boldsymbol{X\beta} + \boldsymbol{\varepsilon}) \\
&= \left((\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T X\beta} + (\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T \varepsilon} + \boldsymbol{Z}(\boldsymbol{X\beta} + \boldsymbol{\varepsilon})\right) \\
&= ((\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T X\beta} + (\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T \varepsilon} + \boldsymbol{ZX\beta} + \boldsymbol{Z\varepsilon}) \\
&= (\boldsymbol{I\beta} + (\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T \varepsilon} + \boldsymbol{ZX\beta} + \boldsymbol{Z\varepsilon})
\end{aligned}$$

sehingga

$$\begin{aligned}
&\boldsymbol{E}(\widehat{\boldsymbol{\beta}}^*) = \boldsymbol{E}(\boldsymbol{\beta} + (\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T \varepsilon} + \boldsymbol{ZX\beta} + \boldsymbol{Z\varepsilon}) \\
&= \boldsymbol{\beta} + \boldsymbol{ZX\beta} \qquad (karena\ E(\varepsilon) = 0) \\
&= (\mathbf{1} + \boldsymbol{ZX})\boldsymbol{\beta}
\end{aligned}$$

agar $\hat{\beta}^*$ estimasi tak bias dari β maka $ZX = 0$, sehingga :

$$Var(\hat{\beta}^*) = E\left((\hat{\beta}^* - \beta)(\hat{\beta}^* - \beta)^T\right)$$

dengan

$$(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}) = (\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T \varepsilon} + \boldsymbol{ZX\beta} + \boldsymbol{Z\varepsilon}$$

dan diasumsikan bahwa $E(\varepsilon\varepsilon^T) = \sigma_u^2 I_u$. Karena $\boldsymbol{ZX\beta} = \mathbf{0}$ sehingga

$$\begin{aligned}
\boldsymbol{Var}(\widehat{\boldsymbol{\beta}}^*) &= \boldsymbol{E}\left(((\boldsymbol{X^TX})^{-1}\boldsymbol{X^T\varepsilon} + \boldsymbol{Z\varepsilon})((\boldsymbol{X^TX})^{-1}\boldsymbol{X^T\varepsilon} + \boldsymbol{Z\varepsilon})^{\boldsymbol{T}}\right) \\
&= \boldsymbol{E}\left(((\boldsymbol{X^TX})^{-1}\boldsymbol{X^T\varepsilon} + \boldsymbol{Z\varepsilon})(\boldsymbol{\varepsilon^TZ^T} + \boldsymbol{\varepsilon^TX}(\boldsymbol{X^TX})^{-1})\right) \\
&= \boldsymbol{E}\left(((\boldsymbol{X^TX})^{-1}\boldsymbol{X^T} + \boldsymbol{Z})\boldsymbol{\varepsilon\varepsilon^T}(\boldsymbol{Z^T} + \boldsymbol{X}(\boldsymbol{X^TX})^{-1})\right) \\
&= \left(((\boldsymbol{X^TX})^{-1}\boldsymbol{X^T} + \boldsymbol{Z})\boldsymbol{\sigma_u^2 I_u}(\boldsymbol{Z^T} + \boldsymbol{X}(\boldsymbol{X^TX})^{-1})\right) \\
&= \boldsymbol{\sigma^2}\left(((\boldsymbol{X^TX})^{-1}\boldsymbol{X^T} + \boldsymbol{Z})(\boldsymbol{Z^T} + \boldsymbol{X}(\boldsymbol{X^TX})^{-1})\right) \\
&= \boldsymbol{\sigma^2}((\boldsymbol{X^TX})^{-1}\boldsymbol{X^TZ^T} + (\boldsymbol{X^TX})^{-1}\boldsymbol{X^TX}(\boldsymbol{X^TX})^{-1} + \boldsymbol{ZZ^T} + \boldsymbol{ZX}(\boldsymbol{X^TX})^{-1}) \\
&= \boldsymbol{\sigma^2}((\boldsymbol{X^TX})^{-1}\boldsymbol{X^TZ^T} + \boldsymbol{I}(\boldsymbol{X^TX})^{-1} + \boldsymbol{ZZ^T} + \boldsymbol{ZX}(\boldsymbol{X^TX})^{-1})
\end{aligned}$$

karena $\boldsymbol{ZX} = \boldsymbol{X^TZ^T} = \mathbf{0}$ maka

$$\boldsymbol{Var}(\widehat{\boldsymbol{\beta}}^*) = \boldsymbol{\sigma^2}((\boldsymbol{X^TX})^{-1} + \boldsymbol{ZZ^T})$$

Matriks $\boldsymbol{ZZ^T}$ adalah definit positif, karena semua unsur diagonalnya berbentuk kuadrat. Jadi terbukti bahwa variansi dari setiap unsur dari vektor $\hat{\beta}^*$ selalu lebih besar, atau paling kecil sama dengan variansi unsur $\hat{\beta}$ yang sesuai. Estimator kuadrat terkecil yang memenuhi sifat linear, tak bias, dan mempunyai variansi minimum ini bersifat *Best Linear Unbiased Estimator* (*BLUE*).

2.8 Ukuran Pemusatan dan Penskalaan (*centering and scaling*)

Menurut Kutner (2005) pemusatan dan penskalaan data merupakan bagian dari membakukan (*standardized*) variabel. Modifikasi sederhana dari pembakuan atau standarisasi variabel ini adalah transformasi korelasi (*correlation transformation*). Pemusatan merupakan perbedaan antara masing-masing pengamatan dan rata-rata dari semua pengamatan untuk variabel. Sedangkan penskalaan meliputi gambaran pengamatan pada kesatuan (unit) standar deviasi dari pengamatan untuk variabel.

Dalam hal ini yang akan dibakukan (distandarisasi) adalah model regresi linear berganda yang ditujukkan pada model di bawah ini

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \tag{2.10}$$

Berikut ini merupakan pembakuan variabel terikat Y dan variabel bebas $X_1, X_2, \ldots, X_k$

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_Y} \text{ dengan } S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}} \tag{2.11}$$

$$X_j^* = \frac{X_{ij} - \bar{X}_j}{S_{X_j}} \text{ dengan } S_{X_j} = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{ij} - \bar{X}_j\right)^2}{n-1}} \tag{2.12}$$

untuk $i = 1,2, \ldots, n$ dan $j = 1,2, \ldots, k$

dengan

$\bar{Y}$ = rata-rata dari Y

$\bar{X}_j$ = rata-rata dari pengamatan X_j

S_Y = standar deviasi dari Y

S_{X_j} = standar deviasi dari X_j

Model regresi berganda terstandarisasi adalah tranformasi dari model regresi berganda (didefinisikan sebagai transformasi korelasi)

$$Y_i^* = \beta_1^* X_{1i}^* + \beta_2^* X_{2i}^* + \cdots + \beta_k^* X_{ki}^* + \varepsilon_i^* \tag{2.13}$$

Model di atas disebut sebagai model regresi yang baku (*standardized regression model*). Diantara parameter $\beta_1^*,\ \beta_2^*, \dots, \beta_k^*$ pada model regresi baku, dengan parameter $\beta_1, \beta_2, \dots, \beta_k$ dengan model regresi linear berganda terdapat suatu hubungan linear. Hubungan antara kedua parameter dari dua model yang berbeda tersebut dijabarkan seperti di bawah ini.

$$\beta_j = \left(\frac{S_Y}{S_{X_j}}\right)\beta_j^* \ , \qquad j = 1,2, \dots , k \tag{2.14}$$

$$\beta_0 = \bar{Y} - \beta_1\bar{X}_1 - \beta_2\bar{X}_2 - \cdots - \beta_k\bar{X}_k \tag{2.15}$$

$$\beta_0 = \bar{Y} - \sum_{j=1}^{k} \beta_j\bar{X}_j \tag{2.16}$$

prosedur ini disebut dengan prosedur penskalaan. Dari persamaan (2.8) di atas dapat dibentuk menjadi :

$$Y_i = \beta_0 + \beta_1(X_{1i} - \bar{X}_1) + \beta_1\bar{X}_1 + \beta_2(X_{2i} - \bar{X}_2) + \beta_2\bar{X}_2 + \cdots + \beta_k(X_{ki} - \bar{X}_i) + \beta_k\bar{X}_k + \varepsilon_i$$

$$Y_i = (\beta_0 + \beta_1\bar{X}_1 + \cdots + \beta_k\bar{X}_k) + \beta_1(X_{1i} - \bar{X}_1) + \cdots + \beta_k(X_{ki} - \bar{X}_i) + \varepsilon_i \tag{2.17}$$

berdasarkan persamaan (2.13) maka berlaku :

$$\bar{Y} = \beta_0 + \beta_1\bar{X}_1 + \beta_2\bar{X}_2 + \cdots + \beta_k\bar{X}_k$$

sehingga

$$Y_i - \bar{Y} = (\beta_0 + \beta_1\bar{X}_1 + \cdots + \beta_k\bar{X}_k) + (\beta_1(X_{1i} - \bar{X}_1) + \cdots + \beta_k(X_{ki} - \bar{X}_i) + \varepsilon_i) - (\beta_0 + \beta_1\bar{X}_1 + \beta_2\bar{X}_2 + \cdots + \beta_k\bar{X}_k)$$

$$Y_i - \bar{Y} = \beta_1(X_{1i} - \bar{X}_1) + \beta_2(X_{2i} - \bar{X}_2) + \cdots + \beta_k(X_{ki} - \bar{X}_i) + \varepsilon_i \tag{2.18}$$

jika

$$y_i = Y_i - \bar{Y}$$

$$x_{1i} = X_{1i} - \bar{X}_1$$

$$x_{2i} = X_{2i} - \bar{X}_2$$

$$x_{ki} = X_{ki} - \bar{X}_k$$

maka didapat model baru yaitu :

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i \qquad (2.19)$$

Prosedur untuk membentuk model pertama menjadi model terakhir disebut dengan prosedur pemusatan. Prosedur ini mengakibatkan hilangnya β_0 (intercept) yang membuat perhitungan untuk mencari model regresi menjadi lebih sederhana. Keseluruhan dari prosedur di atas disebut prosedur pemusatan dan penskalaan.

2.9 Regresi *Robust*

Menurut Chen (2002) regresi *robust* adalah salah satu penduga regresi yang *robust* atau resisten dalam menganalisis data yang menyimpang terhadap asumsi analisis regresi. Beberapa penyimpangan terhadap asumsi yang dimaksud misalnya galat yang tidak berdistribusi normal atau adanya pencilan yang mempengaruhi model. Metode ini dibutuhkan karena metode kuadrat terkecil yang dianggap penduga terbaik dalam analisis regresi ternyata peka terhadap data yang menyimpang dari asumsi. Prosedur *robust* ditujukan untuk memberikan dugaan yang lebih tepat dan cepat terhadap data yang melanggar asumsi dengan cara meniadakan identifikasi adanya data pencilan, serta bersifat otomatis dalam menanggulangi data pencilan.

Menurut Chen (2002) regresi *robust* dapat mengatasi pencilan tanpa menghapus data pencilan tersebut. Regresi *robust* berperan sebagai penurun bobot data pencilan. Dalam mendeteksi pencilan, metode regresi *robust* yang sering digunakan adalah estimasi MM. Metode-metode estimasi dalam regresi robust diantaranya adalah :

1. Estimasi M (*maximum likelyhood type*) yang dikenalkan oleh Huber (1973) adalah metode yang sederhana baik dalam penghitungan maupun secara teoritis. Estimasi ini menganalisis data dengan mengasumsikan bahwa sebagian besar terdeteksi pencilan pada variable independen.
2. Estimasi LTS (*least trimmed squares*) adalah metode dengan *high breakdown point* yang dikenalkan oleh Rousseeuw. *Breakdown point* adalah ukuran proporsi minimal dari banyaknya data yang terkontaminasi pencilan dibandingkan seluruh data pengamatan.
3. Estimasi S (*scale*) juga merupakan metode dengan *high breakdown point* yang dikenalkan oleh Rousseeuw and Yohai. Dengan nilai *breakdown* yang sama, metode ini mempunyai efisiensi yang lebih tinggi dibanding estimasi LTS.
4. Estimasi MM (*method of moment*), dikenalkan oleh Yohai (1987). Metode ini menggabungkan estimasi S (estimasi dengan *high breakdown point*) dan estimasi M.

2.10 Penduga M (*Maximum Likelihood Type*)

Menurut Hampel dalam tulisan Ali dan Qadir (2005) istilah penduga M termasuk jenis penduga *maximum likelihood*. Penduga M menggunakan pendekatan yang sederhana antara komputasi dan teoritis.

Menurut Montgomery (2006), pada prinsipnya estimasi-M merupakan estimasi yang meminimumkan suatu fungsi sisaan ρ

$$\hat{\beta}_M = \min_{\beta} \sum_{i=1}^{n} \rho(e_i) = \min_{\beta} \sum_{i=1}^{n} \rho\left(y_i - \sum\nolimits_{j=0}^{k} x_{ij}\beta_j\right) \qquad (2.20)$$

Untuk memperoleh persamaan (2.20), yaitu dengan menyelesaikan persamaan

$$\min_{\beta} \sum_{i=1}^{n} \rho(u_i) = \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{e_i}{\sigma}\right) = \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\sigma}\right) \qquad (2.21)$$

dengan dipilih estimasi untuk σ adalah

$$\hat{\sigma} = \frac{MAD}{0.6745} = \frac{median|e_i - median(e_i)|}{0.6745}$$

dengan

e_i = residual ke-i.

$\rho(e_i)$ = fungsi simetris dari residual atau fungsi yang memberikan kontribusi pada masing-masing residual pada fungsi objektif.

$\hat{\sigma}$ = skala

Fungsi ρ yang digunakan adalah fungsi objektif *tukey bisquare*

$$\rho(u_i) = \begin{cases} \frac{u_i^2}{2} - \frac{u_i^4}{2c^2} + \frac{u_i^6}{6c^4} & , |u_i| \leq c \\ \frac{c^2}{6} & , |u_i| > c \end{cases}$$

Untuk meminimumkan persamaan (2.20), dicari turunan parsial pertama dari $\hat{\beta}_M$ terhadap β sehingga diperoleh persamaan

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

$$\sum_{i=1}^{n} x_{ij}\psi \left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}} \right) = 0 \quad , j = 0,1, \dots, k \qquad (2.22)$$

dengan $\psi = \rho'$ dan x_{ij} adalah observasi ke-i pada variabel bebas ke-j dan $x_{i0} = 1$.

Draper dan Smith (1992) memberikan penyelesaian persamaan (2.22), yaitu dengan mendefinisikan suatu fungsi pembobot

$$w(e_i) = \frac{\psi \left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}} \right)}{\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}} \right)} \qquad (2.23)$$

karena nilai $u_i = \frac{e_i}{\hat{\sigma}}$ sebagai pengganti e_i , maka persamaan (2.23) menjadi

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

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

dengan demikian persamaan (2.22) menjadi

$$\sum_{i=1}^{n} x_{ij} w_i \left(y_i - \sum\nolimits_{j=0}^{k} x_{ij}\beta_j \right) = 0 \quad , j = 0,1, \dots, k \tag{2.24}$$

Persamaan (2.24) dapat diselesaikan dengan metode MKT terboboti secara iterasi yang dinamakan *Iteratively Reweighted Least Square* (IRLS). Untuk menggunakan IRLS, diasumsikan bahwa suatu estimasi awal $\hat{\beta}^0$ ada dan $\hat{\sigma}_i$ suatu estimasi skala. Maka persamaan (2.24) menjadi

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

Jika dibuat kedalam notasi matriks menjadi :

$$\boldsymbol{X^T W_i X \beta} = \boldsymbol{X^T W_i Y} \tag{2.25}$$

dengan W_i adalah matriks berukuran n x n dengan elemen-elemen diagonal yang berisi pembobot.

Penyelesaian persamaan tersebut akan memberikan estimator untuk $\hat{\beta}$ yaitu :

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{M}} = (\boldsymbol{X^T W_i X})^{-\mathbf{1}} (\boldsymbol{X^T W_i Y}) \tag{2.26}$$

Pada fungsi pembobot *tukey bisquare*, konstanta yang digunakan adalah c=4,685. Pemilihan nilai c=4,685 pada estimasi-M bertujuan menghasilkan estimasi dengan 95% efisiensi dibandingkan metode kuadrat terkecil (Franke, 1984).

Menurut Susanti (2014) untuk mendapatkan dugaan parameter diperlukan solusi iterasi yang disebut IRLS (*iteratively reweighted least squares*), iterasi dilakukan sampai diperoleh suatu nilai yang konvergen. Algoritma perhitungan untuk mendapatkan nilai estimasi *robust-m* yaitu :

a. Menghitung estimasi parameter dengan MKT.
b. Menghitung nilai residual $e_i = y_i - \hat{y}_i$.
c. Menghitung nilai $\hat{\sigma}_i = \frac{median|e_i - median(e_i)|}{0.6745}$.
d. Menghitung nilai $u_i = \frac{e_i}{\hat{\sigma}_i}$.
e. Menghitung pembobot

$$w_i = \begin{cases} \left[1 - \left(\frac{u_i}{4.685}\right)^2\right]^2 & , \quad |u_i| \leq 4.685 \\ 0 & , \quad |u_i| > 4.685 \end{cases}$$

f. Menghitung $\hat{\beta}_M$ dengan metode WLS dengan pembobot w_i.
g. Mengulangi langkah b-f sampai diperoleh nilai $\hat{\beta}_M$ yang konvergen, dengan syarat nilai $\hat{\beta}_{MM}^{(l)} - \hat{\beta}_{MM}^{(l-1)} < 10^{-6}$, dengan l adalah banyaknya iterasi.

2.11 Penduga S (*Scale*)

Jika data terkontaminasi pencilan pada variabel X (prediktor), penduga M tidak dapat bekerja dengan baik. Penduga M tidak dapat mengidentifikasi *bad observation* yang berarti tidak dapat membedakan *good leverage point* dan *bad leverage point*. Untuk mengatasi hal tersebut, estimasi *high breakdown* sangat diperlukan (Chen, 2002).

Menurut Franke (1984), estimasi-S pertama kali diperkenalkan oleh Rousseeuw dan Yohai, dan dinamakan estimasi-S karena estimasi ini berdasarkan pada skala sisaan dari estimasi-M. Estimasi-S didefinisikan sebagai

$$\hat{\beta}_s = \min_{\beta} \ \hat{\sigma}_s\,(e_1, e_2, e_3, \dots, e_i)$$

dengan menentukan nilai estimator skala *robust* $\hat{\sigma}_s$ yang minimum dan memenuhi

$$\min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta}{\hat{\sigma}_s}\right) \tag{2.27}$$

dengan

$$\hat{\sigma}_s = \sqrt{\frac{1}{nK}\sum_{i=1}^{n} w_i e_i^2}$$

untuk nilai K=0.199 , dan dipillih estimasi awal

$$\hat{\sigma}_s = \frac{median|e_i - median(e_i)|}{0.6745}$$

Penyelesaian persamaan (2.27) adalah dengan cara mencari turunannya terhadap β sehingga diperoleh

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

$$\sum_{i=1}^{n} x_{ij}\psi \left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta}{\hat{\sigma}_s} \right) = 0 \qquad ,j = 0,1,\dots,k \tag{2.28}$$

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

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

$$w_i(u_i) = \begin{cases} \left[1 - \left(\dfrac{u_i}{c}\right)^2\right]^2 & ,|u_i| \le c \\ 0 & ,|u_i| > c \end{cases}$$

dengan $u_i = \frac{e_i}{\hat{\sigma}_s}$ dan c=1,547.

Untuk fungsi pembobot *tukey bisquare,* konstanta yang digunakan adalah c=1,547. Nilai konstanta c=1,547 dipilih karena menghasilkan nilai *breakdown* sebesar 50% tetapi menghasilkan nilai efisiensi sebesar 75,9% (Franke, 1984).

Menurut Susanti (2014) penyelesaian persamaan tersebut akan memberikan estimator untuk $\hat{\beta}$ yaitu :

$$\boldsymbol{\hat{\beta}_S} = (\boldsymbol{X^T W_i X})^{-\boldsymbol{1}}(\boldsymbol{X^T W_i Y}) \tag{2.29}$$

Untuk mendapatkan dugaan parameter diperlukan solusi iterasi yang disebut IRLS (*iteratively reweighted least squares*), iterasi dilakukan sampai diperoleh suatu nilai yang konvergen.

Algoritma perhitungan nilai estimasi *robust*-S

a. Menghitung estimasi parameter dengan MKT.
b. Menghitung nilai residual $e_i = y_i - \hat{y}_i$.
c. Menghitung nilai

$$\hat{\sigma}_i = \begin{cases} \dfrac{median|e_i - median(e_i)|}{0.6745} & , \quad iterasi = 1 \\ \sqrt{\dfrac{1}{nK}\sum_{i=1}^{n} w_i e_i^2} & , \quad iterasi > 1 \end{cases}$$

d. Menghitung nilai $u_i = \frac{e_i}{\hat{\sigma}_i}$.
e. Menghitung pembobot

$$w_i = \begin{cases} \left[1 - \left(\dfrac{u_i}{4.685}\right)^2\right]^2 & , \quad |u_i| \leq 1.547 \\ 0 & , \quad |u_i| > 1.547 \end{cases}$$

f. Menghitung $\hat{\beta}_S$ dengan metode WLS dengan pembobot w_i.
g. Mengulangi langkah b-f sampai diperoleh nilai $\hat{\beta}_S$ yang konvergen, dengan syarat nilai $\hat{\beta}_S^{(l)} - \hat{\beta}_S^{(l-1)} < 10^{-6}$, dengan l adalah banyaknya iterasi.

2.12 Penduga MM (*Method of Moment*)

Estimasi MM menggabungkan estimasi *high breakdown point* dan efisiensi statistik yang dikenalkan oleh Yohai (1987). Langkah pertama dalam estimasi ini adalah mencari nilai duga dengan penduga S. Kedua menaksir penduga parameter regresi akhir dengan residual langkah pertama menggunakan metode penduga M. Penduga S menjamin nilai *breakdown point* tinggi dan penduga M membuat estimator mempunyai efisiensi tinggi. Pada umumnya digunakan fungsi *tukey bisquare* baik pada penduga S maupun estimasi M.

Menurut Susanti (2014) estimasi-MM merupakan penyelesaian dari

$$\hat{\beta}_{MM} = \sum_{i=1}^{n} x_{ij}\, \rho_1' \left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}_s} \right) = 0$$

dengan $y_i - \sum_{j=0}^{k} x_{ij}\beta_j$ adalah sisaan yang diperoleh dari estimasi parameter model regresi dengan estimasi-S dan $\hat{\sigma}_s$ merupakan penyelesaian dari

$$\frac{1}{n}\sum_{i=1}^{n} \rho_0 \left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\tilde{\beta}_j}{\hat{\sigma}_s} \right) = K$$

Penyelesaian persamaan tersebut akan memberikan estimator untuk $\hat{\beta}$ yaitu :

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{MM}} = (\boldsymbol{X^T W_i X})^{-\mathbf{1}}(\boldsymbol{X^T W_i Y})$$

Algoritma perhitungan nilai estimasi *robust*-MM

a. Menghitung nilai residual $e_i = y_i - \hat{y}_i$ dari estimasi-S.
b. Menghitung nilai $\hat{\sigma}_i = \hat{\sigma}_s$.

c. Menghitung nilai $u_i = \frac{e_i}{\hat{\sigma}_i}$.

d. Menghitung pembobot

$$w_i = \begin{cases} \left[1 - \left(\frac{u_i}{4.685}\right)^2\right]^2 & , \quad |u_i| \leq 4.685 \\ 0 & , \quad |u_i| > 4.685 \end{cases}$$

e. Menghitung $\hat{\beta}_{MM}$dengan metode WLS dengan pembobot w_i.

f. Mengulangi langkah b-e sampai diperoleh nilai $\hat{\beta}_{MM}$ yang konvergen, dengan syarat nilai $\hat{\beta}_{MM}^{(l)} - \hat{\beta}_{MM}^{(l-1)} < 10^{-6}$, dengan l adalah banyaknya iterasi.

2.13 Regresi *Ridge*

Regresi *ridge* memberikan estimasi koefisien regresi yang bias dengan memodifikasi metode kuadrat terkecil untuk mendapatkan pengurangan varian dengan menambahkan suatu tetapan k dalam menstabilkan koefisien (Mardikyan dan Cetin, 2008).

Menurut Dereny dan Rashwan (2011), *ridge* didasarkan pada penambahan konstanta bias k pada diagonal matrik $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}$, sehingga koefisien penduga *ridge* dipengaruhi oleh besarnya tetapan bias k, dimana nilai k bernilai 0 sampai 1. Dalam regresi *ridge* variabel bebas X dan variabel terikat Y ditransformasikan kedalam bentuk baku (standarisasi).

Metode regresi *ridge* pertama kali dikemukakan oleh Hoerl (1962) dan dikembangkan oleh Hoerl dan Kennard (1970). Penaksir koefisien regresi *ridge* adalah sebagai berikut

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{Ridge}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} + \boldsymbol{kI})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y} \qquad (2.30)$$

dengan

I = matriks identitas berukuran $p \; x \; p$

k = konstanta bias $0 \leq k \leq 1$

$\widehat{\beta}_{Ridge}$ = vektor parameter regresi *ridge*

2.14 Regresi *Ridge Robust*

Menurut Samkar dan Alpu (2010), regresi *ridge robust* merupakan penggabungan dari metode regresi *ridge* dan regresi *robust* yang dilakukan untuk mengatasi masalah multikolinieritas dan pencilan. Penduga regresi *ridge robust* yang dihasilkan akan stabil dan resisten terhadap pencilan. Rumus penduga parameter regresi *ridge robust* adalah sebagai berikut :

$$\widehat{\beta}_{RR} = (X^T X + k\, I)^{-1} X^T X \hat{\beta}_{Robust-MM} \qquad (2.31)$$

dengan,

$\hat{\beta}_{RR}$ = penduga parameter regresi *ridge-robust*

$\hat{\beta}_{Robust-MM}$ = penduga parameter regresi *robust-MM*

k = bilangan positif dimana $0 \leq k \leq 1$

Ada beberapa cara dalam memilih nilai konstanta k. Salah satu cara pemilihan nilai k yaitu dengan menggunakan metode yang diperkenalkan oleh Hoerl, Kennard and Balwin (HKB) (1975), didasarkan pada metode kuadrat terkecil, digunakan untuk memilih nilai k, dibangun dengan menggunakan penduga *robust-MM*,

$$k = \frac{p(\sigma^2_{Robust-MM})}{\beta^T_{Robust-MM}\ \beta_{Robust-MM}}$$

dengan,

$\hat{\beta}_{Robust-MM}$ = penduga parameter regresi *robust-MM*

p = banyaknya variabel bebas

$\sigma^2_{Robust-MM} = \frac{(y-x\widehat{\beta}_{Robust-MM})^T(y-x\widehat{\beta}_{Robust-MM})}{n-p}$

2.15 *Average Mean of Squares Error* (AMSE)

Menurut Ghozali (2006), MSE merupakan salah satu pengukuran kesalahan yang populer dan mudah digunakan. Nilai MSE dihitung dengan mengkuadratkan selisih antara ramalan dengan nilai aktual. Umumnya, semakin kecil MSE semakin akurat nilai suatu peramalan atau suatu pemodelan. Selain itu dalam kasus multikolinearitas metode terbaik diartikan sebagai metode yang dapat melakukan perbaikan masalah multikolinearitas lebih baik dari yang metode yang lain. Efisiensi dari metode untuk menangani multikolinearitas ini akan dievaluasi berdasarkan rata-rata dari *Mean Square Error* (MSE) dan *Average Mean Square Error* (AMSE) dari hasil estimasi parameter $\hat{\beta}$, yang didefinisikan sebagai berikut

$$\text{MSE}(\hat{\beta}) = \frac{1}{m}\sum_{j=1}^{m}(\hat{\beta}_j - \beta)^2\ ; \quad j = 1,2,\ldots,m \tag{2.32}$$

$$\text{AMSE}(\hat{\beta}) = \frac{1}{m}\sum_{l=1}^{m}\left\| \widehat{\boldsymbol{\beta}}_{\boldsymbol{j}} - \boldsymbol{\beta}_{\boldsymbol{j}} \right\|^2 ; \quad j = 1,2,\ldots,p \tag{2.33}$$

dengan

$\widehat{\beta}_j$ = Penduga parameter regresi

β = Parameter regresi

m = Banyaknya ulangan

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester ganjil tahun ajaran 2017/2018 bertempat di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan untuk analisis adalah data simulasi yang telah didesain sedemikian rupa sehingga memenuhi asumsi-asumsi klasik yang dibutuhkan dalam analisis regresi *ridge robust-MM* yaitu asumsi multikolinieritas menggunakan simulasi Monte Carlo dan asumsi normalitas yang didapatkan dengan menambahkan galat yang mengandung pencilan pada variabel terikat.

McDonald dan Galarneau (1975), Wichern dan Churchill (1978), Gibbon (1981), Kibria (2003) dan peneliti lain menggunakan persamaan berikut untuk membangkitkan variabel-variabel bebas yaitu

$$x_{ij} = (1-\gamma^2)^{\frac{1}{2}} z_{ij} + \gamma z_{ip} \quad , i = 1,2,\dots,n \quad , j = 1,2,\dots,p$$

dimana z_{ij} adalah pembangkit bilangan acak independen normal standar, dan γ adalah nilai tetapan sehingga korelasi antara dua variabel penjelas diberikan oleh γ^2.

Sedangkan n pengamatan untuk variabel terikat ditetapkan oleh persamaan

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i \quad , i = 1,2,3,\dots,n$$

dengan menetapkan nilai $\beta_0 = 0$ dan $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 1$

Untuk mendapatkan data yang tidak normal yaitu dengan membangkitkan galat yang berdistribusi $N(0,1)$ lalu dikontaminasi dengan pencilan $N(0,10)$.

$$e_i = G_j + P_k$$

$$G_j = N(0,1)$$

$$P_k = N(0,10)$$

dimana,

$e_i =$ data bangkitan error yang telah dikontaminasi pencilan

$G_j =$ data bangkitan N(0,1) berukuran 1 × j

$P_k =$ data bangkitan pencilan N(0,10) berukuran 1 × k

Data bangkitan e_i diperoleh dengan menggabungkan data bangkitan G_j dan data bangkitan P_k, sehingga menghasilkan data bangkitan e_i yang berukuran $1 \times n$. Simulasi variabel bebas dengan Monte Carlo dibuat tetap atau dibangkitkan satu kali dengan nilai yang sama, sedangkan variabel terikat dibangkitkan sebanyak 100 kali pengulangan.

Tabel 1. Simulasi Monte Carlo

Jumlah Variabel (p)	Jumlah Pengamatan (n)	Simulasi	Keterangan
6	25	$x_{i1} = (1-\gamma^2)^{\frac{1}{2}} z_{i1} + \gamma z_{i7}$ $x_{i2} = (1-\gamma^2)^{\frac{1}{2}} z_{i2} + \gamma z_{i7}$ $\vdots$ $x_{i5} = (1-\gamma^2)^{\frac{1}{2}} z_{i5} + \gamma z_{i7}$ $x_{i6} = (1-\gamma^2)^{\frac{1}{2}} z_{i6} + \gamma z_{i7}$	$i = 1,2,\dots,n$ $j = 1,2,\dots,7$ $\gamma = 0.99$
	50		
	75		

Tabel 2. Simulasi Pencilan

Jumlah Variabel (p)	Jumlah Pencilan (o)	Simulasi	Keterangan
6	10%	$y_i = x_{i1} + x_{i2} + \cdots + x_{i6} + e_i$ $e_i = G_j + P_k$ $G_j = N(0,1)$ $P_k = N(0,10)$	O =floor($o \times n$) O =Jumlah pencilan $i = 1,2,\dots,n$ $j = 1,2,\dots,n-O$ $k = 1,2,\dots,O$
	15%		
	20%		

3.3 Metodologi Penelitian

Penelitian ini dilakukan secara studi pustaka, yaitu dengan mempelajari buku-buku teks penunjang dan karya ilmiah yang disajikan dalam bentuk jurnal. Untuk mempermudah perhitungan dan hasil yang akurat penelitian ini menggunakan software R dan Minitab. Adapun langkah-langkah yang dilakukan pada data penelitian untuk regresi *ridge robust-MM* antara lain :

1. Melakukan simulasi data
   a. Membangkitkan data variabel bebas sebanyak satu kali.
   b. Membangkitkan data variabel terikat sebanyak 100 kali pengulangan yang diikuti dengan analisis regresi *ridge robust-MM* dan MKT.
2. Melakukan uji multikolinearitas dengan melihat nilai korelasi dan nilai VIF (*Variance Inflation Factor*).
3. Melakukan pengujian normalitas menggunakan uji Kolmogorov-Smirnov pada satu sampel data ulangan.
4. Menghitung nilai $\hat{\beta}_{MM}$ dengan menggunakan metode regresi *robust-MM* dengan pembobot *tukey-bisquare*. Dengan iterasi sebagai berikut :
   a. Menghitung nilai $\hat{\beta}$ awal dengan menggunakan MKT.
   b. Menghitung nilai residual $e_i = y_i - \hat{y}_i$.
   c. Menghitung nilai

$$\hat{\sigma}_i = \begin{cases} \dfrac{median|e_i - median(e_i)|}{0.6745} & , \quad iterasi = 1 \\ \sqrt{\dfrac{1}{nK}\sum_{i=1}^{n} w_i e_i^2} & , \quad iterasi > 1 \end{cases}$$

d. Menghitung nilai $u_i = \frac{e_i}{\hat{\sigma}_i}$.

e. Menghitung pembobot

$$w_i = \begin{cases} \left[1 - \left(\dfrac{u_i}{1.547}\right)^2\right]^2 & , \quad |u_i| \leq 1.547 \\ 0 & , \quad |u_i| > 1.547 \end{cases}$$

f. Menghitung nilai $\hat{\beta}_s$ dengan metode WLS dengan pembobot w_i^0.

g. Mengulangi langkah b-f sampai diperoleh nilai $\hat{\beta}_s$ yang konvergen, dengan syarat nilai $\hat{\beta}_s^{(l)} - \hat{\beta}_s^{(l-1)} < 10^{-6}$, dengan l adalah banyaknya iterasi.

h. Nilai $\hat{\beta}_s$ digunakan sebagai nilai awal, lalu menghitung nilai residual $e_i = y_i - \hat{y}_i$.

i. Menghitung nilai $\hat{\sigma}_i = \frac{median|e_i - median(e_i)|}{0.6745}$

j. Menghitung nilai $u_i = \frac{e_i}{\hat{\sigma}_i}$

k. Menghitung pembobot

$$w_i = \begin{cases} \left[1 - \left(\dfrac{u_i}{4.685}\right)^2\right]^2 & , \quad |u_i| \leq 4.685 \\ 0 & , \quad |u_i| > 4.685 \end{cases}$$

l. Menghitung $\hat{\beta}_{MM}$ dengan metode WLS dengan pembobot w_i^0.

m. Mengulangi langkah b-f sampai diperoleh nilai $\hat{\beta}_{mm}$ yang konvergen, dengan syarat nilai $\hat{\beta}_{MM}^{(l)} - \hat{\beta}_{MM}^{(l-1)} < 10^{-6}$, dengan l adalah banyaknya iterasi.

5. Menghitung nilai $\hat{\beta}_{RR}$ pada metode regresi *ridge robust* dengan memasukkan nilai $\hat{\beta}_{MM}$ kedalam rumus penduga *ridge robust.*

$$\widehat{\beta}_{RR} = (X^T X + k\, I)^{-1} X^T X \hat{\beta}_{Robust-MM}$$

$$k = \frac{p(\sigma^2_{Robust-MM})}{\beta^T_{Robust-MM}\ \beta_{Robust-MM}}$$

6. Membandingkan nilai rata-rata $\hat{\beta}_{RR}$ dengan $\hat{\beta}_{MKT}$.
7. Menghitung nilai MSE $\widehat{\beta}_{RR}$ dan $\hat{\beta}_{MKT}$ setelah dilakukan 100 pengulangan (m) dengan menggunakan rumus

$$\text{MSE}(\hat{\beta}) = \frac{1}{m}\sum_{j=1}^{m}\left(\hat{\beta}_j - \beta\right)^2; \quad j = 1,2,\ldots,m$$

8. Membandingkan nilai MSE $\widehat{\beta}_{RR}$ dan MSE $\hat{\beta}_{MKT}$.
9. Menghitung nilai AMSE $\widehat{\beta}_{RR}$ dan $\hat{\beta}_{MKT}$ setelah dilakukan 100 pengulangan (m) dengan menggunakan rumus

$$\text{AMSE}(\hat{\beta}) = \frac{1}{m}\sum_{l=1}^{m}\left\| \widehat{\boldsymbol{\beta}}_{\boldsymbol{j}} - \boldsymbol{\beta}_{\boldsymbol{j}}\right\|^2; \quad l = 1,2,\ldots,100\,, j = 1,2,\ldots,p$$

10. Membandingkan nilai AMSE $\widehat{\beta}_{RR}$ dan AMSE $\hat{\beta}_{MKT}$.

V. KESIMPULAN

Berdasarkan hasil dan pembahasan, maka diperoleh kesimpulan sebagai berikut :

1. Koefisien Regresi *Ridge Robust-MM* lebih baik dibandingkan Metode kuadrat terkecil berdasarkan nilai MSE yang dihasilkan.
2. Semakin banyak data yang digunakan semakin kecil AMSE yang dihasilkan oleh Regresi *Ridge Robust-MM*.
3. Semakin besar persentase pencilan yang terdapat pada variabel terikat semakin kecil AMSE yang dihasilkan oleh Regresi *Ridge Robust-MM*.
4. Regresi *Ridge Robust-MM* lebih baik dibandingkan metode kuadrat terkecil.

DAFTAR PUSTAKA

Ali, A. and Qadir, M.F. 2005. A modified M-estimator for Detection of Outlier. *PJSOR*. **1**: 49-64.

Chen, C. 2002. Robust Regression and Outlier Detection with the ROBUSTREG Procedure. Statistics and Data Analysis. SUGI Paper 265-27. SAS Institute, North Carolina.

Dereny, M. El. and Rashwan, N.I. 2011. Solving Multicollinearity Problem Using Ridge Regression Models. *Int. J. Contemp. Math. Sciences*. **6**(12): 585-600.

Draper, N.R. and Smith, H. 1992. *Analisis Regresi Terapan.* Ed. Ke-2. Diterjemahkan oleh Bambang Sumantri. Gramedia Pustaka Utama, Jakarta.

Franke, J., *et al*. 1984. Robust *and Nonlinear Time Series Analysis*. Springer-Verlag, New York.

Ghozali, I. 2006. *Aplikasi Analisis Multivariate dengan Pogram SPSS.* Ed. Ke-4. Badan Penerbit Universitas Diponegoro, Semarang.

Gujarati, D. 1995. *Ekonometri Dasar.* Diterjemahkan oleh Sumarno Zain. Erlangga, Jakarta.

Hoerl, A.E. and Kennard, R.W. 1970. Ridge Regression: Biased Estimator to Nonorthogonal Problems. *Technometrics*. **12**(1): 68-82.

Kutner, M.H., *et al*. 2005. *Applied linear Statistic Model*. Ed. Ke-5. Mc-Graw-hill, New York.

Mardikyan, S. and Cetin, E. 2008. Efficient Choice of Biasing Constant for Ridge Regression. *Int. J. Contemp. Math. Sciences*. **3**(11): 527-536.

Marona, R.A. 2011. Robust Ridge Regression for High-Dimensional Data. *Technometrics.* **53(1)**: 44-53.

Montgomery, D.C., *et al*. 2006. *Introduction to Linear Regression Analysis.* Wiley and Sons, Inc., New York.

Myers, R.H. 1990. *Clasical and Modern Regression With Application.* PWSKENT publishing Company, Boston.

Samkar, H. and Alpu, O. 2010. Ridge Regression Based on Some Robust Estimators. *Journal of Modern Applied Statistical Methodes*. **9**: 495-501.

Sumodiningrat, G. 1998. *Ekonometrika Pengantar*. BPFE, Yogyakarta.

Susanti, Y., *et al.* 2014. M Estimation, S Estimation and MM Estimation in Robust Regression. *International Journal of Pure and Applied Mathematics*. **3**(91): 349-360.

Usman, M. dan Warsono. 2001. *Teori Model Linear dan Aplikasinya.* C.V. Darmajaya, Bandar Lampung.

Yohai, V.J. 1987. High Breakdown-point and High Efficiency Robust Estimates for Regression. *The Annals of Statistics.* **15**(20): 642-656.