ANALISIS OUTLIER DAN HETEROSKEDASTISITAS DENGAN MENGGUNAKAN REGRESI ROBUST WEIGHT LEAST SQUARE

Skripsi

Disusun sebagai salah satu syarat

untuk memperoleh gelar Sarjana Sains

Program Studi Matematika

oleh

Novian Bayu Hartono

4111412055

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI SEMARANG

2016

PERNYATAAN

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

Semarang, Desember 2016

METERAI TEMPEL

6000

Novian Baru Hartono

4111412055

UNNES
UNIVERSITAS NEGERI SEMARANG

PENGESAHAN

Skripsi yang berjudul

Analisis *Outlier* dan Heteroskedastisitas dengan Menggunakan Regresi Robust *Weight Least Square*

Disusun oleh

Novian Bayu Hartono

4111412055

Sudah dipertahankan di hadapan sidang Panitia Ujian Skripsi FMIPA UNNES pada tanggal 19 - 12 - 2016

Panitia:

Ketua	Sekretaris
Prof. Dr. Zaenuri, S.E., M.Si, Akt	Drs. Arief Agoestanto, M.Si
NIP. 196412231988031001	NIP. 196807221993031005

Ketua Penguji

Dra. Sunarmi, M.Si

NIP. 19550624198832001

Anggota Penguji/	Anggota Penguji/
Pembimbing 1	Pembimbing 2
Prof. Dr. Zaenuri, S.E., M.Si, Akt	Drs. Sugiman, M.Si
NIP. 196412231988031001	NIP. 196401111989011001

UNNES
UNIVERSITAS NEGERI SEMARANG

MOTTO DAN PERSEMBAHAN

MOTTO

- *Cukuplah Allah bagiku, tidak ada Tuhan selain Dia. Hanya kepada-Nya aku bertawakal. (QS. At-Taubah: 129)*
- *Saya tak mau jadi pohon bambu, saya mau jadi pohon oak yang berani menentang angin. (Soe Hok Gie)*
- *Hidup adalah anugerah yang wajib diperjuangkan*

PERSEMBAHAN

Karya ini saya persembahkan untuk:

- *Untuk kedua orang tua tercinta Ibu Sumiyati dan Bapak Wiwin Budi Hartono.*
- *Untuk adik-adikku tersayang, Annas Bagas Hartono dan Yuga Bekti Hartono.*
- *Untuk teman-teman Matematika 2012.*
- *Untuk teman-teman Manuver 2012 dan BEM FMIPA.*

UNNES
UNIVERSITAS NEGERI SEMARANG

KATA PENGANTAR

Alhamdulillah, puji syukur penulis panjatkan kehadirat Allah SWT atas limpahan rahmat serta karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul "Analisis *outlier* dan Heteroskedastisitas dengan Menggunakan Regresi Robust *Weight Least Square"*.

Skripsi ini dapat tersusun dengan baik berkat bantuan dan bimbingan banyak pihak. Oleh Karena itu, penulis menyampaikan terimakasih kepada:

1. Prof. Dr. Fathur Rokhman, M.Hum., Rektor Universitas Negeri Semarang, yang telah memberikan kesempatan kepada penulis untuk menyelesaikan studi strata di Jurusan Matematika FMIPA UNNES.
2. Prof. Dr. Zaenuri, S.E, M.Si, Akt., Dekan FMIPA Universitas Negeri Semarang, yang telah memberikan izin untuk melakukan penelitian.
3. Drs. Arief Agoestanto, M.Si., Ketua Jurusan Matematika Universitas Negeri Semarang yang telah memberikan izin untuk melakukan penelitian.
4. Prof. Dr. Zaenuri, S.E, M.Si, Akt., sebagai pembimbing I, yang telah menuntun, memberikan arahan dan bimbingan dalam penyelesaian skripsi ini.
5. Drs. Sugiman, M.Si sebagai pembimbing II, yang telah menuntun, memberikan arahan dan bimbingan dalam penyelesaian skripsi ini.
6. Dra. Sunarmi, M.Si sebagai penguji, yang telah menguji dan memberikan arahan hasil skripsi ini.
7. Drs. Mashuri M.Si selaku dosen wali, yang telah memberikan bimbingan, arahan dan nasehat selama menempuh kuliah.

8. Bapak Ibu Dosen Jurusan Matematika Universitas Negeri Semarang yang telah membekali penulis dengan berbagai ilmu selama perkuliahan.
9. Ibu dan Bapak tercinta, Ibu Sumiyati dan Bapak Wiwin Budi Hartono yang senantiasa memberikan dukungan, kesempatan, dan doa yang tiada terkira.
10. Adik-adikku tersayang, Annas Bagas Hartono dan Yuga Bekti Hartono yang selalu memberikan semangat dan doa.
11. Keluarga Besar Ibu maupun Bapak yang telah memberikan dukungan dan doa.
12. Teman terbaikku, Eka Desi Setiowati yang memberikan dukungan, semangat dan doa.
13. Teman-teman Manuver 2012, Ulil Albab, Eka Desi, Eli, Wiwik, Rahmat, Apri, Kartika, Belynda, Amalia, Diena, dan Meis yang senantiasa memberikan semangat dan menyertai perjuangan saya.
14. Teman-teman BEM FMIPA UNNES 2013 s.d 2015 yang telah memberikan motivasi dan menyertai perjuangan saya.
15. Teman-teman Mathematics Study Club (MSC) 2013 s.d 2014 yang telah memberikan semangat dan doa.
16. Teman-teman Program Studi Matematika 2012 yang telah setia menemani dan memberikan motivasinya.
17. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah membantu dalam penyelesaian skripsi ini.

Hanya ucapan terima kasih dan doa, semoga apa yang telah diberikan tercatat sebagai amal baik dan mendapatkan balasan dari Allah SWT.

UNNES
UNIVERSITAS NEGERI SEMARANG

Semoga Tugas Akhir ini dapat membawa manfaat bagi penulis sendiri khususnya dan bagi para pembaca pada umumnya.

Semarang, Desember 2016

Penulis

UNNES
UNIVERSITAS NEGERI SEMARANG

ABSTRAK

Hartono, Novian Bayu. 2016. Analisis *Outlier* dan Heteroskedastisitas dengan Menggunaka Regresi Robust *Weight Least Square.* Skripsi, Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang. Pembimbing Utama: Prof. Dr. Zaenuri, S.E, M.Si, Akt. dan Pembimbing Pendamping: Drs. Sugiman, M.Si.

Kata Kunci: Regresi, *Outlier*, Heteroskedastisitas, *Weighted Least Square*, Robust, *Least Trimmed Square*.

Regresi merupakan salah satu teknik statistika yang paling banyak digunakan dalam penelitian kuantitatif. Banyak pula studi kasus yang dapat dianalisis dalam pengembangan regresi. Tujuan dari penelitian ini adalah mengkaji data yang memiliki masalah o*utlier* atau pencilan dan heteroskedastisitas. Metode yang digunakan dalam penelitian ini adalah metode Robust *Weighted Least Square.*

Penelitian ini menggunakan data yang bersumber dari Indonesia *Capital Market Directory* (ICMD) Bank MEGA periode Januari 2012 sampai dengan Juli 2014, dengan menggunakan variabel *stock price close* (X_1), *frequaency* (X_2) dan *market capitalization* (Y). Hasil estimasi parameter dengan menggunakan metode regresi berganda diperoleh persamaan awal data $\hat{Y} = 1{,}381\text{E+7} - 370{,}937X_1 + 11098{,}157X_2$. Hasil pengujian asumsi klasik pada model tersebut dapat diketahui bahwa data berdistribusi normal, terdapat hubungan linier antar variabel bebas dengan tak bebas, koefisien regresi signifikan pada X_2, tidak terjadi multikolinearitas dan autokorelasi serta terdapat heteroskedastisitas dan *outlier* pada data. Analisis metode Robust (*Least Trimmed Square) Weighted Square* menggunakan variabel X_2 dan Y sebagai variabel yang digunakan untuk menjalankan metode ini dikarenakan berdasarkan hasil analisis asumsi awal masalah *outlier* dan heteroskedastisitas berada pada kedua variabel tersebut. Dari hasil analisis dengan metode RWLS diperoleh persamaan regresi $\hat{Y} = 1{,}426\text{E+12} - 6{,}241\text{E+9}X_2$, dengan R sebesar 0,926 dan R^2 sebesar 0,858. Hasil pengujian asumsi klasik pada model tersebut dapat diketahui bahwa data berdistribusi normal, terdapat hubungan linier antar variabel bebas dengan tak bebas, tidak terjadi heteroskedastisitas dan masalah *outlier* telah hilang.

Berdasarkan hasil penelitian dapat disimpulkan bahwa metode Robust *Weighted Least Square* merupakan metode yang dapat digunakan untuk mengatasi masalah *outlier* dan heteroskedastisitas pada suatu data.

DAFTAR ISI

Halaman

1. Halaman *Cover* i
2. Pernyataan Keaslian Tulisan ii
3. Halaman Pengesahan iii
4. Motto dan Persembahan iv
5. Prakata v
6. Abstrak viii
7. Daftar Isi ix
8. Daftar Tabel xi
9. Daftar Gambar.....xiii
10. Daftar Lampiran xiv
11. Daftar Simbol dan Istilah xv
12. Bab 1. Pendahuluan 1
    - 1.1 Latar Belakang 1
    - 1.2 Rumusan Masalah 5
    - 1.3 Tujuan Penelitian 5
    - 1.4 Manfaat Penelitian 5
    - 1.5 Sistematika Penulisan 5
13. Bab 2. Tinjauan Pustaka..... 7
    - 2.1 Regresi Linier 7
    - 2.2 *Outlier* (Pencilan) 10
    - 2.3 Regresi Robust..... 14

2.4 Heteroskedastisitas 18

2.5 *Weighted Least Square* 23

2.6 Robust (*Least Trimmed Square*) *Weighted Least Square*..... 26

2.7 Kerangka Berfikir 27

14. Bab 3. Metode Penelitian 32

3.1 Fokus Penelitian 32

3.2 Klasifikasi Penelitian Berdasarkan Tujuan dan Pendekatan .. 32

3.3 Pengumpulan Data 33

3.4 Penyelesaian Masalah 33

3.5 Penarikan Kesimpulan 34

15. Bab 4 Hasil dan Pembahasan 35

4.1 Hasil Penelitian 35

4.1.1 Asumsi Regresi Linier Berganda 35

4.1.2 Estimasi Regresi Linier Berganda 41

4.1.3 Pendeteksian *Outlier* 41

4.1.4 Metode RWLS 46

4.1.5 Uji Asumsi Estimasi RWLS 51

4.2 Pembahasan 56

16. Bab 5 Penutup 60

5.1. Simpulan 60

5.2. Saran 61

17. Daftar Pustaka 62

18. Lampiran 64

UNNES
UNIVERSITAS NEGERI SEMARANG

DAFTAR TABEL

Halaman

Tabel 4.1 Hasil Uji Kolmogorov-Smirnov data ICMD Bank MEGA 36

Tabel 4.2 Hasil Analisis Regresi: ANOVA 37

Tabel 4.3 Nilai Signifkansi Regresi ICMD Bank MEGA 38

Tabel 4.4 Nilai VIF Variabel Bebas Data ICMD Bank MEGA 38

Tabel 4.5 Model Summary Regresi Data ICMD Bank MEGA 39

Tabel 4.6 Nilai Signifikansi Uji Glejser data ICMD Bank MEGA 40

Tabel 4.7 Koefisien Regresi Linier Berganda Data ICMD Bank MEGA 41

Tabel 4.8 Hasil Pendeteksian *Outlier* dengan Metode *Cook's Distance* 43

Tabel 4.9 Hasil Metode Standardisasi Data 45

Tabel 4.10 Tabel *Near-neighbors* Variabel X_2 47

Tabel 4.11 Hasil perhitungan $MED(X_{(j)})$, $MED(Y_{(j)})$,$MAD(Y_{(j)})$ dan $\{MAD(Y_{(j)})\}^2$ 48

Tabel 4.12 Hasil Perhitungan nilai $\hat{y}$ 50

Tabel 4.13 Hasil perhitungan w_{1i}, *Standard Residuals*, w_{2i} dan w_i 51

Tabel 4.14 Koefisien Regresi Metode *Weighted Least Square* 51

Tabel 4.15 Hasil Uji Kolmogorov Smirnov Hasil Estimasi RWLS 52

Tabel 4.16 Hasil Analisis Regresi: ANOVA 53

Tabel 4.17 Nilai Signifikan WLS 54

Tabel 4.18 Nilai Signifikan Uji Glejser Hasil Estimasi WLS 54

Tabel 4.19 Hasil Pendeteksian *Outlier* Metode *Cook's Distance* 55

Tabel 4.20 Hasil pendeteksian *outlier* dengan metode standardisasi data ... 56

UNNES
UNIVERSITAS NEGERI SEMARANG

DAFTAR GAMBAR

Halaman

Gambar 2.1 Kerangka Berfikir Penulisan 31

Gambar 3.1 Diagram Alir Pemecahan Masalah 34

Gambar 4.1 *Normal* $P - P$ *Plot of Regression Standardized Residual* 35

Gambar 4.2 Graffik *Scatter Plot* variabel X_1 dengan variabel Y 42

Gambar 4.3 Graffik *Scatter Plot* variabel X_3 dengan variabel Y 42

UNNES
UNIVERSITAS NEGERI SEMARANG

DAFTAR LAMPIRAN

Halaman

Lampiran 1 Data ICMD Bank MEGA 64

Lampiran 2 *Output* Uji Kolmogorov-Smirnov Data ICMD Bank MEGA .. 66

Lampiran 3 *Output* Regresi Data ICMD Bank MEGA 67

Lampiran 4 *Output* Uji Glejser Data ICMD Bank MEGA 68

Lampiran 5 *Output* Uji Standardisasi Data ICMD Bank MEGA 69

Lampiran 6 Pengelompokan data berdasarkan variabel X_2 71

Lampiran 7 Pengelompokan variabel X_2 dan Y 72

Lampiran 8 Tabel hasil perhitungan $MED(X_{(j)})$, $MED(Y_{(j)})$, $MAD(Y_{(j)})$, $\{MAD(Y_{(j)})\}^2$ 73

Lampiran 9 Hasil perhitungan metode Robust *Least Trimmed Square* 74

Lampiran 10 Hasil Perhitungan $\hat{y}$ 77

Lampiran 11 Hasil perhitungan w_{1i}, *Standard Residuals*, w_{2i} dan w_i 78

Lampiran 12 Uji Kolmogorov-Smirnov 79

Lampiran 13 *Output* perhitungan *Weighted Least Square* 80

Lampiran 114 *Output* Uji Gleljer RWLS 81

UNNES
UNIVERSITAS NEGERI SEMARANG

DAFTAR SIMBOL DAN ISTILAH

1. X = variabel bebas
2. Y = variabel terikat
3. α = koefisien garis regresi
4. β = koefisien garis regresi
5. e_i = residual
6. p = parameter regresi
7. σ^2 = varians
8. X^T = transpose variabel bebas
9. $\hat{Y}$ = nilai tafsiran hasil tafsiran variabel terikat
10. MSE = nilai rata-rata jumlah kuadrat residual
11. med = median
12. mad = *median absolute deviation*
13. *leverage* = elemen-elemen diagonal dari matriks H
14. $\hat{\beta}^*{}_i$ = tafsiran koefisien regresi dari metode kuadrat tertimbang
15. w_i = penimbang metode kuadrat tertimbang

UNNES
UNIVERSITAS NEGERI SEMARANG

BAB I

PENDAHULUAN

1.1 Latar Belakang

Regresi merupakan salah satu teknik analisis statistika yang paling banyak digunakan. Dari prinsip-prinsip analisis regresi ini kemudian diturunkan berbagai teknik analisis statistika yang dapat digunakan dalam berbagai bidang, mulai dari bidang sains, bidang ekonomi dan bisnis, bidang industri dan lain sebagainya. Regresi sendiri pertama dikenalkan oleh Francis Galtom pada tahun 1877 yang kemudian banyak mengalami perkembangan dari tahun ke tahun.

Analisis regresi merupakan suatu analisis yang mengukur korelasi antar variabel dan menunjukan arah dari hubungan antara variabel dependen dengan variabel independen. Secara umum regresi merupakan analisis terhadap ketergantungan satu variabel dependen (terikat) dengan satu atau lebih variabel independen (bebas) dengan tujuan untuk mencari estimasi atau memperkirakan nilai rata-rata dari populasi. Dari analisis regresi akan diperoleh koefisien regresi untuk masing-masing variabel independen. Koefisien ini diperoleh dengan memprediksi nilai dari variabel dependen dengan suatu persamaan.

Salah satu teknik analisis regresi yang paling sering digunakan adalah *Ordinary Least Square* atau metode estimasi kuadrat terkecil. Dalam metode ini akan diestimasi koefisien garis regresi dengan cara meminimumkan jumlah kuadrat terkecil untuk mendapatkannya, terdapat beberapa asumsi klasik yang harus dipenuhi dalam OLS. Asumsi tersebut adalah terpenuhinya distribusi

UNNES
UNIVERSITAS NEGERI SEMARANG

normal, tidak terdapat multikolinearitas, varian konstan, dan homoskedastisitas serta tidak terdapat autokorelasi.

Dalam penerapannya tidak selalu asumsi-asumsi tersebut dapat terpenuhi, sehingga harus menggunakan metode yang lain selain metode kuadrat terkecil. Salah satu sebab yang mengakibatkan tidak terpenuhinya uji asumsi klasik adalah akibat adanya *outlier* dalam suatu data. *Outlier* sendiri merupakan suatu data yang didalamnya terdapat satu atau beberapa data yang berada jauh dari pola kumpulan data secara keseluruhan.

Dengan adanya *outlier* tersebut berakibat estimasi koefisien garis regresi yang diperoleh tidak tepat. Hal ini berarti estimasi parameter-parameter dalam model regresi linier dapat dipengaruhi oleh satu titik data ekstrim yang merupakan *outlier*. Pada metode kuadrat terkecil perlu dilakukan diagnosis terhadap *outlier* dikarenakan kepekaan dari metode tersebut terhadap kasus *outlier*. Sebaran data yang dihasilkan dari kasus *outlier* pun tidak lagi simetrik tetapi cenderung menjulur kearah *outlier* berada sehingga akan melanggar asumsi normalitas. Salah satu cara yang dilakukan untuk mengatasi kasus *outlier* diantaranya adalah dengan melakukan transformasi pada data agar asumsi dapat terpenuhi.

UNNES
UNIVERSITAS NEGERI SEMARANG

Dalam kasus tertentu terkadang transformasi yang dilakukan terhadap data tidak dapat memperkecil nilai *leverage outlier* yang akhirnya pendugaan menjadi bias. Dalam kasus seperti di atas, metode regresi robust dapat menjadi metode yang dapat digunakan. Regresi robust sendiri merupakan suatu metode regresi yang dapat digunakan pada saat distribusi error tidak normal dan atau adanya beberapa *outlier* yang mempengaruhi suatu model dari data (Wijayanti, 2015).

Metode regresi ini menjadi penting untuk menganalisis data yang dipengaruhi oleh *outlier*, sehingga akan dihasilkan suatu model yang resisten terhadap *outlier*. Estimasi yang resisten sendiri merupakan suatu estimasi yang relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data.

Dalam regresi robust sendiri terdapat beberapa jenis metode penduga menurut Chen (2014), mulai dari *Least Trimmed Square, M-estimsted, MM-estimated, Least Median of Square* dan *S-estimated.* Dalam penelitian kali ini penulis menggunakan *Least Trimmed Square* dalam menyelesaikan kasus *outlier* dengan menggunakan regresi robust. Hal ini dikarenakan menurut beberapa penelitian Heru Nurcahyadi (2010), metode *Least Trimmed Square* dianggap lebih mampu mengurangi pengaruh residual terhadap keakuratan koefisien regresi. Data *outlier* yang ada tidak dibuang begitu saja, tetapi diproses dan dieliminasi melalui sebuah iterasi menggunakan median dari kuadrat residual, penduga yang dihasilkan akan lebih kekar dalam menghadapi kasus *outlier.*

Pada kasus *oulier* ini akan menyebabkan pelanggaran-pelanggaran pada uji asumsi klasik. Salah satu asumsi klasik yang harus dipenuhi dalam estimasi OLS agar hasilnya dapat dipercaya atau diandalkan adalah homoskedstisitas. Homoskedastisitas berarti bahwa varian dari *error* berdifat konstan (Mokosolang, Prang & Mananohas, 2015). Kasus penyimpangan pada homoskedastisitas disebut heteroskedastisitas. Pelanggaran pada asumsi ini akan mengakibatkan penduga Metode Kuadrat Terkecil yang diperoleh tetap memenuhi persyaratan tak bias, tetapi varian yang diperoleh menjadi tidak efisien, artinya varian cenderung

UNNES
UNIVERSITAS NEGERI SEMARANG

membesar sehingga tidak lagi merupakan varian yang kecil. Hal tersebut tentunya mengharuskan model harus diperbaiki agar pengaruh dari heteroskedastisitas hilang.

Secara statistik permasalahan heteroskedastisitas tersebut akan menganggu model yang akan diestimasi. Salah satu metode yang dapat menjadi alternatif untuk menyelesaikan kasus ini adalah metode *Weight Least Square* atau metode kuadrat tertimbang. Metode WLS sama halnya seperti metode OLS dengan meminimumkan jumlah sisaan hanya saja pada metode WLS dilakukan pembobotan suatu faktor yang tepat kemudian baru menggunakan metode OLS terhadap data yang telah diboboti (Maziyya, Sukarsa & Asih, 2015).

Kasus *outlier* dan heteroskedastisitas akan memberikan masalah berbeda apabila berada pada suatu data. Tentunya perlu suatu metode yang dapat digunakan untuk menyelesaikan kedua kasus tersebut. Pada bagian atas penulis telah menyampaikan bahwa untuk kasus *outlier* dapat menggunakan metode regresi robust sementara untuk kasus heteroskedastisitas dapat menggunakan *Weight Least Square.* Dalam penelitian ini akan diuraikan bagaimana menyelesaikan kedua kasus ini dengan menggunakan kedua metode tersebut yaitu metode Robust *Weight Least Square* yang tentunya diharapakan dapat memberikan solusi terbaik terhadap kedua kasus tersebut.

1.2 Rumusan Masalah

Berdasarkan identifikasi masalah di atas, maka dapat dirumuskan permsalahan sebagai berikut:

UNNES
UNIVERSITAS NEGERI SEMARANG

Bagaimana hasil estimasi analisis *outlier* dan heteroskedastisitas dengan metode regresi Robust *Weight Least Square*?

1.3 Tujuan Penelitian

Tujuan dari penulisan penelitian ini adalah mengetahui model estimasi pada *outlier* dan heteroskedastisitas dengan metode regresi Robust *Weight Least Square.*

1.4 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberikan manfaat, antara lain:

1. Bagi peneliti dapat menambah pengetahuan dan memperluas wawasan mengenai analisis *oulier* dan heteroskedastisitas dengan metode regresi *Robust Weight Least Square.*
2. Bagi Jurusan Matematika dapat dijadikan sebagai bahan studi kasus dan referensi bagi pembaca dann acuan bagi mahasiswanya.
3. Bagi Bank MEGA dapat dijadikan acuan untuk menganalisi data keuangan instansinya.

1.5 Sistematika Penulisan

Untuk mengetahui apa saja yang terkandung dalam penulisan proposal skripsi ini, maka penulis membuat sistematika penulisan. Secara garis besar sistematika penulisan skripsi ini dibagi menjadi tiga bagian yaitu bagian awal, bagian isi, bagian akhir yang masing-masing diuraikan sebagai berikut.

1. Bagian Awal

Dalam penulisan skripsi ini, bagian awal berisi halaman judul, halaman pengesahan, motto dan persembahan, abstrak, daftar isi, daftar tabel, daftar gambar, dan daftar lampiran.

2. Bagian Isi

Bagian pokok dari penulisan skripsi ini adalah isi proposal skripsi yang terdiri atas lima bab, yaitu:

a. BAB 1 PENDAHULUAN

Bab ini membahas tentang latar belakang, pembatasan masalah, identifikasi masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penelitian.

b. BAB 2 TINJAUAN PUSTAKA

Bab ini berisi tentang materi-materi yang berkaitan dengan regresi, data *outlier*, kasus heteroskedastisitas, regresi *Robust*, metode *Weight Least Square*, penelitian terdahulu, dan kerangka berpikir.

c. BAB 3 METODE PENELITIAN

Bab ini berisi tentang langkah-langkah yang dilakukan dalam penelitian ini meliputi fokus penelitian, pengumpulan data, penyelesaian masalah, dan penarikan kesimpulan.

d. BAB 4 HASIL DAN PEMBAHASAN

Bab ini berisi tentang hasil analisis *outlier* dan heteroskedastisitas dengan menggunakan metode regresi Robust *Weight Least Square.*

e. BAB 5 PENUTUP

UNNES
UNIVERSITAS NEGERI SEMARANG

Bab ini berisi simpulan hasil penelitian dan saran yang berkaitan dengan hasil penelitian yang diperoleh.

3. Bagian akhir

Bagian ini berisi daftar pustaka sebagai acuan penulisan dan lampiran-lampiran yang melengkapi uraian pada bagian isi.

UNNES
UNIVERSITAS NEGERI SEMARANG

BAB 2

TINJAUAN PUSTAKA

2.1 Regresi Linier

2.1.1 Regresi Linier Sederhana

Analisis regresi adalah suatu metode yang berguna untuk menentukan hubungan suatu variabel yang disebut variabel dependen dengan satu atau lebih variabel yang disebut variabel independen (Gujarati, 1978). Salah satu tujuan atau hasil dari analisis regresi adalah untuk menentukan model regresi yang baik, sehingga model dapat akurat ketika digunakan untuk menerangkan kondisi data dan memprediksi hal-hal yang berhubungan dengan variabel-variabel yang terlibat dalam model regresi tersebut. Konsep regresi linier sederhana dapat digambarakan dalam fungsi sebagai berikut:

$$Y = \alpha + \beta X + \varepsilon.$$

Keterangan: α, β= koefisien garis regresi,
Y= variabel terikat,
X= variabel bebas,
ε = residual

UNNES
UNIVERSITAS NEGERI SEMARANG

2.1.2 Regresi Linier Berganda

Dalam banyak kasus analisis regresi tentu akan menghadirkan lebih dari satu variabel bebas X terhadap satu variabel terikat Y yang nilainya bergantung pada beberapa variabel bebas tersebut. Regresi linier ganda menjelaskan hubungan linier antara kelompok variabel bebas $[X_1, X_2, X_3, \dots, X_n]$ dan variabel

terikat Y. Secara umum hubungan antara kelompok variabel X dengan satu variabel Y dinyatakan sebagai berikut:

$$Y = \alpha + \beta_1 X_1 + \cdots + \beta_n X_n + \varepsilon$$

Keterangan: $X_1 X_2, \ldots, X_n$ = varianel bebas

α, β = koefisien regresi

Y= variabel terikat,

ε = residual untuk pengamatan ke-n yang diasumsiskan berdistribusi normal yang bebas dan identik dengan rata-rata 0 (nol) dan variansi σ^2.

2.1.3 Metode Kuadrat Terkecil (*Ordinary Least Square)*

Salah satu metode yang dapat digunakan untuk menjadi penduga model untuk bentuk regresi linier adalah dengan metode *Ordinary Least Square.* Metode ini memiliki konsep untuk meminimumkan jumlah kuadrat residual (selisih antara data sebenarnya dengan data dugaan) dari model regresi yang terbentuk. Metode ini kemudian menjadi metode yang paling banyak digunakan dibandingkan dengan motode-metode lainnya dalam pembentukan model regresi atau mengestimasi parameter regresi.

Untuk mengestimasi koefisien regresi $\beta_0, \beta_1, \beta_2, \ldots, \beta_n,$ pada n data suatu penelitian adalah

$$j = \sum_{i=n}^{n} e_i^2 = \sum_{i=n}^{n} (Y_i - b_0 - b_1 X_1 - b_2 X_2 - \cdots - b_n X_n)^2$$

dan j harus bernilai minimum. Jika j berubah untuk diturunkan terhadap b_0, $b_1, b_2, \ldots, b_n$. Kemudian menyamakannya dengan nol akan diperoleh

$$\frac{\partial j}{\partial b_0} = -2 \sum_{i=n}^{n} (Y_i - b_0 - b_1 X_1 - b_2 X_2 - \cdots - b_n X_n) = 0$$

$$\frac{\partial j}{\partial b_1} = -2\sum_{i=n}^{n}(Y_i - b_0 - b_1X_1 - b_2X_2 - \cdots - \ b_nX_n) = 0$$

$$\vdots$$

$$\frac{\partial j}{\partial b_n} = -2\sum_{i=n}^{n}(Y_i - b_0 - b_1X_1 - b_2X_2 - \cdots - \ b_nX_n) = 0$$

Persamaan di atas dapat disederhanakan menjadi,

$$nb_0 + b_1\sum_{i=n}^{n}{X_1}^2 + b_2\sum_{i=n}^{n}X_1X_2 + \cdots + \ b_n\sum_{i=n}^{n}X_1X_n = \sum_{i=n}^{n}Y_i$$

$$b_0\sum_{i=n}^{n}X_1 + b_1\sum_{i=n}^{n}{X_1}^2 + b_2\sum_{i=n}^{n}X_1X_2 + \cdots + \ b_n\sum_{i=n}^{n}X_1X_n$$
$$= \sum_{i=n}^{n}Y_2Y_i$$

$$b_0\sum_{i=n}^{n}X_n + b_1\sum_{i=n}^{n}X_1X_n + b_2\sum_{i=n}^{n}X_2^2X_n + \cdots + \ b_n\sum_{i=n}^{n}X_n^2$$
$$= \sum_{i=n}^{n}X_nY_i$$

Apabila dinyatakan dalam bentuk matriks, persamaan di atas menjadi:

$$X^TXb = X^TY$$

Dengan demikian b sebagai penduga β dapat diperoleh melalui rumus:

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

2.1.4 Asumsi Regresi Linier

Asumsi-asumsi yang harus dipenuhi agar OLS dapat menghasilkan estimasi yang tidak bias atau yang terbaik / BLUE (*Best Linear Unbias Estimator)* pada model regresi adalah sebagai berikut.

1) Nilai residual berdistribusi normal.

   $u_i \sim N(0, \sigma^2)$

2) Model regresi adalah linier, yaitu linier dalam parameter.

3) Variabel bebas memberikan pengaruh secara signifikan terhadap variabel terikat.

4) Tidak terdapat multikolinieritas atau hubungan antara variabel bebas X.

5) Homoskedastisitas atau varians dari residual adalah konstan.

   $E[u_i^2] = \sigma^2$ untuk $i = 1, 2, \dots, n$.

6) Tidak terdapat autokorelasi antara nilai residual.

   $E(u_i u_j) = 0$ dengan $i \neq j$.

2.2 *Outlier* (Pencilan)

Pencilan atau *outlier* adalah data yang tidak mengikuti pola umum pada model regresi yang dihasilkan, atau pola data secara keseluruhan. Secara tidak langsung *outlier* merupakan hasil observasi yang menyimpang dari pola yang terbentuk oleh sebagaian besar data. Dalam suatu data biasanya terdapat 10% amatan yang merupakan *outlier*. Jumlah maksimum adanya *outlier* dalam sata yang diperbolehkan adalah 50%. Dalam Ghozali (2009: 40) terdapat empat penyebab timbulnya data *outlier* antara lain.

1) Kesalahan dalam memasukan data;

2) Kegagalan menspesifikasi adanya *missing value* dalam program komputer;

3) *Outlier* bukan berasal dari populasi yang di ambil sebagai sampel; dan

4) Variabel dalam populasi memiliki nilai ekstrim serta tidak berdistribusi normal.

Outlier sendiri menandakan adanya data yang tidak tipikal dengan sebagian besar data. *Outlier* akan memberikan pengaruh terhadap proses analisis data. Oleh karena itu adanya *outlier* pada suatu data haruslah dihindari. *Outlier* dapat pula menyebabkan varians pada data menjadi lebih besar, interval dan *range* menjadi lebar, *mean* tidak dapat menunjukan nilai yang sebenarnya atau bias. Hal tersebut tentunya akan memberikan dampak dalam pengambilan keputusan dan kesimpulan.

Terdapat beberapa hal yang dapat dilakukan untuk mengatasi *outlier,* salah satunya dengan menghilangkan *outlier* dari data, kemudian baru dilanjutkan dengan langkah analisis. Namun hal tersebut dapat dilakukan ketika data tersebut diperoleh dari kesalahan teknis yang peneliti lakukan, seperti kesalahan mencatat amatan dan menyiapkan amatan. Tetapi adakalanya *outlier* justru akan memberikan informasi yang tidak bisa didapatkan dari data yang lain, misalnya *outlier* muncul karena kombinasi keadaan yang tidak biasa yang mungkin saja sangat penting dan perlu dilakukan analisis lebih mendalam.

UNNES
UNIVERSITAS NEGERI SEMARANG

Keberadaan *outlier* tentunya akan memberikan dampak tersendiri pada proses analisis, menurut Paludi (2009: 57) dampak tersebut diantaranya adalah munculnya residual yang besar dari model yang terbentuk, varians pada data menjadi lebih besar, dan taksiran interval memiliki rentang yang lebar. Selain itu adanya *outlier* akan memberikan nilai penduga parameternya bersifat bias sehingga berakibat interpretasi hasil yang diperoleh menjadi tidak valid.

2.2.1 Pendeteksian *Outlier*

Pada saat kita akan mendeteksi suatu *outlier* maka kita harus mampu melihat kemungkinan bahwa data tersebut terkontaminasi *outlier*. Data *outlier* dapat dikenali dengan pemeriksaan visual data mentahnya atau dari diagram pencar dari variabel dependen. Jika terdapat lebih dari dua variabel independen, beberapa *outlier* mungkin tidak akan mudah untuk dideteksi dengan pemeriksaan secara visual. Sehingga perlu adanya alat bantu untuk memudahkan pendeketesksian *oulier*.

Data *oulier* harus dilihat terhadap posisi dan sebaran data yang lainnya sehingga menjadi bahan evaluasi apakah data *outlier* tersebut perlu untuk dihilangkan atau tidak. Terdapat beberapa metode yang dapat digunakan untuk mendeteksi *outlier* yang akan memperngaruhi koefisien regresi, diantaranya adalah metode *boxplot, Leverage Value, Standardized Residual, Breakdown Point,* dan *Cook's Distance.* Namun pada kesempatan ini penulis akan menggunakan metode *Cook's Distance* dan *Standardized Residual* untuk mendeteksi *outlier*.

1) Metode *Cook's Distance*

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

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

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

$$H = X_i(X_i^T X_i)^{-i} X^T{}_i$$

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

2) *Standardized Residual*

Metode *Standardized Residual* merupakan metode yang sederhana dan efektif untuk mendeteksi *outlier*, pada metode ini pendeteksian *outlier* dengan memeriksa residual. Residual ke-i didefinisikan sebagai berikut:

$$e_i = y_i - \hat{y}_i$$

Sesuai dengan residual ke-i di atas, dapat didefiniskan *standardized residual* ke-*I* sebagai

$$e_{is} = \frac{e_i}{\sqrt{MSE}}$$

dengan MSE = $\frac{\sum_{i=1}^{n} e_i^2}{n-2}$

Dengan MSE adalah nilai rata-rata residual kuadrat dan akar dari MSE disebut standar error. Standar error megukur besarnya variansi model regresi, semakin kecil nilainya semakin baik model regresinya. Untuk mengidentifikasi *outlier*, diperhatikan nilai dari *standardized residual*. Menurut Singgih Santoso (2010:35), jika nilai

dari *standardized residual* lebih dari *2,5* atau kurang dari -2,5 maka data tersebut dikatakan sebagai data *outlier*.

2.3 Regresi Robust

Regresi Robust merupakan salah satu metode penting untuk menganalisis data yang terdetekai sebagau data pencilan. Regresi Robust digunakan untuk mendeteksi *outlier* dan memberikan hasil yang resisten terhadap adanya *outlier*. Beberapa peneliti menggunakan metode regresi robust sebagai pengontrol hasil pendugaan menggunakan metode kuadrat terkecil, bila kedua hasil tersebut tidak berbeda jauh maka hasil metode kuadrat terkecil dapat digunakan dengan lebih yakin, namun apabila terdapat perbedaan hasil yang signifikan maka hasil dari metode regresi robust lebih mampu menjelaskan dalam menggambarkan pengamatan mana yang perlu mendapat perhatian lebih lanjut.

Terdapat tiga kelas masalah yang dapat menggunakan teknik regresi robust menurut Chen (2002:1), yaitu:

1) Masalah *outlier* yang terdapat pada peubah *y;*
2) Masalah *outlier* yang terdapat pada peubah *x;* dan
3) Masalah *outlier* yang terdapa pada keduanya, yaitu peubah *x* dan peubah *y*.

Dalam metode robust sendiri terdapat beberapa metode untuk mengatasi masalah *outlier*.

2.3.1 Estimasi-M

Metode Estimasi M merupakan salah satu regresi robust yang paling sering untuk digunakan. *M-Estimation* merupakan estimasi yang meminimumkan jumlah nilai mutlak dan residual.

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

Pada estimasi M meode iterasi sangatlah diperlukan, hal tersebut dikarenakan residualnya tidak dapat dihitung sampai diperoleh model yang cocok dan parameter regresi juga tidak dapat dihitung tanpa mengetahui nilai *iteratively reweighted least square.*

2.3.2 *Least Median Square* (LMS)

Pada 1984 Rousseeuw memperkenalkan LMS yang merupakan metode *High Breakdown Value*. Prinsip metode ini adalah mencocokan sebagian besar data setelah pencilan teridentifikasi sebagai titik yang tidak berhubungan dengan data. Metode LMS adalah suatu metode estimasi parameter regresi robust dengan meminimumkan median dari kuadrat residual.

$$M_j = min\{med\ e_i^2\} = min\{M_1, M_2, \dots, M_s\}$$

dengan e_i^2 adalah kuadrat residual hasil taksiran dengan OLS. dengan menggantikann penjumlahan dengan median, menghasilkan estimasi yang resisten terhadap *outlier*.

2.3.3 *Least Trimmed Square (LTS)*

Serupa dengan metode LMS, metode robust LTS merupakan metode *High Breakdown Value* yang diperkenalkan oleh Rousseeuw pada 1984. Metode LTS

merupakan metode estimasi parameter regresi robust dengan meminimunkan jumlah kuadrat h residual.

$$\min \sum_{i=1}^{h} e_i^2$$

dengan

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

keterangan:

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

n = banyaknya sampel

p = parameter regresi

Jumlah h menunjukan sejumlah subset data dengan kuadrat fungsi objektif terkecil. Nilai h pada persamaan akan membangun *breakdown point* sebesar 0,5. LTS memiliki resisten yang paling tinggi terhadap *outlier*, akan tetapi LTS sangat tidak efisien dan dapat mengakibatkan kesalahan dalam penggambaran model data jika dinilai dari pengelompokan *outlier*, atau jika jumlah data relatif kecil. LTS merupakan metode Robust yang memiliki resisten paling tinggi terhadap *outlier*.

Langkah-langkah melakukan estimasi LTS sebagai berikut

1) Menghitung estimasi parameter a_0 dan b_0.
2) Menentukan n residual $e_i^2 = (y_i - x_i b_0 - a_0)^2$ yang bersesuaian dengan a_0 dan b_0, kemudian menghitung $h = \frac{(3n+p+1)}{4}$ pengamatan dengan nilai e_i^2 terkecil.

3) Menghitung $\sum_{i=1}^{h} e_i^2$.

4) Melakukan estimasi parameter a_0 dan b_0 dari h pengamatan.

5) Menentukan n kuadrat residual yang baru yang bersesuaian dengan b_0 kemudian menghitung sejumlah h_{new} pengamatan dengan nilai e_i^2 terkecil.

6) Menghitung $\sum_{i=1}^{h_{new}} e_i^2$ dan menghitung langkah 4 sampai 6 untuk mendapatkan fungsi obyektif yang kecil dan konvergen.

2.3.4 Estimasi-S

Metode robust S juga merupakan metode *High Breakdown Value* yang diperkenalkan oleh Rousseeuw pada 1984. Estimasi S merupakan solusi dengan kemungkinan terkecil dari penyebaran residual. Estimasi S didefinisikan sebagai berikut

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

Selain meminimumkan varians dari residual, Estimasi S juga meminimumkan skala residual dari estimasi S. Estimasi S memiliki *breakdown point* sebesar 0,5 diperoleh dengan menggunakan fungsi Tukey Bisquare. Estimasi S dianggap kurang efisien meskipun memiliki *breakdown point* 0,5, karena sangat relatif terhadap OLS ketika distribusi error normal.

2.3.5 Estimasi-MM

Metode Estimasi-MM pertama kali diperkenalkan oleh Yohai pada 1987 dengan menggabungkan *High Breakdown Value* dengan tingkat efisiensi yang tinggi. Etimasi-MM diawali dengan mencari estimasi-S yang akan menjadi *high*

breakdown value. Selanjutnya di akhiri dengan menetapkan parameter-parameter regresi menggunakan Estimasi-M. Estimasi-MM didefinisikan sebagai berikut.

$$\hat{\beta} = \min \sum \rho(\frac{e_i}{\hat{\sigma}}) = min \sum_{i=1}^{n} \rho \left(\frac{y_i - \sum_{i=n}^{n} x_{ij}\beta_j}{\hat{\sigma}_s} \right)$$

Seperti pada Estimasi-M, pada Estimasi-MM juga menggunakan *Iteratively Reweighted Least Square* (IRLS) untuk mencari estimasi parameter regresi.

2.4 Heteroskedastisitas

Salah satu asumsi dalam regresi linier adalah semua residual atau *error* memiliki varians yang sama. Asumsi tersebut sering disebut dengan homoskedastisitas, namun apabila asumsi tersebut tidak dipenuhi maka disebut heteroskedastisitas. Heteroskedastisitas tidak akan merusak sifat ketidakbiasan dan konsistensi dari penaksir OLS, namun penaksir ini tidak lagi memiliki varians yang minimum atau efisien. Dengan kata lain, penaksir menjadi tidak BLUE (*Best Linear Unbias Estimator)*.

Suatu contoh dalam kehidupan sehari-hari untuk menggambarkan heteroskedastisitas adalah pada kasus hubungan pendapatan dengan konsumsi. Orang dengan pendapatan rendah tentunya akan memiliki variansi yang rendah dalam menggunakan pendapatannya untuk menabung atau konsumsi. Tetapi orang dengan pendapatan tinggi tentu karena memiliki variansi yang lebih tinggi untuk menabung dan konsumsi. Orang berpendapatan tinggi yang boros tentunya akan memiliki konsumsi yang tinggi, serta berbanding terbalik dengan orang yang tidak boros, karena akan memiliki tabungan yang besar.

Heteroskedastisitas akan banyak ditemui pada data *cross-sectional* namun bukan berarti heteroskedastisitas tidak ditemukan pada data *time series*. Pada data *cross-sectional* pengamatan dilakukan pada suatu populasi dan dilakukan pada suatu saat tertentu. Misalnya pada suatu perusahaan, wilayah, keluarga dan lain sebagainya. Hal tersebut memungkinkan terjadinya ukuran yang berbeda dalam setiap populasi, seperti perusahaan besar, perusahaan sedang dan perusahaan kecil serta wilayah kota besar dan wilayah kota kecil. Sementara pada data *time series* variabelnya cenderung memiliki derajat atau ukuran yang sama karena biasanya pada data ini peneliti mengumpulkan data untuk suatu variabel yang sama selama suatu periode waktu tertentu.

2.4.1 Penyebab terjadinya Heteroskedastisitas

Terdapat beberapa penyebab yang mengakibatkan varians menjadi variabel yang berubah-ubah, antara lain sebagai berikut:

1) Terjadi situasi *error-learning models*, hal ini terjadi karena proses manusia belajar sehingga mengakibatkan semakin kecil kesalahan dalam perilaku yang dilakukan, sehingga dalam kasus ini diharapkan σ_i^2 menurun.
2) *Discretionary income.* Dengan meningkatnya pendapatan, orang akan memiliki kesempatan lebih banyak untuk menggunakan *income*-nya. Sehingga varians akan meningkat sejalan dengan peningkatan *income.*
3) Perbaikan dalam tehnik pengambilan data. Hal tersebut akan menjadikan σ_i^2 menurun. Contoh pada bank yang memilik peralatan yang canggih akan menemui kesalahan yang lebih kecil dalam laporan

UNNES
UNIVERSITAS NEGERI SEMARANG

bulanan atau tahunan untuk pelanggan merka dibandingkan dengan bank dengan peralatan yang lebih buruk.

Kasus hetereoskedastisitas lebih sering terjadi pada data *cross-sectional* dibandingkan dengan data *time series.*

2.4.2 Akibat dari Heteroskedastisitas

Ketika seluruh asumsi model regresi linier dipenuhi, maka penaksir OLS menjadi *BLUE*, yang berarti semua penaksir tak bias linier memiliki varians yang minimum, atau penaksir efisien. Ketika semua asumsi terpenuhi kecuali heteroskedastisitas maka penaksir menjadi tak bias dan konsisten tetapi menjadi tidak efisien baik untuk sampel besar maupun sampel kecil. Dengan kata lain, jika suatu model diperbaharui ulang dengan menambah data atau dengan sampel-sampel yang digunakan berbeda, penaksir OLS secara rata-rata akan sama dengan nilai populasi sebenarnya (sifat tak bias) dan dengan meningkatnya ukuran sampel hingga ke tak terhingga penaksir tersebut akan mengarah ke sifat yang sebenarnya namun variansi tidak lagi minimum, bahkan ketika sampel terus meningkat secara tak terbatas.

2.4.3 Pendeteksian Heteroskedastisitas

UNNES
UNIVERSITAS NEGERI SEMARANG

Terdapat beberapa metode yang dapat digunakan untuk mendeteksi heteroskedastisitas. Beberapa metode formal dan informal untuk mendeteksi heteroskedastisitas sebagai berikut

1) Sifat dasar masalah

Sifat dasar masalah yang ditemui terkadang menjadikan kasus heteroskedastisitas menjadi suatu hal yang harus dijumpai. Dalam survei

untuk memperoleh suatu data pasti akan diharapkan varians yang tak sama diantara *disturbances.* Pada data *cross-sectional* akan menghadirkan data yang sangat heterogen sehingga heteroskedastisitas akan menjadi hal yang sangat lazim ditemui.

2) Metode Grafik

Cara lain yang dapat digunakan untuk mendeteksi heteroskedastisitas adalah dengan melihat ada tidaknya pola tertentu pada grafik yang dihasilkan, dimana sumbu X adalah Y yang telah diprediksi, dan Y adalah residual (Y sesungguhnya – Y prediksi). Dasar pengambilan keputusan metode grafik adalah adanya pola tertentu, seperti titik-titik yang membentuk suatu pola tertentu yang teratur (bergelombang, melebar, kemudian menyempit) maka terjadi heteroskedastisitas dan jika tidak ada pola yang jelas, serta titik-titik menyebar di atas dan di bawah angka 0 pada sumbu Y, maka terjadi heteroskedastisitas.

Salah satu kelemahan pengujian metode grafik adalah tidak jarang kita ragu terhadap pola dari grafik yang dihasilkan. Hal ini bisa saja akan menghasilkan keputusan yang berbeda antara orang yang satu dengan orang yang lain ketika membaca pola pada grafik. Oleh karena itu perlu adanya metode lain selain metode grafik.

3) Pengujian Glejser (*Glejser test*)

Pengujian Glejser dilakukan dengan menggunakan nilai residual. Setelah diperoleh nilai residual e_i dari hasil regresi OLS, dilakukan regresi terhadap nilai mutlak dari e_i, terhadap varieabel X yang diperkirakan

memiliki hubungan yang erat dengan σ_i^2. Bentuk fungsional yang dilakukan dalam pengujian Glejser adalah sebagai berikut.

$$|e_i| = \beta_i X_i + v_i$$

$$|e_i| = \beta_i \sqrt{X_i} + v_i$$

$$|e_i| = \beta_i \frac{1}{X_i} + v_i$$

$$|e_i| = \beta_i \frac{1}{\sqrt{X_i}} + v_i$$

$$|e_i| = \beta_0 + \beta_i X_i + v_i$$

$$|e_i| = \sqrt{\beta_0 + \beta_i X_i} + v_i$$

$$|e_i| = \sqrt{\beta_0 + \beta_i X_i^2} + v_i$$

di mana v_i adalah unsur kesalahan.

Namun dalam Goldfeld dan Quandt menunjukan bahwa v_i mempunyai beberapa masalah dalam hal nilai yang diharapkan tidak sama dengan nol, berkorelasi secara serial dan bersifat heteroskedastisitas. Teknik Glejser dapat digunakan untuk sampel besar dan kecil sebagai alat kualitatif untuk mempelajari heteroskedastisitas.

4) Pengujian rank korelasi dari Spearman

Koefisien korelasi rank Spearmana didefiniskan sebagai berikut:

$$r_s = 1 - 6\left|\frac{\sum d_i^2}{N(N^2 - 1)}\right|$$

di mana d_i= perbedaan penempatan rank untuk dua karakteristik yang berbeda dari sampel ke *i* dan N menyatakan banyaknya sampel di rank.

Langkah-langkah korelasi rank Spearman dengan asumsi $Y_i = \beta_0 + \beta_1 X_i + v_i$, adalah sebagai berikut:

a. Mencocokan hasil regresi pada X dan Y serta mencari residual e_i.

b. Mengambil nilai mutlak dari e_i, meranking nilai mutlak e_i dan X_i dengan urutan besar ke kecil atau kecil ke besar dan menghitung nilai korelasi rank Spearman yang telah diberikan.

c. Dengan mangasumsikan nilai korelasi rank Spearman ρ_s adalah nol dan $N > 8$, nilai signifikan pada r_s dapat dicari dengan menggunakan cara sebagai berikut

$$t = \frac{r_s\sqrt{N-2}}{\sqrt{t - r_s^2}}$$

Dengan derajat kebebasan $N - 2$

Jika nilai t yang dihitung melebihi t kritis maka hipotesis adanya heteroskedastisitas dapat diterima, namun apabila tidak maka dapat ditolak. Jika model data memiliki lebih dari satu variabel independen maka nilai r_s dapat dihitung nilai $|e_i|$ dan tiap-tiap variabel X secara terpisah kemudian dibandingkan nilai t yang diperoleh.

UNNES
UNIVERSITAS NEGERI SEMARANG

2.5 Metode *Weight Least Square*

Terdapat banyak metode yang dapat digunakan untuk menyembuhkan kasus heteroskedastisitas. Beberapa model transformasi dapat digunakan untuk memperbaiki data yang memiliki varians yang berubah-ubah. Akan tetapi alternatif model estimasi yang baik untuk kasus heteroskedastisitas adalah metode *Weight Least Square* atau metode kuadrat tertimbang. Hal ini dikarenakan metode

WLS memiliki kemampuan untuk menetralisir akibat dari pelanggaran asumsi serta dapat menghilangkan sifat ketidakbiasan dan konsistensi dari model estimasi OLS.

Pada model estimasi WLS ini bertumpu pada dua dasar, yaitu melakukan transformasi terhadap data yang dianalisis dan menerapkan model OLS terhadap data yang telah ditransformasikan tersebut. Untuk menggunakan metode WLS pada regresi linier berganda didefinisikan variabel-variabel dalam model regresi sebagai berikut.

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Metode kuadrat terkecil meminimumkan

$$\sum e_i^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

Untuk metode kuadrat tertimbang atau *weighted least square* meminimumkan jumlah kuadrat residual tertimbang.

$$\sum w_i e_i^2 = \sum w_i (Y_i - \hat{\beta}^*{}_0 - \hat{\beta}^*{}_1 X_i)^2$$

Dimana β_0^*, β_1^* adalah nilai tafsiran dari metode kuadrat tertimbang. Penimbang pada metode ini adalah w_i.

$$w_i = \frac{1}{\sigma_i^2}$$

penimbang tersebut merupakan nilai invers dari varians dari u_i atau Y_i atas X_i.

$$var(u_i|X_i) = var(Y_i|X_i) = \sigma_i^2$$

Dengan mendeferensialkan persamaan pada *weighted least square* dengan β_0^* dan β_1^* diperoleh

$$\frac{\partial \sum w_i e_i^2}{\partial \beta_0^*} = 2\sum w_i(Y_i - \hat{\beta}^*{}_0 - \hat{\beta}^*{}_1 X_i)(-1)$$

$$\frac{\partial \sum w_i e_i^2}{\partial \beta_1^*} = 2\sum w_i(Y_i - \hat{\beta}^*{}_0 - \hat{\beta}^*{}_1 X_i)(-X_i)$$

Dengan menyamakan kedua persamaan di atas sama dengan nol, diperoleh persamaan sebagai berikut

$$\sum w_i Y_i = \beta_0^* \sum w_i + \beta_1^* \sum w_i X_i$$

$$\sum w_i X_i Y_i = \beta_0^* \sum w_i X_i + \beta_1^* \sum w_i {X_i}^2$$

Dari kedua persamaan di atas sangatlah mirip dengan persamaan metode kuadrat terkecil, maka dimisalkan

$$\bar{X}^* = \frac{\sum w_i X_i}{\sum w_i}$$

$$\bar{Y}^* = \frac{\sum w_i Y_i}{\sum w_i}$$

$$x_i^* = X_i - \bar{X}^*$$

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

Secara simultan persamaan yang telah disama dengan nol didapatkan

$$\beta_0^* = \bar{Y}^* - \beta_0^* \bar{X}^*$$

$$\beta_1^* = \frac{\sum w_i {x_i}^* y_i^*}{\sum w_i {x_i}^{*2}}$$

Maka diperoleh nilai $w_1 = w_2 = \cdots = w_n = w$ penaksir kuadrat terkecil sama dengan penaksir kuadrat tertimbang.

2.6 Metode Robust (*Least Trimmed Square*) *Weighted Least Square*

Metode Robust *Weighted Least Square* merupakan kombinasi dari metode Robust dan *Weighted Least Square*, tentunya ada suatu langkah tersendiri yang digunakan untuk mengkombinasikan kedua metode tersebut. Menurut Habsah Midi, 2009, langkah yang digunakan dalam metode RWLS adalah sebagai berikut.

1) Mengidentifikasi kelompok *near-neighbors* atau tetangga terdekat dari variabel bebas yang memiliki pengamatan yang cukup dekat satu sama lain atau setidaknya memiliki nilai yang diulang pada variabel bebas (X). Jumlah dari *near-neighbors* atau tetangga terdekat dilambangkan dengan g.
2) Menghitung nilai $MED(X_{(j)}), j = 1, 2, \dots, g$, median dari setiap kelompok variabel penjelas (X) dan menghitung nilai $MAD(Y_j) = Median\{|Y_j - Median(Y_j)|\}$ dari kelompok g yang merupakan variabel respon (Y) sesuai dengan kelompok (X).
3) Menghitung nilai regresi $\{MAD(Y_j)\}^2$ ke $MED(X_{(j)})$ dengan metode *Least Trimmed Square* (LTS) dan menghitung nilai koefisien regresinya.
4) Menghitung nilai tafsiran $\hat{y}$ menggunakan hasil estimasi pada metode LTS.
5) Menghitung nilai mutlak invers $\hat{y}$, dilambangkan dengan w_{1i} sebagai pembobot awal. Nilai pembobot akhir atau w_{2i} diperoleh dari fungsi bobot robust, namun pada penelitian ini akan digunakan bobot fungsi Huber yang didefinisikan sebagai.

$$w_{2i} = \begin{cases} 1 & |e_i| \leq 1.345 \\ \dfrac{1.345}{|e_i|} & |e_i| \geq 1.345 \end{cases}$$

Nilai konstan 1.345 disebut sebagai konstanta tala dan nilai e_i adalah nilai residual standar yang dari metode LTS pada langkah 3. Nilai bobot w_{1i} dikalikan dengan nilai bobot w_{2i} untuk mendapatkan nilai bobot akhir w_i.

6) Langkah terakhir melakukan regresi *Weighted Least Square* menggunakan bobot akhir w_i.

2.7 Kerangka Berfikir

Dalam mengestimasi koefisien regresi dalam suatu model regresi, pada umumnya metode estimasi OLS merupakan metode yang paling sering digunakan. Akan tetapi terdapat kekurangan dalam metode OLS tersebut, salah satunya adalah metode ini sangat sensitif terhadap kehadiran dari pencilan atau *outlier*. Hasil estimasi koefisien regresi pada data yang terkena *outlier* dengan menggunakan metode OLS akan menjadi tidak tepat.

Pendeteksian terhadap *outlier* merupakan tahapan yang harus dilakukan apabila akan dilakukan estimasi model regresi dengan menggunakan metode OLS. Beberapa cara yang dapat dilakukan untuk mendeteksi adanya *outlier* pada data dapat dilakukan dengan metode *cook's distance* dan metode *standardisasi*. Apabila dalam tahapan pendeteksian *outlier* tidak ditemukan adanya *outlier* pada data maka model estimasi dengan metode OLS dapat diterima. Namun apabila

ditemukan *outlier* pada data tersebut, perlu dilakukan satu langkah yang akan membuat keberadaan *outlier* menjadi tidak berarti dalam metode OLS tersebut.

Regresi *robust* merupakan salah satu metode yang dapat digunakan untuk mengatasi kekurangan OLS terhadap data yang memiliki *outlier*. Regresi *robust* merupakan metode yang digunakan ketika distribusi dari residual tidak normal atau adanya beberapa *outlier* yang berpengaruh terhadap model, sehingga akan dihasilkan model yang resisten terhadap *outlier* (Draper dan Smith, 1992). Dalam metode *robust* terdapat beberapa metode yang dapat digunakan untuk mengestimais koefisien regresi, diantaranya estimasi M, estimasi MM, estimasi S, estimasi LMS dan estimasi LTS. Dari penelitian terdahulu, diperoleh bahwa metode LTS lebih baik daripada metode Estimasi-M untuk mengatasi outlier pada data karena memiliki nilai koefisien determinasi lebih besar (Suyanti, 2014).

Dengan adanya *outlier* pada data tentunya akan berakibat pada tidak terpenuhinya asumsi syarat pada OLS. Asumsi-asumsi tersebut adalah nilai residual berdistribusi normal, model regresi linier dalam parameter, tidak terjadi multikolinieritas, tidak terdapat autokorelasi dan varians konstan atau homoskedastisitas. *Outlier* erat kaitannya dengan keberadaan beberapa data yang berada jauh dari kumpulan kebanyakan data, hal ini tentu akan berakibat pada varians menjadi tidak konstan atau cenderung menjadi lebih besar. Dengan varians yang tidak konstan tersebut tentu akan melanggar salah satu asumsi pada OLS yaitu tidak terpenuhinya asumsi homoskedastisitas. Dengan tidak terpenuhinya asumsi homoskedastisitas atau terjadi heteroskedastistas akan mengakibatkan hasil estimasi pada OLS menjadi bias dan tidak efisien.

UNNES
UNIVERSITAS NEGERI SEMARANG

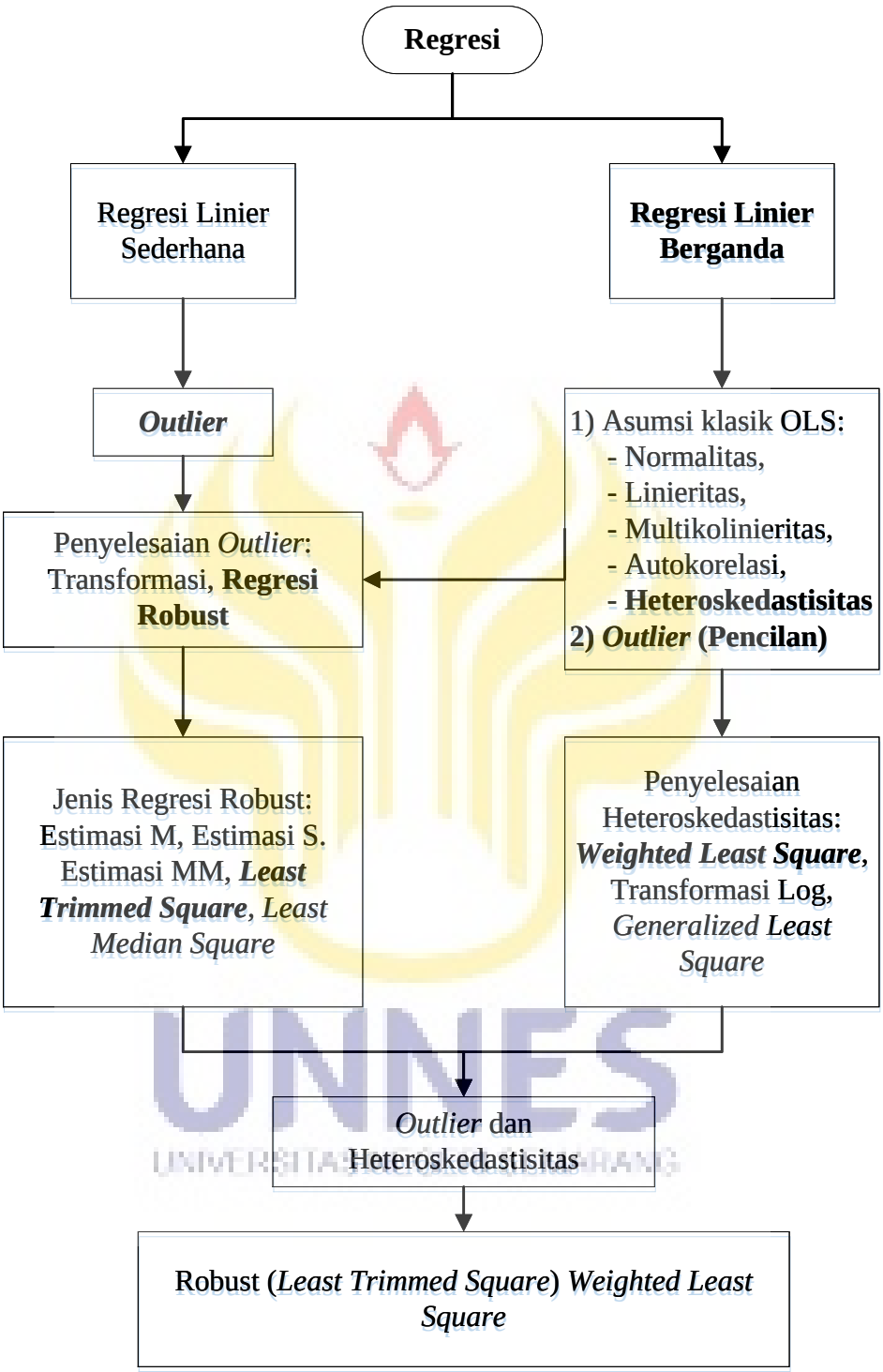
Banyak metode yang dapat dilakukan untuk mengatasi masalah heteroskedastisitas ini, diantaranya adalah transformasi data. Salah satu metode yang diangap paling mampu menangani masalah heteroskedastisitas yaitu dengan menggunajan metode kuadrat tertimbang atau *weighted least square.* metode ini mirip dengan OLS, namun pada WLS menggunakan suatu faktor pembobot untuk melakukan estimasi koefisien regresi. Dengan adanya dua kasus tersebut, yaitu *outlier* dan heteroskedastisitas, tentunya perlu adanya suatu metode yang dapat memberikan pengaruh yang berarti pada kedua kasus tersebut. Oleh karena itu dengan metode *robust* LTS dan *weighted least square* diharapkan mampu memberikan penyelasaian pada kasus *outlier* dan heteroskedastisitas.

Penelitian ini bermula dengan pengumpulan data sekunder, yaitu data ICMD Bank MEGA periode Januari 2012 sampai dengan Juni 2015. Dari data atersebut dilakukan uji asumsi normalitas, linieritas, keberartian stimultan, multikolinieritas, heteroskedastisitas dan autokorelasi. Kemudian dibuat model regresi awal menggunakan metode OLS dan medeteksi keberadaan *outlier* dengan metode *cook's distance* dan *standardisasi*. Karena pada data sudah tercirikan terkena *outlier* dan heteroskedastisitas maka dilakukan pendugaan parameter regresi menggunakan metode RWLS.

Metode RWLS diawali dengan mengelompokan variabel X_2 berdasarkan *near-neighbors* atau kelompok tetangga terdekat, yang diikuti pula pengelompokan variabel Y berdasarkan kelompok pada variabel X_2. Setelah diperoleh sejumlah g kelompok, dihitung nilai $MED(X_{2(j)})$ dan $MAD(Y_{(j)})^2$. Dari kedua hasil tersebut dilakukan estimasi pendugaan parameter menggunakan

metode LTS, dengan $MED(X_{2_{(j)}})$ sebagai X_2 dan $MAD(Y_{(j)})^2$ sebagai Y. Dari hasil estimasi dengan metode LTS dihitung nilai dari $\hat{Y}$, yang kemudian dihitung nilai mutlak invers dari $\hat{Y}$ tersebut yang akan digunakan sebagai pembobot awal 1 atau w_{1i}. Dari hasil metode LTS diperoleh nilai residu standar yang digunakan sebagai pembobot awal 2 atau w_{21} dengan bantuan fungsi Huber. Nilai pembobot akhir atau w_i diperoleh dari hasil perkalian w_{1i} dan w_{2i}. Pembobot akhir w_i digunakan sebagai pembobot pada penghitungan estimasi dengan metode *Weighted Least Square.* Hasil estimasi akhir dengan metode RWLS di uji dengan uji asumsi klasik dan pendeteksian *outlier* untuk mengetahui hasil akhir dari analisis *outlier* dan heteroskedastisitas dengan metode RWLS.

UNNES
UNIVERSITAS NEGERI SEMARANG



Gambar 2.1. Kerangka Berfikir Penulisan

BAB 5

PENUTUP

5.1 Simpulan

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan bahwa penyelesaian masalah *outlier* dan heteroskedastisitas pada suatu data dapat diselesaikan dan dianalisis menggunakan metode Robust (*Least Trimmed Square*) *Weighted Least Square*. Hasil estimasi dengan metode tersebut diperoleh persamaan regresi sebagai berikut:

$$\hat{Y} = 1{,}426\text{E+}12 - 6{,}241\text{E+}9X_2$$

Dengan nilai R sebesar 0,926 dan R^2 sebesar 0,858. Pada hasil estimasi tersebut data berdistribusi normal, terdapat hubungan linier antara variabel bebas dan variabel terikat, tidak terjadi heteroskedastisitas serta data yang bersifat *outlier* telah diiterasi dalam metode Robust *Weight Least Square*.

5.2 Saran

Berdasarkan hasil penelitian dan pembahasan, saran yang dapat penulis berikan sebagai berikut:

1. Bagi peneliti yang akan menyelesaikan masalah *outlier* dan heteroskedastisitas yang ada pada suatu data dapat menggunakan metode Robust (*Least Trimmed Square) Weighted Least Square.*
2. Bagi peneliti yang akan melakukan analisis lanjutan dengan metode Robust *Weighted Least Square* dapat menggunakan variasi pada metode

Robust dikarenakan masih terdapat banyak jenis metode Robust selain *Least Trimmed Square*, seperti *Least Median Square*, *MM-Estimation*, dsb.

3. Penelitian selanjutnya sebaiknya mencoba membandingkan metode-metode lain untuk menyelesaikan dan menganalisis data yang terkena *outlier* dan heteroskedastisitas yang ada pada suatu data.
4. Untuk menambah variasi dalam analisis metode Robust (*Least Trimmed Square) Weighted Least Square* dapat menggunakan bantuan program Matlab atau program lain yang mendukung.

UNNES
UNIVERSITAS NEGERI SEMARANG

DAFTAR PUSTAKA

Ardhiyanti, Hanna. 2011. Perbandingan Keefektifan Metode Regresi Robust Estimasi-M dan Estimasi-MM Karena Pengaruh Outlier Dalam Analisis Regresi Linier. Skripsi. Semarang: FMIPA Universitas Negeri Semarang. Tersedia di http://lib.unnes.ac.id/8132/1/10170.pdf [diakses pada 19-4-2016]

Chen, C. 2002. *Robust Regression and Outlier Detection with the ROBUSTBERG Procedure*. SAS Institute Inc: Cary NC.

Djauhari, Maman A. 1990. *Statistika Matematik*. Bamdung: Institut Teknologi Bandung.

Farida, Lina Suli. 2010. Analisis Regresi Linier Berganda Dengan Heteroskedastisitas Melalui Pendekatan *Weight Least Square* (Studi Kasus Data APBN Tahun 1976-2007). Skripsi. Jakarta: FST Universitas Islam Negeri Syarif Hidyatullah Jakarta. Tersedia di http://repository.uinjkt.ac.id/dspace/bitstream/123456789/851/1/94953 [diakses pada 27-08-2016]

Gujarati, Damodar. 1978. *Ekonometrika Dasar*. Translated by Sumarno Zain. 2010. Jakarta: Penerbit Erlangga.

Hanifah, Nurul., Nar Herrhyanto, & Fitriani Agustina. 2015. Penerapan Metode *Weighted Least Square* Untuk Mengatasi Heteroskedastisitas Pada Analisis Regresi Linier. *EurekaMatika Vol. 3, No. 1*: 105-114. Tersedia di http://journal.fpmipa.upi.edu/index.php/eurekamatika/article/viewFile/605/460 [diakses pada 28-08-2016]

Hendikawati, Putriaji. 2015. *Statistika: Metode dan Aplikasinya Dengan Excel dan SPSS*. Semarang: Universitas Negeri Semarang.

Mazziyya, Putri Ayu., I Komang Gede Sukarsa, & Ni Made Asih. 2015. Mengatasi Heteroskedastisitas pada Regresi Menggunakan Weighted Least Square. *E-Jurnal Matematika Vol. 4(1)*: 20-25.

Midi, Habshah., MD. Sohel Rana, & A.H.M. Rahmatullah Imon. 2009. The Performance of Robust Weighted Least Square in the Presence of Outliers and Heteroscedastic Errors. *Issues 7 Volume 8*: 351-361.

Musarafah, Raupong, & Nasrah Sirajang. Perbandingan Metode Robust Least Trimmed Square Dengan Metode Scale Dalam Mengestimasi Parameter Regresi Linier Berganda Untuk Data Yang Mengandung Pencilan. Fakultas MIPA Universitas Hasanuddin.

Nurcahyadi, Heru. 2010. Analisis Regresi Pada Data Outlier Dengan Menggunakan Least Trimmes Square (LTS) dan MM-Estimasi. Skripsi. Jakarta: FST Universitas Islam Negeri Syarif Hidayatullah Jakarta. Tersedia di http://repository.uinjkt.ac.id/dspace/bitstream/123456789/2309/1/HERU [diakses pada 25-04-2016]

Permana, A.T. 2012. Perbandingan Metode Least Trimmed Square (LTS) dan Penduga-S Sebagai Metode Pendugaan Parameter Regresi Robust. *Jurnal Matematika Universitas Brawijaya.* [diakses pada 25-04-2016]

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

Santoso, Singgih. 2010. *Statistik Multivariat: Konsep dan Aplikasi dengan SPSS*. Jakarta: PT Gramedia.

Sembiring, R. K. 1995. *Analisis Regresi.* Bandung: ITB.

Shin, Hee-Choon. 2013. Weighted Least Squares Estimation with Sampling Weights. *JSM*: 1523-1530. Tersedia di https://www.amstat.org/sections/srms/proceedings/y2013/files/308377_80748.pdf [diakses pada 28-08-2016]

Sungkawa, Iwa. 2009. Pendeteksian Pencilan (Outlier) dan Residual pada Regresi Linier. *Informatika Pertanian Vol. 18 No. 2:* 95-105. Tersedia di http://www.litbang.pertanian.go.id/warta-ip/pdf-file/2.iwa_ipvol18-2-2009.pdf [diakses 8-08-2016]

UNNES
UNIVERSITAS NEGERI SEMARANG