PERBANDINGAN PENDUGA THEIL DAN PENDUGA M (*M-ESTIMATOR*) TUKEY *BISQUARE* UNTUK MENDUGA *SLOPE* REGRESI LINIER SEDERHANA DENGAN METODE RESAMPLING BOOTSTRAP

SKRIPSI

Sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains dalam bidang Statistika

Oleh :
Retno Dewi Priutami
0810950058-95

PROGRAM STUDI STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2012

LEMBAR PENGESAHAN SKRIPSI

PERBANDINGAN PENDUGA THEIL DAN PENDUGA M (*M-ESTIMATOR*) TUKEY *BISQUARE* UNTUK MENDUGA *SLOPE* REGRESI LINIER SEDERHANA DENGAN METODE RESAMPLING BOOTSTRAP

oleh :
RETNO DEWI PRIUTAMI
NIM. 0810950058

Setelah dipertahankan di depan Majelis Penguji
pada tanggal 17 April 2012
dan dinyatakan memenuhi syarat untuk memperoleh gelar Sarjana Sains dalam bidang Statistika

Pembimbing	Ketua Penguji
Eni Sumarminingsih, S.Si, MM.	**Dr. Ir. Ni Wayan Surya W., MS.**
NIP. 19770515 200212 2 009	**NIP. 19551102 198103 2 001**

Mengetahui,
Ketua Jurusan Matematika
Fakultas MIPA Universitas Brawijaya

Dr. Abdul Rouf Alghofari, MSc
NIP. 19670907 199203 1 001

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Retno Dewi Priutami

NIM : 0810950058

Jurusan : Matematika

Penulisan Skripsi berjudul : PERBANDINGAN PENDUGA THEIL DAN PENDUGA M (*M-ESTIMATOR*) TUKEY *BISQUARE* UNTUK MENDUGA *SLOPE* REGRESI LINIER SEDERHANA DENGAN METODE RESAMPLING BOOTSTRAP

Dengan ini menyatakan bahwa :

1. **Isi dari Skripsi yang saya buat adalah benar-benar karya sendiri dan tidak menjiplak karya orang lain selain nama-nama yang termaktub di isi dan tertulis di daftar pustaka dalam Skripsi.**
2. **Apabila di kemudian hari ternyata Skripsi yang saya tulis terbukti hasil jiplakan, maka saya akan bersedia menanggung segala risiko yang akan saya terima.**

Demikian pernyataan ini dibuat dengan segala kesadaran.

Malang, 17 April 2012
Yang menyatakan

(Retno Dewi Priutami)
NIM. 0810950058

PERBANDINGAN PENDUGA THEIL DAN PENDUGA M (*M-ESTIMATOR*) TUKEY *BISQUARE* UNTUK MENDUGA *SLOPE* REGRESI LINIER SEDERHANA DENGAN METODE RESAMPLING BOOTSTRAP

ABSTRAK

Analisis regresi adalah teknik statistika untuk memeriksa dan memodelkan hubungan antara beberapa variabel. Metode kuadrat terkecil (MKT) merupakan metode pendugaan koefisien regresi yang sangat populer, namun pada MKT terdapat asumsi ideal yaitu bahwa vektor galat tersebut berdistribusi normal $[\varepsilon \sim N(0, \sigma^2)]$. Dalam praktek, tidak jarang terjadi penyimpangan terhadap asumsi tersebut di mana terdapat *outlier* pada data. Untuk mengatasi masalah tersebut di atas, terdapat beberapa metode regresi yang *robust* terhadap *outlier*, antara lain regresi nonparametrik yaitu penduga Theil dan regresi *robust* yaitu penduga M Tukey *Bisquare*. Selain kedua metode tersebut, Metode resampling Bootstrap juga mampu untuk mengakomodasi adanya penyimpangan pada data. Dalam skripsi ini, akan dibandingkan Metode Theil dan Tukey *Bisquare* dengan menerapkan Metode resampling Bootstrap di dalamnya. Perbandingan didasarkan KTG dan bias dari *slope* pada beberapa data yang digunakan. Data yang digunakan berupa tiga data sekunder dengan masing-masing kontaminasi *outlier* adalah 3%, 5% , 7% dan 9 set data simulasi dengan kombinasi dari kontaminasi *outlier* 3%, 5% , 7% dan banyak amatan 15, 30, dan 45. Dapat disimpulkan bahwa Metode Penduga M Tukey *Bisquare* merupakan metode pendugaan parameter yang lebih stabil dalam berbagai tingkat kontaminasi *outlier* karena menghasilkan nilai KTG dan bias yang lebih kecil daripada Metode Theil**.** Pola kebaikan dari Metode Theil dan Tukey *Bisquare* dipengaruhi oleh kombinasi antara banyak amatan dan proporsi kontaminasi *outlier*.

Kata kunci : Analisis regresi, MKT, *robust*, Metode Theil, Metode Tukey *Bisquare*, Resampling Bootstrap

A COMPARISON OF REGRESSION SLOPE PARAMETER OF THEIL ESTIMATOR AND M TUKEY BISQUARE ESTIMATION UNDER SIMPLE LINEAR REGRESSION WITH BOOTSRAP RESAMPLING METHOD

ABSTRACT

Regression analysis is a statistical technique to examine and model the relationships between several variables. Ordinary Least Square (OLS) is a very popular method of estimation of regression coefficients, but the OLS there is an ideal assumption that the error vector are normally distributed $[\varepsilon \sim N(0, \sigma^2)]$. In practice, deviation from that assumption is there are outliers in the data. To solve that problem, there are several methods for outlier robust regression, including nonparametric regression that is Theil estimators and *robust* regression which are M estimator Tukey Bisquare. besides the two methods, bootstrap resampling method is also able to accommodate the deviation in the data. In this thesis, there is a compared between Theil method and Tukey Bisquare by applying Bootstrap resampling method in it. MSE and bias of slope comparisons based on some of the data used. The data used in the form are three secondary data with each outlier contamination is 3%, 5%, 7% and 9 sets of simulation data with a combination of outlier contamination of 3%, 5%, 7% and many observations there are 15, 30, and 45. Based on the results, can be concluded that the M Tukey Bisquare estimator method is more stable method in varying degrees of outlier contamination because it produces the smaller MSE and bias than Theil method. The pattern of goodness of Theil method and Tukey Bisquare method are influenced by a combination of many observations and the proportion of outlier contamination.

Keywords : regression analysis, OLS, robust, Theil Method, Tukey Bisquare Method, Bootstrap Resampling

KATA PENGANTAR

Puji Syukur kehadirat Allah SWT atas berkat, rahmat dan hidayah-Nya sehingga skripsi dengan judul Penerapan Metode Resampling Bootstrap dalam Perbandingan Pendugaan Parameter *Slope* Penduga Theil dan Penduga M (*M-Estimator*) Tukey *Bisquare* pada Regresi Linier Sederhana dapat terselesaikan dengan baik.

Dalam penyusunan skripsi ini penulis telah banyak dibantu oleh berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Ibu Eni Sumarminingsih,S.Si,MM selaku dosen pembimbing I atas waktu dan bimbingan yang telah diberikan.
2. Ibu Dr. Ir. Ni Wayan Surya W., MS selaku dosen pembimbing II atas waktu dan bimbingan yang telah diberikan.
3. Ibu Dr. Rahma Fitriani, Msc selaku dosen penguji atas saran dan masukan yang telah diberikan.
4. Dr. Abdul Rouf Alghofari, M.Sc selaku Ketua Jurusan Matematika.
5. Almh.Ibu, Ayah, Prima, Pramu, Vicky Maulana dan seluruh keluarga atas cinta, kasih sayang, doa, semangat dan dukungannya.
6. Sahabat-sahabat Prodi Statistika 2008 atas kebersamaan, persahabatan dan dukungan yang berlimpah.
7. Seluruh staf pengajaran Matematika atas bantuan dan kerjasamanya
8. Semua pihak yang telah membantu hingga selesainya penyusunan skripsi ini.

Dalam penyusunan skripsi ini penulis menyadari masih terdapat banyak kekurangan. Oleh karena itu, berbagai saran ataupun kritik yang membangun akan sangat berguna bagi penulis dalam penulisan ilmiah selanjutnya

Malang, April 2012

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

Hal.

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan metode statistika yang digunakan dalam banyak penelitian baik bidang sosial maupun non sosial untuk mencari hubungan antara peubah respon dengan peubah prediktor. Bila pada teori menyatakan bahwa terdapat hubungan linier antara satu peubah respon dengan satu atau beberapa peubah prediktor, maka hubungan antara beberapa peubah tersebut dapat dimodelkan dengan $Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$, di mana β_0, β_1 adalah parameter persamaan regresi, ε_i adalah galat, p adalah banyak parameter dan i adalah nomor pengamatan (Walpole, 1992).

Salah satu metode pendugaan parameter yang digunakan adalah Metode Kuadrat Terkecil (MKT) yang membutuhkan asumsi ideal tertentu terhadap vektor galat, yaitu galat berdistribusi normal, ragam galat homogen dan non-autokorelasi galat. Sementara dalam beberapa penelitian, tidak jarang terjadi beberapa penyimpangan terhadap asumsi tersebut (Drapper and Smith, 1992).

Pelanggaran asumsi biasa disebabkan oleh galat yang tidak berdistribusi normal atau terdapat *outlier* pada data, apabila MKT diterapkan, maka penduga koefisien regresi menjadi bias dan atau tidak konsisten (Myers, 1990). Hal tersebut disebabkan oleh konsep MKT yang diketahui sangat peka terhadap *outlier* (Drapper and Smith, 1992), sehingga menyebabkan hasil pendekatan parameter dengan MKT kurang memuaskan karena koefisien dugaan parameter yang diperoleh tidak resisten atau bias.

Dalam MKT, untuk mengatasi keadaan tersebut langkah yang diambil adalah dengan tidak mengikutsertakan *outlier,* tetapi diperlukan identifikasi khusus untuk memutuskan apakah diikutsertakan atau tidak dalam analisis karena terkadang terdapat informasi yang tidak dimiliki oleh data lain. Maka. Beberapa saran telah diajukan Andrews,dkk.(1972) untuk mengatasi masalah tersebut dengan metode *robust* sebagai pengganti MKT (Drapper and Smith, 1992).

Untuk mengatasi masalah dalam pemenuhan asumsi, terdapat beberapa metode alternatif yang tahan terhadap pelanggaran asumsi. Metode statistika yang bersifat *robust* dapat mengakomodasi adanya *outlier* pada data dan sekaligus meniadakan pengaruhnya terhadap hasil analisis tanpa terlebih dahulu mengidentifikasi data tersebut sekaligus Metode *robust* lebih bersifat otomatis dalam menanggulangi data dengan *outlier,* sehingga hasil pendugaan parameter yang diperoleh lebih resisten. Beberapa metode regresi *robust* dengan pendekatan parametrik yang diajukan dalam Hoaglin (1983) adalah penduga M (*M-Estimation*) Huber, penduga M (*M-Estimation*) Tukey *bisquare* dan *Least Trimmed Squares* (LTS). Selain menggunakan pendekatan parametrik, adapula pendekatan secara nonparametrik seperti metode Theil, metode Fourier, metode Wavelet serta lainnya. Metode pendugaan parameter regresi dengan pendekatan nonparametrik sering digunakan karena tidak memerlukan asumsi-asumsi seperti pada MKT.

Pada penelitian sebelumnya yang dilakukan oleh Dewi (2011), telah dibandingkan antara penduga Theil dan penduga M (*M-Estimator*) Huber. Dari hasil penelitian tersebut disimpulkan bahwa penduga Theil lebih stabil dari penduga M (*M-Estimator*) Huber untuk data dengan kontaminasi *outlier* dan data dengan sebaran galat asimetris atau memiliki *skewness*. Sementara itu, pada penelitian yang dilakukan oleh Bekti (2009), dibandingkan penggunaan penduga M (*M-Estimator*) Tukey *bisquare*, penduga M (*M-Estimator*) Huber dan metode *Least Trimmed Square* (LTS) pada kasus hubungan antara luas panen padi dan curah hujan terboboti (WRI) di kabupaten Indramayu, Subang dan Karawang, di dapatkan hasil bahwa penduga M (*M-Estimator*) Tukey *bisquare* memberikan hasil yang lebih baik dibandingkan dengan kedua metode lainnya. Penelitian ini akan mengkaji lebih lanjut hasil penelitian Dewi (2011) dan Bekti (2009) dengan membandingkan penduga Theil dan penduga M (*M-Estimator*) Tukey *bisquare* untuk menentukan *slope* pada regresi linier dengan menerapkan Metode Bootstrap. Metode Bootstrap adalah metode resampling dengan pengembalian yang dapat diterapkan pada statistika parametrik dan nonparametrik. Pada statistika nonparametrik, metode ini bisa untuk menghindari asumsi-asumsi yang tidak terpenuhi, sedangkan pada statistika parametrik metode ini bisa memberikan pendugaan galat yang lebih akurat. Metode resampling Bootstrap ini dapat digunakan untuk model regresi yang menggunakan metode pendugaan selain MKT

Perbandingan kedua metode tersebut adalah dengan menghitung nilai bias penduga *slope* regresi dan Kuadrat Tengah Galat (KTG) penduga koefisien *slope* regresi.

1.2 Rumusan Masalah

Berdasarkan latar belakang, dapat dirumuskan permasalahan sebagai berikut:

1. Metode mana yang lebih baik antara penduga Theil dan penduga M (*M-Estimator*) Tukey *bisquare*, berdasarkan nilai bias dan Kuadrat Tengah Galat (KTG)
2. Bagaimana pola dari kebaikan kedua metode dalam mengatasi kasus *outlier* dengan memperhatikan besar kontaminasi *outlier* dan banyak amatan yang berbeda untuk masing-masing kontaminasi.

1.3 Batasan Masalah

Batasan masalah dalam skripsi ini adalah :

1. Masalah dibatasi hanya pada data dengan satu peubah prediktor dan satu peubah respon.
2. Data simulasi dibatasi bahwa data yang digunakan adalah data dengan *outlier* pada peubah Y (*Y-outlier*) dengan jarak *outlier* maksimal 4σ.
3. Data simulasi dan sekunder yang digunakan adalah data yang mengandung *outlier* sebesar 3%, 5% dan 7% dengan n untuk masing-masing proporsi *outlier* pada data sumilasi adalah 15, 30, 45.

1.4 Tujuan

Tujuan dalam penulisan skripsi ini adalah :

1. Mengetahui metode *robust* terbaik untuk mengatasi kasus *outlier*
2. Mengetahui pola kebaikan kedua metode dalam mengatasi kasus *outlier* dengan memperhatikan besar kontaminasi *outlier* dan banyak amatan yang berbeda untuk masing-masing kontaminasi.

1.5 Manfaat

Melalui penelitian ini diharapkan dapat menerapkan metode yang tepat dalam menduga parameter *slope* jika asumsi ideal tidak terpenuhi dan atau terdapat *outlier* pada amatan di berbagai bidang ilmu.

BAB II
TINJAUAN PUSTAKA

2.1 Analisis Regresi

Apabila memperhatikan ketergantungan antara suatu peubah acak Y terhadap peubah bebas X yang bervariasi, maka persamaan yang menghubungkan Y dan X disebut persamaan regresi. Suatu persamaan regresi diasumsikan bahwa peubah bebas (X) bersifat tetap atau *fixed*, sedangkan peubah respon (Y) bersifat acak (Drapper and Smith,1992).

Berikut merupakan persamaan garis lurus antara dua peubah yaitu peubah Y dan X :

$$Y = \beta_0 + \beta_1 X \tag{2.1}$$

persamaan regresi linier (2.1) dengan peubah respon Y dan peubah bebas X, menggambarkan hubungan antara nilai peubah Y dan X, di mana besar nilai peubah Y dipengaruhi oleh besar nilai peubah X dengan β_0 adalah *intercept* dan β_1 adalah *slope* (koefisien regresi). Dalam banyak kasus nilai dari koefisien β_0 tidak selalu diartikan sebagai besar nilai Y apabila X bernilai nol, akan tetapi sebagai koefisien penyesuai. β_1 menyatakan besarnya perubahan nilai Y pada setiap satuan perubahan nilai X, sehingga bisa dinyatakan sebagai berikut :

$$\beta_1 = \frac{\Delta Y}{\Delta X} \tag{2.2}$$

Hubungan antara peubah X dan Y tidak selalu tepat mengikuti persamaan linier (2.1), tetapi terdapat keragaman pada setiap nilai peubah bebas. Pada x_i misalnya, dapat dihasilkan bermacam-macam nilai *output* Y yang pusatnya diperkirakan terletak pada titik $Y_i = \beta_0 + \beta_1 X_i$. Adanya perbedaan nilai pengamatan dari garis regresi, maka persamaan (2.1) harus dimodifikasi. Dimisalkan perbedaan antara nilai amatan dari Y dengan $\hat{Y}$ adalah galat (ε_i) yaitu peubah acak yang menjelaskan kesalahan dari persamaan dalam mencocokkan data dengan tepat. Oleh karena itu, persamaan untuk menjelaskan sebuah garis regresi adalah :

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \tag{2.3}$$

2.2 Metode Kuadrat Terkecil (MKT)

Pendugaan parameter untuk persamaan regresi linier sederhana biasa menggunakan MKT di mana nilai dari Jumlah Kuadrat Galat (JKG) akan menjadi minimum. Drapper and Smith (1992) menyatakan bahwa pola hubungan antara peubah Y dan X yang bersifat linier dan sederhana dapat dimodelkan seperti persamaan (2.3) di mana $i = 1,2,3,\dots,n$, sehingga JKG pada pengamatan-pengamatan garis regresi sebenarnya adalah :

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2 \qquad (2.4)$$

Sebagai nilai dugaan β_0 dan β_1, dipilih nilai b_0 dan b_1. Untuk menentukan nilai b_0 dan b_1, persamaan (2.4) di diferensial parsial terhadap β_0 dan β_1 dan kemudian menyamakan hasil diferensial dengan nol, sehingga:

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i) \qquad (2.5)$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_{i=1}^{n} X_i (Y_i - \beta_0 - \beta_1 X_i) \qquad (2.6)$$

sehingga nilai dugaan bagi b_0 dan b_1 dapat diperoleh dari :

$$\left.\begin{array}{r} \sum_{i=1}^{n} Y_i - n b_0 - b_1 \sum_{i=1}^{n} X_i = 0 \\ \sum_{i=1}^{n} X_i Y_i - b_0 \sum_{i=1}^{n} X_i - b_1 \sum_{i=1}^{n} X_i{}^2 = 0 \end{array}\right] \qquad (2.7)$$

atau

$$\left.\begin{array}{r} b_0 n + b_1 \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i \\ b_0 \sum_{i=1}^{n} X_i + b_1 \sum_{i=1}^{n} X_i{}^2 = \sum_{i=1}^{n} X_i Y_i \end{array}\right] \qquad (2.8)$$

Persamaan (2.8) disebut dengan persamaan normal.

Penyelesaian dari persamaan (2.8) untuk menentukan nilai b_0 dan b_1 adalah :

$$b_0 = \bar{Y} - b_1 \bar{X} \qquad (2.9)$$

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - [(\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)]}{\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2} \qquad (2.10)$$

Model dugaan untuk regresi linier sederhana adalah :

$$\begin{aligned} \hat{Y} &= b_0 + b_1 X \\ &= \bar{Y} - b_1 \bar{X} + b_1 X \\ &= \bar{Y} - b_1 (\bar{X} - X) \end{aligned} \qquad (2.11)$$

Berdasarkan persamaan (2.11), MKT terkait dengan konsep rata-rata, karena garis regresi tepat melalui titik $(\bar{X}, \bar{Y})$.

2.2.1 Asumsi Metode Kuadrat Terkecil (MKT)

Beberapa asumsi yang harus dipenuhi dalam penggunaan MKT untuk menduga parameter regresi adalah :

1. Galat menyebar normal

Galat atau ε_i merupakan variabel acak normal dengan nilai tengah (*mean*) nol dan ragam sebesar σ^2, atau $E(\varepsilon_i) = 0, V(\varepsilon_i) = \sigma^2$. Kenormalan dari galat tersebut berpengaruh terhadap hasil dari pendugaan parameter, yaitu menyebabkan keputusan yang di bawah duga (*under estimate*) atau kelebihan duga (*over estimate*) terhadap taraf nyata percobaan atau α. Untuk mengetahui apakah galat menyebar normal atau tidak, dilakukan pengujian terhadap hipotesis:

H_0: galat menyebar normal

H_1: galat tidak menyebar normal

Salah satu metode yang digunakan adalah uji Anderson-Darling dengan statistik uji :

$$A^2 = -N - (\tfrac{1}{N}) \sum (2i-1) \Big(lnF(Y_i) + ln\big(1 - F(Y_{N+1-i})\big)\Big) \qquad (2.12)$$

di mana A menyebar Anderson Darling dan F adalah fungsi sebaran kumulatif dari sebaran normal dan Y_i adalah nilai pengamatan ke – i.

Kriteria yang digunakan dalam pengambilan keputusan adalah

$$A^2 \begin{cases} \leq A_\alpha \; ; H_0 \; diterima \\ > A_\alpha \; ; H_0 \; ditolak \end{cases} \qquad (2.13)$$

(Drapper and Smith, 1992).

2. Homoskedastisitas Ragam Galat

Berasal dari gabungan kata *homo* yang berarti sama dan *skedastisitas* yang berarti sebaran, sehingga artinya adalah ragam sama atau konstan

$$Var\,(\varepsilon) = \sigma^2$$

Heteroskedastisitas menyebabkan pendugaan parameter regresi dengan MKT menjadi tidak efisien. Apabila pendugaan parameter regresi tetap dilakukan dengan adanya *heteroskedastisitas* maka nilai parameter yang diperoleh tidak bias tetapi *standard error* (ragam) dari parameter yang diperoleh bias yaitu akan memiliki ragam yang lebih besar, sehingga mengakibatkan statistik uji (uji t atau uji F) menjadi tidak menentu. Untuk mengetahui apakah ragam homogen atau tidak, dilakukan pengujian terhadap hipotesis :

H_0: ragam homogen

H_1: ragam tidak homogen

Salah satu metode yang digunakan adalah uji Park (Kutner, 2004). Uji Park memiliki konsep seperti regresi, yaitu membentuk suatu persamaan. Akan tetapi untuk peubah Y, digunakan galat sebagai penggantinya. Adapun langkah-langkah dalam uji Park adalah :

1. Mentransformasi galat dan peubah X dengan tranformasi Logaritma Normal
2. Mengkuadratkan hasil transformasi galat
3. Meregresikan hasil transformasi dengan peubah Y adalah galat
4. Melakukan uji simultan terhadap persamaan regresi yang terbentuk

Statistik uji yang digunakan adalah :

$$F_{hitung} = \frac{Kuadrat\ \ Tengah\ Regresi\ \ (KTR)}{Kuadrat\ \ Tengah\ Galat\ \ (KTG)} \quad (2.14)$$

di mana $KTR = \frac{\boldsymbol{e}'\boldsymbol{e} - n\bar{e}^2}{db_{regresi}}$, dan

$$KTG = \frac{(\boldsymbol{e}'\boldsymbol{e} - n\bar{e}^2) - (\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{e} - n\bar{e}^2)}{db_{galat}}$$

Kriteria pengambilan keputusan adalah :

$$F_{hitung} \begin{cases} \leq\ F^{\alpha}_{p,n-p}\ ;\ H_0\ diterima \\ >\ F^{\alpha}_{p,n-p}\ ;\ H_0\ ditolak \end{cases} \quad (2.15)$$

3. Non-Autokorelasi Galat

Autokorelasi adalah korelasi antara anggota serangkaian amatan yang diurutkan menurut waktu atau ruang, di mana ε_i dan ε_j tidak berkorelasi sehingga $cov(\varepsilon_i \varepsilon_j) = 0$, dengan $i \neq j$

Konsekuensi bila pendugaan parameter dengan MKT adalah penduga parameter tidak lagi efisien (ragamnya tidak lagi minimum).

Selang kepercayaan menjadi lebar dan uji signifikansi kurang kuat atau pengujian dengan uji-t dan uji-F tidak lagi valid sehingga menghasilkan kesimpulan yang tidak valid pula.

Untuk mengetahui apakah galat saling bebas atau tidak, hipotesis yang digunakan adalah:

$H_0: \rho\left(\varepsilon_i \varepsilon_j\right) = 0$

$H_1: \rho\left(\varepsilon_i \varepsilon_j\right) \neq 0$

Salah satu metode yang digunakan adalah uji Durbin Watson (Drapper and Smith, 1992), dengan langkah-langkah sebagai berikut:

1. Regresi antara X dan Y dengan MKT
2. Didapatkan galat e_i
3. Menghitung statistik d dengan rumus:

$$d = \frac{\sum_{t=2}^{t=N}(e_t - e_{t-1})^2}{\sum_{t=1}^{t=N} e_t^2} \quad (2.16)$$

4. Mencari nilai kritis dL dan dU dari tabel Durbin Watson.

Kriteria pengambilan keputusan adalah :

$$d\begin{cases} > dU \ dan < 4 - dU \ ; terima \ H_0 \\ > 4 - \ dL \ dan < dL \ ; tolak \ H_0 \\ \geq 4 - dU \ dan \ \leq 4 - dL \ ; tidak \ ada \ solusi \end{cases} \quad (2.17)$$

4. Linieritas Data

Linieritas dapat diartikan bahwa antar peubah mempunyai pola hubungan linier $Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$. Jika hubungan antara peubah tidak linier, hasil analisis regresi akan *under estimate*. Untuk mendeteksi linieritas dapat digunakan beberapa metode yaitu (Montgomery and Peck, 1992) :

a. Metode Grafik

Melalui metode ini, pemeriksaan dilakukan melalui diagram pencar untuk melihat apakah hubungan antar peubah linier atau tidak. Diagram pencar menunjukkan hubungan antara nilai prediksi dengan galat yang menunjukkan pola acak.

b. Perbandingan nilai R^2

Nilai R^2 adalah persentase kemampuan peubah bebas menerangkan peubah respon. Semakin besar nilai R^2, maka semakin baik persamaan yang terbentuk.

c. Tabel Anova

Nilai F pada tabel Anova dapat dijadikan sebagai salah satu pertimbangan untuk menentukan apakah data yang digunakan linier atau tidak. Jika $F_{hitung} < \alpha$ maka model hubungan yang teridentifikasi adalah linier.

2.3 *Outlier*

Pada data hasil penelitian atau pengamatan, tidak jarang ditemukan satu atau beberapa amatan yang mempunyai nilai jauh dari pola kumpulan data keseluruhan. Sembiring (1995) mengatakan bahwa *outlier* adalah pengamatan yang nilainya jauh dari pusat data yang mungkin mempunyai pengaruh yang besar terhadap koefisien regresi.

Keberadaan *outlier* diindikasikan akan mengganggu proses analisis data dan harus dihindari. Dalam kaitannya dengan analisis regresi, dapat menyebabkan hal-hal berikut (Soemartini,2007) :

1. Galat yang besar dari model yang terbentuk
2. Dugaan parameter memiliki interval yang lebar

Berbagai kaidah telah diajukan untuk menolak atau memutuskan membuang *outlier* dari data. Penolakan atau pembuangan begitu saja bukanlah prosedur yang tepat untuk dilakukan, karena terkadang terdapat informasi yang tidak bisa diberikan oleh titik-titik data yang lain, misalnya timbul karena adanya kombinasi keadaan yang tidak biasa yang mungkin saja sangat penting dan perlu diselidiki lebih jauh. Oleh karena itu, sangat disarankan untuk menguji dahulu apakah *outlier* yang ada benar-benar memiliki pengaruh atau tidak (Drapper and Smith, 1992).

Pendeteksian *Outlier*

Dalam analisis regresi, terdapat asumsi ideal yang harus dipenuhi. Salah satu penyimpangan disebabkan adanya *outlier* pada amatan. Terdapat dua macam *outlier* pada data, yaitu terhadap X (*leverage point*) dan terhadap Y.

Salah satu cara untuk mendeteksian adanya *outlier* terhadap X adalah elemen diagonal dari matriks **H**. Matriks ini dikenal dengan nama *hat matrix* (Rosseeuw and Leroy, 2003). Berdasarkan persamaan normal pada (2.8), diperoleh model regresi linier yang ditulis dalam bentuk matriks:

$$\boldsymbol{Y} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{e}$$

di mana

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{pmatrix}, \widehat{\boldsymbol{\beta}} = \begin{pmatrix} \widehat{\beta_0} \\ \widehat{\beta_1} \\ \vdots \\ \widehat{\beta_p} \end{pmatrix}, \boldsymbol{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

vektor dari nilai duga $\hat{Y}_i$ dalam bentuk matriks dapat ditulis sebagai berikut :

$$\begin{aligned} \widehat{\boldsymbol{Y}} &= \boldsymbol{X}\widehat{\boldsymbol{\beta}} \\ &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{Y} \\ &= \boldsymbol{HY} \end{aligned}$$

matriks $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'$ merupakan matriks berukuran $n \times n$ yang berperan penting dalam analisis regresi. Selisih antara nilai amatan Y_i dengan nilai duga $\hat{Y}_i$ adalah galat $\widehat{e_i} = Y_i - \hat{Y}_i$ dan n galat dapat ditulis dalam bentuk matriks sebagai berikut :

$$\boldsymbol{e} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}} \qquad (2.18)$$

terdapat cara lain untuk menuliskan vektor galat $\boldsymbol{e}$ sebagai berikut :

$$\begin{aligned} \boldsymbol{e} &= \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} \\ &= \boldsymbol{Y} - \boldsymbol{HY} \\ &= (\mathbf{1} - \boldsymbol{H})\boldsymbol{Y} \end{aligned}$$

Menurut Myers (1990), elemen diagonal dari $\boldsymbol{H}$ memberikan informasi tentang data amatan yang mempunyai nilai laverage yang besar. Elemen diagonal ke-i dan matriks $\boldsymbol{H}$ yang dilambangkan dengan h_{ii} diperoleh dari :

$$\boldsymbol{h_{ii}} = \boldsymbol{x_i'}(\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{x_i} \qquad (2.19)$$

Dengan $\boldsymbol{x_i'}$ adalah vektor baris yang berisi nilai-nilai dari peubah bebas dalam pengamatan ke-i. Pada elemen diagonal matriks $\boldsymbol{H}$, diperoleh $\sum_{i=1}^{n} h_{ii} = p$ di mana p adalah banyaknya parameter model. Jika nilai h_{ii} lebih besar dari ${}^{2p}/_{n}$ maka pengamatan ke-i dikatakan sebagai *outlier* terhadap X (*leverage point*).

Untuk mendeteksi ada atau tidak *outlier* terhadap Y, dilakukan pengujian terhadap hipotesis :

H_0 : Pengamatan ke-i bukan *outlier*

H_1 : Pengamatan ke-i merupakan *outlier*

Salah satu statistik uji yang dapat digunakan untuk menguji adalah *studentized residual* atau *studentized deleted residual* yang didefinisikan :

$$|TRES|_i = e_i \left[\frac{n-p-1}{(n-p)MSE(1-h_{ii})-e_i^2}\right]^{1/2} \qquad (2.20)$$

Menurut Montgomery and Peck (1992), kriteria pengambilan keputusan adalah :

$$|TRES|_i \begin{cases} \leq t_{n-p-1}^{\alpha/2} \; ; H_0 \; diterima \\ > t_{n-p-1}^{\alpha/2} \; ; H_0 \; ditolak \end{cases} \qquad (2.21)$$

Untuk semua nilai yang mungkin dari $|TRES|_i$ mengikuti sebaran t dengan derajat bebas $n-p-1$, di mana p adalah banyaknya peubah bebas ditambah satu dan h_{ii} adalah elemen diagonal ke-i dari matriks $\boldsymbol{H}$ (Montgomery and Peck, 1992).

2.4 Pendeteksian Amatan Berpengaruh

Menurut Drapper and Smith (1992), amatan berpengaruh adalah amatan yang mempunyai pengaruh besar dalam pendugaan koefisien regresi. Amatan berpengaruh akan memberikan nilai galat yang besar atau mungkin juga tidak, tergantung model yang digunakan. Terdapat beberapa metode untuk mendeteksi amatan berpengaruh sebagai berikut :

1. *Cook's Distance*

Untuk mendeteksi ada atau tidak amatan berpengaruh terhadap koefisien regresi, dilakukan pengujian terhadap hipotesis:

H_0 :Pengamatan ke-i tidak berpengaruh terhadap koefisien regresi

H_1 : Pengamatan ke-i berpengaruh terhadap koefisien regresi

Menurut Montgomery and Peck (1992), *Cook's Distance* merupakan jarak antara pendugaan parameter dengan MKT yang diperoleh dari n amatan yaitu $\hat{\beta}$ dan pendugaan parameter yang diperoleh dengan terlebih dahulu menghapus pengamatan atau amatan ke-i yaitu $\hat{\beta}_{(i)}$. Jarak tersebut dapat dituliskan sebagai berikut:

$$D_i = \frac{(\widehat{\beta_i} - \hat{\beta})' X'X(\widehat{\beta_i} - \hat{\beta})}{p\ MSE}, i = 1, 2, \dots, n \tag{2.22}$$

dengan

$$MSE = \frac{SSE}{n-p} \tag{2.23}$$

kriteria pengambilan keputusan adalah :

$$D_i \begin{cases} \leq F^{\alpha}_{p,n-p}\ ; terima\ H_0 \\ > F^{\alpha}_{p,n-p}\ ; tolak\ H_0 \end{cases} \tag{2.24}$$

2. *The Difference In Fits Statistic* (DFITS)

Untuk mendeteksi ada atau tidak amatan berpengaruh, dilakukan pengujian terhadap $\hat{Y}_{(i)}$:

H_0 : Pengamatan ke-i tidak berpengaruh terhadap $\hat{Y}_{(i)}$

H_1 : Pengamatan ke-i berpengaruh terhadap $\hat{Y}_{(i)}$

Menurut Montgomery and Peck (1992), DFITS merupakan pengaruh pengamatan atau amatan ke-i pada nilai dugaan $\widehat{Y_i}$ yang didefinisikan sebagai :

$$DFITS_I = \left(\frac{h_{ii}}{1-h_{ii}}\right)^{1/2} e_i \left[\frac{n-p-1}{(n-p)MSE(1-h_{ii})-e_i^2}\right]^{1/2} \tag{2.25}$$

dan kriteria pengambilan keputusan adalah :

$$DFITS_i \begin{cases} \leq 2\sqrt{p/n}\ ;\ H_0\ diterima \\ > 2\sqrt{p/n};\ H_0\ ditolak \end{cases} \tag{2.26}$$

2.5 Model Kontaminasi

Dalam regresi linier sederhana, sering menggunakan MKT untuk menduga nilai dari parameter. Terdapat beberapa asumsi yang melandasi penggunaan MKT yang disebabkan oleh *outlier*, yang berasal dari sebaran normal dengan nilai tengah yang berbeda atau sebaran normal dengan ragam yang lebih besar daripada σ^2 yang diasumsikan. Sebaran yang mengalami hal tersebut disebut dengan sebaran normal yang terkontaminasi (Drapper and Smith, 1992).

Suatu sebaran tertentu $G(x)$ yang terkontaminasi dengan sebaran lain yang tidak diketahui $H(x)$ disebut sebagai suatu populasi yang terkontaminasi yang kemudian modelnya disebut dengan model terkontaminasi. Model terkontaminasi tersebut adalah sebagai berikut:

$$F(x) = (1 - \varepsilon)G(x) + \varepsilon H(x) \quad (2.27)$$

di mana ε adalah proporsi atau tingkat kontaminasi sebaran $H(x)$ pada sebaran $G(x)$, dengan nilai ε berkisar antara $0 \leq \varepsilon \leq 1$ (Hampel, dkk., 1986).

Model kontaminasi yang digunakan pada penelitian ini adalah gabungan dari model populasi yang memiliki sebaran galat normal dengan nilai galat awal yang sudah ditambahkan *outlier*. Model kontaminasinya adalah sebagai berikut:

$$F(x) = \Phi(x) + \varepsilon(x) \quad (2.28)$$

di mana ε adalah *outlier* yang yang sengaja dimasukkan kedalam nilai galat dari model populasi.

2.6 Metode Bootstrap

Metode Bootstrap adalah metode resampling dengan pengembalian yang dapat diterapkan pada statistika parametrik dan nonparametrik. Pada statistika nonparametrik, metode ini bisa untuk menghindari asumsi-asumsi yang tidak terpenuhi, sedangkan pada statistika parametrik metode ini bisa memberikan pendugaan galat yang lebih akurat. Metode resampling Bootstrap ini dapat digunakan untuk model regresi yang menggunakan metode pendugaan selain MKT (Efron and Tibshirani, 1993).

Dikatakan bahwa untuk contoh Bootstrap (B) ideal yang harus diambil adalah $B \to \infty$. Untuk menghitung simpangan baku, jumlah pengulangan contoh Bootstrap adalah antara 25-200. Untuk menghitung kuantil diperlukan jumlah pengulangan contoh Bootstrap sebesar $B \leq 2000$. Setiap contoh Bootstrap merupakan contoh acak yang saling bebas.

Prinsip pendugaan metode Bootstrap dapat dijelaskan dengan menganggap F sebagai sebaran empiris yang terdiri dari n pengamatan $(F \to x_1^*, x_2^*, \dots, x_n^*)$ yang masing-masing nilai pengamatan mempunyai peluang terambil sama besar yaitu 1/n. kemudian dilakukan proses sebagai berikut :

1. Ambil contoh acak Bootstrap sebanyak B pengulangan dari data awal (populasi awal) yang masing-masing terdiri dari n nilai

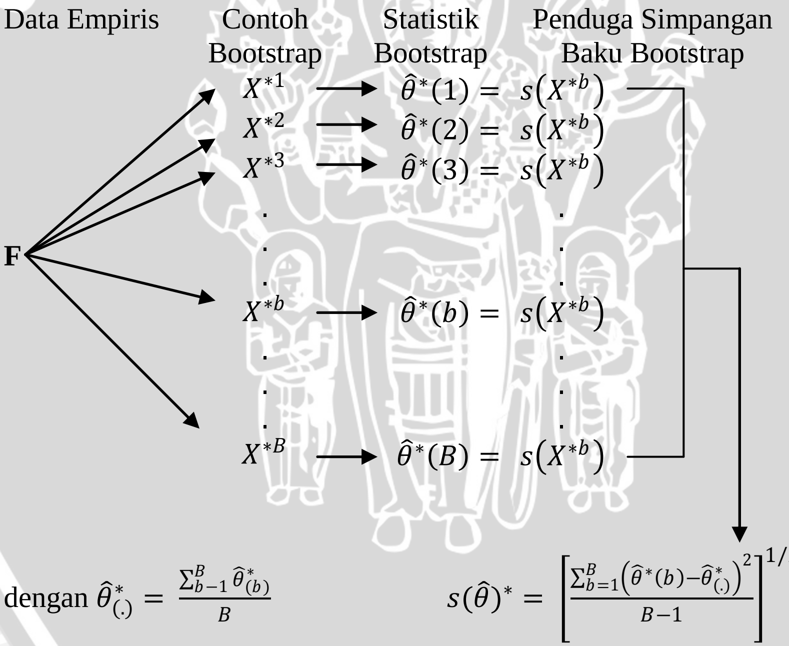
pengamatan, diambil dengan pengembalian dan saling bebas sehingga didapatkan $X^{*1}, X^{*2}, X^{*3}, \dots, X^{*B}$.

2. Hitung statistik Bootstrap untuk masing-masing contoh Bootstrap, $\widehat{\theta^*}(b) = s(X^{*b}); b = 1,2,3,\dots,B$ di mana $s(X^*)$ adalah rata-rata contoh Bootstrap.
3. Dilakukan pendugaan salah baku $se_F(\hat{\theta})$ dengan simpangan baku contoh pada B pengulangan :

$$S(\hat{\theta})^* = \left[\frac{\sum_{b=1}^{B}\left(\hat{\theta}^*(b)-\hat{\theta}^*_{(.)}\right)^2}{B-1}\right]^{1/2} \qquad (2.29)$$

dengan $\hat{\theta}^*_{(.)} = \frac{\sum_{b-1}^{B}\hat{\theta}^*_{(b)}}{B}$ (2.30)

Secara skematis prosedur pendugaan simpangan baku contoh Bootstrap menurut Efron and Tibshirani (1993) ditampilkan pada Gambar 2.1



Gambar 2.1 Skema Prosedur Pendugaan Simpangan Baku contoh Bootstrap (Efron and Tibshirani, 1993)

2.6.1 Resampling pada Pasangan Pengamatan

Resampling Bootstrap pada regresi dapat dilakukan dengan dua cara, yaitu resampling pada galat dan resampling pada pasangan pengamatan. Resampling pada galat dilakukan ketika metode regresi cukup baik dalam memodelkan data, maka galat akan mempunyai ragam yang homogen. Untuk resampling pada pasangan pengamatan dilakukan ketika metode regresi tidak cukup baik dalam memodelkan data, maka galat tidak akan mempunyai ragam yang homogen (Kutner, 2004).

Resampling Bootstrap pada pasangan pengamatan membutuhkan satu asumsi yaitu pasangan (X_i, Y_i)di tarik secara acak dari contoh yang telah ada (Efron and Tibshirani, 1993).

Hasil resampling dari pasangan pengamatan tersebut diberi notasi $(X_i{}^*, Y_i{}^*)$dan didapatkan penduga kuadrat terkecil Bootstrap sebagai berikut :

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

di mana $X^* = (x_1^*, x_2^*, \dots, x_n^*)$ adalah contoh Bootstrap untuk X, dan $Y^* = (y_1^*, y_2^*, \dots, y_n^*)$ adalah contoh Bootstrap untuk Y.

Nilai penduga Bootstrap untuk koefisien regresi adalah :

$$\hat{\beta}_{(.)}^* = \frac{\sum_{b-1}^{B} \widehat{\beta}_{(b)}^*}{B} \qquad (2.31)$$

dan nilai simpangan baku Bootstrap untuk *slope* regresi b_1^* adalah :

$$S(b_1)^* = \left[\frac{\sum_{b=1}^{B}\left(\widehat{\beta_1}^*(b) - \widehat{\beta_1}^*_{(.)}\right)^2}{B-1}\right]^{1/2} \qquad (2.32)$$

di mana $\widehat{\beta_1}^*_{(.)}$: nilai penduga Bootstrap untuk *slope* regresi

$\widehat{\beta_1}^*_{(b)}$: *slope* regresi dari setiap contoh Bootstrap

B : banyaknya pengambilan contoh Bootstrap

2.7 *Breakdown Point*

Breakdown point menunjukkan proporsi kontaminasi *outlier* yang masih dapat diatasi oleh suatu penduga sebelum memberikan hasil pendugaan yang jauh dari parameter sebenarnya. Pada pendugaan parameter dengan menggunakan MKT, apabila terdapat satu *outlier* dalam data, hasil dapat berpengaruh pada pendugaan parameter regresi karena MKT memiliki *breakdown point* 0%. Nilai tertinggi yang mungkin untuk *breakdown point* adalah 50%, karena pada data dengan kontaminasi *outlier* lebih dari 50% akan sulit untuk mendeteksi atau membedakan antara data yang merupakan *outlier* dan bukan. Oleh karena itu, semakin besar nilai *breakdown point* suatu penduga maka semakin *robust* penduga tersebut (Rousseeuw and Leroy, 1987).

2.8 Metode Theil

Menurut Daniel (1989), dalam banyak hal, beberapa pengamatan yang akan dikaji tidak selalu memenuhi asumsi yang mendasari uji parametrik. Dibutuhkan teknik inferensial dengan validitas yang tidak bergantung pada asumsi parametrik. Beberapa teknik dalam regresi nonparametrik memenuhi kebutuhan tersebut, karena tetap valid walaupun tidak dilakukan pemenuhan terhadap keseluruhan asumsi. Hal tersebut disebabkan untuk memperoleh koefisien penduga *slope* regresi, perhitungan dilakukan pada pasangan titik-titik amatan (X,Y) tanpa memperhatikan kondisi galat serta tidak terpengaruh terhadap *outlier* atau amatan berpengaruh. Penggunaan regresi nonparametrik berlandaskan pada asumsi :

1. Contoh yang diambil bersifat acak dan kontinyu
2. Regresi (Y | X) bersifat linier
3. Populasi diasumsikan tidak berdistribusi normal

Metode Theil merupakan suatu metode yang dirancang untuk mengurangi efek dari *outlier* terhadap pendugaan *slope* dan *intercept* dalam model regresi (Glaister, 2005).

Misalkan terdapat n pasangan pengamatan, $(X_{1,}Y_1)$, $(X_{2,}Y_2)$,…, $(X_{n,}Y_n)$, persamaan regresi linier sederhana adalah:

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

Theil (1950) dalam Sprent (1991) mengusulkan perkiraan *slope* garis regresi sebagai median *slope* dari seluruh pasangan garis dari titik-titik dengan nilai X yang berbeda. Untuk satu pasangan $(X_{i,}Y_i)$ dan (X_j, Y_j) *slope*-nya adalah:

$$b_{ij} = \frac{Y_j - Y_i}{X_j - X_i} \qquad (2.33)$$

untuk $i < j$ dan $i, j = 1, 2, \dots, n$

penduga β_1 dinotasikan dengan $\widehat{\beta_1}$ dinyatakan sebagai median dari nilai-nilai b_{ij} sehingga:

$$\widehat{\beta_1} = median\ (\ b_{ij}\ ) \qquad (2.34)$$

2.9 Penduga M (*M-Estimator*) dengan Fungsi Pembobot Tukey *Bisquare*

Regresi *robust* diperkenalkan Andrews (1972) dalam Ryan (1997). Metode ini merupakan alat untuk menganalisis data yang dipengaruhi oleh *outlier*. Suatu estimasi yang *robust* adalah relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada sebagian besar data. Prosedur ini ditujukan untuk mengakomodasi keanehan data, sekaligus menghilangkan dan bersifat otomatis dalam menanggulangi *outlier* (Ainuddin, 1989). Chen (2002) menyebutkan beberapa prosedur estimasi parameter dalam regresi *robust*, dua diantaranya penduga M (*M-Estimator*) dan *Least Trimmed Squares* (LTS).

Penduga M yang dilambangkan $t(x_1, \dots, x_n)$ merupakan penduga yang meminimumkan fungsi objektif

$$z = \sum_{i=1}^{n} \rho(x_i\ ;\ t) \qquad (2.35)$$

Seringkali $\rho(x_i\ ;\ t)$ tergantung pada fungsi x dan t dalam bentuk $x - t$, sehingga dapat ditulis dengan $\rho(x_i - t)$. Penduga t adalah nilai t yang diperoleh dengan menyelesaikan persamaan

$$\frac{dz}{dt} = 0 \qquad (2.36)$$

jika ψ adalah turunan pertama dari ρ, maka

$$\frac{dz}{dt} = \sum_{i=1}^{n} \psi(x_i - t) = 0 \qquad (2.37)$$

(Hoaglin, 1983).

Menurut Ali and Qadir (2005). Fungsi ρ mempunyai beberapa sifat sebagai berikut :

1. $\rho\,(0) = 0$
2. $\rho\,(e_i) = \rho(-e_i)$
3. $\rho\,(e_i) = 0$
4. Untuk $0 < e_1 < e_2 \rightarrow \rho\,(e_1) \leq \rho\,(e_2)$
5. ρ bersifat *continuous* dan *differentiable*

Penduga t jelas tergantung pada sebaran data, karena fungsi $\psi(.)$ diperoleh dari fungsi sebarannya. Penggunaan fungsi $\psi(.)$ yang didasarkan pada asumsi kenormalan akan menghasilkan penduga t yang tidak tepat, meskipun sebarannya mirip dengan normal namun memiliki ekor yang lebih panjang. Penduga *robust* didapatkan dengan memilih bentuk fungsi $\psi(.)$, sehingga menghasilkan penduga yang *robust* dan tidak banyak berubah walaupun terkontaminasi oleh data ekstrim. Adanya kontaminasi data ekstrim menyebabkan setiap pengamatan menerima pembobot w_i yang berbeda. Hettmanspenger and Sheater (1992) menyatakan bentuk fungsi $\psi(.)$ dengan menggunakan fungsi pembobot yaitu

$$\sum_{i=1}^{n} w_i(t)(x_i - t) = 0 \qquad (2.38)$$

Pada penduga *robust* M, terdapat beberapa jenis pembobot yang bisa digunakan, akan tetapi pada penelitian ini yang digunakan adalah jenis pembobot tukey *bisquare* yang disarankan oleh Tukey(1983). Fungsi pembobot yang disarankan oleh Tukey memakai fungsi obyektif

$$\rho\,(e_i^*) = \begin{cases} \dfrac{k^2\left[1-\left(1-\left(\frac{e_i^*}{r}\right)^2\right)^3\right]}{6} & ;\ |e_i^*| \le r \\ \dfrac{r^2}{6} & ;\ |e_i^*| > r \end{cases} \qquad (2.39)$$

dengan fungsi pengaruh

$$\varphi(e_i^*) = \begin{cases} e_i^*\left(1-\left(\frac{e_i^*}{r}\right)^2\right)^2 & ;\ |e_i^*| \le r \\ 0 & ;\ |e_i^*| > r \end{cases} \qquad (2.40)$$

dan fungsi pembobot

$$w(e_i^*) = \begin{cases} \left(1-\left(\frac{e_i^*}{r}\right)^2\right)^2 & ;\ |e_i^*| \le r \\ 0 & ;\ |e_i^*| > r \end{cases} \qquad (2.41)$$

dimana e_i^* adalah galat baku yang didapat dari proses MKT.

Nilai r adalah *tunning constant*. Nilai r muncul pada semua fungsi M-Tukey *bisquare*, baik itu pada fungsi objektif, fungsi pengaruh dan fungsi pembobot. *Tunning constant* adalah sebuah nilai atau konstanta yang dimasukkan dalam fungsi M-Tukey *bisquare*. Kelly (2008) menyatakan permasalahan dalam estimasi regresi *robust* adalah perlunya dilakukan pemilihan *tunning constant* agar estimasi M-Tukey *bisquare* ini menjadi lebih spesifik dan JKG yang dihasilkan dapat minimum. Kuzmic (2004) menyebutkan, pemilihan nilai r sebesar 4.685 pada semua fungsi M-Tukey *bisquare,* 95% efektif dalam pendugaan parameter regresi dan masih memberikan perlindungan yang baik terhadap *outlier*.

Konsep yang telah diuraikan diatas digunakan dalam pendugaan koefisien regresi. Pada pendugaan koefisien regresi sederhana dengan penduga M, dilakukan dengan menggunakan persamaan

$$\sum \psi\left(\frac{y_i-\beta_0-x_i\beta_1}{\sigma}\right)=0 \qquad (2.42)$$

$$\sum \psi\left(\frac{y_i-\beta_0-x_i\beta_1}{\sigma}\right)x_i=0 \qquad (2.43)$$

Jika w_i sebagai fungsi pembobot, maka bentuk persamaan (2.42) dan (2.43) menjadi

$$\sum w_i(\beta_0\beta_1)\left(\frac{y_i-\beta_0-x_i\beta_1}{\sigma}\right)=0 \qquad (2.44)$$

$$\sum w_i(\beta_0\beta_1)\left(\frac{y_i-\beta_0-x_i\beta_1}{\sigma}\right)x_i=0 \qquad (2.45)$$

(Hettmanspenger and Sheather, 1992)

Pendugaan koefisien regresi dengan penduga M dilakukan dengan menggunakan konsep MKT Terboboti dengan pembobot iteratif. Dimana prosedur pendugaan ini membutuhkan proses iterasi pada penentuan w_i yang dipengaruhi oleh nilai pendugaan sebelumnya. Nilai w_i akan berubah pada setiap iterasi sehingga pada akhirnya akan didapatkan nilai akhir w_i untuk memperoleh $\widehat{\beta_0}$ dan $\widehat{\beta_1}$. Koefisien regresi $\widehat{\beta_0}$ dan $\widehat{\beta_1}$ yang dihasilkan, akan memiliki sebaran. Sebaran ini yang berfungsi untuk mengetahui ukuran pemusatan dan penyebaran dari dugaan. Dimana ukuran pemusatan dilihat dari hasil nilai dugaan yang sama dengan nilai sebenarnya, sedangkan ukuran penyebaran dilihat dari nilai ragam dugaannya. Besar kecilnya ragam akan menjadi tolak ukur tentang ketelitian dari dugaan yang diperoleh.

Beberapa peneliti menyarankan untuk mendekati sebaran koefisien regresi β_0 dan β_1 dengan sebaran asimptotiknya, karena adanya proses iterasi dengan pembobot yang bergantung pada nilai galat. Sebaran asimptotik dari dugaan koefisien regresi diperoleh dengan persamaan

$$\sum_{i=1}^{n}\psi\left(\frac{y_i-\beta_0-x_i\beta_1}{\sigma}\right)=\sum_{i=1}^{n}\psi\left(\frac{y_i}{\sigma}\right)-\beta_0\sum_{i=1}^{n}\frac{1}{\sigma}\psi'\left(\frac{y_i}{\sigma}\right)-\beta_1\sum_{i=1}^{n}\frac{1}{\sigma}\psi'\left(\frac{y_i}{\sigma}\right)x_i \qquad (2.46)$$

$$\sum_{i=1}^{n}\psi\left(\frac{y_i-\beta_0-x_i\beta_1}{\sigma}\right)x_i=\sum_{i=1}^{n}\psi\left(\frac{y_i}{\sigma}\right)x_i-\beta_0\sum_{i=1}^{n}\frac{1}{\sigma}\psi'\left(\frac{y_i}{\sigma}\right)x_i-\beta_1\sum_{i=1}^{n}\frac{1}{\sigma}\psi'\left(\frac{y_i}{\sigma}\right){x_i}^2 \qquad (2.47)$$

Teorema 1 : jika nilai $\beta = 0$ adalah benar, maka
$n^{-1/2}\sum_{i=1}^{n}\psi(Y_i - \beta) = n^{-1/2}\sum_{i=1}^{n}\psi(Y_i) - n^{1/2}\beta E\psi'$ (2.48)

Berdasarkan teorema 1, maka persamaan (2.46) dan (2.47) menjadi
$$n^{-1/2}\sum_{i=1}^{n}\psi\left(\frac{y_i-\beta_0-x_i\beta_1}{\sigma}\right) = n^{-1/2}\sum_{i=1}^{n}\psi\left(\frac{y_i}{\sigma}\right) - n^{-1/2}\beta_0\frac{1}{\sigma}E\psi' - n^{-1/2}\beta_1\frac{1}{\sigma}E\psi'\bar{x} \quad (2.49)$$

$$n^{-1/2}\sum_{i=1}^{n}\psi\left(\frac{y_i-\beta_0-x_i\beta_1}{\sigma}\right)x_i = n^{-1/2}\sum_{i=1}^{n}\psi\left(\frac{y_i}{\sigma}\right)x_i - n^{-1/2}\beta_0\frac{1}{\sigma}E\psi'\bar{x} - n^{-1/2}\beta_1\frac{1}{\sigma}E\psi'\frac{1}{n}\sum_{i=1}^{n}x_i^2 \quad (2.50)$$

Dengan asumsi bahwa $\bar{x} = 0$ serta persamaan (2.49) dan (2.50) sama dengan nol, maka
$$\widehat{\beta_0} = \frac{\sigma}{E\psi'}\frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{y_i}{\sigma}\right) \quad (2.51)$$
$$\widehat{\beta_1} = \frac{\sigma}{E\psi'\sum x_i^2}\sum_{i=1}^{n}\psi\left(\frac{y_i}{\sigma}\right)x_i \quad (2.52)$$
$\widehat{\beta_1}$ konvergen terhadap sebaran normal dengan nilai harapan 0 dan ragam sebesar
$$\begin{aligned}
var\left(\widehat{\beta_1}\right) &= var\left(\frac{\sigma}{E\psi'\sum x_i^2}\sum_{i=1}^{n}\psi\left(\frac{y_i}{\sigma}\right)x_i\right)\\
&= \frac{\sigma^2}{(E\psi')^2}\frac{1}{\left(\sum x_i^2\right)^2}var\left(\sum_{i=1}^{n}\psi\left(\frac{y_i}{\sigma}\right)x_i\right)\\
&= \frac{\sigma^2}{(E\psi')^2}\frac{1}{\left(\sum x_i^2\right)^2}\sum var\left(\psi\left(\frac{y_i}{\sigma}\right)x_i\right)\\
&= \frac{\sigma^2}{(E\psi')^2}\frac{1}{\left(\sum x_i^2\right)^2}\sum x_i^2 var\left(\psi\left(\frac{y_i}{\sigma}\right)\right)\\
&= \frac{\sigma^2}{(E\psi')^2}\frac{1}{\left(\sum x_i^2\right)^2}\left[E\psi^2 - (E\psi)^2\right]\\
&= \frac{\sigma^2}{\left(\sum x_i^2\right)^2}\frac{E\psi^2}{(E\psi')^2} \quad (2.53)
\end{aligned}$$
$\widehat{\beta_0}$ konvergen terhadap sebaran normal dengan nilai harapan 0 dan ragam sebesar
$$\frac{\sigma^2 E\psi^2}{n(E\psi')^2} \quad (2.54)$$

Pendugaan koefisien regresi dengan pendekatan sebaran asimptotik mempunyai kekurangan yaitu terlalu sulit dalam perhitungan numerik, karena harus menduga beberapa nilai seperti $E\psi'$, $E\psi^2$, σ yang berkaitan dengan sebaran dari populasi awal dan $\widehat{\beta_0}$ dan $\widehat{\beta_1}$, oleh karena itu digunakan metode lain yang lebih sederhana yaitu dengan mengasumsikan galat mengikuti sebaran normal.

Rey (1983) menyatakan bahwa Penduga M (*M-Estimator*) adalah generalisasi dari penduga *maximum likelihood*. Fungsi *likelihood* yang digunakan adalah sebagai berikut :

$$L = \max \prod_{i=1}^{n} f(e_i) \tag{2.55}$$

atau ekuivalen dengan

$$-lnL = \min - \sum_{i=1}^{n} lnf(e_i) \tag{2.56}$$

sehingga Penduga M (*M-Estimator*) adalah solusi dari struktur yang lebih umum :

$$\min \sum_{i=1}^{n} \rho(e_i) \tag{2.57}$$

di mana $\rho(.) = -lnf(.)$ adalah fungsi obyektif dari galat atau fungsi yang memberikan kontribusi pada masing-masing galat.

Jika galat adalah peubah bebas acak yang mengikuti sebaran normal :

$$f(e_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp - \left(\frac{1}{2}\left[\frac{e_i - \mu}{\sigma^2}\right]^2\right) \tag{2.58}$$

untuk menduga $\widehat{\boldsymbol{\beta}}$ digunakan metode *maximum likelihood* dan fungsi *likelihood* yang digunakan adalah :

$$\begin{aligned} L(\beta_0, \beta_1) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma}\, exp - \left(\frac{1}{2}\left[\frac{(e_i-\mu)^2}{\sigma^2}\right]\right) \\ &= \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} \exp\left(-\sum_{i=1}^{n} \frac{1}{2}\left[\frac{(e_i-\mu)^2}{\sigma^2}\right]\right) \end{aligned} \tag{2.59}$$

Yang ekuivalen dengan

$$\begin{aligned} -l(\beta_0, \beta_1) &= -lnL(\beta_0, \beta_1) \\ &= n\, ln\sigma + \frac{n}{2}\, ln2\pi + \frac{1}{2\sigma^2} \sum_{i=1}^{n} (\varepsilon_i - \mu)^2 \end{aligned} \tag{2.60}$$

Penduga M (*M-Estimator*) adalah penyelesaian dari :

$$\begin{aligned} min \sum_{i=1}^{n} \rho(e_i)^2 &= \min \sum_{i=1}^{n} \rho\left(Y_i - \boldsymbol{x}_{\boldsymbol{i}}'\widehat{\boldsymbol{\beta}}\right)^2 \\ &= \min \sum_{\mathrm{i}=1}^{\mathrm{n}} \rho\left(Y_i - \sum_{j=0}^{k} \widehat{\beta}_j X_{ij}\right)^2 \end{aligned} \tag{2.61}$$

Untuk meminimumkan persamaan (2.61). maka dihitung turunan pertama dari ρ terhadap $\beta_j\,(j = 0,1,\dots,k))$ yang nilainya sama dengan nol (2.62). Misalkan φ adalah hasil *diferensiasi* dari fungsi simetris dan X_{ij} adalah amatan ke-i pada *regressor* ke-j, maka dapat diperoleh sistem persamaan sebagai berikut (Montgomery and Peck, 1992) :

$$\frac{\partial\rho}{\partial\beta_j} = -2\sum_{i=1}^{n}\varphi\left(Y_i - \sum_{j=0}^{k}\widehat{\beta_j}X_{ij}\right)X_{ij} = 0$$

$$\sum_{i=1}^{n} X_{ij}\,\varphi(e_i) = 0\,;\, j = 0,1,\dots,k \qquad (2.62)$$

Menurut Myers (1990), persamaan (2.61) dan (2.62) merupakan hal khusus dari bentuk umum berikut :

$$\sum_{i=1}^{n}\rho(e_i^*) = \sum_{i=1}^{n}\rho\left(\frac{Y_i - x_i'\widehat{\beta}}{s}\right) \qquad (2.63)$$

sehingga diperoleh fungsi pengaruh φ:

$$\sum_{i=1}^{n} X_{ij}\,\varphi(e_i^*) = \sum_{i=1}^{n} X_{ij}\,\varphi\left(\frac{Y_i - x_i'\widehat{\beta}}{s}\right) = 0 \qquad (2.64)$$

Prosedur pendugaan yang *robust* terhadap *outlier* menggunakan persamaan (2.64) dengan memilih fungsi pengaruh $\varphi(.)$ yang menyebabkan data dengan nilai galat yang besar tidak membangkitkan pengaruh yang tidak biasa.

Menurut Montgomery and Peck (1992), persamaan yang harus diselesaikan pada (2.64) bersifat tidak linier dan harus dipecahkan dengan cara iterasi. Nilai σ harus diduga dengan penduga yang *resistant*, yaitu :

$$s = {median|e_i - median(e_i)|}\big/{0.6745} \qquad (2.65)$$

Median digunakan dalam perhitungan penduga *robust* bagi σ karena median bersifat tidak peka terhadap *outlier*. Persamaan (2.64) dapat ditulis menjadi :

$$\sum_{i=1}^{n} X_{ij}\,\varphi\left(\frac{Y_i - x_i'\widehat{\beta}}{s}\right)$$
$$= \sum_{i=1}^{n} X_{ij}\,\varphi(e_i^*)$$
$$= \sum_{i=1}^{n} X_{ij}\,\frac{\varphi(e_i^*)}{e_i^*}\,e_i^* = 0 \qquad (2.66)$$

Jika

$$w_i = \frac{\varphi(e_i^*)}{e_i^*} \qquad (2.67)$$

Maka persamaan (2.66) dapat ditulis menjadi :

$$\sum_{i=1}^{n} X_{ij}\, w_i e_i^* = 0 \qquad (2.68)$$

Persamaan (2.66) juga merupakan penyelesaian dari peminimuman $\sum_{i=1}^{n} w_i \left(Y_i - \widehat{Y_l}\right)^2$ yaitu JKG dari MKT Terboboti. Oleh karena itu, regresi terboboti dapat dijadikan sebagai alat untuk penghitungan penduga M (*M-Estimator*) dengan matrik penduga koefisien regresi sebagai berikut

$$\widehat{\boldsymbol{\beta}_{(1)}} = (\boldsymbol{X}'\boldsymbol{W_o}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W_o}\boldsymbol{Y} \qquad (2.69)$$

di mana $\boldsymbol{W_o}$ adalah matriks diagonal dari pembobot dengan elemen diagonal ke-i adalah $w_{i.0}$ (Myers, 1990).

Penyelesaian koefisien regresi pada persamaan (2.64) disebut Penduga M (*M-Estimator*). Pendekatan yang dapat digunakan dalam pemilihan fungsi objektif dan fungsi pengaruh yang *robust* terhadap *outlier* adalah fungsi Tukey *bisquare* yang didefinisikan pada persamaan (2.39), (2.40) dan (2.41).

Menurut Myers (1990), langkah-langkah penghitungan penduga M (*M-Estimator*) adalah :

1. Mendapatkan vektor penduga awal $\widehat{\boldsymbol{\beta}_{(0)}}$ dari model regresi dengan Metode Kuadrat Terkecil (MKT) dan menghitung galat $(e_{i.0})$.
2. Dari sisaan awal, dihitung penduga *robust* s_0 dan pembobot awal $w_{i.0} = \frac{\varphi(e_{i.0}^*)}{e_{i.0}^*}$
3. Menggunakan Metode Kuadrat Terkecil Terboboti (MKTT) untuk mendapatkan penduga parameter baru yang *robust* : $\widehat{\boldsymbol{\beta}_{(1)}} = (\boldsymbol{X}'\boldsymbol{W_o}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W_o}\boldsymbol{Y}$.
4. Menjadikan penduga parameter pada langkah ke-3 sebagai $\widehat{\boldsymbol{\beta}_{(0)}}$ pada langkah ke-1 dan mendapatkan galat baru, nilai s baru dan pembobot yang baru
5. Kembali ke langkah 3.

Metode ini dilakukan berulang-ulang sampai tercapai kekonvergenan yang mempunyai kriteria sebagai berikut :

a. Perubahan maksimal pembobot dari satu iterasi ke iterasi berikutnya tidak berarti (kurang dari 0.01).
b. Jumlah harga mutlak sisaan dari satu iterasi ke iterasi berikutnya sama.

Metode yang menggunakan MKTT yang diulang ini disebut dengan iterasi kuadrat terkecil terboboti kembali (*Iterativly Reweighted Least Square*).

2.10 Uji Signifikansi Parameter

Uji signifikansi parameter dilakukan untuk menguji apakah $\beta_0 = 0, \beta_1 = 0, \ldots, \beta_p = 0$ atau minimal ada satu $\beta_p \neq 0$. Jika dalam pengujian parameter ternyata minimal ada satu $\beta_p \neq 0$, maka parameter layak untuk diikutsertakan dalam model.

1. Metode Theil

Pengujian signifikansi dalam statistika nonparametrik dilakukan dengan uji simultan (bersama) dengan metode Brown-Mood. Untuk mengetahui apakah peubah bebas mempunyai pengaruh secara bersama-sama atau tidak, dilakukan pengujian terhadap hipotesis (Daniel, 1989) :

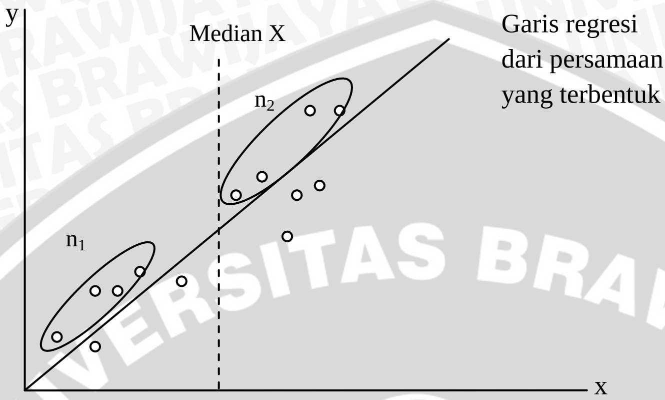
H_0 : $\beta_0 = \cdots = \beta_p = 0$

H_1 : minimal ada satu $\beta_p \neq 0$

Statistik uji yang digunakan untuk menguji adalah uji X^2 dengan pendekatan sebaran χ^2 yang didefinisikan

$$X^2 = \frac{8}{n}\left[\left(n_1 - \frac{n}{4}\right)^2 + \left(n_2 - \frac{n}{4}\right)^2\right] \qquad (2.70)$$

di mana n_1 adalah banyaknya titik data yang berada diatas garis regresi dan berada disebelah kiri garis median, n_2 adalah banyaknya titik data yang berada diatas garis regresi dan berada disebelah kanan garis median. Untuk lebih jelasnya maksud dari n1 dan n2 dapat dilihat pada Gambar 2.2



Gambar 2.2 Lokasi titik data himpunan n_1 dan n_2

Dengan kriteria pengambilan keputusan :

$$X^2 \begin{cases} \leq \chi^2_{p,\alpha/2} \ ; H_0 \ diterima \\ > \chi^2_{p,\alpha/2}; H_0 \ ditolak \end{cases} \qquad (2.71)$$

Untuk nilai p sebagai derajat bebas, tergantung dari banyak data. Apabila n < 20, maka p akan bernilai 2 dan apabila n ≥ 20, maka p akan bernilai 1.

2. Metode Penduga M (*M estimator*) Tukey *Bisquare*

- Uji Parsial (Uji t)

Uji statistik t digunakan untuk mengetahui apakah pengaruh satu peubah bebas secara individual dalam persamaan signifikan atau tidak. Untuk mengetahui apakah peubah bebas signifikan atau tidak, dilakukan pengujian terhadap hipotesis (Ghozali, 2006):

H_0 : $\beta_0 = 0$ H_0 : $\beta_p = 0$

H_1 : $\beta_0 \neq 0$ H_1 : $\beta_p \neq 0$, di mana $p \neq 0$

Statistik uji yang digunakan untuk menguji adalah uji t. Pada metode Tukey *Bisquare*, terdapat pembobot yang digunakan dalam menduga koefisien parameter regresi atau biasa disebut MKT terboboti, maka statistik uji yang digunakan adalah uji t terboboti yang didefinisikan

Untuk H_0 : $\beta_0 = 0$

$$t_{hit} = \frac{b_0}{s(b_0)} \qquad (2.72)$$

Untuk $H_0 : \beta_p = 0$

$$t_{hit} = \frac{b_p}{s(b_p)} \quad (2.73)$$

Dimana

$s(b_0)^2 = s^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum w_i(x_i - \bar{x})^2}\right]$, dan

$s(b_p)^2 = \frac{s^2}{\sum w_i(x_i - \bar{x})^2}$

Dengan kriteria pengambilan keputusan :

$$t_{hit} \begin{cases} \leq t_{n-p}^{\alpha/2} \ ; \ H_0 \ diterima \\ > t_{n-p}^{\alpha/2} \ ; \ H_0 \ ditolak \end{cases} \quad (2.74)$$

- **Uji Simultan (uji F)**

Uji statistik F digunakan untuk mengetahui apakah semua peubah bebas model regresi mempunyai pengaruh secara bersama-sama (simultan) terhadap peubah respon. Untuk mengetahui apakah peubah bebas mempunyai pengaruh secara bersama-sama atau tidak, dilakukan pengujian terhadap hipotesis (Ghozali, 2006) :

$H_0 : \beta_0 = \cdots = \beta_p = 0$

H_1 : minimal ada satu $\beta_p \neq 0$

Statistik uji yang digunakan untuk menguji adalah uji F yang didefinisikan

$$F_{hit} = \frac{Kuadrat \ \ Tengah \ Regresi \ \ (KTR)}{Kuadrat \ \ Tengah \ Galat \ \ (KTG)} \quad (2.75)$$

Dimana

$KTR = \frac{\sum_{i=1}^{n} w_i(\hat{y}_i - \bar{y})^2}{p}$, dan

$KTG = \frac{\sum_{i=1}^{n} w_i(\hat{y}_i - \bar{y})^2}{n-p-1}$

Kriteria pengambilan keputusan adalah :

$$F_{hitung} \begin{cases} \leq F_{p,n-p}^{\alpha} \ ; \ H_0 \ diterima \\ > F_{p,n-p}^{\alpha} \ ; \ H_0 \ ditolak \end{cases} \quad (2.76)$$

2.11 Sifat-sifat Penduga Parameter Yang Baik

Secara umum, suatu penduga dikatakan sebagai penduga yang baik jika rata-rata penduga tersebut sama dengan nilai parameter yang sebenarnya. Sifat-sifat penduga yang baik antara lain adalah :

1. Tidak Bias

Bias dari suatu penduga adalah perbedaan antara nilai harapan penduga dan nilai parameter. Secara metematis dapat ditulis:

$$Bias = E\left(\hat{\theta}\right) - \theta \neq 0 \tag{2.77}$$

Suatu penduga parameter dikatakan tidak bias jika $E\left(\hat{\theta}\right) - \theta = 0$

2. Ragam Penduga Kecil

Suatu penduga dikatakan terbaik apabila penduga tersebut memiliki ragam terkecil dibandingkan dengan penduga lain dengan metode yang berbeda. Dalam bentuk matematis dapat didefinisikan sebagai berikut :

$$E\left(\hat{\theta} - E(\hat{\theta})\right)^2 < E\left(\hat{\theta}^* - E(\hat{\theta}^*)\right)^2 \tag{2.78}$$

di mana : $\hat{\theta}^*$ adalah penduga dengan metode lain

3. Kuadrat Tengah Galat (KTG) Minimum

Kuadrat Tengah Galat (KTG) adalah nilai harapan dari kuadrat perbedaan penduga dengan parameter populasi.

KTG $\left(\hat{\theta}\right) = E\left(\hat{\theta} - \theta\right)^2$

$= E\left(\theta - E\left(\hat{\theta}\right) + E\left(\hat{\theta}\right) - \theta\right)^2$

$= E\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)^2 + E\left(E\left(\hat{\theta}\right) - \theta\right)^2 - 2E[(\hat{\theta} -$

$E\theta)E\theta-\theta$

(2.79)

Karena

$E\left[\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)\left(E\left(\hat{\theta}\right) - \hat{\theta}\right)\right]$

$= E\left[\hat{\theta}.E\left(\hat{\theta}\right) - \left(E\left(\hat{\theta}\right)\right)^2 - \hat{\theta}.\theta + \theta.E\left(\hat{\theta}\right)\right]$

$= \left(E\left(\hat{\theta}\right)\right)^2 - \left(E\left(\hat{\theta}\right)\right)^2 - \theta.E\left(\hat{\theta}\right) +$

$\theta.E\left(\hat{\theta}\right) = 0$

(2.80)

Sehingga KTG $\left(\hat{\theta}\right)$ = Var $\left(\hat{\theta}\right)$ + (Bias$\left(\hat{\theta}\right)$)^2

KTG adalah jumlah dari dua kuantitas yaitu ragam dan bias kuadrat. Oleh karena itu penduga yang memiliki KTG terkecil lebih baik dari kriteria minimum dari salah satu komponen KTG (Mendenhall, Scheaffer and Wackerly, 1981).

4. Konsisten

Suatu penduga dikatakan konsisten apabila memenuhi syarat berikut :

1. Jika ukuran contoh semakin bertambah maka penduga akan mendekati parameternya. Jika besarnya contoh menjadi tak terhingga maka penduga konsisten harus dapat memberi suatu pendugaan titik yang sempurna terhadap parameternya.
2. Jika ukuran contoh bertambah tak terhingga maka sebaran sampling penduga akan mengecil menjadi suatu garis tegak lurus di atas parameter yang sebenarnya dengan probabilitas sama dengan 1.

Jadi $\hat{\theta}$ merupakan penduga konsisten, jika dan hanya jika $E\left(\hat{\theta}-\theta\right)^2 \rightarrow 0, jika\ n \rightarrow \infty$. Suatu penduga konsisten belum tentu merupakan penduga yang baik, karena konsisten hanya merupakan salah satu syarat.

2.12 Perbandingan Penduga Theil dan Penduga M (*M-Estimator*) Tukey *Bisquare* untuk Menduga *Slope* Regresi

Dalam pendugaan parameter dengan kondisi data yang tidak ideal atau tidak memenuhi asumsi, menggunakan penduga Theil dan penduga M (*M-Estimator*) Tukey *bisquare* dapat diperoleh hasil yang lebih baik daripada menggunakan MKT, hal tersebut dikarenakan pada penduga Theil dan M (*M-Estimator*) Tukey *bisquare* memiliki persamaan yaitu tahan terhadap adanya *outlier* pada data, sehingga bisa dijadikan metode alternatif untuk menduga parameter regresi. Kedua penduga tersebut mempunyai perbedaan mendasar yaitu penduga Theil didasarkan pada pendekatan nonparametrik karena sebaran galat tidak didasarkan pada asumsi yang mengikuti suatu sebaran tertentu dan M (*M-Estimator*) Tukey *bisquare* didasarkan pada pendekatan parametrik karena masih melibatkan galat yang didapat dari perhitungan dengan MKT.

Perhitungan untuk memperoleh koefisien penduga *slope* regresi pada penduga Theil dilakukan pada pasangan titik-titik amatan (X, Y). Koefisien penduga *slope* didapatkan dari median *slope* dari seluruh *slope* yang diperoleh dari pasangan titik-titik amatan. Sementara itu untuk M (*M-Estimator*) Tukey *bisquare* menggunakan konsep MKT dengan menambahkan pembobot pada proses pembentukan matrik penduga.

Pada data sekunder, Metode resampling Bootstrap diterapkan pada pasangan peubah X dan Y serta pada proses analisis pada kedua metode regresi. Sementara untuk data simulasi, Metode resampling Bootstrap diterapkan pada proses analisis data. Pembandingan sifat penduga Theil dan M (*M-Estimator*) Tukey *bisquare* untuk data simulasi dilakukan dengan menghitung nilai bias dan KTG dari dugaan parameter yang dihasilkan, dengan rumus sebagai berikut :

$$Bias = \frac{1}{m}\sum_{i=1}^{m}\left|\beta - \hat{\beta}_i\right| \qquad (2.81)$$

$$KTG = \frac{1}{m}\sum_{i=1}^{m}\left(\beta - \hat{\beta}_i\right)^2 \qquad (2.82)$$

di mana m adalah banyaknya replikasi pada studi simulasi.

Untuk Pembandingan sifat penduga Theil dan M (*M-Estimator*) Tukey *bisquare* untuk data sekunder dilakukan dengan menghitung nilai KTG dari persamaan regresi yang terbentuk. Nilai dapat dihitung menggunakan rumus sebagai berikut :

$$KTG = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (2.86)$$

di mana n adalah banyaknya amatan pada data.

BAB III
METODE PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini berupa data sekunder dan data simulasi. Untuk data sekunder, diambil dari penelitian Hidayat (2009) dengan judul penelitian Pengaruh Kadar *Immunoglobulin Gamma* (IgG) Kolostrum Induk terhadap IgG Serum Darah Anak pada Sapi *Friesiean Holstein*, Aida (2009) dengan judul penelitian Pengaruh Jumlah Pasien terhadap Produksi Limbah Medis Padat Diruang Rawat Inap dan Unit Gawat Darurat dan Aliudin (2006) dengan judul penelitian Korelasi antara Bobot Badan Awal dengan Bobot Badan Akhir dan Bobot Badan Karkad Ayam Pedaging dengan Penambahan *Choline Chloride* dalam Pakan.

Data simulasi yang digunakan adalah data simulasi dengan tiga level ukuran sampel (n = 15, 30, 45). Pada studi simulasi ini, pembangkitan data berbagai ukuran sampel dilakukan pada tiga tipe distribusi galat yaitu data dengan distribusi galat terkontaminasi *outlier* (3% Y-*outlier*, 5% Y-*outlier*, 7% Y-*outlier*). Pada masing-masing tipe distribusi galat dan ukuran sampel tersebut dilakukan replikasi sebanyak 1000 kali (m = 1000).

Pembandingan sifat penduga Metode Theil dan Penduga M (*M-Estimator*) Tukey *bisquare* untuk data simulasi dilakukan dengan menggunakan nilai bias dan KTG dari dugaan nilai parameter yang dihasilkan. Semakin kecil nilai bias dan KTG menunjukkan bahwa suatu metode semakin baik, yaitu tak bias dan efisien. Untuk data sekunder, pembandingan sifat penduga dilakukan dengan menggunakan nilai KTG dari dugaan nilai peubah Y dari persamaan regresi yang terbentuk.

3.2 Metode Analisis

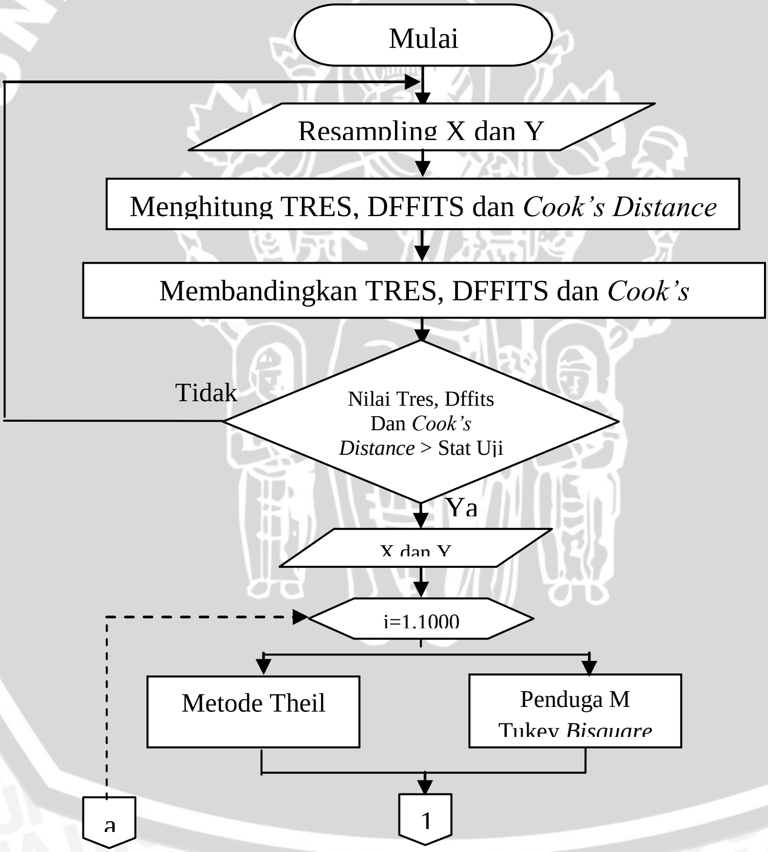
3.2.1 Data Sekunder

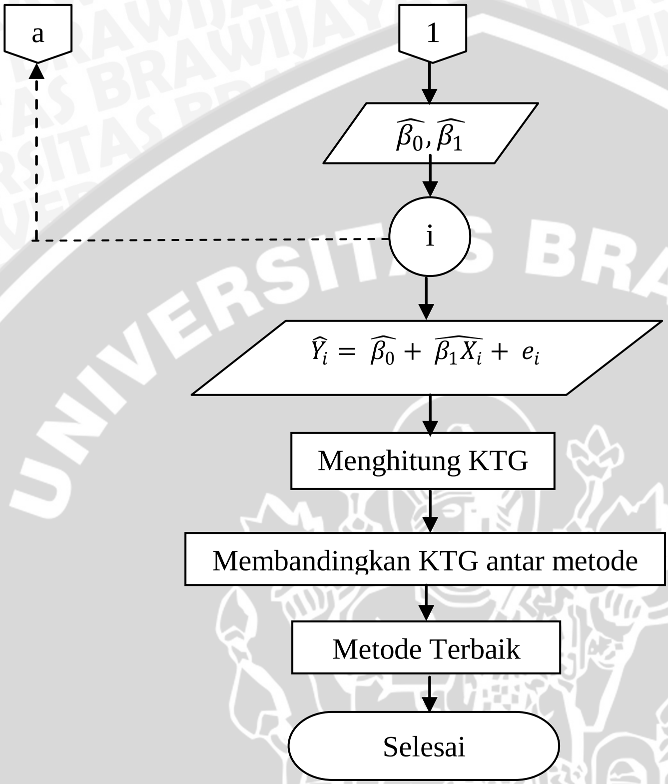
Ada beberapa pengujian awal yang harus dilakukan sebelum melakukan analisis dengan Metode Theil dan Penduga M (*M Estimator*) Tukey *Bisquare*. Beberapa pengujian awal tersebut adalah mendeteksi *outlier* dengan menghitung nilai TRES (2.22), mendeteksi amatan berpengaruh dengan menghitung nilai *Cook's Distance* (2.24) dan DFITS (2.27).

Setelah pengujian awal dilakukan dan mendapatkan hasil sesuai yang diinginkan yaitu terdapat *outlier* dan terdapat amatan berpengaruh, maka langkah selanjutnya adalah sebagai berikut:

1. Melakukan resampling Bootstrap pada masing-masing pasangan data sebanyak 1000 kali.
2. Melakukan analisis pada data dengan menggunakan metode Theil dan penduga M (*M Estimator*) Tukey *bisquare* sebanyak 1000 kali.
3. Menghitung nilai KTG dari dugaan nilai peubah Y yang dihasilkan dari persamaan regresi yang terbentuk.
4. Membandingkan nilai KTG untuk masing-masing metode.

Secara sistematis langkah – langkah analisis untuk data sekunder dalam penelitian ini dapat dilihat pada Gambar 3.1





Gambar 3.1 Diagram Alir Metode Analisis pada Data Sekunder.

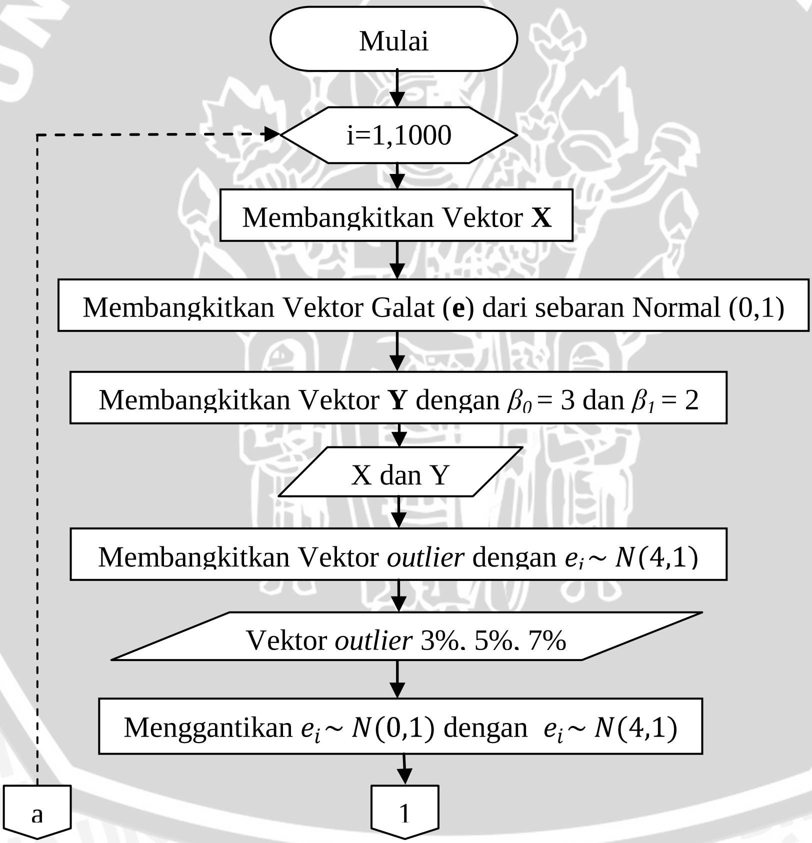
3.2.2 Data Simulasi

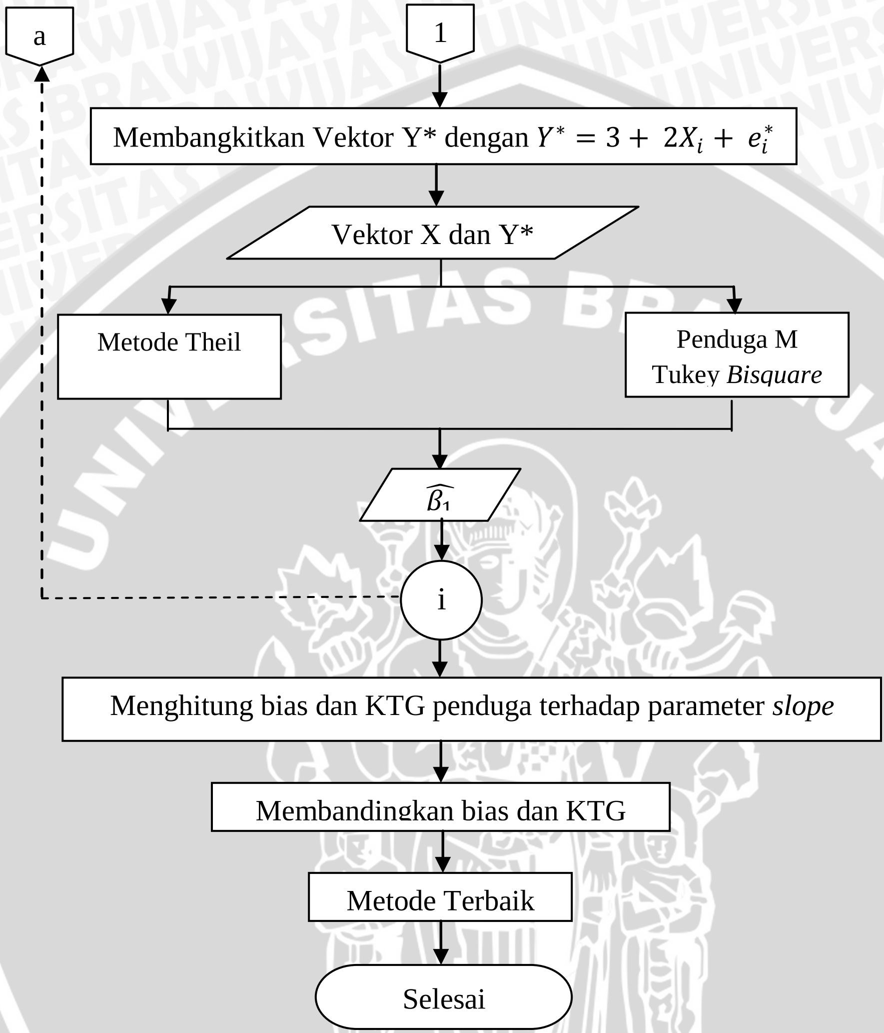
Dalam simulasi ini dibangkitkan data sampel berukuran (n=15, 30, 45) dengan replikasi masing-masing sebanyak 1000 kali. Langkah-langkah simulasi yang dilakukan adalah sebagai berikut:

1. Membangkitkan vektor galat (**e**) dari sebaran normal baku N (0,1) berukuran $n \times 1$.
2. Menetapkan vektor **X** dengan ukuran $n \times 1$, dimana n = 15, 30, 45.
3. Menetapkan β_0 = 3 dan β_1 = 2, dibangkitkan vektor **Y**, yaitu $Y_i = 3 + 2X_{\boldsymbol{i}} + e_i$.

4. Dari data bangkitan tersebut kemudian dikontaminasi *outlier* dengan cara menggantikan 3%, 5%, 7% variabel respon data tersebut dengan *outlier* pada variabel Y dengan jarak pencilan sejauh 4σ yang dibangkitkan dari $e_i \sim N(4,1)$.
5. Menduga koefisien regresi dengan menggunakan metode Theil dan penduga M (*M-Estimator*) Tukey *bisquare*. Menyimpan nilai $\hat{\beta}$ yang diperoleh dari masing-masing metode.
6. Mengulangi langkah 1 sampai 5 sebanyak 1000 kali.
7. Menghitung nilai bias dan KTG ketika sudah mendapatkan nilai $\hat{\beta}$ sebanyak 1000.
8. Membandingkan nilai bias dan KTG untuk masing-masing penduga.

Secara sistematis dapat dilihat pada Gambar 3.2





Gambar 3.2 Diagram Alir Metode Analisis pada Data Simulasi

BAB IV
HASIL DAN PEMBAHASAN

4.1 Pendeteksian *Outlier* dan Amatan Berpengaruh pada Data Sekunder

Metode Theil dan M Tukey *Bisquare* tidak bisa diterapkan pada semua jenis data. Data yang digunakan adalah data yang mengandung *outlier* dan atau amatan berpengaruh. Oleh karena itu, sebelum mulai melakukan analisis, terlebih dahulu dilakukan pendeteksian ada tidaknya *outlier* dengan menggunakan uji TRES, dan amatan berpengaruh menggunakan uji *Cook's Distance* dan DFITS.

Pada Tabel 4.1 didapatkan hasil bahwa terdapat *outlier* maupun amatan berpengaruh pada ketiga data tersebut.

Tabel 4.1 Hasil Deteksi *Outlier* dan Amatan Berpengaruh

	% *outlier*	% amatan berpengaruh	
		COOK	DFITS
Data 1	3.333%	0%	3.333%
Data 2	5%	0%	6%
Data 3	7.142%	0%	20%

Berdasarkan Tabel 4.1, diperoleh beberapa tingkat prosentase dari *outlier* dan amatan berpengaruh. Pada amatan berpengaruh dengan uji *Cook's Distance*, ketiga data mempunyai prosentase sebesar 0%, hal tersebut berarti bahwa tidak ada amatan yang berpengaruh terhadap koefisien regresi. Akan tetapi berdasarkan uji DFITS, didapatkan bahwa ada amatan yang berpengaruh terhadap nilai Y(i). Apabila *outlier* dan amatan berpengaruh tersebut diabaikan dalam analisis, serta MKT tetap digunakan untuk menduga parameter regresi, maka hasil pendugaan yang diperoleh menjadi bias dan atau tidak konsisten.

Tindakan dalam mengatasi adanya *outlier* dan amatan berpengaruh dengan cara dibuang begitu saja, akan merugikan peneliti itu sendiri. Hal tersebut dikarenakan informasi yang terkandung akan hilang, padahal belum tentu informasi tersebut tidak penting dalam proses pengambilan keputusan atau kesimpulan. Oleh karena itu, pada skripsi ini metode pendugaan parameter yaitu Metode Theil dan Metode M Tukey *Bisquare* dengan penggunaan

Metode resampling Bootstrap diterapkan pada kasus dengan kondisi data tidak memenuhi asumsi MKT, serta membandingkan metode mana yang lebih baik dalam mengatasi hal tersebut.

4.2 Pendugaan Parameter

Pendugaan parameter koefisien regresi dilakukan pada ketiga jenis data dengan kontaminasi *outlier* yang berbeda. Hal tersebut dilakukan karena ingin mengetahui sampai sejauh mana metode yang digunakan bisa mengatasi *outlier* serta metode mana yang memberikan hasil pendugaan yang lebih baik. Pendugaan koefisien parameter regresi dilakukan dengan menerapkan metode resampling Bootstrap.

Metode Bootstrap diterapkan sebanyak dua kali. Yang pertama digunakan untuk meresampling ulang amatan pada data, dengan jumlah pengambilan untuk mendapatkan satu amatan baru adalah 1000 kali. Hal ini dimaksudkan untuk menghindari asumsi-asumsi yang tidak terpenuhi pada data tersebut. Yang kedua adalah untuk mendapatkan nilai pendugaan koefisien regresi, proses analisis dilakukan sebanyak 1000 kali untuk mendapatkan nilai pendugaan yang benar-benar konsisten dan tidak bias. Hasil pendugaan parameter regresi dapat dilihat pada Tabel 4.2.

Tabel 4.2 Hasil Pendugaan Koefisien Parameter Regresi

Outlier	Metode koefisien regresi			
	Theil		Tukey *Bisquare*	
	b0	b1	b0	b1
3.333%	7.965634	0.556068	8.131265	0.610869
5%	541.1422	22.67746	540.531	22.89636
7.142%	-4.99622	0.227207	-4.80884	0.235311

Dari hasil pendugaan parameter pada Tabel 4.2, dilakukan uji signifikansi untuk masing-masing metode. Uji signifikansi dilakukan untuk mengetahui apakah hasil pendugaan parameter koefisien regresi menunjukkan hubungan yang nyata antara peubah bebas dan respon. Selain itu, juga untuk mengetahui kelayakan parameter dalam menjelaskan model. Uji signifikansi untuk Metode Theil hanya menggunakan uji simultan, sementara untuk Metode Tukey *Bisquare* dilakukan uji parsial dan simultan.

4.2.1 Uji Signifikansi Parameter

- **Metode Theil**

Uji signifikansi untuk Metode Theil, hanya dilakukan uji simultan dengan menggunakan uji Brown-Mood. Hal tersebut disebabkan karena tidak ada uji parsial untuk β_0. Hipotesis yang digunakan adalah

$H_0 : \beta_0 = \beta_1 = 0$

H_1 :minimal ada satu $\beta_p \neq 0$

Dengan statistik uji

$$X^2 = \frac{8}{n}\left[\left(n_1 - \frac{n}{4}\right)^2 + \left(n_2 - \frac{n}{4}\right)^2\right]$$

Dari ketiga data yang digunakan didapatkan hasil pengujian signifikansi sebagai berikut

Tabel 4.3 Uji Signifikansi Metode Theil

% *Outlier*	Statistik uji (X^2)	$\chi^2_{2,0.025}$
3.333%	0.13333	0.0506356
5%	2.08	0.0506356
7.142%	0.571429	0.0506356

Berdasarkan Tabel 4.3, nilai untuk masing-masing statistik uji dari ketiga data yang digunakan adalah lebih besar daripada nilai $\chi^2_{2,0.025}$, sehingga dapat disimpulkan bahwa berdasarkan uji simultan, hasil pendugaan parameter regresi pada ketiga data dengan Metode Theil adalah tolak H_0 atau minimal satu nilai dugaan dari parameter regresi berpengaruh secara nyata terhadap peubah respon.

- **Metode M Tukey *Bisquare***

Uji yang digunakan adalah uji parsial dan uji simultan. Untuk uji parsial hipotesis yang digunakan adalah

$H_0 : \beta_0 = 0$ $\qquad$ $H_0 : \beta_1 = 0$

$H_1 : \beta_0 \neq 0$ $\qquad$ dan $\qquad$ $H_1 : \beta_1 \neq 0$

Dengan statistik uji

Untuk $H_0 : \beta_0 = 0$

$$t_{hit} = \frac{b_0}{s(b_0)}$$

Untuk $H_0 : \beta_1 = 0$

$$t_{hit} = \frac{b_1}{s(b_1)}$$

Sedangkan untuk uji simultan, hipotesis yang digunakan adalah

$H_0 : \beta_0 = \beta_1 = 0$

H_1 : minimal ada satu $\beta_p \neq 0$

Dengan statistik uji

$$F_{hit} = \frac{Kuadrat\ Tengah\ Regresi\ (KTR)}{Kuadrat\ Tengah\ Galat\ (KTG)}$$

Hasil dari uji signifikansi untuk ketiga data, baik uji parsial maupun simultan dapat dilihat di Tabel 4.4

Tabel 4.4 Uji Signifikansi Metode M Tukey *Bisquare*

%*Outlier*		Uji Parsial		Uji Simultan	
		Statistik uji (t)	$t_{n-p}^{0.025}$	Statistik Uji (F)	$F_{p,n-p}^{0.05}$
3%	b0	3.96	2.04841	349.16	3.34039
	b1	18.69			
5%	b0	2.01	1.98447	109.56	3.08920
	b1	10.47			
7%	b0	-2.13	2.05553	26.7	3.36902
	b1	5.17			

Berdasarkan Tabel 4.4, nilai untuk masing-masing statistik uji pada uji parsial dan uji simultan dari ketiga data yang digunakan adalah lebih besar daripada nilai $t_{n-p}^{0.025}$ dan $F_{p,n-p}^{0.05}$, sehingga dapat disimpulkan bahwa berdasarkan uji parsial adalah tolak H_0 atau nilai dugaan dari parameter regresi berpengaruh secara nyata terhadap peubah respon dan untuk uji simultan adalah tolak H_0 atau minimal satu nilai dugaan dari parameter regresi berpengaruh secara nyata terhadap peubah respon.

Dari hasil uji signifikansi pada kedua metode yang digunakan yaitu Metode Theil dan M Tukey *Bisquare*, memberikan keputusan yang sama yaitu nilai dugaan dari parameter regresi berpengaruh secara nyata terhadap peubah respon. Berdasarkan hal tersebut, dapat dikatakan bahwa pada data yang mengandung *outlier* sebesar 3%, 5% dan 7%, Metode Theil dan M Tukey *Bisquare* masih mampu mengatasi *outlier* dengan memberikan hasil pendugaan parameter regresi yang signifikan. Sehingga bisa dijadikan sebagai salah satu alternatif pilihan dalam melakukan analisis terhadap data dengan prosentase *outlier* tertentu.

4.3 Perbandingan Nilai KTG Berdasarkan Tingkat Kontaminasi *Outlier*

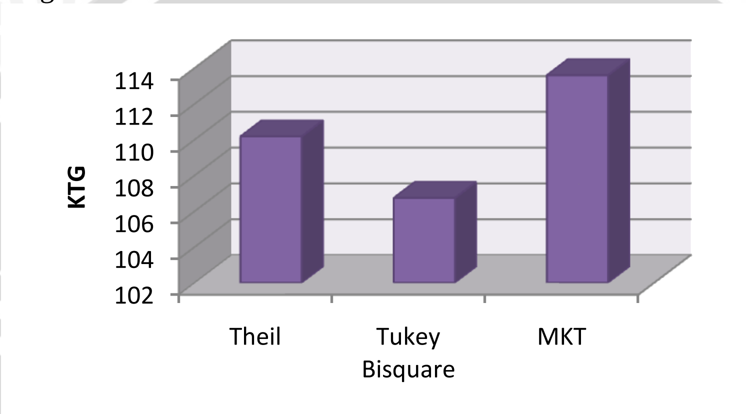
Berdasarkan Tabel 4.3 dan 4.4 yang menunjukkan bahwa hasil pendugaan koefisien parameter regresi berpengaruh secara nyata terhadap peubah respon, belum bisa menunjukkan metode mana yang lebih baik diantara keduanya. Oleh karena itu, perlu dibandingkan nilai KTG dari kedua metode tersebut untuk mengetahui metode mana yang terbaik. Metode pendugaan parameter yang terbaik ditentukan dari nilai KTG yang paling kecil. Hal tersebut didasarkan pada persamaan 2.82, di mana KTG merupakan selisih antara nilai dugaan dengan nilai sebenarnya (galat), semakin kecil galat maka akan semakin kecil pula nilai KTG. Untuk membuktikan bahwa nilai KTG yang dihasilkan oleh kedua metode yang digunakan adalah lebih baik daripada metode pendugaan yang biasa digunakan yaitu MKT, maka pada Tabel 4.5 disajikan pula nilai KTG dari MKT.

Tabel 4.5 Nilai KTG Masing – Masing Metode

% *Outlier*	Theil	Tukey *Bisquare*	MKT
3%	110.1781	106.7385	113.6
5%	11843.1	11823.32	12048
7%	16.24861	15.47845	16.54

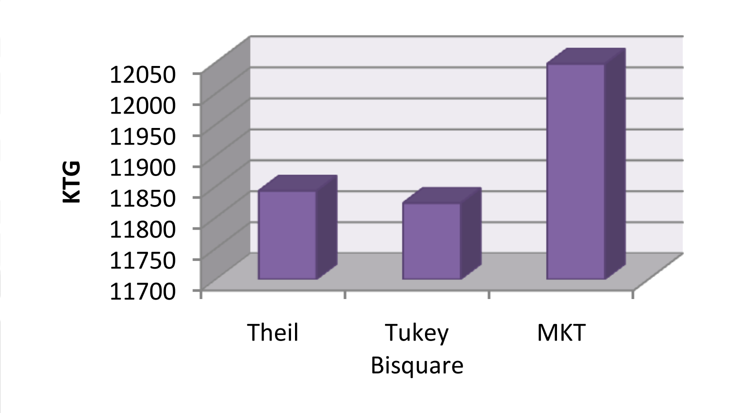
Untuk memudahkan dalam membandingkan nilai KTG, Tabel 4.5 dapat disajikan dalam bentuk grafik yang diklasifikasikan dalam tiga kelompok berdasarkan prosentase *outlier* pada data. Berdasarkan Gambar 4.1, dapat dilihat bahwa pada semua prosentase *outlier*, MKT merupakan metode yang lebih dulu menghasilkan nilai KTG yang paling besar atau lebih dulu mengalami *breakdown*. Hal ini disebabkan MKT membutuhkan asumsi-asumsi yang harus dipenuhi, sehingga hasil pendugaan parameter menjadi bias. Hal tersebut mengindikasikan bahwa MKT merupakan metode yang tidak lebih baik dari Theil dan Tukey *Bisquare*,serta penggunaannya tidak disarankan.

Metode Theil adalah metode selanjutnya yang mengalami *breakdown* setelah MKT, dan yang terakhir adalah Metode Tukey *Bisquare*. Sehingga di antara metode Theil dan Tukey *Bisquare*, Metode Tukey *Bisquare* memberikan hasil pendugaan yang lebih baik pada prosentase kontaminasi *outlier* 3% bila dibandingkan dengan Metode Theil.

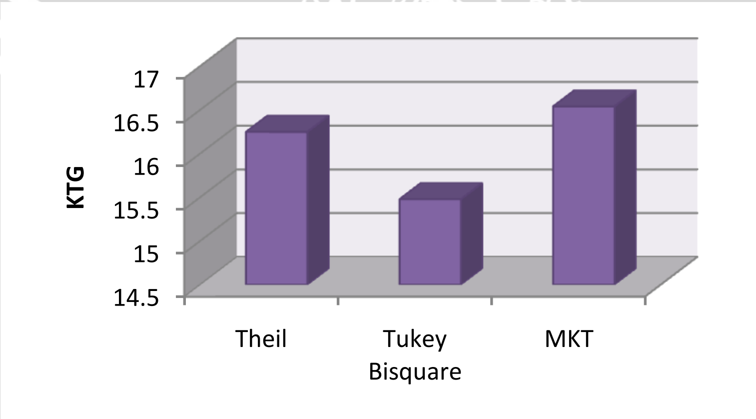


Gambar 4.1 Nilai KTG untuk Kontaminasi Outlier 3%

Untuk nilai KTG pada data dengan prosentase *outlier* 5% dan 7%, dapat dilihat pada Gambar 4.2 dan 4.3. Secara umum, hasil yang dapat disimpulkan adalah sama, yaitu Metode Tukey *Bisquare* menghasilkan nilai KTG yang lebih kecil daripada Metode Theil. Sehingga secara keseluruhan dapat disimpulkan bahwa Metode Tukey *Bisquare* merupakan metode *robust* yang memberikan hasil pendugaan parameter yang lebih baik daripada Metode Theil pada tingkat kontaminasi *outlier* 3%, 5% dan 7%. Hal ini disebabkan karena Metode Tukey *Bisquare* menggunakan pembobot dalam menentukan koefisien parameter regresi. Fungsi dari pembobot adalah untuk menghasilkan galat yang homogen, sehingga nilai pembobot setiap amatan berbeda disesuaikan dengan seberapa jauh nilai amatan terhadap rata-rata.



Gambar 4.2 Nilai KTG untuk Kontaminasi Outlier 5%



Gambar 4.3 Nilai KTG untuk Kontaminasi Outlier 7%

4.4 Perbandingan Bias dan KTG Berdasarkan Tingkat Kontaminasi *Outlier* serta Banyak Amatan Pada Data Simulasi

Pada pembahasan sebelumnya, didapatkan bahwa Metode Tukey *Bisquare* merupakan metode yang mampu memberikan hasil pendugaan parameter yang lebih baik daripada Metode Theil pada kontaminasi *outlier* 3%, 5% dan 7%. Untuk mengetahui bagaimana pola dari ketahanan atau *robust* kedua metode tersebut dalam mengatasi *outlier* apabila banyak amatan yang dianalisis berbeda-beda, maka perlu dibandingkan nilai KTG dan bias dari kedua metode untuk data dengan kombinasi dari kontaminasi *outlier* 3%, 5% dan 7% dengan amatan sebanyak 15, 30 dan 45. Data yang digunakan dalam pembandingan ini adalah data hasil simulasi dengan menerapkan Metode Bootstrap pada proses mendapatkan nilai dugaan parameter regresi. Banyak perulangan yang digunakan adalah 1000.

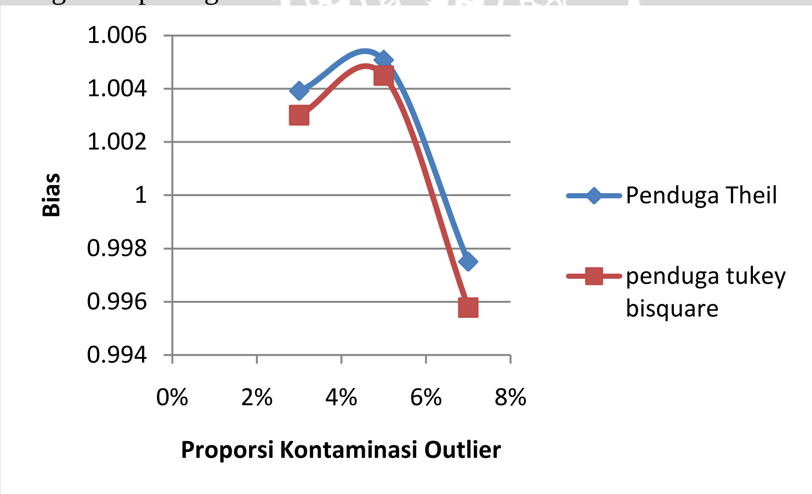
Pada Tabel 4.6, dapat dilihat nilai KTG dan bias dari Metode Theil dan Tukey *Bisquare* dan selanjutnya didapatkan pola ketahanan masing-masing metode. Pada Tabel 4.6, Metode M Tukey *Bisquare* menghasilkan nilai KTG dan bias yang lebih kecil pada semua kombinasi data dengan tingkat kontaminasi *outlier* 3%, 5% dan 7% dengan amatan sebanyak 15, 30 dan 45. Sehingga dapat disimpulkan bahwa Metode M Tukey *Bisquare* merupakan metode yang lebih baik dan mempunyai *breakdown point* yang lebih tinggi daripada Metode Theil.

Tabel 4.6 Nilai KTG dan Bias Masing – Masing Metode

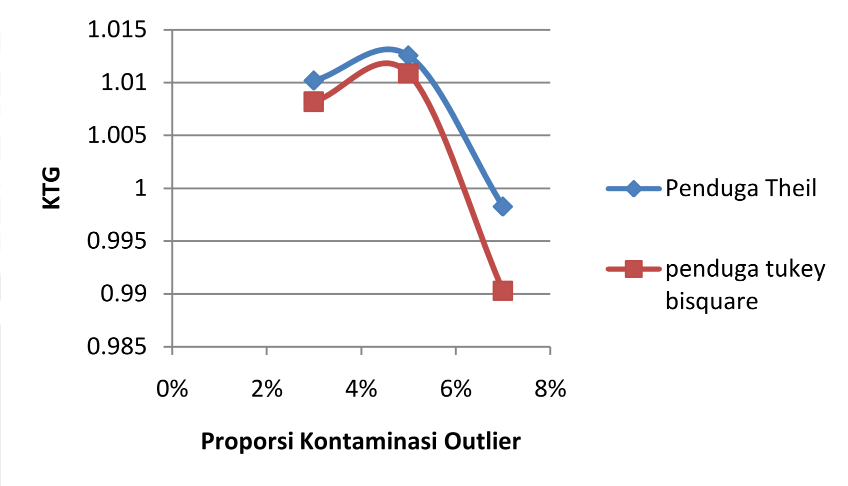
N	Tipe Distribusi Galat	Bias		KTG	
		Metode Theil	Metode M Tukey *Bisquare*	Metode Theil	Metode M Tukey *Bisquare*
15	3%	1.003918	1.003005	1.010175	1.008198
	5%	1.005081	1.004485	1.012546	1.010864
	7%	0.997506	0.995788	0.99827	0.990312
30	3%	0.992724	0.99211	0.988176	0.997453
	5%	0.996464	0.992031	0.989505	0.988133
	7%	0.990271	0.989827	0.984366	0.98566
45	3%	1.004402	1.002481	1.010739	1.006478
	5%	0.998352	0.99787	0.998435	0.997818
	7%	1.002155	1.00067	1.006778	1.003255

Untuk mengidentifikasi pola kebaikan dari kedua metode tersebut, dilihat dari bias dan KTG masing-masing metode. Dimana sifat dari kebaikan metode adalah berbanding terbalik dengan nilai bias dan KTG. Semakin kecil nilai bias dan KTG maka semakin baik metode tersebut, begitu pula sebaliknya semaikn besar nilai bias dan KTG maka semakin menurun kebaikan dari metode tersebut. Untuk memudahkan dalam mengidentifikasi pola dari kebaikan kedua metode tersebut, hasil analisis pada Tabel 4.6 dapat disajikan dalam bentuk grafik yang diklasifikasikan dalam tiga kelompok berdasarkan banyak amatan.

Grafik nilai bias dan KTG dari data dengan banyak amatan 15, dapat dilihat pada Gambar 4.4 dan 4.5. Bila dilihat dari pola grafik untuk masing-masing metode, pada saat kontaminasi *outlier* 5%, nilai bias dan KTG yang dihasilkan lebih tinggi daripada 3%. Hal ini mengindikasikan bahwa pada saat kontaminasi *outlier* meningkat sebesar 2%, maka kebaikan dari masing-masing metode mengalami penurunan. Akan tetapi, pada saat kontaminasi kembali dinaikkan sebesar 2%, kebaikan dari kedua metode tersebut mengalami peningkatan.

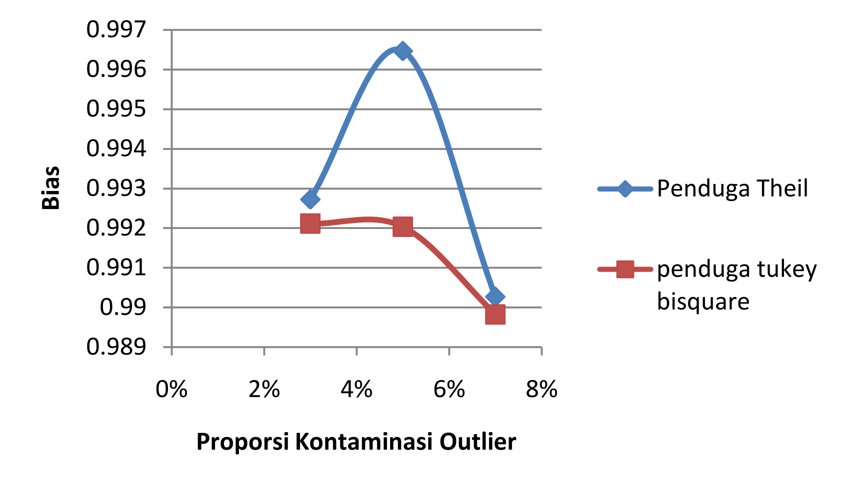


Gambar 4.4 Nilai Bias pada n = 15

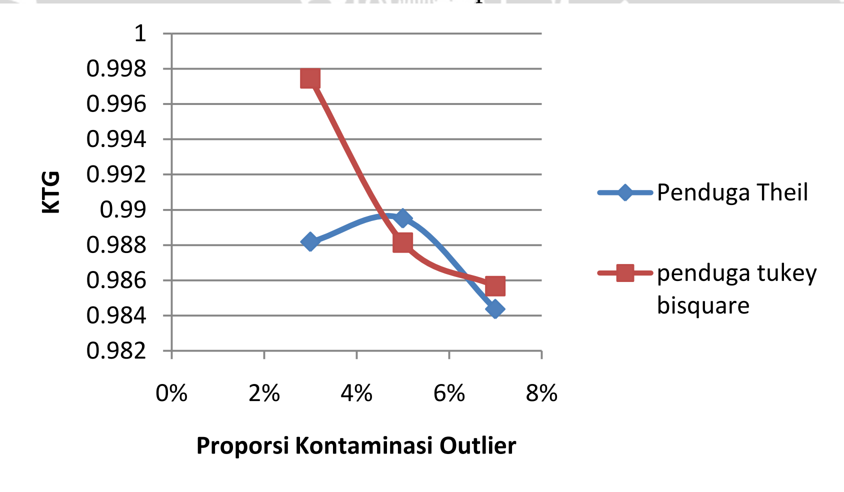


Gambar 4.5 Nilai KTG pada n = 15

Grafik nilai bias dan KTG dari data dengan amatan sebanyak 30, dapat dilihat pada Gambar 4.6 dan 4.7. Bila dilihat dari pola grafik untuk Metode Theil, pada saat kontaminasi *outlier* 5%, nilai bias dan KTG yang dihasilkan lebih tinggi daripada *outlier* 3%. Hal ini mengindikasikan bahwa pada saat kontaminasi meningkat sebesar 2%, maka kebaikan Metode Theil mengalami penurunan. Pada saat kontaminasi kembali dinaikkan sebesar 2%, kebaikan Metode Theil mengalami peningkatan. Sementara itu untuk Metode Tukey *Bisquare*, dapat dilihat bahwa semakin besar kontaminasi *outlier* yang diberikan, maka kebaikan dari Metode Tukey *Bisquare* semakin meningkat. Hal tersebut dapat dilihat dari nilai bias dan KTG yang dihasilkan, semakin besar kontaminasi *outlier* maka semakin kecil nilai bias dan KTG yang dihasilkan.

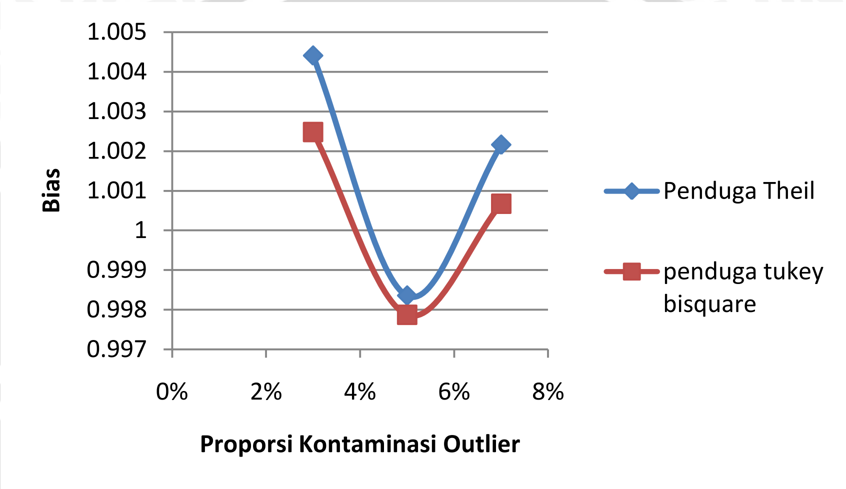


Gambar 4.6 Nilai Bias pada n = 30

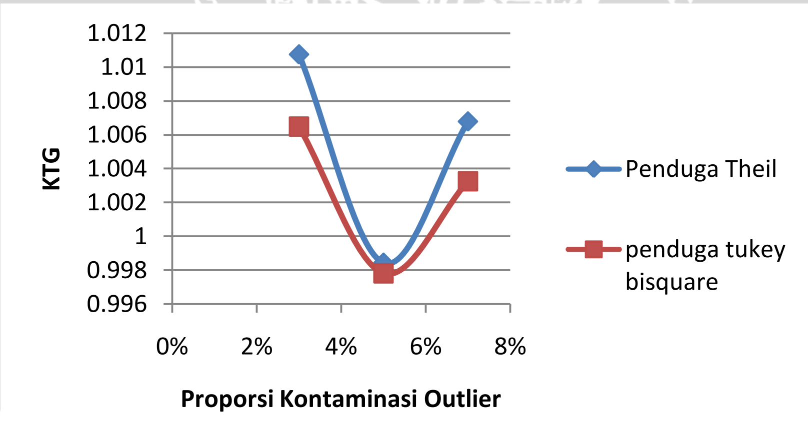


Gambar 4.7 Nilai KTG pada n = 30

Grafik nilai bias dan KTG dari data dengan amatan sebanyak 45, dapat dilihat pada Gambar 4.8 dan 4.9. Bila dilihat dari pola grafik untuk Metode Theil dan Tukey *Bisquare*, pada saat kontaminasi *outlier* 5%, nilai bias dan KTG yang dihasilkan lebih kecil daripada *outlier* 3%. Hal ini mengindikasikan bahwa pada saat kontaminasi meningkat sebesar 2%, maka kebaikan dari kedua metode tersebut mengalami peningkatan dan pada saat kontaminasi kembali dinaikkan sebesar 2%, kebaikan dari metode tersebut mengalami penurunan.



Gambar 4.8 Nilai Bias pada n = 45



Gambar 4.9 Nilai KTG pada n = 45

Berdasarkan hasil analisis pada data sekunder dan simulasi, hasil yang didapatkan adalah sama, yaitu Metode Tukey *Bisquare* merupakan metode yang paling baik dalam pendugaan parameter pada data dengan kontaminasi *outlier*. Metode Tukey *Bisquare* mampu memberikan hasil penduga yang *robust* pada tingkat kontaminasi *outlier* 3%, 5% dan 7% serta pada jumlah amatan yang berbeda yaitu 15, 30 dan 45. Metode Tukey *Bisquare* dapat dijadikan salah satu alternatif pilihan untuk menangani data dengan kontaminasi *outlier* tersebut.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian dapat diperoleh kesimpulan sebagai berikut:

1. Berdasarkan hasil pendugaan parameter pada data dengan kontaminasi *outlier* 3%, 5% dan 7%, terlihat bahwa Metode Tukey *Bisquare* memberikan hasil pendugaan yang lebih baik daripada Metode Theil. Hal tersebut dapat dilihat dari nilai KTG dan bias yang dihasilkan oleh metode Tukey *Bisquare* lebih kecil daripada metode Theil.
2. Pola kebaikan dari Metode Theil dan Tukey *Bisquare* dipengaruhi oleh kombinasi antara banyak amatan dengan proporsi kontaminasi *outlier*

5.2 Saran

1. Untuk aplikasi pada penelitian berbagai bidang, sebelum menggunakan MKT data harus dipastikan memenuhi asumsi regresi. Namun, bila terdapat salah satu asumsi yang tidak terpenuhi, metode penduga M Tukey *Bisquare* dapat diterapkan.
2. Pada penelitian selanjutnya, sebaiknya membandingkan metode penduga M Tukey *Bisquare* dengan metode regresi *robust* yang lain seperti penduga M Hampel, penduga MM, penduga S dan lainnya.
3. Menambahkan jumlah peubah bebas dan tingkat kontaminasi *outlier* sampai maksimal 50% serta banyak amatan untuk lebih memastikan pola kebaikan dari metode yang digunakan.
4. Menggunakan metode resampling yang lain selain Metode resampling Bootstrap, misalnya Metode Fisher.

DAFTAR PUSTAKA

Aida, R.N. 2009. *Pengaruh Jumlah Pasien terhadap Produksi Limbah Medis Padat di Ruang Rawat Inap dan Unit Gawat Darurat*. Tidak dipublikasikan.

Ainuddin. 1989. *Analisis Data*. IPB. Bogor

Ali, A. and M. F. Qadir. 2005. *A Modified M-estimator for The Detection of Outliers*. PJSOR, Vol. 1, pp : 49-64.

Aliudin, F.R. 2006. *Korelasi antara Bobot Badan Awal dengan Bobot Badan Akhir dan Bobot Badan Karkad Ayam Pedaging dengan Penambahan Choline Chloride dalam Pakan*. Tidak dipublikasikan.

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H. and J.W., Tukey. 1972. *Robust Estimates of Location : Survey and Advance*. Princeston University Press. Princeston.

Bekti, R.D. 2009. *Model Hubungan Anomali Luas Panen Padi dan Curah Hujan Terboboti (Weighted Rainfall Index) dengan Regresi Robust*. Tidak dipublikasikan.

Chen, C. 2002. *Robust Regression and Outlier Detection with The Robustreg Procedure.* http://www2.sas.com/proceding/pdf. Diakses tanggal 2 November 2011.

Daniel, W.W. 1989. *Statistika Nonparametrik Terapan*. Gramedia Pustaka Utama, Jakarta. Terjemahan oleh : A. T. Kantjono W.

Dewi, S.P. 2011. *Perbandingan Penduga Theil dan Penduga M Huber untuk Slope Regresi pada Regresi Linier Sederhana.* Tidak dipublikasikan.

Drapper, N. and H. Smith. 1992. *Analisis Regresi Terapan Edisi Kedua*. Gramedia Pustaka Utama, Jakarta. Terjemahan oleh : Ir. Bambang Sumantri.

Efron, B. and R.J. Tibshirani. 1993. *An Introduction to The Bootstrap*. Chapman dan Hall. New York.

Glaister, P. 2005. *Robust Linear Regression Using Theil's Method*. Journal of Chemical Education, Vol. 82, pp : 1472-1473.

Ghozali, I. 2006. *Aplikasi Analisis Multivariat dengan Program SPSS*. Badan Penerbit Universitas Diponegoro. Semarang.

Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw and A.S., Werner. 1986. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley and Sons, New York.

Hettmansperger, T.P., S.J. Sheather. 1992. *A Cautionary Note on the Method of Least Median Squares*. American Statiscial Association.

Hidayat, S. 2009. *Pengaruh Kadar Immunoglobulin Gamma (IgG) Kolostrum Induk terhadap IgG Serum Darah Anak pada Sapi Friesian Holstein.* Tidak dipublikasikan.

Hoaglin, D.C., F., Mosteller, and J.W. Tukey. 1983. *Understanding Robust and Explanatory Data Analysis*. John Wiley and Sons, Inc, New York.

Kutner, M.H., C.J. Nachtsheim dan J. Neter. 2004. *Applied Linier Regression Models*. Fourth Ed. McGraw Hill. New York.

Kuzmic, P. 2004. *Practical Robust Fit of Enzyme Inhibition Data*. Journal of Methods in Enzymology, pp : 383;366-381.

Mendenhall, W., R. L. Scheaffer and D. D. Wackerly. 1981. *Mathematical Statistic with Application Fourth Edition*. Wasworth. California.

Montgomery, D.C and E.A Peck. 1992. *Introduction to Linear Regression Analysis (2nd edition)*. John Wiley and Sons, Inc, New York.

Myers, R. H. 1990. *Classical and Modern Regression with Aplications*. Duxbury Press, California.

Rey, W.J.J. 1983. *Introduction to Robust and Quasi-Robust Statistical Methods*. Spinger-Verlag, Berlin.

Rosseeuw, P.J and A.M Leroy. 1987. *Robust Regression and Outlier Detection*. John Wiley and Sons, New York.

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

Soemartini. 2007. *Pencilan (Outlier)*.Universitas Padjajaran. Jatinangor.

Ryan, T.P. 1997. *Modern Regression Method Volume 1*. John Wiley and Sons, New York.

Walpole, R. E. 1992. *Pengantar Statistika Edisi Ketiga*.. Gramedia Pustaka Utama, Jakarta. Terjemahan oleh : Ir. Bambang Sumantri.

Lampiran 1. Data dengan kontaminasi *outlier* 3%

KADAR IgG KOLOS-TRUM	KADAR IgG SERUM	TRES	COOK	DFIT	UJI TRES 2.05	UJI COOK 3.34	UJI DFIT 0.52
70.61	69.64	2.17	0.11	0.53	YES	NOT	YES
76.15	47.42	-0.36	0.01	-0.11	NOT	NOT	NOT
69.31	34.65	-1.28	0.04	-0.28	NOT	NOT	NOT
71.32	63.88	1.48	0.06	0.36	NOT	NOT	NOT
72.73	56.60	0.68	0.02	0.18	NOT	NOT	NOT
69.91	38.78	-0.90	0.02	-0.20	NOT	NOT	NOT
61.76	43.73	-0.05	0.00	-0.01	NOT	NOT	NOT
56.22	54.05	1.20	0.04	0.27	NOT	NOT	NOT
57.14	27.65	-1.41	0.04	-0.30	NOT	NOT	NOT
49.65	45.90	0.72	0.03	0.23	NOT	NOT	NOT
82.45	51.76	-0.24	0.01	-0.10	NOT	NOT	NOT
70.07	32.75	-1.52	0.06	-0.35	NOT	NOT	NOT
55.08	36.56	-0.43	0.01	-0.10	NOT	NOT	NOT
71.16	56.49	0.74	0.02	0.18	NOT	NOT	NOT
58.12	30.36	-1.18	0.03	-0.24	NOT	NOT	NOT
66.38	39.65	-0.65	0.01	-0.13	NOT	NOT	NOT
44.76	31.67	-0.45	0.02	-0.18	NOT	NOT	NOT
46.33	46.06	0.91	0.06	0.34	NOT	NOT	NOT
39.16	36.56	0.32	0.01	0.17	NOT	NOT	NOT
58.88	32.70	-0.98	0.02	-0.20	NOT	NOT	NOT
80.72	47.53	-0.58	0.03	-0.23	NOT	NOT	NOT
71.05	45.14	-0.34	0.00	-0.08	NOT	NOT	NOT
68.93	63.17	1.52	0.05	0.33	NOT	NOT	NOT
66.76	61.76	1.47	0.04	0.29	NOT	NOT	NOT
60.51	57.90	1.38	0.03	0.26	NOT	NOT	NOT
63.82	44.32	-0.09	0.00	-0.02	NOT	NOT	NOT

Lampiran 1. Data dengan kontaminasi *outlier* 3% (Lanjutan)

55.85	48.78	0.70	0.01	0.16	NOT	NOT	NOT
64.39	37.81	-0.73	0.01	-0.14	NOT	NOT	NOT
57.45	32.21	-0.96	0.02	-0.20	NOT	NOT	NOT
54.15	30.91	-0.94	0.03	-0.23	NOT	NOT	NOT
Prosentase					3.33		3.33

Lampiran 2. Data dengan kontaminasi *outlier* 5%

Bobot Akhir	Bobot Awal	TRES	UJI TRES	COOK	DFIT	Uji COOK	Uji DFIT
1645	43.9		1.98			3.09	0.28
1870	53.3	0.96	NOT	0.01	0.14	NOT	NOT
2030	54.3	1.04	NOT	0.02	0.18	NOT	NOT
1650	46.2	2.37	YES	0.10	0.45	NOT	YES
1725	50.8	0.51	NOT	0.00	0.06	NOT	NOT
1840	48.7	0.23	NOT	0.00	0.03	NOT	NOT
1885	51.8	1.74	NOT	0.02	0.18	NOT	NOT
1530	43.6	1.50	NOT	0.02	0.21	NOT	NOT
1570	45.7	-0.04	NOT	0.00	-0.01	NOT	NOT
1610	48.6	-0.11	NOT	0.00	-0.01	NOT	NOT
1600	45.7	-0.36	NOT	0.00	-0.04	NOT	NOT
1922	51.8	0.16	NOT	0.00	0.02	NOT	NOT
1598	40.9	1.85	NOT	0.03	0.26	NOT	NOT
1710	46.4	1.18	NOT	0.03	0.24	NOT	NOT
1720	48.9	1.02	NOT	0.01	0.11	NOT	NOT
1676	51.4	0.58	NOT	0.00	0.06	NOT	NOT
1649	50.9	-0.35	NOT	0.00	-0.05	NOT	NOT
1562	50.8	-0.49	NOT	0.00	-0.06	NOT	NOT
1460	41.9	-1.27	NOT	0.01	-0.16	NOT	NOT
1472	50.7	-0.32	NOT	0.00	-0.06	NOT	NOT
1608	45.8	-2.11	NOT	0.03	-0.26	NOT	NOT
1700	54.8	0.21	NOT	0.00	0.02	NOT	NOT
1583.5	44.7	-0.86	NOT	0.02	-0.17	NOT	NOT
1400	42.6	0.22	NOT	0.00	0.03	NOT	NOT
1625	52.9	-1.03	NOT	0.02	-0.17	NOT	NOT
1914	51	-1.14	NOT	0.02	-0.18	NOT	NOT
1904	47.2	1.95	NOT	0.03	0.25	NOT	NOT
1728	47	2.71	YES	0.04	0.28	NOT	NOT

Lampiran 2. Data dengan kontaminasi *outlier* 5% (Lanjutan)

1498	44.7	1.06	NOT	0.01	0.11	NOT	NOT
1596	44.7	-0.56	NOT	0.00	-0.07	NOT	NOT
1460	42.2	0.33	NOT	0.00	0.04	NOT	NOT
1574	48.4	-0.39	NOT	0.00	-0.07	NOT	NOT
1510	45	-0.65	NOT	0.00	-0.07	NOT	NOT
1810	52.6	-0.52	NOT	0.00	-0.06	NOT	NOT
1516	47.3	0.63	NOT	0.00	0.10	NOT	NOT
1660	49.7	-0.95	NOT	0.00	-0.10	NOT	NOT
1600	46.9	-0.13	NOT	0.00	-0.01	NOT	NOT
1660	49.4	-0.09	NOT	0.00	-0.01	NOT	NOT
2060	51.5	2.00	YES	0.00	-0.01	NOT	NOT
1520	44.9	3.32	YES	0.09	0.45	NOT	YES
1540	42.5	-0.40	NOT	0.00	-0.05	NOT	NOT
1625	44.7	0.29	NOT	0.00	0.05	NOT	NOT
1814	52.9	0.60	NOT	0.00	0.08	NOT	NOT
1688	49.9	0.60	NOT	0.00	0.10	NOT	NOT
1670	48.5	0.08	NOT	0.00	0.01	NOT	NOT
1680	49.1	0.21	NOT	0.00	0.02	NOT	NOT
1890	52.4	0.17	NOT	0.00	0.02	NOT	NOT
1642.5	48.3	1.42	NOT	0.02	0.21	NOT	NOT
1460	39	0.00	NOT	0.00	0.00	NOT	NOT
1540	47.8	0.30	NOT	0.00	0.08	NOT	NOT
1910	48.2	-0.83	NOT	0.00	-0.08	NOT	NOT
1650	44.2	2.54	YES	0.03	0.26	NOT	NOT
1435	43.7	0.94	NOT	0.01	0.13	NOT	NOT
1681.25	45.5	-0.93	NOT	0.01	-0.14	NOT	NOT
1730	45.9	0.95	NOT	0.01	0.11	NOT	NOT
1570	42.8	1.31	NOT	0.01	0.15	NOT	NOT
1695	48.8	0.50	NOT	0.00	0.08	NOT	NOT

Lampiran 2. Data dengan kontaminasi *outlier* 5% (Lanjutan)

1715	51.2	0.37	NOT	0.00	0.04	NOT	NOT
1638.75	48.5	0.05	NOT	0.00	0.01	NOT	NOT
1575	44.8	-0.08	NOT	0.00	-0.01	NOT	NOT
1702	48.3	0.12	NOT	0.00	0.02	NOT	NOT
1592	48	0.54	NOT	0.00	0.05	NOT	NOT
1568	46.2	-0.40	NOT	0.00	-0.04	NOT	NOT
1500	40.8	-0.24	NOT	0.00	-0.03	NOT	NOT
1478	40	0.28	NOT	0.00	0.06	NOT	NOT
1530	44.4	0.25	NOT	0.00	0.06	NOT	NOT
1550	48.5	-0.21	NOT	0.00	-0.03	NOT	NOT
1554	51.4	-0.89	NOT	0.00	-0.09	NOT	NOT
1630	52.6	-1.48	NOT	0.02	-0.20	NOT	NOT
1506	43.9	-1.03	NOT	0.01	-0.16	NOT	NOT
1590	49.1	-0.32	NOT	0.00	-0.05	NOT	NOT
1580	47.9	-0.65	NOT	0.00	-0.07	NOT	NOT
1615	53.1	-0.49	NOT	0.00	-0.05	NOT	NOT
1460	40	-1.28	NOT	0.02	-0.21	NOT	NOT
1700	58.1	0.08	NOT	0.00	0.02	NOT	NOT
1542	52.9	-1.61	NOT	0.10	-0.46	NOT	YES
1533	49.2	-1.93	NOT	0.05	-0.31	NOT	YES
1308	41.5	-1.20	NOT	0.01	-0.13	NOT	NOT
1484	49	-1.66	NOT	0.05	-0.32	NOT	YES
1798	54	-1.61	NOT	0.01	-0.17	NOT	NOT
1464	49	0.22	NOT	0.00	0.04	NOT	NOT
1705.5	51.2	-1.80	NOT	0.02	-0.19	NOT	NOT
1792	51.4	-0.03	NOT	0.00	0.00	NOT	NOT
1640	49	0.72	NOT	0.00	0.10	NOT	NOT
1926	54.3	-0.17	NOT	0.00	-0.02	NOT	NOT
1770	58.4	1.35	NOT	0.03	0.26	NOT	NOT

Lampiran 2. Data dengan kontaminasi *outlier* 5% (Lanjutan)

1588	47.5	-1.00	NOT	0.04	-0.29	NOT	YES
1650	47.5	-0.33	NOT	0.00	-0.03	NOT	NOT
1486	43.6	0.24	NOT	0.00	0.02	NOT	NOT
1738	52.7	-0.44	NOT	0.00	-0.07	NOT	NOT
1470	45.2	-0.05	NOT	0.00	-0.01	NOT	NOT
1598	47.6	-0.93	NOT	0.01	-0.11	NOT	NOT
1422	44.5	-0.26	NOT	0.00	-0.03	NOT	NOT
1390	43.9	-1.22	NOT	0.01	-0.16	NOT	NOT
1470	44.9	-1.40	NOT	0.02	-0.20	NOT	NOT
1720	51.6	-0.86	NOT	0.01	-0.11	NOT	NOT
1615	48	0.01	NOT	0.00	0.00	NOT	NOT
1666.25	51.1	-0.19	NOT	0.00	-0.02	NOT	NOT
1600	44.6	-0.37	NOT	0.00	-0.05	NOT	NOT
1730	52.8	0.39	NOT	0.00	0.05	NOT	NOT
		-0.15	NOT	0.00	-0.02	NOT	NOT
Prosentase			5				6

Lampiran 3. Data dengan kontaminasi *outlier* 7%

					TRES UJI	COOK UJI	DFIT UJI
Jumlah Pasien	Jumlah Limbah	TRES	COOK	DFIT	2.06	3.37	0.53
64	7.5	-0.73	0.01	-0.14	NOT	NOT	NOT
58	10	0.29	0.00	0.07	NOT	NOT	NOT
56	6.25	-0.51	0.01	-0.13	NOT	NOT	NOT
69	8.75	-0.77	0.02	-0.19	NOT	NOT	NOT
66	12.5	0.37	0.00	0.08	NOT	NOT	NOT
55	5	-0.76	0.02	-0.21	NOT	NOT	NOT
70	23.75	3.56	0.29	0.91	YES	NOT	YES
59	10	0.22	0.00	0.05	NOT	NOT	NOT
74	20	1.85	0.17	0.61	NOT	NOT	YES
71	10	-0.59	0.01	-0.16	NOT	NOT	NOT
65	8.75	-0.49	0.00	-0.10	NOT	NOT	NOT
63	6.25	-0.99	0.02	-0.19	NOT	NOT	NOT
58	10	0.29	0.00	0.07	NOT	NOT	NOT
54	7.5	-0.06	0.00	-0.02	NOT	NOT	NOT
56	7.5	-0.20	0.00	-0.05	NOT	NOT	NOT
59	8.75	-0.09	0.00	-0.02	NOT	NOT	NOT
70	10	-0.52	0.01	-0.13	NOT	NOT	NOT
63	21.25	3.23	0.14	0.62	YES	NOT	YES
62	10	0.02	0.00	0.00	NOT	NOT	NOT
63	10	-0.05	0.00	-0.01	NOT	NOT	NOT
70	12.5	0.10	0.00	0.03	NOT	NOT	NOT
58	6.25	-0.64	0.01	-0.15	NOT	NOT	NOT
70	7.5	-1.17	0.04	-0.30	NOT	NOT	NOT
45	5	-0.08	0.00	-0.04	NOT	NOT	NOT
51	6.25	-0.17	0.00	-0.06	NOT	NOT	NOT
87	12.5	-1.29	0.41	-0.92	NOT	NOT	YES

Lampiran 3. Data dengan kontaminasi *outlier* 7% (Lanjutan)

57	11.25	0.67	0.01	0.16	NOT	NOT	NOT
70	10	-0.52	0.01	-0.13	NOT	NOT	NOT
Prosentase					7.14		20.00

Lampiran 4. Macro Analisis Data dengan Metode Penduga M Tukey *Bisquare*

```
macro
boot1 x y id b0 b1 par nilai
mcolumn x y x1 y1 cn cn1 coef b0 b1 id b0s b1s par nilai z
mconstant B n i j pos b0rata b1rata sdb0 sdb1 a bb1 ba1 bb2 ba2 bb3
          ba3 bb3 ba3 b0h b1h text1 text2 text3 text4 text5 text6
          cb1 cb0
let b=2
let n=count(x)
do i=1:n
let cn(i)=i
enddo
do i=1:b
let id(i)=i
sample n cn cn1;
replace.
do j=1:n
let pos=cn1(j)
let x1(j)=x(pos)
let y1(j)=y(pos)
enddo
call robust x1 y1 cb0 cb1
let b0(i)=cb0
let b1(i)=cb1
enddo
let b0rata=average(b0)
let sdb0=stdev(b0)
let b1rata=average(b1)
let sdb1=stdev(b1)
let a=0.05
sort b0 b0s;
BY b0.
let pos=a/2*b
let bb3=b0s(pos)
let pos=(1-a/2)*b
let ba3=b0s(pos)
do i=1:b
```

```
let b0h=b3s(i)
let z(i)=(b0h=b0rata)/sdb0
enddo
let pos=(1-a/2)*B
let bb4=b0rata-z(pos)*sdb0
let pos=a/2*B
let ba4=b0rata-z(pos)*sdb0
let a=0.05
sort b1 b1s;
BY b1.
let pos=a/2*b
let bb1=b1s(pos)
let pos=(1-a/2)*b
let ba1=b1s(pos)
do i=1:b
let b1h=b1s(i)
let z(i)=(b1h=b1rata)/sdb1
enddo
let pos=(1-a/2)*B
let bb2=b1rata-z(pos)*sdb1
let pos=a/2*B
let ba2=b1rata-z(pos)*sdb1
let text1="b1(.)"
let text2="se(b1)"
let text3="batas bawah bootstrap"
let text4="batas atas bootstrap"
let text5="batas bawah bootstrap-t"
let text6="batas atas bootstrap-t"
let text7="b0(.)"
let text8="se(b0)"
let text9="batas bawah bootstrap"
let text10="batas atas bootstrap"
let text11="batas bawah bootstrap-t"
let text12="batas atas bootstrap-t"
let par(1)=text1
let par(2)=text2
```

Lampiran 4. Macro Analisis Data dengan Metode Penduga M Tukey *Bisquare* (Lanjutan)

```
let par(3)=text3
let par(4)=text4
let par(5)=text5
let par(6)=text6
let par(7)=text7
let par(8)=text8
let par(9)=text9
let par(10)=text10
let par(11)=text11
let par(12)=text12
let nilai(1)=b1rata
let nilai(2)=sdb1
let nilai(3)=bb1
let nilai(4)=ba1
let nilai(5)=bb2
let nilai(6)=ba2
let nilai(7)=b0rata
let nilai(8)=sdb0
let nilai(9)=bb3
let nilai(10)=ba3
let nilai(11)=bb4
let nilai(12)=ba4
print par nilai
endmacro
#
#
macro
robust x1 y1 cb0 cb1
mcolumn x1 y1 e e1 ae b w
mconstant med mad s se1 se2 se i n j cb0 cb1
let n=count(y1)
regress y1 1 x1;
residuals e;
constant;
brief 2.
let se=1
```

Lampiran 4. Macro Analisis Data dengan Metode Penduga M Tukey *Bisquare* (Lanjutan)

```
let i=0
while se>0.01
let i=i+1
let ae=abso(e)
let se1=sum(ae)
let med=median(e)
let e1=abso(e-med)
let mad=median(e1)
let s=mad/0.6745
do j=1:n
 if ae(j)/s<=4.685
  let w(j)=(1-((ae(j)/4.685)^2))^2
 else
  let w(j)=0
 endif
enddo
regress y1 1 x1;
 weights w;
 residuals e;
 coefficients b;
 constant;
 brief 2.
let ae=abso(e)
let se2=sum(ae)
let se=abso(se2-se1)
if i=10
let se=0.01
endif
let se1=se2
endwhile
let cb0=b(1)
let cb1=b(2)
endmacro
```

Lampiran 5. Macro Analisis Data dengan Metode Penduga Theil

```
macro
boot1 x y id b0 b1 par nilai
mcolumn x y x1 y1 cn cn1 coef b0 b1 id b1s b0s par nilai z
mconstant B n i j pos b1rata b0rata sdb1 sdb0 a bb1 ba1 bb2 ba2 bb3
          ba3 bb4 ba4 b1h b0h text1 text2 text3 text4 text5 text6
          text7 text8 text9 text10 text11 text12 cb1 cb0
let b=100
let n=count(x)
do i=1:n
let cn(i)=i
enddo
do i=1:b
let id(i)=i
sample n cn cn1;
replace.
do j=1:n
let pos=cn1(j)
let x1(j)=x(pos)
let y1(j)=y(pos)
enddo
call theil x1 y1 cb0 cb1
let b0(i)=cb0
let b1(i)=cb1
enddo
let b1rata=average(b1)
let sdb1=stdev(b1)
let b0rata=average(b0)
let sdb0=stdev(b0)
let a=0.05
sort b1 b1s;
BY b1.
let pos=a/2*b
let bb1=b1s(pos)
let pos=(1-a/2)*b
let ba1=b1s(pos)
do i=1:b
let b1h=b1s(i)
```

Lampiran 5. Macro Analisis Data dengan Metode Penduga Theil (Lanjutan)

```
let z(i)=(b1h=b1rata)/sdb1
enddo
let pos=(1-a/2)*B
let bb2=b1rata-z(pos)*sdb1
let pos=a/2*B
let ba2=b1rata-z(pos)*sdb1
let a=0.05
sort b0 b0s;
BY b0.
let pos=a/2*b
let bb3=b0s(pos)
let pos=(1-a/2)*b
let ba3=b0s(pos)
do i=1:b
let b0h=b0s(i)
let z(i)=(b0h=b0rata)/sdb0
enddo
let pos=(1-a/2)*B
let bb4=b0rata-z(pos)*sdb0
let pos=a/2*B
let ba4=b0rata-z(pos)*sdb0
let text1="b1(.)"
let text2="se(b1)"
let text3="batas bawah bootstrap"
let text4="batas atas bootstrap"
let text5="batas bawah bootstrap-t"
let text6="batas atas bootstrap-t"
let text7="b0(.)"
let text8="se(b0)"
let text9="batas bawah bootstrap"
let text10="batas atas bootstrap"
let text11="batas bawah bootstrap-t"
let text12="batas atas bootstrap-t"
let par(1)=text1
let par(2)=text2
let par(3)=text3
```

Lampiran 5. Macro Analisis Data dengan Metode Penduga Theil (Lanjutan)

```
let par(4)=text4
let par(5)=text5
let par(6)=text6
let par(7)=text1
let par(8)=text2
let par(9)=text3
let par(10)=text4
let par(11)=text5
let par(12)=text6
let nilai(1)=b1rata
let nilai(2)=sdb1
let nilai(3)=bb1
let nilai(4)=ba1
let nilai(5)=bb2
let nilai(6)=ba2
let nilai(7)=b0rata
let nilai(8)=sdb0
let nilai(9)=bb3
let nilai(10)=ba3
let nilai(11)=bb4
let nilai(12)=ba4
print par nilai
endmacro
#
#
macro
theil x1 y1 cb0 cb1
mcolumn x1 y1 b a
mconstant i j n k cb0 cb1
sort x1 y1 x1 y1;
by x1.
let n=count(y1)
let k=1
do i=1:n-1
 do j=i+1:n
  if x1(i) <= x1(j)
```

Lampiran 5. Macro Analisis Data dengan Metode Penduga Theil (Lanjutan)

```
let b(k)=(y1(j)-y1(i))/(x1(j)-x1(i))
   let k=k+1
  endif
 enddo
enddo
let cb1=median(b)
do i=1:n
let a(i)=y1(i)-cb1*x1(i)
enddo
let cb0=median(a)
print cb0 cb1
endmacro
```

Lampiran 6. Macro Simulasi dan Analisis Data dengan Galat Terkontaminasi *Outlier*

```
macro
simulasi es xs ys eu xu yu nilai
mcolumn es xs ys eu xu yu ek1 xk1 yk1 ek2 xk2 yk2 ek3 xk3 yk3 o os1 os2 os3 nol1 nol2 nol3 bls1k1 bls1k2 bls1k3  bth1k1 bth1k2 bth1k3  brb1k1 brb1k2 brb1k3 bls1k1b bth1k1b brb1k1b bls1k2b bth1k2b brb1k2b bls1k3b bth1k3b brb1k3b bls1k1b2 bth1k1b2 brb1k1b2 bls1k2b2 bth1k2b2 brb1k2b2 bls1k3b2 bth1k3b2 brb1k3b2 par nilai
mconstant k n m p pc pr q r l u t v w y z cb1 bth b1m  bls1k1r bls1k2r bls1k3r bth1k1r bth1k2r bth1k3r brb1k1r brb1k2r brb1k3r bls1k1m bls1k2m bls1k3m bth1k1m bth1k2m bth1k3m brb1k1m brb1k2m brb1k3m
mconstant bls1k1br bth1k1br brb1k1br bls1k2br bth1k2br brb1k2br bls1k3br bth1k3br brb1k3br bls1k1v bls1k2v bls1k3v bth1k1v bth1k2v bth1k3v brb1k1v brb1k2v brb1k3v
mconstant text1 text2 text3 text4 text5 text6 text7 text8 text9 text10 text11 text12 text13 text14 text15 text16 text17 text18 text19 text20 text21 text22 text23 text24 text25 text26 text27

do k=1:1000
let n=count(xs)
random n es;
normal 0 1.

do u=1:n
 let ys(u)=3+2*xs(u)+es(u)
enddo
sort es xs ys eu xu yu;
by xs.
copy es xs ek1 xk1
let pc=3/100
let p=pc*n
let pr=round(p)
if pr<p
let pr=pr+1
else
```

Lampiran 6. Macro Simulasi dan Analisis Data dengan Galat Terkontaminasi *Outlier* (Lanjutan)

```
 let pr=pr
endif
random pr o;
normal 4 1.
do q=1:n-pr
 let nol1(q)=0
enddo
stack nol1 o os1
sample n os1 os1
do l=1:n
 if os1(l)=0
  let ek1(l)=ek1(l)
 else
  let ek1(l)=os1(l)
 endif
enddo
do t=1:n
 let yk1(t)=3+2*xk1(t)+ek1(t)
enddo
sort ek1 xk1 yk1 ek1 xk1 yk1;
by xk1.
print ek1 xk1 yk1
call mkt xk1 yk1 cb1
do k=1:n
 let bls1k1(k)=cb1
 let bls1k1b(k)=abso(1-bls1k1(k))
 let bls1k1b2(k)=(bls1k1b(k))**2
enddo
call theil xk1 yk1 bth
do k=1:n
 let bth1k1(k)=bth
 let bth1k1b(k)=abso(1-bth1k1(k))
 let bth1k1b2(k)=(bth1k1b(k))**2
enddo
call mkt xk2 yk2 cb1
 let bls1k2(k)=cb1
```

Lampiran 6. Macro Simulasi dan Analisis Data dengan Galat Terkontaminasi *Outlier* (Lanjutan)

```
call rob xk1 yk1 b1m
do k=1:n
 let brb1k1(k)=b1m
let brb1k1b(k)=abso(1-brb1k1(k))
 let brb1k1b2(k)=(brb1k1b(k))**2
enddo
copy es xs ek2 xk2
let pc=5/100
let p=pc*n
let pr=round(p)
if pr<p
 let pr=pr+1
else
 let pr=pr
endif
random pr o;
normal 4 1.
do r=1:n-pr
 let nol2(r)=0
enddo
stack nol2 o os2
sample n os2 os2
do m=1:n
 if os2(m)=0
  let ek2(m)=ek2(m)
 else
 let ek2(m)=os2(m)
 endif
enddo
do v=1:n
 let yk2(v)=3+2*xk2(v)+ek2(v)
enddo
sort ek2 xk2 yk2 ek2 xk2 yk2;
by xk2.
print ek2 xk2 yk2
let bls1k2b(k)=abso(1-bls1k2(k))
```

Lampiran 6. Macro Simulasi dan Analisis Data dengan Galat Terkontaminasi *Outlier* (Lanjutan)

```
let bls1k2b2(k)=(bls1k2b(k))**2
call theil xk2 yk2 bth
 let bth1k2(k)=bth
 let bth1k2b(k)=abso(1-bth1k2(k))
 let bth1k2b2(k)=(bth1k2b(k))**2
call rob xk2 yk2 b1m
 let brb1k2(k)=b1m
 let brb1k2b(k)=abso(1-brb1k2(k))
 let brb1k2b2(k)=(brb1k2b(k))**2
copy es xs ek3 xk3
let pc=7/100
let p=pc*n
let pr=round(p)
if pr<p
 let pr=pr+1
else
 let pr=pr
endif
random pr o;
normal 4 1.
do w=1:n-pr
 let nol3(w)=0
enddo
stack nol3 o os3
sample n os3 os3
do y=1:n
 if os3(y)=0
  let ek3(y)=ek3(y)
 else
  let ek3(y)=os3(y)
 endif
enddo
do z=1:n
 let yk3(z)=3+2*xk3(z)+ek3(z)
enddo
sort ek3 xk3 yk3 ek3 xk3 yk3;
```

Lampiran 6. Macro Simulasi dan Analisis Data dengan Galat Terkontaminasi *Outlier* (Lanjutan)

```
by xk3.
print ek3 xk3 yk3
call mkt xk3 yk3 cb1
 let bls1k3(k)=cb1
 let bls1k3b(k)=abso(1-bls1k3(k))
 let bls1k3b2(k)=(bls1k3b(k))**2
call theil xk3 yk3 bth
 let bth1k3(k)=bth
 let bth1k3b(k)=abso(1-bth1k3(k))
 let bth1k3b2(k)=(bth1k3b(k))**2
call rob xk3 yk3 b1m
 let brb1k3(k)=b1m
 let brb1k3b(k)=abso(1-brb1k3(k))
 let brb1k3b2(k)=(brb1k3b(k))**2
enddo
let bls1k1r=average(bls1k1)
let bls1k2r=average(bls1k2)
let bls1k3r=average(bls1k3)
let bth1k1r=average(bth1k1)
let bth1k2r=average(bth1k2)
let bth1k3r=average(bth1k3)
let brb1k1r=average(brb1k1)
let brb1k2r=average(brb1k2)
let brb1k3r=average(brb1k3)
let bls1k1br=average(bls1k1b)
let bls1k2br=average(bls1k2b)
let bls1k3br=average(bls1k3b)
let bth1k1br=average(bth1k1b)
let bth1k2br=average(bth1k2b)
let bth1k3br=average(bth1k3b)
let brb1k1br=average(brb1k1b)
let brb1k2br=average(brb1k2b)
let brb1k3br=average(brb1k3b)
let bls1k1m=average(bls1k1b2)
let bls1k2m=average(bls1k2b2)
let bls1k3m=average(bls1k3b2)
```

Lampiran 6. Macro Simulasi dan Analisis Data dengan Galat Terkontaminasi *Outlier* (Lanjutan)

```
let bth1k1m=average(bth1k1b2)
let bth1k2m=average(bth1k2b2)
let bth1k3m=average(bth1k3b2)
let brb1k1m=average(brb1k1b2)
let brb1k2m=average(brb1k2b2)
let brb1k3m=average(brb1k3b2)
let text1="bias penduga b1 ols data kontaminasi 3%"
let text2="bias penduga b1 ols data kontaminasi 5%"
let text3="bias penduga b1 ols data kontaminasi 7%"
let text4="bias penduga b1 theil data kontaminasi 3%"
let text5="bias penduga b1 theil data kontaminasi 5%"
let text6="bias penduga b1 theil data kontaminasi 7%"
let text7="bias penduga b1 m data kontaminasi 3%"
let text8="bias penduga b1 m data kontaminasi 5%"
let text9="bias penduga b1 m data kontaminasi 7%"
let text10="mse penduga b1 ols data kontaminasi 3%"
let text11="mse penduga b1 ols data kontaminasi 5%"
let text12="mse penduga b1 ols data kontaminasi 7%"
let text13="mse penduga b1 theil data kontaminasi 3%"
let text14="mse penduga b1 theil data kontaminasi 5%"
let text15="mse penduga b1 theil data kontaminasi 7%"
let text16="mse penduga b1 m data kontaminasi 3%"
let text17="mse penduga b1 m data kontaminasi 5%"
let text18="mse penduga b1 m data kontaminasi 7%"
let par(1)=text1
let par(2)=text2
let par(3)=text3
let par(4)=text4
let par(5)=text5
let par(6)=text6
let par(7)=text7
let par(8)=text8
let par(9)=text9
let par(10)=text10
let par(11)=text11
let par(12)=text12
```

Lampiran 6. Macro Simulasi dan Analisis Data dengan Galat Terkontaminasi *Outlier* (Lanjutan)

```
let par(13)=text13
let par(14)=text14
let par(15)=text15
let par(16)=text16
let par(17)=text17
let par(18)=text18
let nilai(1)=bls1k1br
let nilai(2)=bls1k2br
let nilai(3)=bls1k3br
let nilai(4)=bth1k1br
let nilai(5)=bth1k2br
let nilai(6)=bth1k3br
let nilai(7)=brb1k1br
let nilai(8)=brb1k2br
let nilai(9)=brb1k3br
let nilai(10)=bls1k1m
let nilai(11)=bls1k2m
let nilai(12)=bls1k3m
let nilai(13)=bth1k1m
let nilai(14)=bth1k2m
let nilai(15)=bth1k3m
let nilai(16)=brb1k1m
let nilai(17)=brb1k2m
let nilai(18)=brb1k3m
print par nilai
endmacro
#
#
macro
mkt x y cb1
mcolumn x y bls e
mconstant cb1
regress y 1 x;
coefficients bls;
residuals e;
brief 2.
```

Lampiran 6. Macro Simulasi dan Analisis Data dengan Galat Terkontaminasi *Outlier* (Lanjutan)

```
let cb1=bls(2)
endmacro
#
#
macro
theil x y bth
mcolumn x y b a
mconstant i j n k ath bth
sort x y x y;
by x.
let n=count(y)
let k=1
do i=1:n-1
 do j=i+1:n
  if x(i) <= x(j)
   let b(k)=(y(j)-y(i))/(x(j)-x(i))
   let k=k+1
  endif
 enddo
enddo
let bth=median(b)
do i=1:n
let a(i)=y(i)-bth*x(i)
enddo
let ath=median(a)
endmacro
#
#
macro
rob x y b1m
mcolumn x y e e1 ae b w
mconstant med mad s se1 se2 se i n j b1m
let n=count(y)
regress y 1 x;
residuals e;
constant;
```

Lampiran 6. Macro Simulasi dan Analisis Data dengan Galat Terkontaminasi *Outlier* (Lanjutan)

```
brief 2.
let se=1
let i=0
while se>0.01
let i=i+1
let ae=abso(e)
let se1=sum(ae)
let med=median(e)
let e1=abso(e-med)
let mad=median(e1)
let s=mad/0.6745
do j=1:n
 if ae(j)/s<=4.685
  let w(j)=(1-((ae(j)/4.685)**2))**2
 else
  let w(j)=0
 endif
enddo
regress y 1 x;
 weights w;
 residuals e;
 coefficients b;
 constant;
 brief 2.
let ae=abso(e)
let se2=sum(ae)
let se=abso(se2-se1)
let se1=se2
endwhile
let b1m=b(2)
endmacro
```