Kajian penaksiran parameter regresi robust untuk data *outlier* (pencilan) dengan estimasi-mm

Citra Setya Ningrum[1]
Swasono Rahardjo[2]

Program Studi Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Malang

Abstrak: Hubungan linear variabel terikat dan variabel bebas dapat diwujudkan dalam suatu model regresi. Dalam memodelkan suatu regresi dibutuhkan metode penaksiran. Metode penaksiran parameter model dengan Metode Kuadrat Terkecil (MKT) akan menghasilkan kesimpulan yang tidak sempurna apabila diketahui data yang memuat pencilan. Karena adanya pencilan menyebabkan penaksiran koefisien regresi yang dihasilkan tidak tepat. Oleh karena itu, diperlukan metode penaksiran parameter regresi yang kekar terhadap keberadaan pencilan, yaitu regresi *robust.* Regresi *robust* merupakan salah satu analisis statistik terutama analisis regresi yang digunakan untuk mengatasi pencilan yang tidak perlu dihapus dari data. Regresi ini memiliki beberapa metode penaksiran salah satunya adalah estimasi-MM yang diperkenalkan oleh Yohai (1987). Estimasi ini merupakan gabungan dari metode estimasi-S (*High Breakdown*) dan metode estimasi-M. Sebelum menaksir dengan estimasi-MM, data diidentifikasi untuk melihat pencilan dengan menggunakan metode grafis (*Scatterplot*), *boxplot*, dan sebagainya. Pada kasus ini dalam mengestimasi parameter regresi dengan software SAS 9.1, Minitab 16, dan SPSS 21.

Analisis ini bertujuan untuk mengetahui dampak pencilan terhadap hasil penaksiran dari estimasi-MM dan MKT dengan penghapusan pencilan serta menentukan metode terbaik antara keduanya. Data yang digunakan adalah data sekunder yang memuat pencilan. Hasil analisis menunjukkan adanya pengaruh terhadap hasil analisis regresi yang terlihat pada perubahan nilai dan tanda koefisien regresi, serta nilai R^2. Berdasarkan kriteria keakuratan model R^2 yang dihasilkan menunjukkan bahwa nilai R^2 MKT dengan menghapus pencilan lebih tinggi dibanding nilai R^2 estimasi-MM. Meskipun nilai keakuratannya lebih tinggi MKT dengan menghapus pencilan tetap saja lebih baik menggunakan estimasi-MM untuk menaksir parameter model regresi pada data yang memuat pencilan. Hal ini karena estimasi-MM dapat menaksir parameter pada data yang memuat pencilan tanpa menghapus pencilan tersebut, tetapi hanya menurunkan bobot dari pencilan tersebut. Berbeda dengan MKT, apabila data memuat pencilan untuk mendapatkan model regresi yang baik data pencilan tersebut dihapus. Padahal menghapus data bukan tindakan baik, dengan menghapus sebagian data berarti mengubah data aslinya sehingga kebenaran hasil prediksi masih dipertanyakan.

Kata Kunci: Pencilan, Regresi *Robust*, Estimasi-MM.

Abstract: Linear relationship dependent variable and independent variables can be realized in a regression model . Method of model parameter estimation with Least Squares Method (MKT) would lead to the conclusion that the data is

---

[1]*Mahasiswa Program Studi Matematika Universitas Negeri Malang*
[2]*Dosen Universitas Negeri Malang*

not perfect if known load outliers . Due to the presence of outliers cause the resulting regression coefficients assessment is not appropriate . Therefore, the regression parameter estimation methods are needed to the existence of outliers stocky , namely robust regression . Robust Regression analysis is one of the statistical regression analysis is used primarily to address the outliers that can not be removed from the data . This regression has several valuation methods one of which is MM estimation introduced by Yohai (1987) . This estimate is a combination of the method of S estimation (High Breakdown) and M estimation methods . Before assessing the MM - estimation , the data identified for see the outliers with using graphical methods (scatterplot) , boxplot , and so on. In this case the estimate of regression parameters with the software SAS 9.1, Minitab 16, and SPSS 21 .

This analysis aims to determine the impact of outliers on the results of an assessment of the MM estimation and MKT with the elimination of outliers and than determine the best method between the two. The data used are secondary data that contains outliers . The analysis revealed the existence of an influence on the results of the regression analysis that looks at changes in the value and sign of the regression coefficients , as well as value R^2 . Based on the criteria of the accuracy of the resulting model R^2 shows that the value R^2 MKT with the elimination of outliers taller than value R^2 MM estimation. Though, value of the accuracy taller than MKT with the elimination of outliers nevertheless better used MM estimation for estimate of regression parameters models of the data containing outliers. This is because the MM - estimation can estimate the parameters of the data containing outliers without removing the outliers , but only reduces the weight of outliers . In contrast to the MKT, if the data contains outliers to obtain a good regression model is deleted the data outliers. Though not act either delete the data , by deleting some of the data means changing the original data so that the truth of the prediction results are questionable .

Keywords: Outlier, Robust Regression, MM Estimation.

Dalam menentukan estimator terbaik sangat dipengaruhi oleh metode yang digunakan. Metode yang biasa digunakan untuk mengestimasi parameter regresi adalah Metode Kuadrat Terkecil (MKT). Tetapi Metode ini apabila terdapat data yang memuat pencilan akan menghasilkan model yang kurang baik atau nilai estimasi parameter bersifat bias sehingga berakibat interpretasi hasil tidak valid.

Adanya pencilan menandakan suatu titik data yang sama sekali tidak tipikal dibanding data lainnya. *Outlier* tidak dapat dibuang atau dihapus begitu saja dari pengamatan. Adakalanya *outlier* memberikan informasi yang tidak bisa diberikan oleh titik data lainnya, misalnya karena *outlier* timbul dari kombinasi keadaan yang tidak biasa yang mungkin saja sangat penting dan perlu diselidiki lebih jauh. *Outlier* dapat diabaikan apabila setelah ditelusuri ternyata merupakan akibat dari kesalahan mencatat amatan yang bersangkutan atau kesalahan ketika menyiapkan peralatan (Draper dan Smith,1992:146) sehingga untuk menghilangkan pencilan harus dilandasi alasan yang kuat.

Untuk mengidentifikasi adanya pencilan yang berpengaruh dalam koefisien regresi, terdapat beberapa metode untuk menentukan batasan pencilan dalam sebuah analisisnya, yaitu : Metode Grafis, *Boxplot*, *Leverage Values*, *DfFITS*, *Cook's Distance*, dan *DfBETA(s)*, dan *Internal Studenization* (Residu Yang Distudentkan).

Dalam menganalisis data pencilan ini bukan pencilannya yang diolah, tetapi menggunakan analisis regresi yang lain yang cocok untuk data yang memuat pencilan sehingga untuk mengatasinya diantaranya dilakukan dengan menggunakan regresi *robust*. Metode ini untuk mengatasi pencilan tanpa menghapusnya. Regresi robust digunakan ketika distribusi dari residu tidak normal artinya residual menyebar tidak mengikuti pola sebaran normal dengan kata lain tidak berdistribusi normal (W. Mendenhall& Sincich, T.,1996) dan adanya beberapa pencilan yang berpengaruh pada model tersebut.

Apabila suatu data pencilan tidak dihapus atau tidak menggunakan metode yang mengatasi masalah data pencilan misalnya menggunakan regresi *robust*, maka suatu pencilan akan memberikan dampak pada proses analisis data yang dihasilkan dan harus dihindari dalam banyak hal. Sehingga dampak dari pencilan menurut Soemartini (2007:7), yaitu kaitannya dengan analisis regresi adalah Residual yang besar dari model yang terbentuk atau $[\mathbf{e}] \neq \mathbf{0}$, Varians pada data tersebut menjadi lebih besar, dan Taksiran interval memiliki rentang yang lebar

Regresi *Robust* ini juga mempunyai kriteria penting, yaitu *Breakdown point* dan *efficiency* merupakan dalam regresi robust. *Breakdown point* adalah ukuran umum proporsi dari pencilan yang dapat ditangani sebelum pengamatan tersebut mempengaruhi model, sedangkan *efficiency* adalah taksiran parameter regresi yang mempunyai simpangan baku minimum (Wulansari, 2012:382).

Dengan menggunakan pendekatan regresi *robust* maka adanya pencilan tidak akan mempengaruhi pendugaan parameter. Sehingga diharapkan melalui metode regresi *robust* dapat diperoleh penduga parameter yang lebih baik yang bersifat bias dan mempunyai ragam minimum yang menghasilkan model yang lebih baik dari model hasil MKT. Akibatnya, model estimasi dalam regresi *robust* yang digunakan untuk menganalisis parameter yang digunakan. Metode estimasi dalam regresi *robust* salah satunya estimasi MM, dikenalkan oleh Yohai (1987). Metode ini menggabungkan estimasi S (estimasi scale dengan *high breakdown point*) dan estimasi M (SAS Inc, 2004). Sehingga estimasi-M dan estimasi-S dijelaskan sebagai berikut:

1. Estimasi-M

Estimasi-M diperkenalkan oleh Huber (1973) yang merupakan pengembangan dari estimasi kemungkinan maksimum (*Maximum Likelihood Estimator*). Estimasi ini menganalis data dengan mengasumsikan bahwa sebagian besar yang memuat pencilan (*outlier*) pada variabel x dan memiliki *breakdown point* $\frac{1}{\mathrm{n}}$, sehingga metode ini juga tidak dapat ditentukan secara signifikan. Estimasi-M diperoleh dengan meminimumkan fungsi ρ dari sisaannya (Montgomery dan Peck, 1982:367) yaitu:

$$\sum_{i=1}^{n} \rho(e_i^*) = \sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}}\right) = \sum_{i=1}^{n} \rho\left(\frac{y_i - \boldsymbol{x_i b}}{\hat{\sigma}}\right) \tag{4.1}$$

Nilai $\hat{\sigma}$ diperoleh melalui iterasi:

$$\hat{\sigma}^{(l)} = \frac{med_{i=1}^{n}\left|y_i - \boldsymbol{x_i b^{(l-1)}}\right|}{\beta_0} \tag{4.2}$$

Dengan $l(l = 1,2,\ldots)$ adalah iterasi, $\beta_0 = \Phi^{-1}(0,75)$ dan Φ^{-1} adalah invers fungsi komulatif normal standart (Bekti, 2009:9).
Sehingga (4.1) ekuivalen dengan

$$\min_{\beta} \sum_{i=1}^{n} \rho(e_i) = \min_{\beta} \sum_{i=1}^{n} \rho\left(y_i - \sum_{j=0}^{k} x_{ij}\hat{\beta}_j\right) \tag{4.3}$$

Dimana

$\hat{\beta}_j$: penaksir estimasi-M
$\rho(e_i^*) = \rho(e_i)$: fungsi obyektif

Akibatnya, meminimumkan persamaan (4.1) dengan diturunkan terhadap b sehingga

$$\sum_{i=1}^{n} \rho'\left(\frac{y_i - x_i b}{\hat{\sigma}}\right) = \sum_{i=1}^{n} \psi\left(\frac{y_i - x_i b}{\hat{\sigma}}\right) x_i = 0 \tag{4.4}$$

Dengan didefinisikan suatu fungsi pembobot

$$w(e_i) = \frac{\psi\left(\frac{y_i - x_i b}{\hat{\sigma}}\right) x_i}{\left(\frac{y_i - x_i b}{\hat{\sigma}}\right) x_i} \tag{4.5}$$

Dan misal fungsi pembobot $w_i = w(e_i) = \frac{\psi(e_i)}{e_i}$ maka persamaan (4.4) diperoleh

$$\sum_{i=1}^{n} w_i \left(\frac{y_i - x_i b}{\hat{\sigma}}\right) x_i = 0 \tag{4.6}$$

Apabila dinotasikan dalam bentuk matriks menjadi

$$X^T W X b = X^T W y \tag{4.7}$$

Persamaan ini disebut *weighted least square* yang meminimumkan $\sum_{i=1}^{n} w_i (y_i - \hat{y}_i)^2$. *Weighted least square* digunakan untuk memperoleh estimasi-M, sehingga estimasi parameternya menjadi

$$b = (X^T W X b)^{-1} X^T W y \tag{4.8}$$

Pembobot estimasi-M bergantung pada residual dan koefisien. Estimasi-M ini mempunyai tiga bentuk diantaranya estimasi *least square*, Huber, dan *Tukey bisquare* (*beweight*). Bentuk fungsi *objektif*, fungsi *influence*, dan fungsi pembobot untuk ketiga bentuk estimasi-M pada Tabel 4.1 sebagai berikut:

Tabel 4.1 Fungsi objektif, fungsi *influence*, dan fungsi pembobot pada estimasi-M

Metode	*Least square*	*Huber*	*Tukey bisquare*
Fungsi Objektif	$\rho_{LS}(e) = (e_i)^2$	$\rho_H(e) = \begin{cases} \frac{(e_i)^2}{2} & , \lvert e_i \rvert \le k \\ k\lvert e_i \rvert - \frac{k^2}{2} & , \lvert e_i \rvert > k \end{cases}$	$\rho_B(e) = \begin{cases} \frac{k^2}{6}\left[1 - \left(1 - \left(\frac{e_i}{k}\right)^2\right)^3\right] & , \lvert e_i \rvert \le k \\ \frac{k^2}{6} & , \lvert e_i \rvert > k \end{cases}$
Fungsi *Influence*	$\psi_{LS}(e) = e_i$	$\psi_H(e) = \begin{cases} e_i & , \lvert e_i \rvert \le k \\ k & , e_i > k \\ -k & , e_i < -k \end{cases}$	$\psi_B(e) = \begin{cases} e_i\left(1 - \left(\frac{e_i}{k}\right)^2\right)^2 & , \lvert e_i \rvert \le k \\ 0 & , \lvert e_i \rvert > k \end{cases}$

Fungsi Pembobot	$w_{LS}(e) = 1$	$w_H(e) = \begin{cases} 1 & , \lvert e_i \rvert \leq k \\ \frac{k}{\lvert e_i \rvert} & , \lvert e_i \rvert > k \end{cases}$	$w_B(e) = \begin{cases} \left(1 - \left(\frac{e_i}{k}\right)^2\right)^2 & , \lvert e_i \rvert \leq k \\ 0 & , \lvert e_i \rvert > k \end{cases}$

Sumber: Fox (2002), Montgomery&Peck (1992)

Nilai k dalam fungsi *objektif*, fungsi *influence*, dan fungsi pembobot pada Tabel 4.1 adalah nilai konstan. Menurut Kizmic (2004) dalam Bekti (2009:10) menyebutkan estimasi-M bentuk Huber efektif digunakan pada $\alpha = 5\%$ dengan $k = 1{,}345$. Sedangkan bentuk Tukey bisquare dengan $k = 4{,}685$.

Algoritma estimasi-M yang digunakan adalah IRLS (*iteratively reweighted least square*) yaitu untuk menyelesaikan masalah pada estimasi-M sehingga perlu dilakukan prosedur iterasi ini. Langkah-langkahnya, yaitu:

i. Menaksir parameter regresi dan diperoleh residual $e_{i,0}$.
ii. Menentukan $\hat{\sigma}^{(0)}$ dan fungsi pembobot $w_{i,0}$.
iii. Mencari estimasi pada iterasi l ($l = 1, 2, \ldots$) dengan *weighted least square*

$$\boldsymbol{b_l} = \left(\boldsymbol{X^T W_{l-1} X}\right)^{-1} \boldsymbol{X^T W_{l-1}} y$$

dengan w_{l-1} merupakan matrik diagonal dengan elemen diagonalnya adalah $w_{i,l-1}$. Sehingga estimasi parameter pada iterasi pertama ($l = 1$) menggunakan $e_{i,0}$ dan $w_{i,0}$.

iv. Mengulang tahap (ii) dan (iii) sehingga diperoleh penaksir parameter yang konvergen.

2. Estimasi-S

Estimasi-S diperkenalkan oleh Rousseeuw dan Yohai (1984). Estimasi ini dengan *high breakdown point* yang nilainya sama. Dan mempunyai efisiensi yang tinggi. Sehingga tujuan estimasi ini untuk memperoleh penafsir dengan nilai simpangan baku terkecil, yaitu (Ryan, 2009:434):

$$\min \hat{s}(e_1, e_2, \ldots, e_n) \qquad (4.12)$$

Dengan $\hat{s}(e)$ diperoleh dari estimasi-M skala sisaan yag merupakan solusi dari

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}_M}\right) = K \qquad (4.13)$$

ekuivalen dengan

$$\frac{1}{n}\sum_{i=1}^{n} \rho(u_i) = K \qquad (4.14)$$

Dimana $\rho(u_i)$ = fungsi kriteria dan nilai K = 0,5. Untuk mencari nilai K = 0,5 dapat diselesaikan dengan (Rousseeuw & Leroy, 1987:135):

$$K = E_{\Phi}(\rho) \qquad (4.15)$$

dimana Φ : standart normal

fungsi ρ harus memenuhi kondisi berikut:

(i) ρ simetris dan mempunyai turunan yang kontinu, dan $\rho(0) = 0$.
(ii) Terdapat $c > 0$ sehingga ρ naik pada $[0, c]$ dan konstan pada $[c, \infty)$.
(iii) Untuk menunjukkan nilai breakdown point dari estimasi-S yaitu

$$\frac{K}{\rho(c)} = \frac{1}{2}$$

Sisaan awal yang digunakan pada estimasi-S adalah sisaan yang diperoleh dari estimasi-M. Selanjutnya, untuk IRLS merupakan proses penaksiran melalui MKT terboboti dilanjutkan dengan menghitung sisaan dan pembobot w_i yang baru dan dilakukan penaksiran secara berulang-ulang sampai konvergen. Kekonvergenan tercapai jika perubahan jumlah mutlak sisaan dari iterasi terakhir ke iterasi sebelumnya kurang dari 0,01 (Salibian dan Yohai, 2006).

$$\sum |e_i|_n - \sum |e_i|_{n-1} < 0{,}01$$

dimana

$\sum |e_i|_n$ = jumlah mutlak sisaan pada saat iterasi ke-n

$\sum |e_i|_{n-1}$ = jumlah mutlak sisaan pada saat iterasi ke $n-1$

Untuk Algoritma Estimasi-S sebagai berikut:

1. Mengambil *m* subset contoh berdasarkan kombinasi banyaknya parameter $p+1$ dari n banyaknya pengamatan
$$m = {}_nC_{p+1} = \frac{n!}{(p+1)(n-(p+1))!}$$
2. Membentuk persamaan dengan $Y = b_0 + b_2X_1 + b_3X_3 + \cdots + b_nX_n + e$
3. Menaksir koefisien regresi $\hat{\boldsymbol{b}}$ pada persamaan langkah (2) menggunakan metode kuadrat terkecil (MKT)
4. Menaksir parameter model regresi dengan *Iteratively Reweighted Least Square* (IRLS) untuk setiap subset contoh dengan langkah-langkah, yaitu:
   i. Berdasarkan metode awal langkah (3), menghitung u_i dan S_M. u_i yang didefinisikan sebagai $u_i = \frac{e_i}{\hat{\sigma}}$
   dimana
   e_i : sisaan
   $\hat{\sigma}$: simpanan baku dari estimasi-M yang diperoleh berdasarkan
   $$\hat{\sigma}_M = med\,\frac{|e_i - med\ (e_i)|}{0{,}6475}, \qquad i = 1,2,\dots,n$$
   dengan med = median
   ii. Menghitung nilai w_i yaitu pembobot pengamatan ke-i dengan $k = 1{,}547$ yang didefinisikan sebagai
   $$w_i = \frac{\psi(u_i)}{u_i} = \begin{cases} \frac{6}{(1{,}547)^2}\left[1-\left(\frac{u_i}{1{,}547}\right)^2\right]^2 & , |u_i| \le 1{,}547 \\ 0 & , |u_i| > 1{,}547 \end{cases}$$
   iii. Menggunakan MKT terboboti untuk mendapatkan estimasi kuadrat terkecil terboboti $\boldsymbol{b}^* = (\boldsymbol{X'WX})^{-\mathbf{1}}\boldsymbol{X'WY}$
   dimana $\boldsymbol{W}$ (matriks diagonal nxn) = diagonal utama $[w_1, w_2, \dots, w_2]$ sehingga diperoleh sisaan e_i yang baru.
   iv. Menjadikan sisaan langkah (iii) sebagai sisaan awal pada langkah (i) sehingga diperoleh nilai $\hat{\sigma}_S, u_m, w_i$ yang baru.
   v. Iterasi diulang sampai diperoleh kekonvergenan sehingga didapatkan $\hat{\boldsymbol{b}}_{\boldsymbol{M}}$ yang merupakan estimasi-M dan diperoleh sisaan e_i yang baru.
   vi. Berdasarkan sisaan pada langkah (v) menghitung $\hat{\sigma}_s$ sesuai $\hat{\sigma}_M = med\,\frac{|e_i - med\ (e_i)|}{0{,}6475}$ untuk mendapatkan u_k sesuai persamaan $u_i = \frac{e_i}{\hat{\sigma}}$

vii. Menghitung w_i sesuai persamaan $w_i = \frac{\psi(u_i)}{u_i} =$

$$\begin{cases} \frac{6}{(1{,}547)^2}\left[1-\left(\frac{u_i}{1{,}547}\right)^2\right]^2 & ,|u_i| \leq 1{,}547 \\ 0 & ,|u_i| > 1{,}547 \end{cases}$$

viii. Menggunakan MKT terboboti untuk mendapatkan estimasi kuadrat terkecil terboboti $\boldsymbol{b}^* = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{Y}$ sehingga diperoleh sisaan e_i yang baru.

ix. Menjadikan sisaan langkah (viii) sebagai sisaan awal pada langkah (vi) sehingga diperoleh nilai $\hat{\sigma}_S$ dan pembobot w_i yang baru.

x. Iterasi diulang sampai diperoleh kekonvergenan sehingga didapatkan $\widehat{\boldsymbol{b}}_{\boldsymbol{i}}$ yang merupakan estimasi-M dan simpangan baku $\hat{\sigma}_S$.

5. Menyatakan $\widehat{\boldsymbol{b}}_{\boldsymbol{i}}$ dari subset contoh dengan nilai simpangan baku $\hat{\sigma}_S$ terkecil sebagai penaksir parameter model dari estimasi-S.

METODE

Pada kajian ini, langkah yang dilakukan terbagi menjadi dua bagian, yaitu:

Metode Kajian

Identifikasi Pencilan

Terdapat beberapa metode untuk menentukan batasan pencilan dalam sebuah analisis, yaitu :

a. Metode Grafis (Scatter Plot)
b. Boxplot
c. Leverage Values, DfFITS, Cook's Distance, dan DfBETA(s)
d. Internal Studenization (Residu Yang Distudentkan)

Estimasi-MM

Metode estimasi-MM adalah menaksir parameter regresi dengan estimasi-S dan dilanjutkan dengan estimasi-M. Langkah-langkah penaksiran, yaitu:

1. Menaksir parameter model regresi dengan estimasi-S. Langkah-langkah penaksiran sebagai berikut:

a. Mengambil N subset contoh berukuran ${}_nC_{p+1}$ dengan

$$ {}_nC_{p+1} = \frac{n!}{(p+1)(n-p-1)!} \tag{4.23}$$

sehingga terbentuk subset contoh I_r $(r = 1,2,\dots,N)$ yang berisi $(p+1)$ pengamatan dimana p adalah banyakya peubah prediktor.

b. Membentuk model untuk setiap subset contoh I_r yaitu $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}$ dimana

$$\underset{(p+1)\text{x}1}{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{p+1} \end{bmatrix}$$

$$\underset{(p+1)\text{x}(p+1)}{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{(p+1)1} & \cdots & X_{(p+1)p} \end{bmatrix}$$

$$\underset{(p+1)\text{x}1}{\beta} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix}$$

$$\underset{(p+1)\text{x}1}{y} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{p+1} \end{bmatrix}$$

c. Menghitung $\widehat{\boldsymbol{\beta}}$ dengan MKT sehingga diperoleh sisaan e_i dimana $m = 1,2, \dots, p+1$.

d. Menaksir parameter model regresi dengan MKT terboboti (*Iteratively Reweighted Least Square*/IRLS) untuk setiap subset contoh dengan langkah-langkah berikut ini:

   i. Menghitung u_m berdasarkan persamaan $u_i = \frac{e_i}{\hat{\sigma}}$ (4.24)
   dengan $\hat{\sigma}$ merupakan simpangan baku estimasi-M yang diperoleh berdasarkan persamaan sebgai berikut:
   $$\hat{\sigma}_M = med \frac{|e_i - med\ (e_i)|}{0{,}6475} \tag{4.25}$$
   dimana med : median

   ii. Menghitung nilai w_M dengan persamaan
   $$w_m = \begin{cases} \frac{6}{(1{,}547)^2}\left[1 - \left(\frac{u_i}{1{,}547}\right)^2\right]^2 & , |u_i| \leq 1{,}547 \\ 0 & , |u_i| > 1{,}547 \end{cases} \tag{4.26}$$

   iii. Menaksir parameter model regresi dengan MKT terboboti $\boldsymbol{\beta}^* = (\boldsymbol{X'WX})^{-1}\boldsymbol{X'WY}$ sehingga diperoleh sisaan e_m yang baru.

   iv. Menjadikan sisaan langkah (iii) sebagai sisaan awal pada langkah (i) sehingga diperoleh nilai $\hat{\sigma}_M, u_m, w_m$ yang baru.

   v. Iterasi diulang sampai diperoleh kekonvergenan sehingga didapatkan $\widehat{\boldsymbol{\beta}}_{\boldsymbol{M}}$ yang merupakan estimasi-M dan diperoleh sisaan e_i yang baru.

e. Menaksir parameter model regresi dengan metode IRLS dan sisaan yang diperoleh pada langkah d (v). Langkah-langkah penaksiran yaitu:

   i. Menghitung nilai $\hat{\sigma}_S$ berdasarkan persamaan
   $$\frac{1}{n}\sum_{i=1}^{n} \rho_0'\left(\frac{y_i - X_{ij}\hat{\beta}_j}{\hat{\sigma}_S}\right) = K \tag{4.27}$$
   untuk mendapatkan nilai u_m berdasarkan persamaan $u_i = \frac{e_i}{\hat{\sigma}}$ dengan pembobot $\hat{\sigma}_S$.

   ii. Menghitung nilai w_m berdasarkan persamaan
   $$w_m = \begin{cases} \frac{6}{(1{,}547)^2}\left[1 - \left(\frac{u_i}{1{,}547}\right)^2\right]^2 & , |u_i| \leq 1{,}547 \\ 0 & , |u_i| > 1{,}547 \end{cases}$$

   iii. Menaksir parameter model regresi dengan MKT terboboti $\boldsymbol{\beta}^* = (\boldsymbol{X'WX})^{-1}\boldsymbol{X'WY}$ sehingga diperoleh sisaan e_i yang baru.

   iv. Menjadikan sisaan langkah (iii) sebagai sisaan awal pada langkah (i) sehingga diperoleh nilai $\hat{\sigma}_S, u_m, w_m$ yang baru.

v. Iterasi diulang sampai diperoleh kekonvergenan sehingga didapatkan $\widehat{\boldsymbol{\beta}}_{\boldsymbol{S}}$ dan simpangan baku $\hat{\sigma}_S$.

f. Menyatakan $\widehat{\boldsymbol{\beta}}_{\boldsymbol{S}}$ dari subset contoh dengan nilai simpangan baku $\hat{\sigma}_S$ terkecil sebagai penaksir parameter model dari estimasi-S.

2. Menaksir parameter model regresi dengan estimasi-MM
   a. Menghitung sisaan e_i $(i = 1,2,\dots,n)$ dari estimasi-S.
   b. Menghitung u_i berdasarkan persamaan $u_i = \frac{e_i}{\hat{\sigma}}$ dengan simpangan baku estimasi-MM yang diperoleh dari estimasi-S dan bernilai tetap.
   c. Menghitung nilai w_i berdasarkan persamaan

$$w_i = \begin{cases} \dfrac{6}{(4{,}685)^2}\left[1-\left(\dfrac{u_i}{4{,}685}\right)^2\right]^2 & , |u_i| \leq 4{,}685 \\ 0 & , |u_i| > 4{,}685 \end{cases} \qquad (4.28)$$

   d. Menaksir parameter model regresi dengan MKT terboboti $\boldsymbol{\beta}^* = (\boldsymbol{X'WX})^{-1}\boldsymbol{X'WY}$ sehingga diperoleh sisaan e_i yang baru.
   e. Menjadikan sisaan langkah (d) sebagai sisaan awal pada langkah (a) sehingga diperoleh nilai u_i dan w_i yang baru.
   f. Iterasi diulang sampai diperoleh kekonvergenan sehingga didapatkan $\widehat{\boldsymbol{\beta}}_{\boldsymbol{MM}}$ yang merupakan estimasi-MM.

Studi Kasus

Data yang digunakan adalah data sekunder yaitu data pada skripsi Universitas Brawijaya Jurusan Manajemen Konsentrasi Bidang Keuangan Fakultas Ekonomi tahun 2008 oleh Weiguna Adhe Djatmika.

Alat Bantu yang digunakan dalam menganalisa data yaitu Minitab 16, SPSS 21, dan SAS 9.1 for Windows. Semua alat bantu ini merupakan salah satu *software* yang digunakan untuk membantu menganalisa permasalahan statistika.

Untuk kriteria pemilihan model dalam analisis regresi diperlukan suatu ukuran yang dapat digunakan untuk mengetahui sampai seberapa jauh ketepatan atau kecocokan garis regresi yang terbentuk dapat menjelaskan kondisi yang sebenarnya. Ukuran tersebut dinamakan koefisien determinasi. Nilai koefisien determinasi R^2 mempunyai nilai dalam batas $0 \leq \mathrm{R}^2 \leq 1$ dan diperoleh melalui:

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \hat{y})^2}$$

dimana $i = 1,2,\dots,n$

Pada regresi *robust* koefisien, nilai koefisien determinasi R^2 didefinisikan sebagai berikut:

$$\mathrm{R}^2 = \frac{\sum_{i=1}^{n}\rho\left(\frac{y_i-\hat{y}}{\hat{s}}\right) - \sum_{i=1}^{n}\rho\left(\frac{y_i-\hat{y}_i}{\hat{s}}\right)}{\sum_{i=1}^{n}\rho\left(\frac{y_i-\hat{y}}{\hat{s}}\right)}$$

dimana $i = 1,2,\dots,n$ (SAS Institute Inc, 2004)

Koefisien determinasi yang memperhitungkan banyaknya variabel bebas dalam model disebut koefisien determinasi terkoreksi (R^2_{adj}) didefinisikan:

$$\mathrm{R}^2_{adj} = \left(\mathrm{R}^2 - \frac{p}{n-1}\right)\left(\frac{n-1}{n-p-1}\right)$$

dimana n : banyaknya pengamatan

p : banyaknya variabel bebas

HASIL DAN PEMBAHASAN

Estimasi-MM

Estimasi-MM merupakan metode penaksiran parameter regresi *robust* yang dikenalkan oleh Yohai (1987). Metode estimasi-MM adalah menaksir parameter regresi dengan meminimumkan simpangan baku estimasi-M dan dilanjutkan dengan menaksir parameter regresi menggunakan estimasi-M. Estimator yang diperoleh dengan meminimumkan simpangan baku estimasi-M yang disebut estimasi-S. Dengan demikian estimasi-MM merupakan kombinasi antara metode penaksiran regresi *robust High breakdown point* sebesar 50% dengan *efisiensi* yang tinggi sebesar 95%. Karena estimasi ini merupakan gabungan antara estimasi-S yang mempunyai nilai *breakdown* tinggi dengan estimasi-M yang mempunyai *efisiensi* yang tinggi, maka estimasi-MM ini memenuhi kreteria yang diharapkan untuk suatu regresi *robust*.

Estimasi ini bertujuan untuk mendapatkan estimator yang mempunyai tingkat kekekaran yang tinggi dan memiliki tingkat efisiensi tinggi yaitu sebesar 95% saat sisaan berdistribusi normal. Estimasi-MM diperoleh dengan menyelesaikan:

$$\arg\min \sum_{i=0}^{n} \rho\left(\frac{e_i}{\hat{\sigma}_S}\right) = \arg\min \sum_{i=0}^{n} \rho\left(\frac{y_i - \sum_{j=1}^{p} X_{ij}\hat{\beta}_j}{\hat{\sigma}_S}\right) \tag{4.18}$$

dimana

$\hat{\beta}_j$: estimasi-MM

$\hat{\sigma}_S$: simpangan baku estimasi-MM yang bernilai tetap dan diperoleh dari simpangan baku estimasi-S

Pada persamaan (4.18) untuk meminimumkan ρ (fungsi objektif) dari galatnya, dicari turunan parsial pertama dari ρ terhadap $\hat{\beta}_j$, $j = 0,1,2,\ldots,k$ kemudian disama dengankan nol. Sehingga

$$\sum_{i=0}^{n} \rho'\left(\frac{y_i - \sum_{j=1}^{p} X_{ij}\hat{\beta}_j}{\hat{\sigma}_S}\right) = \sum_{i=0}^{n} \psi\left(\frac{y_i - \sum_{j=1}^{p} X_{ij}\hat{\beta}_j}{\hat{\sigma}_S}\right) X_{ij} = 0$$

Ini memberikan $p = k + 1$ sistem persamaan

$$\sum_{i=0}^{n} X_{ij}\, \psi\left(\frac{y_i - \sum_{j=1}^{p} X_{ij}\hat{\beta}_j}{\hat{\sigma}_S}\right) = 0 \tag{4.19}$$

dengan X_{ij} adalah pengamatan ke-i pada variabel x ke-j dan $x_{i0} = 1$

Pada estimasi-MM ini didefinisikan fungsi pembobotnya sebagai berikut:

$$w(e_i) = \frac{\psi\left(\frac{y_i - \sum_{j=1}^{p} X_{ij}\hat{\beta}_j}{\hat{\sigma}_S}\right)}{\left(\frac{y_i - \sum_{j=1}^{p} X_{ij}\hat{\beta}_j}{\hat{\sigma}_S}\right)} \tag{4.20}$$

Dan misal $w_i = w(e_i)$, maka persamaan (4.19) dapat ditulis:

$$\sum_{i=0}^{n} X_{ij}\, w_i \left(y_i - \sum_{j=1}^{p} X_{ij}\hat{\beta}_j\right) = 0 \tag{4.21}$$

Pada persamaan (4.21) dapat diselesaikan dengan IRLS, penaksir awal koefisien $\hat{\beta}^{(1)}$ dan Residual $e_i^{(1)}$ diambil dari regresi robust dengan *high breakdown point* (estimasi S), untuk bobot permulaan $w_i^{(1)} = w\left(e_i^{(1)}\right)$, maka $p = k + 1$ persamaan (4.21) ditulis:

$$\sum_{i=1}^{n} x_{ij}\, w_i^{(1)} \left(y_i - \sum_{j=0}^{k} x_{ij} \hat{\beta}_j \right) = 0$$

Dimana,

$$w_i^{(1)} = \begin{cases} \dfrac{\psi\left[\dfrac{y_i - \sum_{j=0}^{k} x_{ij}\hat{\beta}_j}{\hat{\sigma}_s}\right]}{\dfrac{y_i - \sum_{j=0}^{k} x_{ij}\hat{\beta}_j}{\hat{\sigma}_s}}, & \text{jika } y_i \neq \sum_{j=0}^{k} x_{ij}\hat{\beta}_j^{(1)} \\ 1 & , \text{ jika } y_i = \sum_{j=0}^{k} x_{ij}\hat{\beta}_j^{(1)} \end{cases}$$

Untuk regresi berganda, persamaan (4.21) menjadi:

$$\sum_{i=1}^{n} x_i\, w_i^{(1)} \left(y_i - \sum_{j=0}^{k} x_{ij} \hat{\beta}_j \right) = 0$$

$$\sum_{i=1}^{n} w_i^{(1)}\, x_i y_i - \sum_{i=0}^{n} \sum_{j=0}^{k} x_{ij}^2 w_i^{(1)} \hat{\beta}_j = 0$$

$$\sum_{i=0}^{n} \sum_{j=0}^{k} x_{ij}^2 w_i^{(1)} \hat{\beta}_j = \sum_{i=1}^{n} w_i^{(1)}\, x_i y_i$$

Dalam bentuk matriks dapat ditulis:

$$\underset{p\,x\,n}{X'}\ \underset{n\,x\,n}{W^{(1)}}\ \underset{n\,x\,p}{X}\ \underset{p\,x\,1}{\beta} = \underset{p\,x\,n}{X'}\ \underset{n\,x\,n}{W^{(1)}}\ \underset{n\,x\,1}{y}$$

Dimana

$W^{(1)}$ adalah matriks diagonal yang berukuran n x n dengan elemen diagonalnya $w_1, w_2, \ldots, w_n$ (banyaknya observasi)

w_i merupakan pembobot pengamatan

Penaksir satu langkah dapat ditulis:

$$X'W^{(1)}X\hat{\beta}_j^{(2)} = X'W^{(1)}y$$

$$(X'W^{(1)}X)^{-1}X'W^{(1)}X\hat{\beta}_j^{(2)} = (X'W^{(1)}X)^{-1}X'W^{(1)}y$$

$$\underset{p\,x\,1}{\hat{\beta}_j^{(2)}} = \underset{p\,x\,p}{(X'W^{(1)}X)^{-1}}\ \ \underset{p\,x\,1}{X'W^{(1)}y}$$

Pada langkah selanjutnya dihitung kembali bobot $W_i^{(2)}$ menggunakan $\hat{\beta}_j^{(2)}$ dan skala parameter $\hat{\sigma}_s$. Untuk $w^{(m)}$ bobot yang diberikan, dapat diperoleh dengan $\hat{\beta}^{m+1} = (X'W^{(m)}X)^{-1}\, X'W^{(m)}y$ sampai $\sum_{i=0}^{n}|e_i^m|$ konvergen ($\sum|e_i|_m - \sum|e_i|_{m-1} < 0{,}01$) dan selisih nilai $\hat{\beta}^{m+1}$ dan $\hat{\beta}^m$ mendekati 0, dengan m banyaknya iterasi.

Analisa Hasil

1. Identifikasi Pencilan

Mengidentifikasi suatu pencilan dapat dilakukan dengan cara sebagai berikut:

a. Diagam Pancar (*Scatter Plot*)

Berdasarkan *output* SPSS 21 diperoleh plot data dengan tiap variabel menunjukkan bahwa terdapat beberapa data yang terletak jauh dari kumpulan data. Data tersebut yang disebut dengan pencilan (*outlier*). Dan diperoleh juga hasil *Scatter Plot* antara residual (e) dengan nilai prediksi y ($\hat{y}$), menunjukkan bahwa terdapat beberapa data yang terletak jauh dari kumpulan data.

b. Boxplot

Untuk menguji secara *boxplot* terlebih dahulu dihitung nilai Quartil yaitu $Q_1, Q_2, dan\ Q_3$ serta jangkauan (IQR, *Interquartile Range*). Atau apabila disajikan dalam *boxplot* diperoleh berdasarkan *output* minitab 16 plot data tiap variabel, menunjukkan bahwa terdapat data yang keluar dari *range* dan batas bukan pencilan yaitu terletak pada data ke-19 untuk variabel x_1 dan data ke-17 untuk variabel x_5. Data tersebut yang disebut dengan pencilan (*outlier*). Untuk lebih jelasnya data mana yang teridentifikasi pencilan dapat dilihat pada hasil *DfFITS*.

c. DfFITS

Nilai *DfFITS* yang diidentifikasi sebagai pencilan adalah datang yang nilai *DfFITS*-nya lebih besar dari $2\sqrt{\frac{p}{n}} = 2\sqrt{\frac{5}{23}} = 0{,}9325048$. Berdasarkan *output* Minitab 16 diperoleh nilai *DfFITS*, data yang diindikasikan sebagai pencilan yaitu data ke 17, 19, 21, dan 22.

d. *Leverage Values, Cook's Dstance,* dan *DfBETA(s)*

Dari perhitungan diperoleh nilai-nilai sebagai berikut:

- *Leverage Values* : $\frac{2p-1}{n} = \frac{2(5)-1}{23} = \frac{10-1}{23} = \frac{9}{23} = 0{,}3913$
- *Cook's Dstance* : $F(0{,}5; p, n-p) = F(0{,}5; 5{,}18) = 0{,}903807$
- *DfBETA(s)* : $\frac{2}{\sqrt{n}} = \frac{2}{\sqrt{23}} = \frac{2}{4{,}7958} = 0{,}4170$

Dengan kriteria diatas, akan diidentifikasi keberadaan pencilan pada masing-masing variabel sehingga diperoleh pada lampiran 1.

Pendeteksian pencilan pada data yang pencilannya terletak pada variabel y menunjukkan beberapa yang termasuk pencilan yaitu data ke-17, 19, 20, dan 22. Sedangkan pendeteksian pencilan pada data yang pencilannya terletak pada variabel x menunjukkan beberapa yang termasuk pencilan yaitu data ke- 17 dan 22 untuk variabel x_1, data ke- 17, 18, dan 19 untuk variabel x_3, data ke- 21 dan 22 untuk variabel x_4, serta data ke- 17 untuk variabel x_5.

Sehingga dari analisis diatas terbukti bahwa data memuat pencilan, dengan demikian metode yang bisa digunakan untuk mengestimasi parameter yaitu regresi *robust* dengan estimasi-MM. Karena apabila mengestimasi menggunakan MKT harus menghapus atau membuang pencilan agar tidak mengganggu proses analisis data dan menghasilkan model yang kurang baik. Meskipun pencilan identik dengan data yang tidak bagus, akan tetapi ini merupakan bagian terpenting dari data, karena menyimpan informasi tertentu.

2. Penaksiran Parameter

Dalam menganalisis data dengan menggunakan regresi linear, dilakukan beberapa tahapan untuk mencari hubungan antara variabel terikat dengan variabel bebas, malalui hubungan antara perputaran kas, perputaran piutang, perputaran persediaan, perputaran aktiva tetap, dan marjin laba operasi terhadap ROI. Berdasarkan output SAS 9.1 dengan sintaks untuk estimasi-MM diperoleh hasil estimasi regresi robust dengan estimasi-MM pada lampiran 2. Sehingga persamaan modelnya adalah $\mathrm{y} = -3{,}6951 - 0{,}1776\,\mathrm{x}_1 + 0{,}4851\,\mathrm{x}_2 + 0{,}1208\,\mathrm{x}_3 - 1{,}2279\,\mathrm{x}_4 + 0{,}1169\mathrm{x}_5$. Interpretasi dari model tersebut adalah sebagai berikut:

- Nilai $-3{,}6951$ merupakan nilai konstanta, yang menunjukkan bahwa jika tidak ada variabel rasio perputaran kas (x_1), rasio perputaran piutang (x_2), rasio perputaran persediaan (x_3), rasio perputaran aktiva tetap (x_4), dan marjin laba operasi (x_5) maka akan diperoleh nilai ROI sebesar $-3{,}6951$.
- Setiap penambahan rasio perputaran kas (x_1) akan menurukan rata-rata y sebesar $0{,}1776$ apabila x_2, x_3, x_4 dan $x_5$tetap.
- Setiap penambahan rasio perputaran piutang (x_2) akan meningkatkan rata-rata y sebesar $0{,}4851$ apabila x_1, x_3, x_4 dan $x_5$tetap.
- Setiap penambahan rasio perputaran persediaan (x_3) akan meningkatkan rata-rata y sebesar $0{,}1208$ apabila x_2, x_1, x_4 dan $x_5$tetap.
- Setiap penambahan rasio perputaran aktiva tetap (x_4) akan menurukan rata-rata y sebesar $1{,}2279$ apabila x_2, x_3, x_1 dan $x_5$tetap.
- Setiap penambahan marjin laba operasi (x_5) akan meningkatkan rata-rata y sebesar $0{,}1169$ apabila x_2, x_3, x_4 dan $x_1$tetap.

Selain dengan regresi robust menggunakan MM, terdapat cara untuk mengestimasi parameter adalah MKT dengan data pencilan di hapus. Berdasarkan output SAS 9.1 diperoleh pada lampiran 3. Sehingga persamaan modelnya adalah $\mathrm{y} = -3{,}98324 - 0{,}19581\,\mathrm{x}_1 + 0{,}43627\,\mathrm{x}_2 + 0{,}10095\,\mathrm{x}_3 + 0{,}43983\,\mathrm{x}_4 + 0{,}15773\mathrm{x}_5$. Interpretasi dari model tersebut adalah sebagai berikut:

- Nilai $-3{,}98324$ merupakan nilai konstanta, yang menunjukkan bahwa jika tidak ada variabel rasio perputaran kas (x_1), rasio perputaran piutang (x_2), rasio perputaran persediaan (x_3), rasio perputaran aktiva tetap (x_4), dan marjin laba operasi (x_5) maka akan diperoleh nilai ROI sebesar $-3{,}98324$.
- Setiap penambahan rasio perputaran kas (x_1) akan menurukan rata-rata y sebesar $0{,}19581$ apabila x_2, x_3, x_4 dan $x_5$tetap.
- Setiap penambahan rasio perputaran piutang (x_2) akan meningkatkan rata-rata y sebesar $0{,}43627$ apabila x_1, x_3, x_4 dan $x_5$tetap.
- Setiap penambahan rasio perputaran persediaan (x_3) akan meningkatkan rata-rata y sebesar $0{,}10095$ apabila x_2, x_1, x_4 dan x_5 tetap.
- Setiap penambahan rasio perputaran aktiva tetap (x_4) akan meningkatkan rata-rata y sebesar $0{,}43983$ apabila x_2, x_3, x_1 dan x_5 tetap.
- Setiap penambahan marjin laba operasi (x_5) akan meningkatkan rata-rata y sebesar $0{,}15773$ apabila x_2, x_3, x_4 dan x_1 tetap.

Berdasarkan hasil kedua model tersebut, terlihat bahwa hasil estimasi-MM dan MKT dengan data pencilan dihapus hampir sama untuk nilainya. Tetapi menghapus data pencilan bukanlah tindakan yang baik, karena adakalanya data yang mengandung pencilan merupakan datang yang berpengaruh terhadap keseluruhan data, selain itu juga dengan menghapus sebagian data berarti merubah data asli yang sudah ada yang mungkin dapat memberikan resiko kesalahan yang besar pada hasil penaksiran. Dengan demikian, metode estimasi-MM merupakan metode yang digunakan untuk data yang memuat pencilan tanpa menghapus data pencilan tersebut.

3. Pengujian Asumsi

Dalam analisis regresi, terdapat asumsi-asumsi yang berhubungan dengan sisaan dan variabel bebas. Asumsi-asumsi yang melandasi analisis regresi yaitu (Drapper, 1992):

a. Kenormalan Sisaan

Asumsi ini dapat dideteksi menggunakan *Normal Probability plot of residual.* Berdasarkan *output* SPSS 21 diperoleh hasil analisis menggunakan estimasi-MM. Dengan menggunakan uji *Anderson Darling* diperoleh nilai *p-value* sebesar $0{,}183$. Sehingga nilai tersebut lebih dari nilai $\alpha = 0{,}05$. Jadi dapat disimpulkan bahwa menerima H_0 dan sisaan berdistribusi normal. Dan untuk hasil analisis menggunakan MKT dengan menghapus pencilan. Dengan menggunakan uji *Anderson Darling* diperoleh nilai *p-value* $< 0{,}05 = \alpha$. Jadi dapat disimpulkan bahwa menolak H_0 dan sisaan tidak berdistribusi normal.

b. Kehomogenan Ragam Sisaan

Asumsi kehomogenan ragam sisaan dapat dideteksi dengan menggunakan plot sisaan terhadap nilai duga y (*fiited value*). Berdasarkan *output* Minitab 16 diperoleh hasil analisis menggunakan estimasi-MM sehingga dapat dilihat bahwa terlihat membentuk pita mendatar yang berada diantara nilai (-2,2) dengan standart sisa 95%. Dengan demikian, menunjukkan ragam sisaan berada dalam sebaran sehingga sisanya mempunyai keragaman yang bersifat homogen. Dan untuk hasil analisis menggunakan MKT dengan menghapus pencilan sehingga dapat dilihat bahwa terlihat membentuk pita mendatar yang melebihi nilai (-2,2), yaitu pada data ke-20 dengan nilai $3{,}42885$ dengan standart sisa 95%. Dengan demikian, menunjukkan ragam sisaan tidak berada dalam sebaran sehingga sisa keragaman bersifat tidak homogen.

c. Kebebasan Antar Sisaan

Asumsi ini dapat dideteksi melalui *plot* antara residual dengan observasi order data atau langsung dengan *autocorelation* pada Minitab 16. Berdasarkan *output* Minitab 16 diperoleh bahwa tidak terdapat garis biru yang melewati garis merah, sehingga dapat disimpulkan bahwa sisaan bersifat bebas. Ini berarti untuk hasil analisis menggunakan estimasi-MM, variabel bebas tidak saling bergantung dan mempengaruhi.Begitu juga untuk MKT dengan menghapus pencilan tidak terdapat garis biru yang melewati garis merah, sehingga dapat disimpulkan bahwa sisaan bersifat bebas. Ini berarti untuk hasil analisis menggunakan MKT dengan menghapus pencilan, variabel bebas tidak saling bergantung dan mempengaruhi.

d. Multikolinearitas

Asumsi ini yang diinginkan tidak terjadi multikolearitas antar variabel bebas. Untuk mendeteksinya menggunakan kriteria *variance inflation factor* (VIF) atau dengan membandingkan nilai *tolerance* $< 0{,}1$ maka terjadi multikolinearitas. Berdasarkan *output* SPSS 21 nilai VIF dan nilai *tolerance* masing-masing variabel bebas untuk keseluruhan data hasil analisis menggunakan estimasi-MM, diperoleh semua nilai VIF tidak ada yang lebih dari 10 maka tidak terjadi multikolearitas atau menggunakan nilai *tolerance* tidak ada yang kurang dari 0,1 maka tidak terjadi multikolinearitas. Dan untuk hasil analisis menggunakan MKT dengan menghapus pencilan, diperoleh semua nilai VIF tidak ada yang lebih dari 10 maka tidak terjadi multikolearitas atau menggunakan nilai *tolerance* tidak ada yang kurang dari 0,1 maka tidak terjadi multikolinearitas.

4. Perbandingan Penaksir Terbaik antara Estimasi-MM dan MKT Dengan Menghapus Pencilan

Koefisien determinasi R^2 menunjukkan proporsi dari total keragaman yang dijelaskan oleh model yang terbentuk, semakin besar nilainya menunjukkan semakin kesil nilai residualnya. Yang pada prinsipnya model mendekati data sebenarnya, dan dikatakan semakin akurat.

Untuk estimasi-MM, terlihat bahwa nilai R^2 (koefisien determinasi) adalah 0,7167. Nilai tersebut mendekati 1 maka dapat disimpulkan bahwa variabel-variabel x memberikan hampir semua informasi yang dibutuhkan untuk memprediksi variabel y.

Dan untuk MKT dengan menghapus pencilan, terlihat bahwa nilai R^2 (koefisien determinasi) adalah 0,7481. Nilai tersebut mendekati 1 maka dapat disimpulkan bahwa variabel-variabel x memberikan hampir semua informasi yang dibutuhkan untuk memprediksi variabel y.

Meskipun untuk nilai R^2 nya lebih baik untuk MKT dengan menghapus pencilan tetapi metode ini kurang baik apabila harus menghapus data pencilan. Jadi, metode yang seharusnya digunakan untuk data yang memuat pencilan adalah metode robust, salah satunya estimasi-MM.

KESIMPULAN

Berdasarkan analisis data yang telah dilakukan pada bab sebelumnya dapat disimpulkan bahwa:

1. Regresi *robust* dengan estimasi-MM adalah menaksir parameter regresi dengan estimasi-S dan dilanjutkan dengan estimasi-M. Dalam mengestimasi MM menunjukkan nilai sisaan e_i dan $\hat{\sigma}, u, w$ berdasarkan persamaan $\frac{1}{n}\sum_{i=1}^{n} \rho_0' \left(\frac{y_i - X_{ij}\hat{\beta}_j}{\hat{\sigma}}\right) = K.$
2. Keberadaan pencilan dapat mempengaruhi hasil analisis regresi, baik pada nilai dan tanda koefisisen regresi juga pada koefisien determinasi (R^2) sebagai salah satu kriteria keakuratan model. Hasil analisis menunjukkan adanya pengaruh terhadap hasil analisis regresi yang terlihat pada perubahan nilai dan tanda koefisien regresi, serta nilai R^2. Berdasarkan kriteria keakuratan model R^2 yang dihasilkan menunjukkan bahwa nilai R^2 MKT dengan menghapus pencilan lebih tinggi dibanding nilai R^2 estimasi-MM. Tetapi dari sini meskipun nilai keakuratannya lebih tinggi MKT dengan menghapus pencilan tetap saja lebih baik menggunakan estimasi-MM untuk menaksir parameter model regresi pada data yang memuat pencilan. Hal ini karena estimasi-MM dapat menaksir parameter pada data yang memuat pencilan tanpa menghapus pencilan tersebut, tetapi hanya menurunkan bobot dari pencilan tersebut. Berbeda dengan MKT, apabila data memuat

pencilan untuk mendapatkan model regresi yang baik data pencilan tersebut dihapus. Padahal menghapus data bukan tindakan baik, dengan menghapus sebagian data berarti mengubah data aslinya sehingga kebenaran hasil prediksi masih dipertanyakan. Dan berdasarkan data yang telah dianalisis juga menunjukkan bahwa estimasi-MM menghasilkan model regresi yang tidak jauh berbeda dengan Metode Kuadrat Terkecil dengan menghapus pencilan.

SARAN

Bagi pengkaji ataupun peneliti selanjutnya diharapkan dapat:

1. Untuk peneliti selanjutnya dapat digunakan metode penaksiran regresi *robust* pada data yang tidak memenuhi asumsi analisis regresi misalnya seperti *Heteroscedasticity and Autocorrelation Consistent* (HAC) *estimates of the standart errors for robust estimator* dan *Significance Regression-Weighted Least Square* (SR-WLS) pada data yang tidak memenuhi asumsi seperti sisaannya tidak normal, terdapat multikolinearitas antara variabel bebas, tidak homogen antar sisaan, dan sebagaimya.
2. Untuk pengkaji selanjutnya dapat digunakan metode penaksiran regresi *robust* dengan estimasi lain yang lebih baru beserta algoritmanya. Karena selama ini kebanyakan masih penjabaran implementasinya belum penjabaran kajian secara teoritis seperti estimasi Welsch pada Jurnal Penelitian Sains oleh Cahyani, Dian.

DAFTAR PUSTAKA

Bekti, R. D. 2009. *Model Hubungan Anomali Luas Panen Padi dan Curah Hujan Terboboti (Weighted Rainfall Index) dengan Regresi Robust.* Skripsi. Surabaya: Jurusan Statistik FMIPA Institut Teknologi Surabaya.

Cahyawati, D., Tanuji, H. & Abdianti, R. 2009. *Efektivitas Metode Regresi Robust Penduga Welsch Dalam Pencilan Pada Pemodelan Regresi Linear Berganda*. Jurnal Penelitian Sains, (online), Volume 12, Nomer 1(A) 12101, http://digilib.unsri.ac.id/download/jpsmipaunsri-v12-no1-01-a-dian.pdf diakses pada tanggal 17 Maret 2013

Chatterjee, S. & Hadi, S. A. 1938. *Regression Analysis By Example Fourth Edition.* New Jersey: John Wiley & Sons Inc.

Djatmika, W. A. 2008. *Analisis Pengaruh Perputasan Kas, Rasio Perputasan Piutang Usaha, Rasio Perputaran Persediaan, Rasio Perputaran Aktiva Tetap, dan Rasio Marjin Laba Operasi Terhadap Pengembalian Investasi (ROI) Perusahaan.* Skripsi. Malang: Jurusan Managemen Konsentrasi Bidang Keuangan Fakultas Ekonomi Universitas Brawijaya.

Draper, N.R & Smith. H. 1992. *Analisis Regresi Terapan Edisi Kedua*. Terjemahan Bambang Sumantri. Jakarta: PT. Gramedia Pustaka Utama.

Draper, R. N. & Smith, H. 1998. *Applied Regression Analysis Third Edition*. New York: John Wiley & Sons Inc.

Eduardus, T. 2001. *Analisis Investasi&Manajemen portofolio.* Yogjakarta: BRFE.

Gujarati, D. N. 2004. *Basic Econometrics*. 4^{th}. Ed. New York: McGraw-Hill.

Hasanah, Isma & Tripena, Agustini, Br. Sb. ___. *Regresi Robust Untuk Mengatasi Outlier Pada Regresi Linear Berganda.* Artikel. Univ. Jenderal Soedirman.

Hocking. R. R. 1932. *Methods And Applications Of Linear Models Regression And The Analysis Of Variance.* New York: John Wiley & Sons Inc.

Huber, Peter,J. 2009.*Robust Statistic Second Edition.* New York: John Wiley&Sons Inc.

Mendenhall, W. & Sincich, T. 1996. *A Second Course In Statistics Regression Analysis Seventh Edition.* USA: Pearson Education Inc.

Marrazi, A.1993. *Algorithms, Rautines, and S function for robust Statistics.* Chapman&Hall: New York

Munawir, S. 2002. *Analisis Laporan Keuangan.* Yogjakarta: Liberty.

Myers, R. H. 1990. *Classical and Modern Regression with Aplication.* PWS. KENT: Boston.

Permadi, H.____.*Teknik Analisis Regresi, Individual Textbook.* Malang:UMPress.

Rousseeuw, Peter. J. & Leroy, Annick. M. 1987. *Robust Regression And Outlier Detection.* New York: John Wiley & Sons Inc.

Ryan, Thomas. P. 2009. *Modern Regression Methods.* New Jersey: John Wiley & Sons Inc.

Salibian, B, M and V. J. Yohai. 2006. *A fast Algorith for S Regression.* Es. Jurnal of Computational and Graphical Statistic, Vol. 15, No. 2, Page 414-427.

SAS. 2004. *SAS STAT® 9.1 User's Guide*. Cary, NC: SAS Institute Inc., USA.

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

Setiawan, E. 2010. *KBBI Offline Versi 1.5.*

Soemartini. 2007. *Pencilan* (*Outlier*). Universitas Padjadjaran, Bandung. (online). http://resources.unpad.ac.id/unpadcontent/uploads/publikasi_dosen/ OUTLIER% 28PENCILAN%29.pdf diakses pada tanggal 11 Maret 2013

Weisberg, Sanford. 1947. *Applied Linear Regression Second Edition.* New York: John Wiley & Sons Inc.

Weston, J. F & Thomas, E. C. 1992. *Managerial Finance* 9^{th} *ed.* Terjemahan oleh A. Jaka, W. MSM & Kibrandoko, MSM. 1995. Jakarta: Binapura Aksara.

Wulansari, Y. , Susanti, Y. , & Roswitha, M. 2012. *Regresi Robust Dengan Generalized S-Estimation Pada Penjualan Tenaga Listrik Di Jawa Tengah Tahun 2010.* Disajikan pada Seminar Nasional Matematika 2012. Surakarta: Jurusan Matematika FMIPA Universitas Sebelas Maret Surakarta. (online). http://math.mipa.uns.ac.id/assets/proceeding/382-388_Revisi%20Yurista.pdf. diakses pada tanggal 14 April 2013.

Yohai, V. J. 1987. *High Breakdown Point and High Efficiency Robust Estimates for Regression, Annals of Statistics.* Vol. 15, No. 20, 642-656.

Lampiran 1 Tabel nilai *Leverage Values, Cook's Dstance,* dan *DfBETA(s)*

No.	Cook's Distance	Centered Leverage Value	Standardized DFBETA Intercept	Standardized DFBETA x_1	Standardized DFBETA x_2	Standardized DFBETA x_3	Standardized DFBETA x_4	Standardized DFBETA x_5
1	0.00388	0.18194	-0.04732	0.05809	0.00459	0.08045	-0.08229	0.03397
2	0.00000	0.07648	0.00027	-0.00010	-0.00010	-0.00016	0.00015	-0.00021
3	0.00443	0.13059	0.10516	-0.00856	-0.05812	-0.02464	0.07226	-0.10678
4	0.13865	0.35260	0.74473	-0.03904	0.10804	-0.30282	-0.10698	-0.69738
5	0.02546	0.13895	-0.19462	-0.05143	0.13993	-0.06592	0.18906	0.11024
6	0.01342	0.10524	0.03440	-0.15448	0.20684	-0.09139	0.09551	-0.07553
7	0.03575	0.14449	0.13707	-0.24065	0.31215	-0.27547	0.09829	-0.13821
8	0.01206	0.26982	0.07350	0.03241	-0.14928	0.07824	0.08805	-0.09713
9	0.01145	0.10010	-0.10897	0.01789	0.03831	0.01875	0.07601	0.05078
10	0.00848	0.04969	0.09040	0.02764	0.03239	-0.01683	-0.00487	-0.11906
11	0.00094	0.09884	-0.03645	-0.01509	0.00494	-0.01786	0.04274	0.03393
12	0.01549	0.14632	-0.01404	-0.06105	0.04894	-0.08200	0.22138	-0.03072
13	0.01000	0.11904	-0.13287	-0.01414	0.04103	0.00807	0.09695	0.08067
14	0.00361	0.03821	0.01201	0.05799	-0.06473	0.05481	-0.01876	-0.03039
15	0.00227	0.13976	0.01280	0.08090	-0.08113	0.04837	-0.06368	-0.00798
16	0.00013	0.18255	-0.00539	0.00093	0.01371	-0.01378	0.01211	0.00250
17	**1.04764**	**0.51320**	-1.89427	**0.45122**	0.39418	**0.53862**	-1.60557	**2.32507**
18	0.12459	0.26335	0.43192	0.09261	-0.25863	**0.60999**	-0.58393	-0.34927
19	0.75114	**0.77532**	-0.26300	-1.31800	-0.15074	**1.13876**	0.04675	0.22833
20	0.06516	**0.41257**	0.02474	0.33592	-0.17253	0.39156	-0.32847	-0.04739
21	0.14519	0.06642	0.07292	-0.32635	0.18891	-0.69740	**0.44709**	0.07692
22	0.51130	**0.40957**	-0.20548	**1.04487**	-1.46213	-0.24674	**0.45324**	0.19627
23	0.05331	0.28493	0.20738	-0.33838	0.37688	0.00677	0.04327	-0.23720

Lampiran 2 Hasil estimasi regresi robust dengan estimasi-MM

```
                    The SAS System         23:33 Thursday, March 3, 2014

                  The ROBUSTREG Procedure

                     Model Information

      Data Set                            WORK.INVESTASI
      Dependent Variable                               y
      Number of Independent Variables                  5
      Number of Observations                          23
      Method                                MM Estimation


           Number of Observations Read          23
           Number of Observations Used          23


                     Summary Statistics

                                                            Standard
Variable            Q1      Median          Q3        Mean    Deviation        MAD

x1              0.9710      1.5190      2.3150      1.7320       1.0494     0.9711
x2              2.2420      4.4970      7.1990      5.0030       2.8867     3.3433
x3             25.3990     50.3900     73.4640     50.7220      25.0136    34.2096
x4              0.3130      0.3430      0.5010      0.3798       0.1491     0.1720
x5             28.5810     30.9840     32.5980     31.2301       3.1301     2.9296
y               5.2830      8.8800     11.7500      8.4514       4.1110     4.6376


             Profile for the Initial LTS Estimate

      Total Number of Observations                  23
      Number of Squares Minimized                   19
      Number of Coefficients                         6
      Highest Possible Breakdown Value          0.2174


                        MM Profile

                 Chi Function           Tukey
                 K1                    3.4400
                 Efficiency            0.8500


                    Parameter Estimates

                           Standard   95% Confidence     Chi-
        Parameter DF Estimate   Error       Limits       Square Pr > ChiSq

       Intercept  1  -3.6951   3.3989 -10.3569   2.9667    1.18     0.2770
       x1         1  -0.1776   0.4089  -0.9791   0.6239    0.19     0.6641
       x2         1   0.4851   0.1579   0.1757   0.7945    9.44     0.0021
       x3         1   0.1208   0.0178   0.0859   0.1557   46.04     <.0001
       x4         1  -1.2279   3.2847  -7.6658   5.2100    0.14     0.7085
       x5         1   0.1169   0.1111  -0.1007   0.3346    1.11     0.2924
       Scale      0   1.7746


                    Diagnostics Summary

                         Observation
            Type           Proportion      Cutoff

            Outlier            0.0870      3.0000


                      Goodness-of-Fit

                    Statistic       Value

                    R-Square       0.7167
                    AICR          17.4319
                    BICR          33.4886
                    Deviance      46.2165
```

Lampiran 3 Hasil estimasi MKT dengan menghapus pencilan

The SAS System 19:51 Thursday, March 23, 2014
The REG Procedure
Model: MODEL1
Dependent Variable: y

Number of Observations Read 22
Number of Observations Used 22

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	238.26540	47.65308	9.50	0.0002
Error	16	80.22326	5.01395		
Corrected Total	21	318.48867			

Root MSE	2.23919	R-Square	0.7481
Dependent Mean	8.12677	Adj R-Sq	0.6694
Coeff Var	27.55320		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-3.98324	6.63657	-0.60	0.5568
x1	1	-0.19581	0.77752	-0.25	0.8044
x2	1	0.43627	0.28387	1.54	0.1439
x3	1	0.10095	0.03354	3.01	0.0083
x4	1	0.43983	6.23714	0.07	0.9447
x5	1	0.15773	0.21672	0.73	0.4773