



METODE LEAST TRIMMED SQUARE (LTS) DAN MM-ESTIMATION UNTUK MENGESTIMASI PARAMETER REGRESI KETIKA TERDAPAT OUTLIER

Elok Tri Kusuma Dewi ✉, Arief Agoestanto, Sunarmi

Jurusan Matematika, FMIPA, Universitas Negeri Semarang, Indonesia
Gedung D7 lantai 1 Kampus Sekaran, Gunungpati, Semarang, 50229

Info Artikel

Sejarah Artikel:
Diterima Maret 2015
Disetujui Januari 2016
Dipublikasikan Mei 2016

Keywords :
Outlier;
Robust Regression;
Least Trimmed Square (LTS);
MM-Estimation;

Abstrak

Artikel ini membahas kajian teoritis dan aplikasi menggunakan Microsoft Excel dan SPSS 19 dari metode Least Trimmed Square (LTS) dan metode MM-Estimation. Kajian teoritis difokuskan pada penjabaran konsep outlier, metode Least Trimmed Square (LTS), metode MM-estimation dan pemilihan model terbaik menggunakan kriteria R^2 dan residualnya. Pencilan (outlier) adalah data yang tidak mengikuti pola umum pada model regresi yang dihasilkan, atau tidak mengikuti pola data secara keseluruhan. Keberadaan outlier dalam data dapat mengganggu proses analisis data, sehingga mengakibatkan residual dan varians pada data menjadi lebih besar. Penelitian ini bertujuan untuk mengetahui keefektifan regresi robust dengan metode Least Trimmed Square (LTS) dan MM-estimation pada regresi linier berganda. Data ini terdiri dari usia (X_1) dan indeks massa tubuh (X_2) sebagai variabel independent sedangkan tekanan darah sistolik (Y) sebagai variabel dependent. Model yang dihasilkan menggunakan metode LTS yaitu $\hat{y} = 67,141 + 0,649X_1 + 0,587X_2$. Sedangkan model yang dihasilkan menggunakan metode MM-estimation yaitu $\hat{y} = 65,308 + 0,666X_1 + 0,618X_2$. Karena R^2 pada metode Least Trimmed Square (LTS) memperoleh nilai lebih besar dan residual lebih kecil dibandingkan metode MM-estimation maka dapat disimpulkan bahwa metode Least Trimmed Square (LTS) lebih efisien dalam mengestimasi parameter regresi dibandingkan metode MM-estimation.

Abstract

This article discusses the theoretical study and use excel and SPSS 19 application of Least Trimmed Square (LTS) methods and MM-estimation methods. Theoretical study focused on the elaboration of the concept of outlier, least trimmed square methods and MM-estimation methods and selection best model use the criteria R^2 and resid value. Outlier is data on who did not attend a pattern common regression on the model produced, or not follow as a pattern data as a whole. The existence of outlier in the data can be disrupt the process of data analysis, that led to the data on residual and variance become larger. This research aims to know the effectiveness of robust regression method with Least Trimmed Square (LTS) and MM-estimation in multiple linear regression. This data consisting of age (X_1) and body mass index (X_2) as variable independent while systolik blood pressure (Y) as dependent variables. The model produced using Least Trimmed Square methods that is $\hat{y} = 67,141 + 0,649X_1 + 0,587X_2$. Regarding the resulting uses the method MM-estimation that is $\hat{y} = 65,308 + 0,666X_1 + 0,618X_2$. Because at Least Trimmed Square method (LTS) obtained the R^2 value of is bigger and smaller than the residual method of MM-estimation then it can be concluded that the method of Least Square Trimmed (LTS) is more efficient in the estimate parameter of the regression compared the methods of MM-estimation.

Analisis regresi linier berganda merupakan salah satu metode statistik yang digunakan untuk memodelkan dan menyelidiki hubungan antara satu variabel terikat dengan satu variabel bebas. Hubungan yang menggambarkan antara variabel-variabel dalam regresi sering disebut sebagai model regresi klasik. Variabel sendiri dibedakan menjadi 2 macam, yakni variabel bebas (variabel independent) dan variabel terikat (variabel dependent). Variabel bebas adalah variabel yang tidak dipengaruhi variabel lain, sedangkan variabel terikat adalah variabel yang keberadaannya dipengaruhi oleh variabel lainnya.

Regresi linear berganda mempunyai bentuk persamaan :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \quad (1)$$

dimana β_0 , β_1 dan β_2 merupakan parameter yang belum diketahui nilainya. Dengan menggunakan metode kuadrat terkecil, maka akan diperoleh nilai parameter tersebut.

Dalam menentukan estimator terbaik dari analisis regresi sangat dipengaruhi oleh penggunaan metode. Metode untuk menaksir koefisien regresi klasik salah satunya dengan cara meminimumkan jumlah kuadrat terkecil, yang disebut dengan metode kuadrat terkecil atau metode *Ordinary Least Square* (OLS). Namun metode ini tidak dapat bekerja dengan baik apabila terdapat data outlier.

Berbagai kaidah telah diajukan untuk menolak pencilan (memutuskan untuk menghilangkan data yang ada outliernya, setelah itu data dianalisis ulang tanpa outlier). Penolakan outlier yang begitu saja bukanlah langkah yang bijaksana. Adakalanya outlier dapat 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 diselidik lebih jauh. Sebagai kaidah umum, outlier baru ditolak jika setelah ditelusuri ternyata merupakan akibat dari kesalahan-kesalahan seperti memasukkan ukuran atau analisis yang salah, ketidaktepatan pencatatan data, dan terjadi kerusakan alat pengukuran. Bila ternyata bukan akibat dari kesalahan-kesalahan semacam itu, penyelidikan yang seksama harus dilakukan (Drapper and Smith, 1992).

Data outlier adalah data pengamatan yang berada jauh (ekstrim) dari pengamatan-pengamatan lainnya (Makkulau et al., 2010).

Keberadaan data outlier akan mengganggu proses analisis data dan harus dihindari dari beberapa hal. Dalam kaitannya dalam analisis regresi, outlier dapat menyebabkan hal-hal berikut (Soemartini, 2007: 7): (1) Residual yang besar dari model yang terbentuk $E(e_i) \neq 0$ (2) Varians pada data tersebut menjadi lebih besar (3) Taksiran interval memiliki rentang yang lebar

Untuk mengidentifikasi apakah terdapat data outlier atau tidak, dapat menggunakan beberapa metode salah satunya yang akan dibahas adalah metode DfFITS (*Difference fitted value FITS*). *Difference fitted value FITS* merupakan metode yang menampilkan nilai perubahan dalam harga yang diprediksi bilamana kasus tertentu dikeluarkan dan yang sudah distandarkan. Selain itu juga bisa menggunakan metode *Cook's Distance*. *Cook's Distance* merupakan salah satu metode pendeteksian outlier dengan cara menampilkan nilai jarak *cook* atau dengan kata lain menunjukkan besarnya pengaruh adanya data outlier terhadap semua estimator koefisien regresi.

Salah satu metode untuk menganalisa data yang dipengaruhi oleh outlier sehingga dihasilkan model yang kekar adalah regresi robust. Regresi robust merupakan metode regresi yang tepat digunakan ketika ada beberapa outlier pada model. Suatu estimator yang kekar adalah relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data.

Prosedur regresi robust cenderung mengabaikan sisaan-sisaan (*error*) yang berhubungan dengan outlier yang besar. Ada beberapa metode dalam regresi robust yang dapat digunakan untuk mengatasi data pencilan, diantaranya *Least Median of Square (LMS)*, *Least Trimmed Squares (LTS)*, *M-estimation*, *S-Estimation* dan *MM-estimation*. Namun pada skripsi hanya akan dibahas metode *Least Trimmed Squares (LTS)* dan *MM-estimation*.

LTS diusulkan oleh Rousseuw (1998) sebagai alternatif robust untuk mengatasi kelemahan ordinary least squares (OLS), yaitu dengan menggunakan sebanyak kuadrat residual yang diturunkan nilainya. Metode *Least Trimmed Square* sebagai salah satu metode penaksiran parameter model regresi robust terhadap kehadiran nilai pencilan.

Adapun tujuan yang ingin dicapai yakni mendapatkan nilai parameter model regresi yang robust terhadap kehadiran nilai pencilan. Metode ini tidak membuang bagian data pencilan tapi menemukan model fit dari mayoritas data.

MM-Estimation adalah metode yang pertama kali diperkenalkan oleh Yohai pada tahun 1987 yakni dengan yang menggabungkan estimasi *high breakdown point* dan efisiensi statistik. Langkah pertama dalam estimasi ini adalah mencari estimator S, kemudian menetapkan parameter-parameter regresi menggunakan estimasi M. Pada umumnya digunakan fungsi *Tukey Bisquare* baik pada estimasi S maupun estimasi M. Metode ini juga mempertahankan kekekaran dari metode estimasi S, serta efisiensi dari metode estimasi M. Diharapkan melalui kedua metode regresi robust tersebut dapat diperoleh estimator yang baik sehingga menghasilkan model yang lebih baik dari model hasil *Ordinary Least Square* (OLS).

Fungsi obyektif adalah fungsi yang digunakan untuk mencari fungsi pembobot pada regresi robust. Fungsi pembobot yang digunakan antara lain adalah fungsi pembobot *Huber* dan fungsi pembobot *Tukey Bisquare* (Montgomery & Peck, 1992: 369).

(1) Fungsi Pembobot *Huber*

$$\rho(e_i) = \begin{cases} \frac{1}{2}e_i^2, & |e_i| \leq c \\ c|e_i| - \frac{1}{2}c^2, & |e_i| > c \end{cases}$$

dengan,

$$\psi(e_i) = \rho'(e_i) = \frac{\partial(\rho(e_i))}{\partial e_i} = \begin{cases} e_i, & |e_i| \leq c \\ c, & e_i > c \\ -c, & e_i < -c \end{cases}$$

dan fungsi pembobot,

$$w_i = w(e_i) = \frac{\psi(e_i)}{e_i} = \begin{cases} 1, & |e_i| \leq c \\ \frac{c}{|e_i|}, & |e_i| > c \end{cases}$$

(2) Fungsi Pembobot *Tukey Bisquare*

$$\psi(e_i) = \rho'(e_i) = \frac{\partial(\rho(e_i))}{\partial e_i} = \begin{cases} e_i \left[1 - \left(\frac{e_i}{c}\right)^2\right]^2, & |e_i| \leq c \\ 0, & |e_i| > c \end{cases}$$

dengan,

$$\rho(e_i) = \begin{cases} \frac{c^2}{6} \left\{1 - \left[1 - \left(\frac{e_i}{c}\right)^2\right]^3\right\}, & |e_i| \leq c \\ \frac{c^2}{6}, & |e_i| > c \end{cases}$$

dan fungsi pembobot,

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

Konstanta yang menghasilkan efisiensi tinggi dengan residual berdistribusi normal dan dapat memberikan perlindungan terhadap outlier yaitu konstanta dengan nilai $c=1.345$ untuk fungsi pembobot *Huber* dan $c=4,685$ untuk pembobot *Tukey Bisquare*.

Pada penelitian sebelumnya didapatkan kesimpulan bahwa S-Estimation lebih baik dari pada *Least Trimmed Squares* (LTS) (Andhika, 2013). Selain itu, pada penelitian lain yang ditulis sebelumnya oleh dengan memodelkan regresi linier berganda data “Ketahanan Pangan di Jawa Tengah Tahun 2007” menggunakan regresi robust estimasi M-IRLS dengan fungsi pembobot *Huber* dan *Tukey Bisquare* (Elen, 2012).

Penelitian ini bertujuan untuk mengetahui keefektifan regresi robust dengan metode *Least Trimmed Squares* (LTS) dan *MM-estimation* pada regresi linier berganda. Ada berbagai kriteria yang bisa ditetapkan sebagai acuan perbandingan kedua metode tersebut, namun pada penelitian ini akan dilihat dari kriteria nilai R^2 dan residualnya. Jika nilai R^2 besar atau mendekati satu berarti variabel-variabel independen memberikan hampir semua informasi yang dibutuhkan untuk memprediksi variansi variabel dependen dan jika semakin kecil nilai residualnya maka semakin baik kecocokan suatu persamaan dengan data nilai duga y semakin mendekati sebenarnya. Dengan melihat nilai R^2 dan residual yang didapat dengan menggunakan rumus ataupun dengan bantuan software SPSS 19 akan dapat terlihat metode mana yang lebih efektif.

Metode Penelitian

Dalam penelitian ini pengumpulan pustaka diperoleh dari berbagai sumber berupa buku-buku referensi, skripsi, jurnal, dan literatur lainnya. Pengumpulan data yang digunakan merupakan data primer. Metode yang digunakan adalah kuantitatif. Data yang digunakan adalah usia dan indeks masa tubuh (IMT) sebagai variabel independen sedangkantekanan darah sistolik sebagai variabel dependen.

Adapun langkah-langkah pada metode LTS adalah menggunakan gabungan FAST LTS, C-Step dan FWLS yaitu : (1)Menghitung estimasi parameter awal (2)Menentukan n residual yang bersesuaian dengan parameter regresi. (3)Kemudian menghitung sejumlah h_0 pengamatan dengan nilai terkecil. (4) Menghitung $\sum_{i=1}^{h_0} \varepsilon_i^2$. (5) Melakukan estimasi parameter dari h_0 pengamatan. (6) Menentukan n kuadrat residual kemudian menghitung sejumlah h_{baru} pengamatan dengan nilai terkecil. (7)Melakukan C-step yaitu tahap (4) sampai (6) untuk mendapatkan fungsi objektif yang kecil dan konvergen.

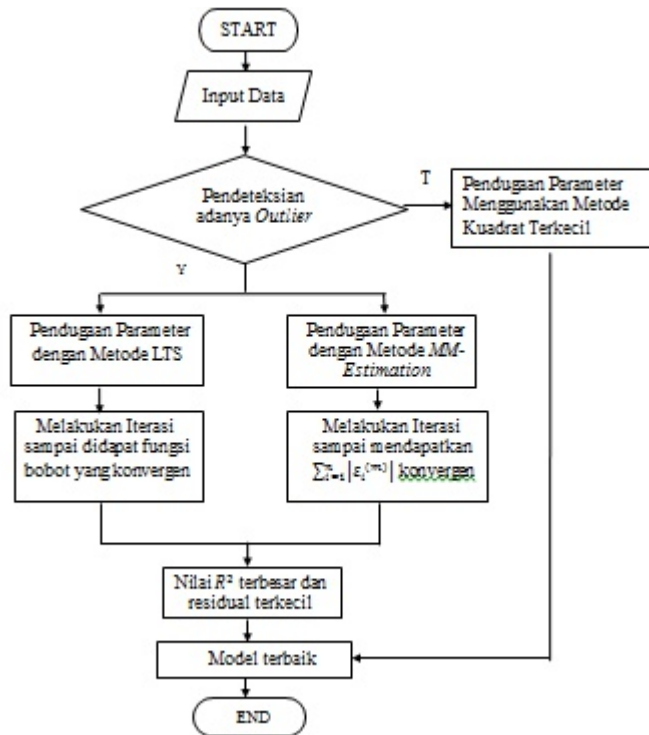
Langkah-langkah dalam proses *MM-estimation* adalah: (1) Menghitung estimator awal koefisien dan residual dari regresi robust

(estimasi S) dengan bobot *huber / bisquare*. (2)Residual pada langkah pertama digunakan untuk menghitung skala estimasi dan dihitung pula pembobot awal w_i^1 . (3) Residual dengan skala estimasi pada langkah kedua digunakan dalam iterasi awal sebagai penaksir WLS untuk menghitung koefisien regresi yang merupakan pembobot *Huber/bisquare*. (4)Menghitung bobot baru w_i^2 dengan skala estimasi dari iterasi awal WLS. (5)Mengulang langkah (2), (3), (4) (dengan skala estimasi tetap konstan) sampai mendapatkan konvergen (selisih dan mendekati 0, dengan banyak m iterasi).

Dalam mengestimasi parameter *MM-estimation* metode iterasi sangat diperlukan, karena residual tidak dapat dihitung sampai diperoleh model terbaik dan parameter regresi juga tidak dapat dihitung tanpa mengetahui nilai bobotnya.

Setelah mendapat parameter regresi masing-masing metode, maka langkah selanjutnya adalah memilih metode pendugaan terbaik berdasarkan nilai R^2 dan residualnya.

Diagram Alir untuk langkah-langkah estimasi model metode Least Trimmed Square dan *MM-Estimation* akan diperlihatkan pada Gambar 1.



Gambar 1. Diagram Alir Estimasi Model

Hasil dan Pembahasan

Contoh kasus dalam penelitian ini menggunakan data yang merupakan hasil penelitian dari 40 responden yaitu pengaruh usia dan indeks massa tubuh (IMT) terhadap tekanan darah sistolik. Responden yang dituju adalah pasien dari Puskesmas Sekaran Kota Semarang. Data ini terdiri dari data yaitu : usia dan indeks massa tubuh (IMT) sebagai variabel independent dan tekanan darah sistolik sebagai variabel dependent.

Berdasarkan hasil penelitian uji *Kolmogorov Smirnov* diperoleh hasil nilai signifikan untuk uji distribusi normal yaitu $0,961 > 0,05$. Karena nilai signifikansi untuk uji test distribusi normal lebih dari 0,05 maka dapat disimpulkan bahwa data mempunyai distribusi normal. Hasil uji nya bisa dilihat pada Gambar 2 berikut.

Tabel 1. Hasil Uji Multikolinier

Model	Collinearity Statistics	
	Tolerance	VIF
(Constant)		
Usia	0,602	1,660
IMT	0,602	1,660

Setelah memenuhi semua asumsi dalam regresi berganda, dilakukan pengecekan adanya outlier pada data tersebut.

Ketika peneliti mendeteksi outlier, perlakuan pertamanya adalah melihat kemungkinan bahwa outlier merupakan data yang terkontaminasi. Data outlier dapat dikenali dengan pemeriksaan visual dari data mentahnya (raw) atau dari diagram pencar dari variabel dependen.

One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual
N		40
Normal Parameters ^{a, b}	Mean	.0000000
	Std. Deviation	5.84902572
Most Extreme Differences	Absolute	.080
	Positive	.045
	Negative	-.080
Kolmogorov-Smirnov Z		.504
Asymp. Sig. (2-tailed)		.961

a. Test distribution is Normal.

b. Calculated from data.

Gambar 2. Hasil Uji Distribusi Normal

Hasil uji selanjutnya adalah uji asumsi multikolinieritas. Uji ini bertujuan untuk mengetahui bahwa data mengalami multikolinieritas atau tidak. Hal ini dilihat dari output VIF dan tolerance pada tabel 1. Dari tabel 1 terlihat bahwa semua nilai VIF kurang dari 10 dan nilai tolerance lebih dari 0,1 karena data tidak mengalami multikolinieritas maka semua variabel prediktor masuk ke dalam model.

Pendeteksian Pecilan pada penelitian ini menggunakan metode *cook's distance* dan metode DfFITS. Perhitungan *cook's distance* di rumuskan sebagai berikut :

$$(\text{cook's distance})_i = \left[\frac{R_{standardi}^2}{2} \right] \times \left[\frac{h_{ii}}{1 - h_{ii}} \right]$$

dimana h_{ii} adalah nilai pengaruh untuk kasus ke- i . Pada metode *cook's distance* berlaku ketentuan jika nilai *cook's distance* dari masing-masing data $> F(0,5;k,n-k)$, maka data tersebut merupakan pecilan. Diperoleh bahwa data tidak terdeteksian pecilan menggunakan metode ini.

Sedangkan perhitungan DfFITS di rumuskan sebagai berikut :

$$(DfFITS)_i = t_i \left(\frac{h_{ii}}{1 - h_{ii}} \right)^{\frac{1}{2}}$$

dimana t, adalah studentized deleted residual untuk kasus ke-i dan h_{ii} adalah nilai pengaruh untuk kasus ke-i.

dengan,

$$t_i = e_i \sqrt{\frac{n - p - 1}{JKG(1 - h_{ii}) - e_i^2}}$$

e_i adalah residual ke-i dan JKG adalah jumlah kuadrat galat. Pada metode DfFIT, jika nilai masing-masing data yang $> 2\sqrt{(k+1)/n}$ maka dikategorikan sebagai outlier. Diperoleh bahwa data yang terdeteksi outlier adalah data ke-1, ke-23 dan ke-28. Dengan nilai DfFITS data ke-1 = 6,42822, data ke-23 = 0,71722 dan data ke-28 = 1,11782. Oleh karena itu, perlu dilakukan analisis regresi menggunakan metode robust untuk data yang mengandung pencilan. Pada metode yang pertama dalam proses regresi robust Least Trimmed Square dengan rumus perhitungannya seperti rumus (2).

$$\min_b \sum_{i=1}^h \varepsilon_i^2 \quad (2)$$

dihasilkan model regresi . Persamaan itu diperoleh dari beberapa iterasi. Iterasi yang terjadi pada data tersebut sebanyak 2 iterasi. Hal ini karena pada iterasi ke-3, data pencilan tidak termasuk didalamnya, ini tidak sesuai dengan konsep regresi robust yaitu tetap mengikut sertakan data pencilan dalam menemukan model persamaan regresi. Pada metode Least Trimmed Square juga terjadi pemangkasan sejumlah data sebesar h. Inilah yang menyebabkan nilai jumlah kuadrat residual pada metode ini semakin kecil dari iterasi 1 sampai iterasi akhir. Nilai R^2 yang didapatkan dalam metode ini adalah 0,919. Ini menunjukkan bahwa variabel independent memberikan pengaruh yang cukup besar terhadap variabel dependen.

Jika ditulis dalam satu tabel,

penyelesaian menggunakan metode Least Trimmed Square (LTS) akan ditunjukkan pada Tabel 2.

Tabel 2. Hasil Iterasi LTS

Tahap	N	H	$\hat{\beta}_0^{(n)}$	$\hat{\beta}_1^{(n)}$	$\hat{\beta}_2^{(n)}$	$\sum_{i=1}^{n_0} \varepsilon_i^2$
1	40	21	65,323	0,665	0,618	1334,233
2	21	12	67,141	0,649	0,587	199,2524

Proses regresi selanjutnya menggunakan metode MM-Estimation dengan rumus perhitungan seperti rumus (3).

$$\begin{aligned} \hat{\beta}_{MM} &= \arg \min \sum_{i=1}^n \rho \left(\frac{\varepsilon_i}{\hat{\sigma}} \right) \\ &= \arg \min \sum_{i=1}^n \rho \left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{\hat{\sigma}} \right) \end{aligned} \quad (3)$$

Model regresi yang dihasilkan adalah $\hat{y} = 67,141 + 0,649x_1 + 0,587x_2$. Pada metode ini juga mengalami 2 iterasi untuk sampai pada model regresi terbaik metode MM-Estimation. Nilai residual yang di dapat dari metode ini termasuk residual yang cukup besar yaitu 186,6879. Sedangkan nilai R^2 yang didapatkan dalam metode MM-Estimationa adalah 0,566. Ini menunjukkan bahwa model regresi yang diperoleh tadi lemah atau kurang baik digunakan untuk memprediksi.

Pada penelitian ini, peneliti membandingkan nilai R^2 dari masing-masing model regresi pada metode Least Trimmed Square (LTS) dan metode MM-Estimation. Jika ditulis dalam satu tabel, penyelesaian menggunakan metode MM-Estimator akan ditunjukkan pada Tabel 3.

Tabel 3. Hasil Iterasi MM-Estimation

Tahap	$\hat{\beta}_0^{(n)}$	$\hat{\beta}_1^{(n)}$	$\hat{\beta}_2^{(n)}$	$\sum_{i=1}^n \varepsilon_i^{(m)} $
Estimator S	65,323	0,665	0,618	186,687
1	65,308	0,666	0,618	186,688
2	65,308	0,666	0,618	186,688

Tahapan pemilihan model regresi terbaik dimulai dengan melihat nilai R^2 dan nilai residual dari model regresinya.

Tabel 4. Perbandingan Nilai R^2

No	Metode Regresi	$\hat{\beta}_1^{(n)}$	$\hat{\beta}_2^{(n)}$	$\hat{\beta}_2^{(n)}$	R^2
1	LTS	67,141	0,649	0,587	0,919
2	MM-Estimation	65,308	0,666	0,618	0,566

Dari tabel 4 dapat dilihat bahwa pada model regresi pada metode *Least Trimmed Square* nilai R^2 nya adalah 0,919. Sedangkan pada metode *MM-Estimation* nilai R^2 nya adalah 0,566. Ini berarti model regresi pada metode *Least Trimmed Square* memberikan pengaruh yang lebih besar yaitu sebanyak 91,9% dibandingkan dengan metode *MM-Estimation* yang hanya memberikan pengaruh sebanyak 56,6%. Dengan kata lain, metode *Least Trimmed Square* (LTS) merupakan metode terbaik untuk mengestimasi parameter pada saat data terdeteksi mengandung outlier karena memiliki R^2 yang lebih banyak.

Untuk nilai residual, semakin besar nilai residualnya maka garis regresi semakin kurang tepat digunakan untuk memprediksi. Yang diharapkan adalah total residualnya kecil sehingga garis regresi cukup baik untuk digunakan. Hasil dari nilai residual metode *Least Trimmed Square* dan *MM-Estimation* akan ditunjukkan pada Tabel 5.

Tabel 5. Nilai residual

No	Metode Regresi	$\sum_{i=1}^n \varepsilon_i $
1	LTS	44,89375
2	MM-Estimation	186,6879

Dari tabel 5 terlihat bahwa metode *Least Trimmed Square* (LTS) mempunyai nilai residual yang lebih kecil, hal ini disebabkan adanya pemangkasan (*trimmed*) data. Jadi, sama halnya dengan nilai R^2 , nilai residual metode *Least Trimmed Square* (LTS) juga lebih baik jika dibandingkan dengan metode *MM-Estimation*.

Kesimpulan dan Saran

Dari analisis data dan pembahasan, maka dapat ditarik kesimpulan antara lain pengecekan outlier pada penelitian ini menggunakan metode *Cook's Distance* dan metode DfFIT (*Difference in fit Standardized*). DfFIT menampilkan nilai perubahan dari variabel yang diprediksikan bila kasus tertentu dikeluarkan yang sudah distandardkan. Sedangkan *Cook's Distance* menampilkan nilai

jarak *cook* atau dengan kata lain menunjukkan besarnya pengaruh adanya data pencilan terhadap semua estimator koefisien regresi. Kriteria data dikategorikan pencilan jika nilai *Cook's Distance* $> F(0,5;k;n-k)$ dan nilai DfFIT $> 2\sqrt{((k+1)/n)}$.

Model regresi robust terbaik dengan adanya data pencilan (outlier) dengan menggunakan metode *Least Trimmed Squares* (LTS) yaitu, sedangkan model regresi robust terbaik dengan adanya data pencilan (outlier) dengan menggunakan metode *MM-estimation* yaitu . Pada regresi robust, metode *Least Trimmed Square* (LTS) lebih efektif jika dibandingkan dengan metode *MM-Estimation* dilihat dari nilai R^2 dan residualnya. Hal ini disebabkan adanya pemangkasan (*trimmed*) terhadap data yang mempunyai residual besar.

Saran pada penelitian ini adalah pada penelitian ini hanya mengkaji pada metode *Least Trimmed Square* dan metode *MM-Estimation* dalam regresi robust, sehingga ada baiknya dilakukan pengujian untuk mengatasi adanya data pencilan dengan menggunakan metode lain yang ada pada regresi robust.

Daftar Pustaka

- Candraningtyas, S., Safitri, D. & Ispriyanti, D. 2013. Regresi Robust MM-Estimator Untuk Menangani Pencilan Pada Regresi Linear Berganda. *Jurnal Gaussian*, 2(4): 395-404.
- Chen, C. 2002. *Robust Regression and Outlier Detection with the ROBUSTREG Procedure*. SAS Institute Inc: Cary NC.
- Draper, N.R., & H. Smith. 1992. *Analisis Regresi Terapan, Diterjemahkan oleh Bambang Sumantri*. Gramedia: Jakarta.
- Hampel. 1986. *Multivariate aregression S-estimation for Robust Estimation and Inference*. Ghent University.
- Herawati, N., Khoirin, N. & Eri, S. 2011. Analisis Ketegaran Regresi Robust Terhadap Letak Outlier: Studi Perbandingan. *Bulletin of Mathematics*, 3(1): 49-60.
- Huber, P. J. 1973. Robust Regression: Asymptotics, conjecture and Monte Carlo. *Ann Stat.* 1(5): 799-821.
- Larson, J. 2011. Our statistical intuitions may be misleading us: Why we need robust statistics. *Cambridge University Press*. 45(4). 460-474.
- Maharani, I. F., N. Satyahadewi, & D. Kusnandar. 2014. Metode Ordinary Least Squares dan Least Trimmed Squares Dalam Mengestimasi Parameter Regresi Ketika Terdapat Outlier. *Buletin Ilmiah Mat. Stat. Dan Terapannya*. 3(3): 163-168.

- Makkulau., Susanti, L., & Muhammad, M. 2010. Pendeteksian Outlier Dan Penentuan Faktor-Faktor Yang Mempengaruhi Produksi Gula Dan Tetes Tebu Dengan Metode Likelihood Displacement Statistic-Lagrange. *Jurnal TI*. 12(2): 25-100.
- Montgomery, D. C., & Peck, L. A. 1992. *Introduction to Linear Regression Analysis (2th ed)*. John Wiley & Sons Inc: New York.
- Permana, A. T. 2012. Perbandingan Metode Least Trimmed Square (LTS) dan Penduga-S Sebagai Metode Pendugaan Parameter Regresi Robust. *Jurnal Matematika Universitas Brawijaya*. Diakses tanggal 8 Maret 2015.
- Pradewi, E. D. & Sudarno. 2012. Kajian Estimasi-M IRLS Menggunakan Fungsi Pembobot Huber Dan Bisquare Tukey Pada Data Ketahanan Pangan di Jawa Tengah. *Journal Media Statistik*. 5(1): 1-10.
- Rousseeuw, P.J., & Leroy, A. M. 1987. *Robust Regression and Outlier Detection*. John Wiley & Sons Inc: New York.
- Sembiring, R. K. 1995. *Analisis Regresi*. Bandung: ITB.
- Sembiring, R. K. 2003. *Analisis Regresi (2th ed)*. Bandung: ITB.
- Soemarti. 2007. *Pencilan (Outlier)*. Bandung: Universitas Padjadjaran wordpress.
- Sukestiyarno. 2013. *Olah Data Penelitian Berbantuan SPSS (4th ed.)*. Semarang: Universitas Negeri Semarang.
- Yohai, V. J. 1987. *Hight Breakdown Point and Hight Efficiency Robust Estimates for Regression*. The Annals of Statistics.