J M S K
Jurnal Matematika, Statistika, & Komputasi
http://journal.unhas.ac.id/index.php/jmsk
Vol. 15, No. 1, 88-96, Juli 2018

p-ISSN: 1858-1382
e-ISSN: 2614-8811

Perbandingan Penduga M, S, dan MM pada Regresi Linier dalam Menangani Keberadaan *Outlier*

Hanifah Lainun[1], Georgina M Tinungki[2], Amran[3]

Abstrak

Metode Kuadrat Terkecil (MKT) merupakan metode penduga parameter yang paling banyak digunakan pada analisis regresi. MKT merupakan metode penduga parameter tak bias yang baik selama asumsi komponen galatnya terpenuhi. Namun dalam aplikasinya sering ditemui terjadinya pelanggaran asumsi. Diantaranya, pelanggaran asumsi galat berdistribusi normal disebabkan adanya *outlier* pada data amatan. Oleh karena itu, dibutuhkan suatu metode yang kekar terhadap keberadaan *outlier*. Metode pendugaan parameter yang kekar terhadap keberadaan *outlier* pada regresi linier diantaranya ialah penduga M, penduga S, dan penduga MM yang masing-masing memiliki keunggulan dari segi efisiensi dan *breakdown point* yang tinggi. Tujuan dari penelitian ini adalah untuk membandingkan penduga M, S, dan MM dalam menduga parameter regresi pada analisis regresi linier sederhana terhadap keberadaan *outlier* menggunakan data simulasi. Simulasi dilakukan untuk ukuran sampel yang berbeda (20, 60, dan 120) ketika terdapat 20% dan 45% *outlier* pada variabel bebas dan variabel terikat. Metode terbaik ialah metode dengan *Standard Error* (SE) dan *Mean Square Error* (MSE) terkecil. Hasil yang diperoleh menunjukkan bahwa penduga MM lebih baik dibandingkan penduga M dan penduga S.

Kata Kunci: Regresi *Robust*, Penduga M, Penduga S, Penduga MM, *Outlier*

1. Pendahuluan

Analisis regresi merupakan suatu metode statistika yang digunakan untuk mengkaji hubungan antara variabel terikat dan satu atau lebih variabel bebas. Metode yang banyak digunakan untuk mengestimasi parameter model regresi ialah Metode Kuadrat Terkecil (MKT) yang meminimumkan jumlah kuadrat galat. Disamping kemudahan dalam komputasinya, MKT merupakan penduga tak bias yang baik selama asumsi komponen galat (e_i) dalam model yang diberikan terpenuhi. Namun dalam aplikasinya seringkali ditemui terjadi pelanggaran asumsi. Diantaranya ialah terlanggarnya asumsi galat berdistribusi $N{\sim}(0,\sigma^2)$ yang disebabkan adanya *outlier* pada data amatan sehingga diperlukan suatu metode yang kekar terhadap keberadaan *outlier*.

Regresi *robust* adalah metode regresi yang digunakan ketika data pengamatan mengandung *outlier* yang mempengaruhi model. Metode ini penting untuk menganalisis data yang mengandung *outlier* agar model yang dihasilkan kekar terhadap keberadaan *outlier* [6]. Kekekaran suatu penduga dapat diukur dari *breakdown point* dan efesiensi penduga. Metode penduga parameter yang *robust* terhadap keberadaan outlier pada regresi linier diantaranya ialah penduga M yang mana efisiensinya dapat mencapai hingga 95%, penduga S dengan *breakdown point* yang dapat mencapai hingga 50%, dan penduga MM yang merupakan kombinasi dari efisiensi dan *breakdown point* yang tinggi.

Penelitian yang mengkaji mengenai penggunaan regresi robust pada data yang mengandung *outlier* ialah Gad dan Qura [9] dan Cankaya dan Abaer [5] yang membandingkan beberapa metode *robust* dan MKT apabila terdapat *outlier* pada variabel terikat (y). Penelitian ini akan membandingkan

hanifahlainun311@gmail.com[1], *ina_matematika@yahoo.co.id*[2], *amranihsani@gmail.com*[3]
[1,2,3]*Program Studi Statistika, Departemen Matematika, FMIPA, Universitas Hasanuddin*

penduga M, penduga S, dan penduga MM ketika data amatan mengandung 20% dan 45% *outlier* pada variabel terikat, *outlier* pada variabel bebas *(good leverage point)*, dan *outlier* pada variabel bebas dan variabel terikat *(bad leverage point)* untuk tiga ukuran sampel berbeda pada regresi linier sederhana.

2. Tinjuan Pustaka

2.1 Analisis Regresi

Model regresi linier secara umum dapat dinyatakan sebagai

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

dengan,

y_i	=	Variabel terikat pada pengamatan ke-i
x_{ij}	=	Variabel bebas pada pengamatan ke-i variabel ke j *(j=1,2,…,p)*
β_0	=	Konstanta (Nilai y ketika $x_1, x_2, \dots, x_p = 0$)
β_j	=	Koefisien regresi *(j=1,2,…,p)*
e	=	Galat

Persamaan (1) dapat dituliskan kedalam notasi matriks sebagaimana pada persamaan (2).

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e} \tag{2}$$

Metode Kuadrat Terkecil (MKT) adalah suatu metode penaksiran parameter pada analisis regresi linier yang paling banyak digunakan dengan tujuan untuk meminimumkan jumlah kuadrat galat.

$$min \sum_{i=1}^{n} e_i{}^2 \tag{3}$$

Persamaan normal untuk memperoleh nilai penduga β menggunakan metode kuadrat terkecil adalah dengan memperoleh turunan parsial dari persamaan (3) terhadap β dan menyamakannya dengan nol.

$$\frac{\partial}{\partial \boldsymbol{\beta}}[\boldsymbol{e}'\boldsymbol{e}] = \frac{\partial}{\partial \boldsymbol{\beta}}[(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})] = 0$$
$$\frac{\partial}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}}\ [\boldsymbol{y}'\boldsymbol{y} - 2(\boldsymbol{X}'\boldsymbol{y})\boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})\boldsymbol{\beta}^2] = 0$$
$$-2\boldsymbol{X}'\boldsymbol{y} + 2(\boldsymbol{X}'\boldsymbol{X})\widehat{\boldsymbol{\beta}} = 0$$
$$(\boldsymbol{X}'\boldsymbol{X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$$
$$(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{X})\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$
$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

2.2 *Outlier*

Salah satu langkah awal untuk mendapatkan analisis yang tepat ialah mendeteksi keberadaan *outlier*. Johnson (1992) mendefinisikan *outlier* sebagai pengamatan dalam kumpulan data yang tampaknya tidak konsisten dengan data lainnya. Penghapusan *outlier* dari kumpulan data yang dianalisis dapat secara dramatis mempengaruhi kinerja model regresi [3]. *Outlier* harus dihilangkan jika ada alasan yang tepat. Alternatif lainnya ialah menggunakan regresi *robust*, yang mana algoritmanya memberi bobot lebih sedikit pada *outlier* namun tidak membuangnya [2]. Chen [6] menyebutkan bahwa secara historis, tiga jenis permasalahan *outlier* yang ditangani dengan regresi *robust*: 1) *Outlier* di variabel terikat (y), 2) *Outlier* di variabel bebas (x) yang juga disebut dengan *leverage point* dan 3) *Outlier* di variabel terikat (y) dan variabel bebas (x).

Ada banyak kesulitan dalam proses identifikasi *outlier*. Bengal [3] menyebutkan bahwa data yang mengandung beberapa *outlier* atau *outlier* yang bergerombol rentan terhadap efek tumpang tindih (*masking and swamping effect*). Uji statistik yang dapat digunakan untuk mengidentifikasi keberadaan *multiple outlier* ialah uji Hampel [8]. Uji ini menggunakan median dan MAD yang mana lebih kekar terhadap keberadaan *outlier* dibandingkan rata-rata dan standar deviasi. Pada uji hampel, pengamatan x_i teridentifikasi sebagai *outlier* apabila

$$x_i - median(x_i) > 4.5\ MAD(x_i) \tag{4}$$

2.3 Metode *Robust*

Metode Kuadrat Terkecil dengan n observasi, untuk model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}$ dengan p parameter mengharuskan terpenuhinya beberapa asumsi agar hasil yang diperoleh dapat mewakili populasi yang sebenarnya. Dalam aplikasinya, seringkali ditemui terjadi pelanggaran asumsi tersebut. Diantaranya ialah terlanggarnya asumsi galat berdistribusi $N \sim (0, \sigma^2)$ yang disebabkan adanya *outlier* pada data amatan. Regresi robust adalah metode regresi yang digunakan ketika terdapat beberapa *outlier* yang mempengaruhi model. Agar dapat mencapai hal ini regresi *robust* membatasi pengaruh dari *outlier* [6]. Selain itu, diwaktu yang bersamaan metode ini dapat digunakan untuk mengidentifikasi data mana yang merupakan *outlier*.

Andersen [1] menyebutkan dua kriteria kekekaran suatu penduga yang patut diperhitungkan yaitu *breakdown point* dan efesiensi penduga. *Breakdown point* adalah proporsi atau persentase data yang terkontaminasi yang masih dapat ditangani suatu penduga sebelum menghasilkan dugaan yang keliru. Sementara itu, suatu penduga dapat dikatakan efisien jika penduga tersebut yang paling akurat dibandingkan penduga tak bias lainnya.

2.3.1 Penduga M

Penduga M dalam konteks regresi pertama kali diperkenalkan oleh Huber (1973) sebagai hasil pendekatan dari penduga kuadrat terkecil yang *robust*. Meskipun penduga M tidak dapat menaksir parameter dengan baik apabila terdapat *leverage point*, penduga M populer ketika dalam aplikasinya *leverage point* bukanlah suatu masalah [10]. Penduga M meminimumkan fungsi objektif dari galat:

$$\widehat{\boldsymbol{\beta}}_m = \min \sum_{i=1}^{n} \rho(\frac{y_i - \boldsymbol{x}_i{}'\boldsymbol{\beta}}{\sigma})$$

Langkah-langkah yang dilakukan dalam mengestimasi parameter dengan penduga M adalah:

1. Menaksir **β** awal yaitu $\widehat{\boldsymbol{\beta}}_{(0)}$ dengan menggunakan metode kuadrat terkecil.
2. Menghitung nilai residual $e_i = y_i - \hat{y}_i$.
3. Menghitung nilai
$$\hat{\sigma}_i = \frac{median|e_i - median(e_i)|}{0.6745}$$
4. Menghitung nilai $u_i = e_i / \hat{\sigma}_i$
5. Menghitung nilai boboti (W_i) menggunakan fungsi pembobot Tukey's *bisquare* dengan nilai *tuning constant* $c = 4.685$ sehingga diperoleh efisiensi sebesar 95%.
$$W(u,c) = \begin{cases} \left(1 - \dfrac{u^2}{4.685^2}\right)^2 & , \ jika\ |u| < 4.685 \\ 0 & , \ lainnya \end{cases}$$
6. Menghitung $\widehat{\boldsymbol{\beta}}_M$ menggunakan kuadrat terkecil terboboti berdasarkan nilai bobot W_i.
$$\widehat{\boldsymbol{\beta}}_M = (\boldsymbol{X'WX})^{-1}\boldsymbol{X'Wy}$$
7. Mengulangi langkah 2-6 sampai diperoleh nilai $\widehat{\boldsymbol{\beta}}_M$ yang konvergen.

2.3.2 Penduga S

Penduga S pertama kali diperkenalkan oleh Rousseeuw dan Yohai (1984) dengan *breakdown point* yang dapat mencapai hingga 50%. Dengan cara yang sama metode kuadrat terkecil meminimumkan ragam dari galat, penduga S meminimumkan dispersi dari galat,

$$\widehat{\boldsymbol{\beta}}_s = \min \sigma(\boldsymbol{\beta})$$

dimana dispersi $\sigma(\beta)$ adalah solusi dari

$$\frac{1}{n}\sum_{i=1}^{n}\rho(\frac{y_i-\boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma})=K$$

dengan K adalah nilai konstan yang tepat untuk memastikan kekonsistenan.

Langkah-langkah yang dilakukan dalam mengestimasi parameter dengan penduga S adalah:

1. Menaksir **β** awal yaitu $\widehat{\boldsymbol{\beta}}_{(0)}$ dengan menggunakan metode kuadrat terkecil.
2. Menghitung nilai residual $e_i = y_i - \hat{y}_i$.
3. Menghitung nilai

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

4. Menghitung nilai $u_i = e_i/\hat{\sigma}_i$
5. Menghitung nilai boboti (W_i) menggunakan fungsi pembobot Tukey's *bisquare* dengan nilai *tuning constant* $c = 1.547$ sehingga diperoleh *breakdown point* sebesar 50%.

$$W(u,c) = \begin{cases} \left(\left(1-\dfrac{u^2}{1.547^2}\right)\right)^2, & jika\ |u| < 1.547 \\ 0, & lainnya \end{cases}$$

6. Menghitung $\hat{\beta}_s$ menggunakan kuadrat terkecil terboboti berdasarkan nilai bobot W_i.

$$\widehat{\boldsymbol{\beta}}_s = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{y}$$

7. Mengulangi langkah 2-6 sampai diperoleh nilai $\widehat{\boldsymbol{\beta}}_s$ yang konvergen.

2.3.3 Penduga MM

Penduga MM adalah prosedur untuk menaksir parameter regresi menggunakan penduga S yang meminimumkan *scale* dari galat pada penduga M dengan *breakdown point* hingga 50%, kemudian melanjutkan proses dengan penduga M yang efisien. Penduga MM adalah solusi dari

$$\widehat{\boldsymbol{\beta}}_{mm} = \min\sum_{i=1}^{n}\rho(\frac{y_i-\boldsymbol{X}_i{}'\boldsymbol{\beta}}{\sigma})$$

Langkah-langkah yang dilakukan dalam mengestimasi parameter dengan penduga MM adalah:

1. Menaksir **β** awal yaitu $\widehat{\boldsymbol{\beta}}_{(0)}$ dengan menggunakan penduga S sehingga diperoleh *breakdown point* sebesar 50%.
2. Menghitung nilai residual $e_i = y_i - \hat{y}_i$.
3. Menghitung nilai

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

4. Menghitung nilai $u_i = e_i/\hat{\sigma}_i$
5. Menghitung nilai boboti (W_i) menggunakan fungsi pembobot Tukey's *bisquare* dengan nilai *tuning constant* $c = 4.685$ sehingga diperoleh efisiensi sebesar 95%.

$$W(u,c) = \begin{cases} \left(\left(1-\dfrac{u^2}{4.685^2}\right)\right)^2, & jika\ |u| < 4.685 \\ 0, & lainnya \end{cases}$$

6. Menghitung $\widehat{\boldsymbol{\beta}}_{MM}$ menggunakan kuadrat terkecil terboboti berdasarkan nilai bobot W_i.

$$\widehat{\boldsymbol{\beta}}_{MM} = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{y}$$

7. Mengulangi langkah 2-6 sampai diperoleh nilai $\widehat{\boldsymbol{\beta}}_{MM}$ yang konvergen.

3. Data Simulasi

Prosedur simulasi yang dilakukan adalah menggunakan *Model Based Simulation* dengan algoritma sebagai berikut.

1. Menentukan parameter bagi populasi yaitu $\beta_0 = \beta_1 = 4$.
2. Membangkitkan e_i, X_i, dan y_i untuk $i = 1,2,\ldots,N$ menggunakan softwere SAS 9.2 dengan ketentuan sebagaimana pada Tabel 3.1.

Tabel 3. 1 Kriteria Data Simulasi

	Data 'normal'	***Outlier* di y**	***Outlier* di x**	***Outlier* di x dan y**
N	1250	250	250	250
e_i	$N(0,1)$	$N(0,1)$	$N(0,1)$	$N(0,1)$
x_i	$N(5,2)$	$N(5,2)$	$N(100,20)$	$N(100,20)$
y_i	$\beta_0 + \beta_1 X_i + e_i$	$200 + e_i$	$\beta_0 + \beta_1 X_i + e_i$	$200 + e_i$

3. Mengambil sampel secara acak untuk $n = 20$, $n = 60$, dan $n = 120$. Adapun persentase data 'normal' dan data *outlier* (baik pada variabel bebas (x), variabel terikat (y), dan variabel bebas (x) dan variabel terikat (y)) ditetapkan: 1) 80% dan 20% dari banyaknya sampel dan 2) 55% dan 45% dari banyaknya sampel.
4. Melakukan penaksiran parameter menggunakan metode kuadrat terkecil (MKT), penduga M, penduga S, dan penduga MM.
5. Menghitung *Mean Square Error (MSE)* berdasarkan model regresi yang diperoleh pada MKT, penduga M, penduga S dan penduga MM.

$$MSE = (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

dengan $\widehat{\boldsymbol{\beta}}$ adalah vektor parameter dugaan dan $\boldsymbol{\beta}$ merupakan vektor koefisien model yang sebenarnya.

6. Mengulangi langkah 3-5 sebanyak lima kali.
7. Menentukan metode yang paling baik yaitu metode dengan MSE terkecil.

4. Hasil dan Pembahasan

Sebelum melakukan penaksiran parameter, identifikasi outlier dilakukan pada data simulasi untuk tiga jumlah sampel dan tiga jenis *outlier* yang berbeda untuk memastikan bahwa data yang diperoleh telah sesuai dengan tujuan penelitian. Statistik uji yang digunakan untuk melihat pengamatan keberapa yang teridentifikasi sebagai *outlier* ialah uji hampel pada Persamaan (4). Hasil yang diperoleh menunjukkan bahwa 20% dan 45% data terakhir dari sampel teridentifiaksi sebagai *outlier*.

Pada berbagai ukuran sampel dan jenis *outlier* yang berbeda diperoleh nilai *Mean Square Error* (MSE) ketika menggunakan MKT, penduga M, penduga S, dan penduga MM. Berdasarkan hasil yang diperoleh dapat diambil kesimpulan sebagai berikut.

Skenario 1: 20% *outlier* di variabel terikat y

Hasil dugaan parameter MKT adalah yang terburuk berdasarkan MSE yang diperoleh. Selain itu, MKT juga memiliki *Standard Error* (SE) yang terbesar terutama pada sampel kecil yaitu *n=20* dibanding metode lainnya. Hal ini disebabkan model yang diperoleh dari MKT bias terhadap keberadaan *outlier* menyebabkan interval kepercayaan lebih lebar. Secara keseluruhan, SE semakin kecil seiring bertambah besarnya ukuran sampel. Sementara itu, metode *robust* tetap tidak dipengaruhi oleh keberadaan *outlier* dengan memberikan bobot yang kecil pada data yang diidentifikasi sebagai *outlier*. Diantara ketiga metode *robust*, penduga MM memiliki MSE yang paling minimum yang kemudian disusul oleh penguga M, dan terakhir penduga S.

Skenario 2: 45% *outlier* di variabel terikat y

Diperoleh bahwa penduga M menghasilkan dugaan yang keliru sebagaimana MKT pada berbagai ukuran sampel. Hal ini disebabkan pada penduga M, semakin tinggi efisiensinya maka kemampuan dalam mendeteksi *outlier* akan berkurang. Pada penelitian ditetapkan nilai c=4.685 yang menghasilkan efisiensi sebesar 95%. Oleh karena itu, disarankan untuk menggunakan nilai *tuning constat c* yang lebih kecil apabila proporsi dari *outlier* besar dan mendekati 50% sehingga kemampuan penduga M dalam mendeteksi *outlier* lebih baik meski efisiensinya berkurang. Sementara itu, penduga S dan penduga MM tetap dapat mengestimasi parameter β dengan baik dengan penduga MM menghasilkan MSE yang lebih kecil dibandingkan penduga S.

Skenario 3: 20% *outlier* di variabel bebas x

MKT merupakan metode terbaik dengan MSE paling minimum yang kemudian disusul oleh penduga MM, peduga M, dan terakhir penduga S.

Skenario 4: 45% *outlier* di variabel bebas x

Hasil yang diperoleh sama halnya ketika tedapat 20% outlier pada variabel bebas. MKT merupakan metode terbaik dengan MSE paling minimum yang kemudian disusul oleh penduga MM, peduga M, dan terakhir penduga S.

Skenario 5: 20% *outlier* di variabel terikat y dan variabel bebas x

Diperoleh bahwa penduga M menghasilkan dugaan yang keliru sebagaimana MKT pada berbagai ukuran sampel. Dugaan parameter yang diperoleh bias terhadap keberadaan *outlier* menyebabkan nilai MSE yang besar. Hal ini disebabkan karena *bad leverage* point memberi pengaruh yang besar terhadap dugaan kemiringan garis regresi apabila metode dengan *breakdown point* yang rendah digunakan. Sementara itu, penduga S dan penduga MM tetap dapat mengestimasi parameter regresi dengan baik dengan penduga MM memiliki MSE yang lebih kecil dibandingkan penduga S.

Skenario 6: 45% *outlier* di variabel terikat y dan variabel bebas x

Kinerja dari keempat metode menurun untuk $n = 20$ dimana MSE yang dihasilkan lebih besar dibadingkan MSE yang diperoleh pada kondisi *outlier* lainnya. Sementara itu, untuk $n = 60$ dan $n = 60$ hasil yang diperoleh sama halnya ketika terdapat *outlier* di variabel terikat dan variabel bebas sebanyak 20%. MKT dan penduga M menghasilkan dugaan parameter yang keliru sehingga diperoleh nilai MSE yang besar, sementara penduga S dan penduga MM tetap dapat mengestimasi parameter regresi dengan baik dengan penduga MM memiliki MSE yang lebih kecil dibandingkan penduga S.

5. Kesimpulan

Pada penelitian ini berdasarkan MSE yang diperoleh, penduga MM lebih baik dibandingkan penduga M dan penduga S ketika terdapat *outlier* pada variabel terikat dan *outlier* ada variabel terikat dan variabel bebas. Namun, kelemahan dari penduga MM ialah ketika ukuran sampel kecil dan proporsi *outlier* besar yaitu mendekati nilai *breakdown point*. Adapun ketika terdapat *outlier* pada variabel bebas, MKT tetap dapat menaksir parameter regresi dengan baik, bahkan dapat meningkatkan keakuratan koefisien regresi.

Daftar Pustaka

[1] Andersen, R., 2008. *Modern Method for Robust Regression*. London: Sage Publication.

[2] Anonymous. tahun. *Outliers*. (http://web.sonoma.edu/users/c/cuellar/econ317/Outliers.pdf diakses pada taggal 27 September 2017)

[3] Ben-Gal, Irad. 2005. Outlier detection. *Data mining and knowledge discovery handbook*, 131-146.

[4] Blatná, Dagmar. 2006. Outliers in regression. *Trutnov*, *30*, 2006-03.

[5] Cankaya, Soner dan Samet Hasan Abaci. 2015. Comparative Study of Some Estimation Methods in Simple Linier Regression Model for Different Sample Sizes in Presence of Outliers. *Turkish Journal of Agliculture – Food Science and Thechnology*, 3(6): 380-386.

[6] Chen, Colin. 2002. Paper 256-27 Robust Regression and Outlier Detection with the ROBUSTREG Procedure. In *Proceedings of the Proceedings of the Twenty-Seventh Annual SAS Users Group International Conference.*

[7] Draper, N. R. dan K. Smith (1998). *Applied Regression Analysis. Third edition*. New York: Wiley.

[8] Farazi, Manzur Rahman. 2015. Identification of Outlier in Gene Expression Data [Thesis]. Indiana: Ball State University.

[9] Gad, Ahmed M. dan Maha E. Qura. 2016. Regression Estimation in Presence of Outliers: A Comparative Study. *International Journal of Probability and Statistics 2016*, 5(3): 65-72.

[10] SAS Institute Inc, 2009. SAS/STAT ® 9.2, User's Guide, Chapter 74: The ROBUSTREG Procedure, Second Edition, Cary, NC: USA

[11] Stuart, Chaterine. 2011. Robust regression. *Department of Mathematical Sciences, Durham Universit*y, 169.

Lampiran: Rata-rata MSE pada berbagai kondisi

Ulangan	20% *outlier* pada variabel terikat											
	n20				n60				n120			
	ols	m	s	mm	ols	m	s	mm	ols	m	s	mm
1	349.072	0.608	0.288	0.587	3926.442	0.264	0.782	0.279	1814.770	0.001	0.003	0.001
2	2794.541	0.411	0.686	0.410	479.173	0.015	0.080	0.011	407.131	0.001	0.011	0.001
3	1748.917	1.815	3.833	1.848	5725.855	0.568	0.567	0.546	46.731	0.000	0.030	0.000
4	5020.405	0.007	0.066	0.008	2197.514	0.029	0.090	0.030	607.483	0.006	0.261	0.001
5	3851.017	0.548	0.045	0.508	2243.063	0.069	0.323	0.081	356.067	0.002	0.001	0.002
	2752.791	0.678	0.983	0.672	2914.410	0.189	0.368	0.189	646.437	0.002	0.061	0.001
MSE min	**Penduga MM**				**Penduga M**				**Penduga MM**			

Ulangan	45% *outlier* pada variabel terikat											
	n20				n60				n120			
	ols	m	s	mm	ols	m	s	mm	ols	m	s	mm
1	3543.169	2957.329	0.186	0.188	6553.030	6088.962	0.238	0.213	4978.162	4509.475	0.004	0.002
2	16992.020	17496.020	0.535	0.532	7837.398	7622.011	0.011	0.009	6176.370	5785.087	0.000	0.004
3	11171.837	11245.498	0.513	0.512	18619.193	19424.090	0.009	0.005	6053.223	5687.225	0.386	0.343
4	11250.259	11358.500	4.452	4.458	3986.085	3489.459	0.012	0.012	6735.370	6370.205	0.022	0.021
5	992067.783	9084.900	2.537	2.551	1397.869	1021.645	2.267	2.259	9875.117	9870.007	0.013	0.013
	207005.014	10428.449	1.645	1.648	7678.715	7529.233	0.507	0.500	6763.648	6444.400	0.085	0.077
MSE min	**Penduga S**				**Penduga MM**				**Penduga MM**			

Ulangan	20% *outlier* pada variabel bebas											
	n20				n60				n120			
	ols	m	s	mm	ols	m	s	mm	ols	m	s	mm
1	0.000	0.002	0.034	0.005	0.003	0.002	0.048	0.000	0.003	0.005	0.000	0.008
2	0.168	0.452	0.603	0.494	0.127	0.124	0.300	0.122	0.021	0.016	0.002	0.012
3	0.106	0.385	0.884	0.208	0.003	0.006	0.226	0.015	0.006	0.006	0.004	0.005
4	0.269	0.378	0.679	0.456	0.017	0.012	0.051	0.006	0.001	0.000	0.105	0.001
5	0.009	0.013	0.081	0.020	0.002	0.002	0.002	0.002	0.000	0.000	0.031	0.000
MSE min	**OLS**				**Penduga M**				**Penduga MM**			

Ulangan	45% *outlier* pada variabel bebas											
	n20				n60				n120			
	ols	m	s	mm	ols	m	s	mm	ols	m	s	mm
1	0.040	0.080	0.083	0.118	0.008	0.014	0.100	0.022	0.002	0.002	0.012	0.003
2	0.018	0.027	0.240	0.047	0.022	0.033	0.232	0.042	0.012	0.032	0.476	0.072
3	0.025	0.015	0.257	0.016	0.029	0.022	0.015	0.016	0.001	0.001	0.080	0.005
4	0.266	0.254	0.000	0.168	0.156	0.151	0.400	0.147	0.013	0.006	0.067	0.003

5	0.070	0.188	0.655	0.295	0.005	0.009	0.041	0.012	0.007	0.008	0.138	0.010
	0.084	0.113	0.247	0.129	0.044	0.046	0.158	0.048	0.007	0.010	0.155	0.019
MSE min	**OLS**				**OLS**				**OLS**			

Ulangan	**20% *outlier* pada variabel terikat dan variabel bebas**											
	n20				n60				n120			
	ols	m	s	mm	ols	m	s	mm	ols	m	s	mm
1	159.475	121.265	0.288	0.587	135.952	102.662	0.782	0.279	170.330	135.282	0.003	0.001
2	90.433	92.182	0.686	0.410	178.942	155.108	0.080	0.011	172.846	132.627	0.011	0.001
3	150.780	124.925	3.833	1.848	182.404	137.586	0.567	0.546	173.728	124.671	0.030	0.000
4	283.296	191.565	0.066	0.008	210.583	189.847	0.090	0.030	183.717	120.706	0.261	0.001
5	204.802	171.381	0.045	0.508	172.269	150.550	0.323	0.081	159.828	107.239	0.001	0.003
	177.757	140.263	0.983	0.672	176.030	147.151	0.368	0.189	172.090	124.105	0.061	0.001
MSE min	**Penduga MM**				**Penduga MM**				**Penduga MM**			

Ulangan	**45% *outlier* pada variabel terikat dan variabel bebas**											
	n20				n60				n120			
	ols	m	s	mm	ols	m	s	mm	ols	m	s	mm
1	287.449	166.349	49.108	56.195	321.302	157.773	0.238	0.213	247.508	137.187	0.004	0.002
2	203.387	101.055	77.236	95.444	349.387	133.489	0.011	0.009	289.996	160.894	0.000	0.004
3	292.252	159.141	34.865	35.681	221.549	135.718	0.009	0.005	324.248	103.477	0.386	0.343
4	175.073	127.158	69.325	70.127	402.687	176.612	0.012	0.011	232.601	123.298	0.022	0.021
5	243.136	124.362	125.467	124.833	338.189	88.891	15.457	15.417	349.972	154.534	0.013	0.013
	240.259	135.613	71.200	76.456	326.623	138.496	3.145	3.131	288.865	135.878	0.085	0.077
MSE min	**Penduga S**				**Penduga MM**				**Penduga MM**			