E-Jurnal Matematika Vol. 12(2), Mei 2023, pp. 132-139 ISSN: 2303-1751
DOI: https://doi.org/10.24843/MTK.2023.v12.i02.p410

PENDUGAAN PARAMETER REGRESI *ROBUST* METODE *MINIMUM COVARIANCE DETERMINANT* DAN METODE TELBS

Ni Ketut Zelina Yeriska[1§], I Gusti Ayu Made Srinadi[2], I Komang Gde Sukarsa[3]

[1]Program Studi Matematika Fakultas MIPA – Universitas Udayana [Email:zelinayeriskaa@gmail.com]
[2]Program StudiMatematika Fakultas MIPA – Universitas Udayana [Email:gedesukarsa@unud.ac.id]
[3]Program Studi Matematika Fakultas MIPA – Universitas Udayana [Email:srinadi@unud.ac.id]
[§]*Corresponding Author*

ABSTRACT

The parameter estimator on the regression model can be obtained through the ordinary least square (OLS). When there are outliers in the data, OLS cannot be applied because it will produce an unbiased estimator that is not the best linear estimator. Another alternative to addressing the presence of outlier data without deleting the data is robust regression. Robust regression methods include the minimum covariance determinant (MCD) and the TELBS method. This study aims to determine the estimation of regression parameters produced using the MCD and TELBS methods when entering outlier data. The data used are simulation data with various levels of outliers, namely 5%, 10%, and 20%. The outliers inserted are the outliers on variable X, variable Y, and variables X and Y. The result of this study is that the robust regression methods of MCD and TELBS both produce unbiased parameter estimators when there are outlier data.

Keywords*: Robust regression, Outlier, MCD, TELBS*

1. PENDAHULUAN

Analisis regresi merupakan alat statistika yang digunakan untuk menduga atau memprediksi nilai peubah terikat berdasarkan nilai-nilai peubah bebasnya (Harlan, 2018). Selain itu, analisis regresi juga dapat digunakan untuk mengetahui seberapa besar pengaruh satu atau beberapa peubah bebas terhadap peubah terikat. Model regresi yang terdiri dari satu peubah bebas dan satu peubah terikat disebut model regresi linear sederhana (Harlan, 2018). Salah satu metode yang sering dipakai untuk mengestimasi parameter-parameter pada model regresi linear adalah metode kuadrat terkecil (MKT). Estimator yang diperoleh dengan MKT adalah *best linear unbiased estimator*, apabila terpenuhi asumsi-asumsi dalam analisis regresi. Beberapa asumsi yang harus dipenuhi diantaranya: residual mengikuti distribusi normal; varians dari residual adalah konstan; tidak ada autokorelasi; dan tidak ada multikolinearitas di antara peubah bebas.

Adanya pencilan atau *outlier* pada data dapat menyebabkan tidak terpenuhinya asumsi klasik (Febrianto dkk., 2018). Metode regresi *robust* dapat mengatasi pengaruh pencilan dengan tetap memakai seluruh data namun pada data pencilan diberikan bobot yang kecil. Terdapat beberapa metode pendugaan parameter dalam regresi *robust* yaitu, metode *minimum covariance determinant* (MCD) dan metode TELBS.

MCD adalah metode regresi *robust* yang dapat mengatasi pencilan pada peubah bebas (X) dan peubah terikat (Y), mendeteksi seluruh pencilan serta memberikan proses yang relatif cepat. Prinsip dari metode MCD yaitu menggunakan vektor rata-rata dan matriks kovarians yang diperoleh untuk menentukan bobot dari setiap data, sehingga diperoleh penduga parameter model MCD (Hubert et al., 2018).

Metode regresi *robust* lainnya yaitu metode TELBS. Serupa dengan metode MCD, metode TELBS juga dapat mengatasi adanya pencilan pada peubah bebas (X) dan peubah terikat (Y) (Gusriani & Firdaniza, 2018). Menurut Tabatabai (2012) metode TELBS dilakukan dengan meminimumkan fungsi objektif.

Penelitian ini membandingkan model yang dihasilkan oleh regresi *robust* metode MCD dan

TELBS. Penduga selang akan menjadi acuan untuk melihat ketakbiasan penduga yang dihasilkan oleh kedua metode dan nilai *mean square error* (MSE) digunakan untuk melihat metode yang terbaik.

2. METODE PENELITIAN

2.1. Sumber Data

Data yang digunakan dalam penelitian ini adalah data simulasi berdistribusi normal. Data dibangkitkan menggunakan bantuan program R 4.1.2. dengan ukuran sampel *20* dan 100. Data yang dibangkitkan merupakan data regresi linear sederhana yaitu satu peubah bebas (X) dan satu peubah terikat (Y).

2.2. Metode Analisis data

a. Membangkitkan data

Membangkitkan data simulasi dengan ketentuan, peubah bebas (X) adalah data berdistribusi normal dengan rataan nol dan simpangan baku lima. Nilai sisaan juga merupakan data berdistribusi normal dengan rataan nol dan ragam satu.

b. Memasukkan Pencilan

Memasukkan pencilan yang dilakukan secara bertahap pada peubah (X), peubah (Y), dan keduanya sebanyak 5%, 10%, dan 20% (Irfagautami et al., 2014). Nilai pencilan yang digunakan dalam penelitian ini diperoleh dengan cara menambahkan nilai maksimum dari peubah bebas (X) dan atau peubah terikat (Y) dengan lebih dari tiga kali simpangan baku yang bersesuaian (Atmagenta, 2016). Sedangkan untuk memperoleh nilai dari peubah terikat (Y) terlebih dahulu menentukan bentuk hubungannya yaitu $Y = \beta_0 + \beta_1 X + \varepsilon$. Nilai koefisien regresi yang digunakan adalah $\beta_0 = 1$ dan $\beta_1 = 2$. Berdasarkan nilai peubah bebas (X) dan nilai sisaan yang telah dibangkitkan serta koefesien regresinya maka dapat diperoleh nilai (Y) sesuai bentuk hubungannya.

c. Mendeteksi Pencilan

Mendeteksi pencilan dilakukan dengan metode grafis, yaitu: (1) Membuat *Scatter Plot* untuk melihat posisi pencilan; (2) Membuat *Box-Plot* untuk memeriksa kembali adanya data pencilan.

d. Uji Kenormalan

Uji Kenormalan dilakukan dengan terlebih dahulu melakukan analisis regresi antara masing-masing kelompok data yang mengandung pencilan dengan menggunakan MKT. MKT merupakan metode pendugaan parameter yang dilakukan dengan meminimumkan jumlah kuadrat galat (Lainun dkk., 2018).

Nilai-nilai pendugaan parameternya diperoleh sebagai berikut

$$b_0 = \frac{\sum_{i=1}^{n} Y_i}{n} - b_1 \frac{\sum_{i=1}^{n} X_i}{n}$$

$$b_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Atau dalam bentuk matriks dapat ditulis sebagai berikut:

$$b = (X^T X)^{-1} X^T Y$$

dengan

$$b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix};$$

$$X^T X = \begin{bmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} {X_i}^2 \end{bmatrix};$$

$$X^T Y = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_i Y_i \end{bmatrix}.$$

Pada penelitian ini digunakan uji Anderson-Darling, yang mana uji kenormalan dilakukan pada data sisaan untuk masing-masing kelompok data yang mengandung pencilan. Hipotesis yang digunakan adalah sebagai berikut:

H_0: sisaan mengikuti sebaran normal,

H_1: sisaan tidak mengikuti sebaran normal.

Misalkan data sisaan yang akan diuji diasumsikan berdistribusi normal dengan tingkat signifikan α, maka statistik uji yang digunakan adalah

$$A^2 = -n - S$$

dengan

$$S = \frac{1}{n}\sum_{i=1}^{n}[2i - 1][\ln(F(Z_i)) + \ln(1 - F(Z_{n+1-i}))]$$

dengan

$$Z_i = \frac{x_i - \bar{x}}{s}$$

Kriteria pengujian yang digunakan yaitu, jika $A^2 >$ nilai kritis uji Anderson-Darling maka tolak H_0 dan jika sebaliknya maka H_0 gagal ditolak. Apabila nilai *p-value* lebih kecil dari α maka tolak H_0, apabila sebaliknya, terima H_0 (Fallo dkk., 2013).

e. Regresi *robust* metode MCD

Pendugaan parameter metode MCD dilakukan dengan mencari h pengamatan yang minumum berdasarkan determinan matriks kovariannya.

E-Jurnal Matematika Vol. 12(2), Mei 2023, pp. 132-139 ISSN: 2303-1751
DOI: https://doi.org/10.24843/MTK.2023.v12.i02.p410

$$h = \frac{(n+p+1)}{2} \qquad (1)$$

dengan $n > p$, dimana n menyatakan banyak data, dan p banyak peubah.

Misalkan terdapat sampel acak

$$X^* = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix}^T$$

Dari persamaan (1) diperoleh kombinasi matriks subhimpunan dari matriks pengamatan X^* sebanyak a, dengan a didefinisikan sebagai berikut:

$$a = C_h^n \qquad (2)$$

Sehingga diperoleh subhimpunan matriks $\mathrm{H_b}$, dengan $b = 1, \dots, a$. Lalu, untuk setiap $\mathrm{H_b}$ sebut sebagai H_{bl} akan dihitung nilai dari vektor rata-rata (t_l) dan matriks kovarians (C_l) sebagai berikut:

$$t_l = \frac{1}{h}.(H_b)^T.V^* \qquad (3)$$

$$C_l = \frac{1}{h}(H_b - V^*(t_l)^T)^T((H_b - V^*(t_l)^T) \qquad (4)$$

dengan V^* merupakan matriks berukuran $h \times 1$, yakni:

$$V^* = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Berdasarkan persamaan (4) diperoleh nilai $det(C_l)$. Jika $\det(C_l) \neq 0$ maka dapat diketahui nilai jarak Mahalanobis dengan rumus:

$$d^2{}_{MD} = (X^* - V(\bar{X}^*{}_{RD})^T)C_{RD}^{-1}(X^* - V(\bar{X}^*{}_{RD})^T)^T \qquad (5)$$

Selanjutnya ambil sebanyak h nilai $d^2{}_{MD}$ yang paling minimum. Identifikasi nomor pengamatan terpilih, lalu buat matriks X^* yang baru sejumlah h, sesuai dengan identifikasi nomor pengamatan. Selanjutnya matriks X^* baru, disebut H_{bl} dengan $l = l + 1$.

Sesuai dengan bentuk persamaan (3) dan (4) didapat nilai t_l dan C_l yang baru, lalu dicari nilai determinannya.

Bandingkan nilai $\det(C_l)$ dan $\det(C_{l-1})$, jika:

1. $\det(C_l) \neq \det(C_{l-1})$, dengan memakai persamaan (5) diperoleh jarak Mahalanobis yang baru, lalu buat matriks X^* baru sehingga diperoleh matriks H_{bl} yang baru.
2. $\det(C_l) = \det(C_{l-1})$, perhitungan dikerjakan dari awal, dengan subhimpunan data H_b selanjutnya, sehingga diperoleh nilai det (H_b) lainnya.

Selanjutnya setelah diperoleh nilai determinan matriks kovarians dari subhimpunan data H_b, dipilih nilai determinan yang paling minimum. Nilai determinan minimum dari H_b disebut sebagai $H_{MCD.}$ Nilai t_l dan C_l yang diperoleh dari H_{MCD} disebut t_{MCD} dan C_{MCD}.

bobot w_{ii} dirumuskan dengan ketentuan berikut:

$$\left(w_{ii} = \begin{cases} 1, & Jika\; d^2{}_{MCD} \leq C \\ 0, & lainnya \end{cases} \right) \qquad (6)$$

dengan

$d^2{}_{MCD} = (X^* - t_{MCD})(C_{MCD})^{-1}(X^* - t_{MCD})^T$

$C = \chi^2_{\rho;\alpha}$, C merupakan nilai *cut-off*. Nilai ini dipakai untuk mendeteksi apakah suatu pengamatan pencilan atau bukan.

Berdasarkan persamaan (6) diperoleh matriks pembobot (W) berukuran $n \times n$ dengan entri matriks $w_{ij} = 0$, dimana $i \neq j$

Sehingga pendugaan parameter regresi MCD dimodelkan sebagai berikut:

$\hat{\beta}_{MCD} = (X^TWX)^{-1}(X^TWY)$

f. Regresi *robust* metode TELBS

Regresi *robust* estimasi TELBS dilakukan dengan meminimumkan fungsi objektif (Tabatabai *et al*., 2012):

$$min_\beta \sum_{i=1}^n \frac{\rho(t_i)}{L_i} = min_\beta \sum_{i=1}^n \frac{1 - sech(\omega t_i)}{L_i}$$

dengan:

$t_i = \frac{(y_i - x_i^T\hat{\beta})(1-h_{ii})}{\sigma}$,

$L_i = \sum_{j=1}^k max\,(M_{j,}|x_{ij}|)$,

$M_j = median\{|x_{1j}|, |x_{2j}|, \dots |x_{ij}|\}$,

dengan

k menyatakan banyak peubah bebas, $i = 1,2,\dots,n, j = 1,2,\dots,k$, dan ω merupakan bilangan real positif yang disebut konstanta kesesuaian. Konstanta kesesuaian ω bernilai 0,405; 0,525; 0,628; 0,721 masing-masing sesuai dengan tingkat kepercayaan 95%, 90%, 85%, 80% (Tabatabai et al., 2012).

Nilai estimator $\hat{\sigma}$ dapat diperoleh dari persamaan sebagai berikut:

$\hat{\sigma} = 1{,}1926\; median_{i;1\leq i\leq n}(median_{j;1\leq j\leq n}|e_i - e_j|)$

dengan e menyatakan sisaan (residual).

Agar suatu estimasi mendekati tak bias maka dipilih konstanta 1,1926. Untuk meminimumkan fungsi objektif, turunan dari ρ terhadap β_0 dan β_j disamakan dengan nol, sehingga menghasilkan persamaan:

$$\sum_{i=1}^n \frac{\psi(t_i)}{L_i}\frac{(1-h_{ii})}{\sigma}\frac{\partial(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik})}{\partial \beta_j} \qquad (7)$$

dengan

$\psi(x) = \frac{\partial\rho(x)}{\partial x} = \omega\; Sech\;(\omega x)\; Tanh\;(\omega x)$

Didefinisikan fungsi pembobot w_{ii} adalah:

$$w_{ii}^* = \frac{\psi(t_i)(1-h_{ii})}{\sigma e_i L_i} \qquad (8)$$

Maka persamaan (7) dapat ditulis menjadi:

$$\sum_{i=1}^{n} w_{ii}^* e_i \frac{\partial(y_i-\beta_0-\beta_1 x_{i1}-\beta_2 x_{i2}-\cdots-\beta_k x_{ik})}{\partial \beta_j}$$

Sehingga diperoleh pendugaan parameter metode TELBS yaitu:

$$\hat{\beta}_{TELBS} = (X^T W^* X)^{-1} X^T W^* Y$$

dengan W^* merupakan matriks berukuran $n \times n$ dengan elemen diagonalnya adalah w_{ii}^* pada persamaan (8) dan entri matriks $w_{ij}^* = 0, i \neq j$.

Pada metode TELBS nilai pendugaan parameternya tidak langsung diperoleh dalam sekali proses, tetapi dengan melakukan iterasi. Iterasi berhenti ketika kekonvergenan telah tercapai yaitu saat $\left\|\beta^{(t)} - \beta^{(t+1)}\right\| = 10^{-5}$.

2.3. Pemilihan Model Terbaik

Berdasarkan penelitian Marzuki dkk. (2010) selang kepercayaan $(1-\alpha)100\%$ untuk masing-masing parameter β_0 dan β_1 yang diambil dari Myers (1990) adalah

$$\beta_0 \pm t_{\alpha/2,v}\, s\sqrt{\frac{1}{n} + \frac{\bar{X}}{S_{XX}}}$$

$$\beta_1 \pm t_{\alpha/2,v}\, s\sqrt{\frac{1}{S_{XX}}}$$

dengan $S_{XX} = \sum(X_i - \bar{X})^2$, $t_{\alpha/2}$ adalah titik $\alpha/2$ persen pada distribusi$-t$, v adalah derajat bebas, dan s adalah simpangan baku galat.

Kebaikan model juga dapat dilihat dari nilai MSE. Nilai MSE diperoleh dari persamaan:

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

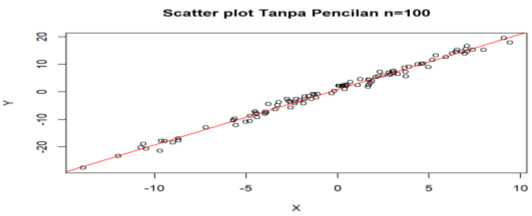
dengan $n =$ banyaknya sampel, $\hat{y}_i =$ nilai y dugaan ke$-i, y_i$ nilai y sebenarnya ke$-i$.

3. HASIL DAN PEMBAHASAN

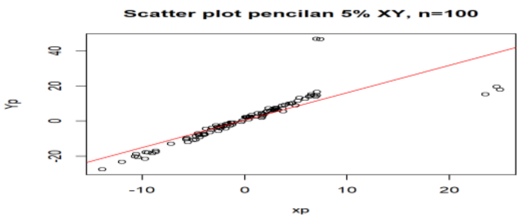
3.1 Mendeteksi Pencilan

a. *Scatter Plot*

Identifikasi pencilan perlu dilakukan untuk melihat keberadaan pencilan. Hal ini dilakukan dengan menggunakan *scatter plot*, yakni plot antara peubah bebas dan peubah terikat. Contoh *Scatter plot* antara peubah bebas dan peubah terikat tanpa pencilan dengan $n = 100$ dapat dilihat pada Gambar 1 dan *Scatter plot* antara peubah bebas dan peubah terikat yang mengandung pencilan pada peubah terikat (Y) sebesar 5% dengan $n = 100$ dapat dilihat pada Gambar 2.



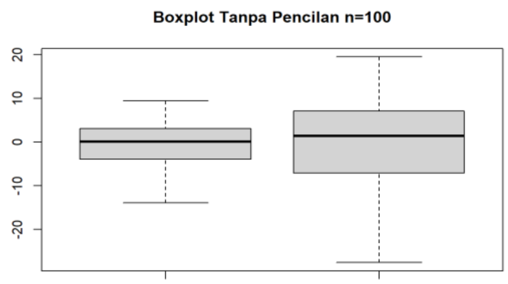
Gambar 1. *Scatter plot* tanpa pencilan



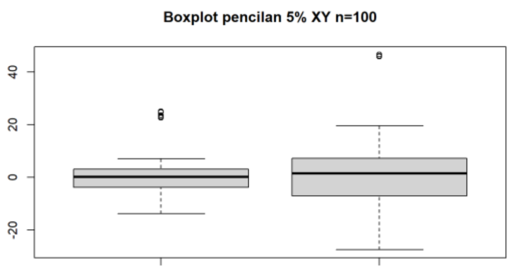
Gambar 2. *Scatter plot* pencilan peubah XY (5%)

b. *Box Plot*

Tahapan ini dimasukkan untuk memeriksa kembali apakah data tersebut mengandung pencilan atau tidak. Pencilan dalam *box plot* dilambangkan dengan tanda (°). Nilai yang akan dianalisis dengan menggunakan *box plot* adalah data awal dan data yang telah diberi pencilan pada peubah bebas (X) dan peubah terikat (Y).



Gambar 3. *Box Plot* tanpa pencilan



Gambar 3. *Box plot* pencilan peubah XY (5%)

Box plot data awal tanpa memasukkan pencilan tidak terdapat lambang (°) yang mengidentifikasikan bahwa data tersebut tidak mengandung pencilan. Sedangkan pada data yang dimasukkan pencilan (pencilan pada peubah X, peubah Y, dan keduanya) terdapat lambang (°) yang mengidentifikasikan bahwa data tersebut mengandung pencilan.

E-Jurnal Matematika Vol. 12(2), Mei 2023, pp. 132-139 ISSN: 2303-1751
DOI: https://doi.org/10.24843/MTK.2023.v12.i02.p410

3.2 Pemeriksaan Kenormalan Data

Setelah dilakukan identifikasi terhadap keberadaan pencilan pada masing-masing kelompok data yang mengandung pencilan, maka selanjutnya dilakukan pengujian terhadap kenormalan data. Pada penelitian ini digunakan uji Anderson-Darling, yang mana uji kenormalan dilakukan pada data sisaan untuk masing-masing kelompok data yang mengandung pencilan. Data sisaan ini didapat dengan terlebih dahulu melakukan analisis regresi antara masing-masing kelompok data yang mengandung pencilan dengan menggunakan MKT. Hasil dugaan koefisien regresi dari MKT dapat dilihat pada Tabel 1. untuk $n = 20$ dan Tabel 2. untuk $n = 100$.

Tabel 1. Hasil Dugaan Koefisien Regresi dengan Menggunakan MKT $n = 20$

Posisi Pencilan	Jumlah Pencilan	Intersep (b_0)	Slope (b_1)
Tanpa pencilan	-	0,9598	1,9843
X	5%	0,5228	1,2814
	10%	0,1211	1,0354
	20%	-0,7829	0,8681
Y	5%	2,036	2,516
	10%	3,129	3,025
	20%	5,462	3,833
X dan Y	10%	1,641	1,514
	20%	2,564	1,249

Tabel 2. Hasil Dugaan Koefisien Regresi dengan Menggunakan MKT $n = 100$

Posisi Pencilan	Jumlah Pencilan	Intersep (b_0)	Slope (b_1)
Tanpa pencilan	-	0,8861	2,0105
X	5%	-0,4722	1,2873
	10%	-1,2390	1,0680
	20%	-2,4395	0,8927
Y	5%	2,7030	2,5130
	10%	4,4900	2,9550
	20%	7,9610	3,6320
X dan Y	5%	0,5630	1,5650
	10%	0,9856	1,6806
	20%	-4,8920	1,5600

Kemudian dilihat nilai sisaan berdasarkan model regresi yang dihasilkan. Hipotesis yang digunakan adalah sebagai berikut:
H_0: sisaan mengikuti sebaran normal,
H_1: sisaan tidak mengikuti sebaran normal.

Sisaan dikatakan menyebar normal jika P-*value* lebih besar dari α, sebaliknya jika P-*value* lebih kecil dari α maka sisaan dikatakan tidak menyebar normal. Nilai α yang digunakan adalah 0,05. Uji kenormalan sisaan dengan menggunakan uji Anderson-Darling dapat dilihat pada Tabel 3 dan Tabel 4.

Tabel 3. Uji Kenormalan Sisaan $n = 20$

PP	JP	P-*value*	Keterangan
TP	-	0,446	Normal
X	5%	0,258	Normal
	10%	0,8538	Normal
	20%	0,5807	Normal
Y	5%	3,086e-06	Tidak Normal
	10%	0,007204	Tidak Normal
	20%	0,004886	Tidak Normal
X dan Y	10%	3,701e-07	Tidak Normal
	20%	3,897e-05	Tidak Normal

Tabel 4. Uji Kenormalan Sisaan $n = 100$

PP	JP	P-*value*	Keterangan
TP	-	0,08246	Normal
X	5%	2,048e-05	Tidak Normal
	10%	0,00083	Tidak Normal
	20%	0,04387	Tidak Normal
Y	5%	< 2,2e-16	Tidak Normal
	10%	1,656e-13	Tidak Normal
	20%	2,165e-05	Tidak Normal
X dan Y	5%	< 2,2e-16	Tidak Normal
	10%	< 2,2e-16	Tidak Normal
	20%	< 2,2e-16	Tidak Normal

Setelah dilakukan uji kenormalan sisaan dapat disimpulkan bahwa, pemberian pencilan pada data awal yaitu pada peubah bebas (X) dengan $n = 20$ tetap memberikan hasil yang normal. Namun, saat $n = 100$ pemberian pencilan pada peubah bebas (X) memengaruhi asumsi kenormalan, dan ketika diberikan pencilan pada peubah terikat (Y) serta pada peubah (X dan Y) dengan $n = 20$ dan 100 seluruhnya memengaruhi data yaitu tidak terpenuhinya asumsi kenormalan. Berdasarkan hasil uji kenormalan dengan ukuran sampel lebih besar terlihat hasil uji kenormalan lebih valid, sehingga untuk bahasan selanjutnya hanya menggunakan ukuran sampel 100.

3.3 Aplikasi Regresi *Robust* dengan MCD

Nilai penduga parameter regresi pada metode MCD dengan ukuran sampel 20 menghasilkan nilai dugaan yang sama pada pemberian pencilan 5% dan 10%. Serta menghasilkan nilai dugaan yang sama pula saat pemberian pencilan 20%. Sedangkan, pada ukuran sampel 100 metode MCD menghasilkan nilai dugaan yang berbeda-beda disetiap pemberian pencilan. Hasil dugaan koefisien regresi menggunakan metode MCD dengan ukuran sampel 100 dapat dilihat pada Tabel 5.

Tabel 5. Hasil Dugaan Koefisien Regresi dengan Metode MCD

Posisi Pencilan	Jumlah Pencilan	Intersep (b_0)	Slope (b_1)
Tanpa pencilan	-	1,0474	1,9972
X	5%	1,0344	2,0225
	10%	1,0126	2,0165
	20%	0,9639	2,0130
Y	5%	0,9748	2,0051
	10%	0,9860	2,0092
	20%	0,9639	2,0130
X dan Y	5%	1,0205	2,0186
	10%	1,0151	2,0174
	20%	0,9835	2,0114

3.4 Aplikasi Regresi *Robust* dengan TELBS

Nilai penduga parameter regresi *robust* metode TELBS menghasilkan nilai yang berbeda-beda disetiap pemberian pencilan pada ukuran sampel 20 dan 100. Namun, pada ukuran sampel kecil menghasilkan rentang atau selisih antara nilai dugaan dengan nilai sebenarnya yang lebih besar. Sehingga pada pembahasan hanya diuraikan untuk ukuran sampel besar.

Hasil dugaan koefisien regresi metode TELBS dengan ukuran sampel 100 dapat dilihat pada Tabel 6.

Tabel 6. Hasil Dugaan Koefisien Regresi dengan Metode TELBS

Posisi Pencilan	Jumlah Pencilan	Intersep (b_0)	Slope (b_1)
Tanpa pencilan	-	1,0503	1,9984
X	5%	1,0579	2,0099
	10%	1,0442	2,0040
	20%	1,0262	1,9985
Y	5%	1,0435	1,9975
	10%	1,0359	1,9980
	20%	1,0278	1,9986
X dan Y	5%	1,0556	2,0067
	10%	1,0462	2,0002
	20%	1,0475	1,9832

3.5 Perbandingan Metode

Untuk memperjelas perbedaan dari MKT, metode MCD, dan metode TELBS diperlukan perbandingan dari ketiga metode tersebut. Perbandingan ini dilakukan dengan menggunakan selang kepercayaan dan nilai MSE. Perbandingan ini didasarkan pada nilai-nilai koefisien regresi yang dihasilkan dari masing-masing metode. Nilai penduga yang baik merupakan nilai penduga yang memiliki sifat takbias, yang berarti nilai penduga yang dihasilkan harus mencakup nilai penduga yang sebenarnya.

Sifat ketakbiasan penduga parameter yang dihasilkan kedua metode dapat dilihat dengan selang kepercayaan masing-masing penduganya. Apabila selang kepercayaan mencakup nilai parameter yang sebenarnya, maka penduga parameter yang dihasilkan memiliki sifat takbias. Pada penelitian ini digunakan selang kepercayaan 95% untuk melihat ketakbiasan penduga. Data awal tanpa pencilan memperlihatkan bahwa hasil penduga parameter regresi yang dihasilkan oleh ketiga metode tidak berbias pada selang kepercayaan 95%. Hasil dugaan dikatakan tidak berbias karena selang kepercayaan dari β_0 dan β_1 mencakup nilai koefisien yang sebenarnya. Hasil pengujian selang kepercayaan 95% pada metode MKT dapat dilihat pada Tabel 7

Berdasarkan hasil analisis menggunakan selang kepercayaan 95%, MKT menghasilkan penduga parameter b_1 yang berbias di setiap pemberian pencilan, meskipun beberapa menghasilkan pendugaan yang tak bias bagi b_0 namun hal itu tidak cukup untuk membuat model yang baik. Oleh sebab itu, dilakukan analisis lebih lanjut dengan menggunakan regresi *robust* metode MCD dan TELBS.

Pendugaan parameter regresi dengan menggunakan selang kepercayaan 95% pada metode MCD menghasilkan pendugaan β_0 dan β_1 yang tak berbias seiring meningkatnya jumlah pencilan yang diberikan (5%, 10%, 20%) pada peubah (X), peubah (Y), dan keduanya. Hal ini menunjukkan bahwa pemberian pencilan pada data tidak memengaruhi hasil dugaan koefisien regresi *robust* metode MCD. Hasil pengujian dapat dilihat pada Tabel 8

Tabel 7. Selang kepercayaan 95% β_0 dan β_1 metode MKT

PP	JP	b_0		Keterangan	b_1		Keterangan
-	-	0,6797	1,0924	Tak bias	1,9702	2,0507	Tak bias
X	5%	-1,2448	0,3004	Bias	1,1798	1,3947	Bias
	10%	-2,2188	-0,2591	Bias	0,8766	1,2593	Bias
	20%	-3,5941	-1,2848	Bias	0,6671	1,1182	Bias
Y	5%	2,1522	3,2537	Bias	2,4054	2,6205	Bias
	10%	3,5081	5,4718	Bias	2,7632	3,1467	Bias
	20%	6,3001	9,6218	Bias	3,3075	3,9564	Bias
X dan Y	5%	0,0655	1,0604	Tak bias	1,4678	1,6621	Bias
	10%	0,5920	1,3791	Tak bias	1,6037	1,7574	Bias
	20%	-5,3940	-4,3899	Bias	1,4619	1,6580	Bias

Tabel 8. Selang kepercayaan 95% β_0 dan β_1 metode MCD

PP	JP	b_0		Keterangan	b_1		Keterangan
-	-	0,8409	1,2538	Tak bias	1,9568	2,0375	Tak bias
X	5%	0,8280	1,2407	Tak bias	1,9821	2,0628	Tak bias
	10%	0,8064	1,2187	Tak bias	1,9762	2,0567	Tak bias
	20%	0,7578	1,1699	Tak bias	1,9727	2,0532	Tak bias
Y	5%	0,7687	1,1808	Tak bias	1,9648	2,0453	Tak bias
	10%	0,7799	1,1920	Tak bias	1,9689	2,0494	Tak bias
	20%	0,7578	1,1699	Tak bias	1,9727	2,0532	Tak bias
X dan Y	5%	0,8143	1,2266	Tak bias	1,9783	2,0588	Tak bias
	10%	0,8090	1,2213	Tak bias	1,9771	2,0576	Tak bias
	20%	0,7775	1,1895	Tak bias	1,9711	2,0516	Tak bias

Pendugaan parameter regresi yang diperoleh pada metode TELBS dengan pemberian pencilan (5%, 10%, 20%) pada kondisi pencilan peubah (X), peubah (Y) dan keduanya, TELBS tetap memberikan hasil pendugaan parameter yang tak berbias pada selang kepercayaan 95%. Hal ini menunjukkan bahwa pemberian pencilan pada data tidak memengaruhi hasil dugaan koefisien pada pendugaan parameter regresi *robust* metode TELBS. Hasil pengujian dapat dilihat pada Tabel 9.

Tabel 9. Selang kepercayaan 95% β_0 dan β_1 metode TELBS

PP	JP	b_0		Keterangan	b_1		Keterangan
-	-	0,8439	1,2566	Tak bias	1,9701	2,0508	Tak bias
X	5%	0,8518	1,2639	Tak bias	1,9696	2,0501	Tak bias
	10%	0,8380	1,2503	Tak bias	1,9637	2,0442	Tak bias
	20%	0,8198	1,2325	Tak bias	1,9581	2,0388	Tak bias
Y	5%	0,8370	1,2499	Tak bias	1,9571	2,0378	Tak bias
	10%	0,8294	1,2423	Tak bias	1,9576	2,0383	Tak bias
	20%	0,8214	1,2341	Tak bias	1,9582	2,0389	Tak bias
X dan Y	5%	0,8495	1,2616	Tak bias	1,9664	2,0469	Tak bias
	10%	0,8399	1,2524	Tak bias	1,9599	2,0405	Tak bias
	20%	0,8396	1,2553	Tak bias	1,9426	2,0238	Tak bias

Berdasarkan hasil pendugaan parameter MKT, metode MCD, dan metode TELBS diperoleh nilai MSE masing-masing. Nilai MSE yang diperoleh akan digunakan untuk melihat metode manakah yang paling baik diantara ketiga metode.

Semakin kecil nilai MSE suatu estimator, maka hasil estimasinya akan semakin baik.

Tabel 10. Nilai MSE $n = 100$

PP		-	MKT	MCD	TELBS
TP		-	1,0873	1,12019	1,12017
X		5%	15,9242	1,11138	1,11694
		10%	27,3361	1,10352	1,11446
		20%	41,8198	1,09334	1,11246
Y		5%	10,2239	1,09644	1,11864
		10%	172,5712	1,09747	1,11579
		20%	549,7434	1,09334	1,11286
X dan Y		5%	6,3495	1,10605	1,11707
		10%	4,0368	1,10437	1,11738
		20%	37,3031	1,09675	1,13765

Pada Tabel 10. menunjukkan bahwa nilai MSE penduga parameter regresi *robust* metode MCD seluruhnya lebih kecil dibandingkan dengan MKT dan TELBS. Namun, pada saat data tanpa pencilan MKT tetap menghasilkan nilai MSE yang terkecil yaitu 1,0873.

4. KESIMPULAN DAN SARAN

Adanya pencilan pada peubah (X), peubah (Y), dan keduanya sebesar (5%, 10%, 20%) tidak terlalu memengaruhi proses pendugaan parameter pada metode MCD dan TELBS. Hal ini dilihat dari hasil pendugaan regresi *robust* metode MCD dan TELBS menghasilkan nilai pendugaan yang bersifat tak bias saat terdapat pencilan. Jika dilihat berdasarkan nilai MSE dengan ukuran sampel 100 metode yang lebih baik adalah metode MCD, karena metode ini menghasilkan nilai MSE terkecil disetiap variasi pencilan.

Adapun saran dari penelitian ini yaitu: (1) Data peubah bebas (X) yang digunakan dalam penelitian ini merupakan data simulasi berdistribusi normal dengan rataan nol dan simpangan baku lima, penelitian selanjutnya dapat mencoba menggunakan data peubah bebas yang lebih beragam; (2) Data pencilan yang dimasukkan pada penelitian ini merupakan data awal ditambah dengan tiga kali standar deviasi data yang bersesuaian, penelitian selanjutnya dapat mencoba memasukkan pencilan yang lebih bervariasi.

DAFTAR PUSTAKA

Atmagenta, H. A. (2016). *Penduga Least Trimmed Square (LTS) Pada Data Yang Mengandung Outlier*. Skripsi. Jurusan Matematika Fakultas MIPA Universitas Jember.

Fallo, J. O., Setiawan, A., & Susanto, B. (2013). Uji Normalitas Berdasarkan Metode Anderson-Darling, Cramer-von Mises dan Lilliefors Menggunakan Metode Bootstrap. *Jurnal Seminar Nasional Matematika Dan Pendidikan Matematika FMIPA UNY*.

Febrianto, L. S., Dwidayati, N. K., & Hendikawati, P. (2018). Perbandingan Metode Robust Least Median of Square (LMS) dan Penduga S Untuk Menangani Outlier Pada Regresi Linier Berganda. *UNNES Journal of Mathematics*, *7*(1), 83–95.

Gusriani, N., & Firdaniza. (2018). Linear regression based on Minimum Covariance Determinant (MCD) and TELBS Methods on The Productivity of Phytoplankton. *IOP Conference Series: Materials Science and Engineering*, *332*(1).

Harlan, J. (2018). *Analisis Regresi Linear*. Jakarta: Gunadarma

Hubert, M., Debruyne, M., & Rousseeuw, P. J. (2018). Minimum Covariance Determinant and Extensions. *WIREs Comput Stat*, 10, 1-11.

Irfagautami, N. P. N., Srinadi, I. G. A. M., & Sumarjaya, I. W. (2014). Perbandingan Regresi Robust Penduga MM dengan Metode Random Sampel Consensus dalam Menangani Pencilan. *E-Jurnal Matematika*, 3(2), 45–52.

Lainun, H., Tinungki, G. M., & Amran. (2018). Perbandingan Penduga M, S, dan MM pada Regresi Linier dalam Menangani Keberadaan Outlier. *Jurnal Matematika, Statistika & Komputasi*, 5(1), 88-96.

Marzuki, Sofyan, H., & Rusyana, A. (2010). Pendugaan Selang Kepercayaan Persentil Bootstrap Nonparametrik untuk Parameter Regresi. *Statistika*, 10(1), 13–23.

Myers. (1990). *Classical and Modern Regression with Applications* (2nd ed.). PWS-KENT Publishing.

Tabatabai, Eby, W., Li, H., Bae, S., & Singh, K. (2012). TELBS robust linear regression method. *Open Access Medical Statistics*, *2*, 65–84