

KAJIAN REGRESI KEKAR MENGGUNAKAN METODE PENDUGA-MM DAN KUADRAT MEDIAN TERKECIL*

Khusnul Khotimah¹, Kusman Sadik², and Akbar Rizki³

¹Department of Statistics, IPB University, Indonesia, khusnulrfa@gmail.com

²Department of Statistics, IPB University, Indonesia, kusmansadik@gmail.com

³Department of Statistics, IPB University, Indonesia, akbar.ritzki@gmail.com

Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)

Vol 4 No 1 (2020), 97 - 115

Copyright © 2020 Khusnul Khotimah, Kusman Sadik, and Akbar Rizki. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Regression is a statistical method that is used to obtain a pattern of relations between two or more variables presented in the regression line equation. This line equation is derived from estimation using ordinary least squares (OLS). However, OLS has limitations that are highly dependent on outliers data. One solution to the outliers problem in regression analysis is to use the robust regression method. This study used the least median squares (LMS) and multi-stage method (MM) robust regression for analysis of data containing outliers. Data analysis was carried out on generation data simulation and actual data. The simulation results of regression analysis in various scenarios are concluded that the LMS and MM methods have better performance compared to the OLS on data containing outliers. MM method has the lowest average parameter estimation bias, followed by the LMS, then OLS. The LMS has the smallest average root mean squares error (RMSE) and the highest average R^2 is followed by the MM then the OLS. The results of the regression analysis comparison of the three methods on Indonesian rice production data in 2017 which contains 10% outliers were concluded that the LMS is the best method. The LMS produces the smallest RMSE of 4.44 and the highest R^2 that is 98%. MM's method is in the second-best position with RMSE of 6.78 and R^2 of 96%. OLS method produces the largest RMSE and lowest R^2 that is 23.15 and 58% respectively.

Keywords: least median squares, multi-stage method, outlier, robust regression, root mean square error.

*Received Jul 2019; Accepted Feb 2020; Published online on Feb 2020

1 Pendahuluan

Metode regresi linear merupakan metode statistika yang banyak digunakan untuk mengevaluasi hubungan linear antara peubah respon yang bersifat kuantitatif dengan satu atau lebih peubah penjelas. Metode kuadrat terkecil (MKT) merupakan pendekatan yang paling populer digunakan dalam pendugaan parameter model regresi linear (D'Urso & Massari, 2013). Namun demikian, MKT ini memiliki keterbatasan, yaitu sangat dipengaruhi oleh data pencilan (Atilgan & Gunay, 2011). Data pencilan merupakan suatu amatan dengan nilai ekstrim yang sangat jauh berbeda dengan kebanyakan data lainnya (Montgomery *et al.*, 2012).

Pembuangan pencilan dari himpunan data hanya dapat dilakukan apabila nilai tersebut diketahui penyebabnya, seperti kesalahan pengukuran atau analisis, kesalahan merekam data, atau kegagalan alat ukur. Pencilan terkadang mengandung informasi yang lebih penting daripada nilai amatan lainnya sehingga akan berpengaruh besar pada model yang dibentuk. Oleh karena itu, membuang data pencilan dengan tujuan memperbaiki kesesuaian persamaan regresi tidak bisa sembarangan dilakukan karena akan memberikan presisi pendugaan tidak tepat (Montgomery *et al.*, 2012).

Penanganan pencilan pada analisis regresi dapat dilakukan dengan menggunakan metode regresi kekar. Regresi kekar merupakan suatu metode yang digunakan untuk keperluan analisis regresi data yang mengandung amatan pencilan (Chen *et al.*, 2002). Beberapa prosedur metode regresi kekar yang dikemukakan dalam Chen *et al.* (2002) adalah penduga-M, penduga-S, penduga-*least trimmed squares* (LTS), dan penduga-*multi-stage method* (MM). Metode regresi robust lainnya adalah penduga-*least median squares* (LMS) yang diperkenalkan oleh Rousseeuw pada tahun 1984. Ehab Mohamed A dan Hisham Mohamed A (Almetwally & Almongy, 2018) serta Arista Oktarinanda (Oktarinanda, 2014) meneliti tentang perbandingan beberapa metode kekar. Almetwally dan Almongy (2018) membandingkan metode M, S, dan MM dan diperoleh hasil kesimpulan bahwa metode MM merupakan metode terbaik berdasarkan kriteria nilai bias dan *mean square error* (MSE) terkecil. Penelitian Oktarinanda (2014) membandingkan metode LMS dan LTS dan dihasilkan kesimpulan bahwa metode LMS lebih akurat dalam menduga model berdasarkan nilai RMSE yang lebih kecil.

Hasil penelitian tersebut membuat penulis tertarik untuk membandingkan regresi kekar penduga-MM dan penduga-LMS serta menentukan metode terbaik. Pemilihan metode terbaik akan didasarkan pada nilai bias pendugaan parameter regresi, *root mean square error* (RMSE), dan R^2 . Oleh karena itu, penulis mengangkat topik mengenai "Kajian Pemodelan Regresi Kekar Menggunakan Metode Penduga-MM dan Kuadrat Median Terkecil". Perbandingan dilakukan melalui simulasi data dengan berbagai ukuran data, tipe pencilan, dan persentase pencilan, kemudian menerapkannya terhadap data pertanian yaitu data produksi padi (dalam juta ton) tahun 2017.

2 Metodologi

2.1 Bahan dan Data

Data Simulasi

Penelitian ini melibatkan data hasil pembangkitan yang mengandung nilai pencilan untuk keperluan simulasi regresi linear kekar. Proses simulasi yang dilakukan menggunakan ukuran data (n) sebanyak 50 data untuk ukuran kecil, 200 data untuk ukuran sedang, dan 1000 data untuk ukuran besar. Terdapat tiga tipe pencilan dalam

penelitian ini yaitu *vertical outlier*, *good leverage point*, dan *bad leverage point* dengan persentase pencilan (m) sebesar 0%, 5%, 10%, 15%, 20%, dan 30% pada setiap tipe pencilan. Data hasil pembangkitan tersebut kemudian digunakan dalam proses pendugaan parameter menggunakan metode MKT, MM, dan LMS. Parameter β_0 dan β_1 yang digunakan dalam penelitian ini secara berturut-turut sebesar 5 dan 2 sehingga model regresi linear yang terbentuk sebagai berikut:

$$\mathbf{y} = 5 + 2\mathbf{x} + \boldsymbol{\varepsilon} \quad (1)$$

dengan,

\mathbf{y} : vektor data respon berukuran $n \times 1$.

\mathbf{x} : vektor data peubah penjelas berukuran $n \times 1$

$\boldsymbol{\varepsilon}$: vektor galat berukuran $n \times 1$

Seluruh hasil simulasi didasarkan pada 1000 kali pengulangan yang dilakukan dengan bantuan *software R*. Proses perbandingan kebaikan metode dilakukan dengan melihat nilai bias pendugaan parameter ($bias(\hat{\beta}_0)$ dan $bias(\hat{\beta}_1)$), RMSE, dan R^2 .

Data Aktual

Penelitian ini juga menggunakan data aktual sebagai aplikasi perbandingan metode MKT, MM dan LMS. Data aktual yang digunakan adalah data produksi padi (dalam juta ton) dan data jumlah penggunaan pupuk organik (dalam ribu ton) pada tahun 2017 yang diperoleh dari Pusat Data dan Sistem Informasi Kementerian Pertanian Republik Indonesia. Data tersebut akan dianalisis menggunakan regresi linear sederhana untuk mengetahui hubungan linear antara peubah penjelas yaitu data jumlah pupuk organik dengan peubah respon yaitu data produksi padi pada tahun 2017. Data terdiri atas 34 amatan yang merupakan provinsi-provinsi di Indonesia.

2.2 Metode Penelitian

Data Simulasi

Proses analisis data dalam penelitian ini menggunakan *software R* dengan bantuan package "*MASS*" dan "*robustbase*". Prosedur simulasi yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Menetapkan $\beta_0 = 5$ dan $\beta_1 = 2$.
2. Membangkitkan vektor peubah penjelas (\mathbf{x}) yang menyebar normal sebanyak n data dengan rataan dan ragam sebagai berikut:
 - (a) $\mathbf{x} \sim N(5, 1)$ untuk data bukan pencilan.
 - (b) $\mathbf{x}^* \sim N(30, 1)$ untuk data pencilan.
3. Membangkitkan vektor nilai error ($\boldsymbol{\varepsilon}$) yang menyebar normal sebanyak n data berikut:
 - (a) $\boldsymbol{\varepsilon} \sim N(0, 1)$ untuk data bukan pencilan.
 - (b) $\boldsymbol{\varepsilon}^* \sim N(30, 1)$ untuk data pencilan.
4. Menentukan vektor nilai \mathbf{y} berdasarkan skenario berikut:

(a) *Vertical Outlier*

Menentukan vektor \mathbf{y} berdasarkan model $\mathbf{y} = 5 + 2\mathbf{x} + \mathbf{r}$,

dengan \mathbf{r} merupakan vektor komponen galat yang diperoleh dari gabungan penarikan sampel data $\boldsymbol{\varepsilon} \sim N(0, 1)$ sebanyak $(1 - m) \times n$ dan data $\boldsymbol{\varepsilon}^* \sim N(30, 1)$ sebanyak $m \times n$ (m merupakan persentase pencilan).

Vektor \mathbf{y} kemudian akan diregresikan dengan vektor \mathbf{x} pada tahap berikutnya.

(b) *Good leverage point*

Menentukan vektor \mathbf{y} berdasarkan model $\mathbf{y} = 5 + 2\mathbf{s} + \boldsymbol{\varepsilon}$,

dengan \mathbf{s} merupakan vektor peubah penjelas yang diperoleh dari gabungan penarikan sampel data $\mathbf{x} \sim N(5, 1)$ sebanyak $(1 - m) \times n$ dan data $\mathbf{x}^* \sim N(30, 1)$ sebanyak $m \times n$.

Vektor \mathbf{y} kemudian akan diregresikan dengan vektor \mathbf{s} pada tahap berikutnya

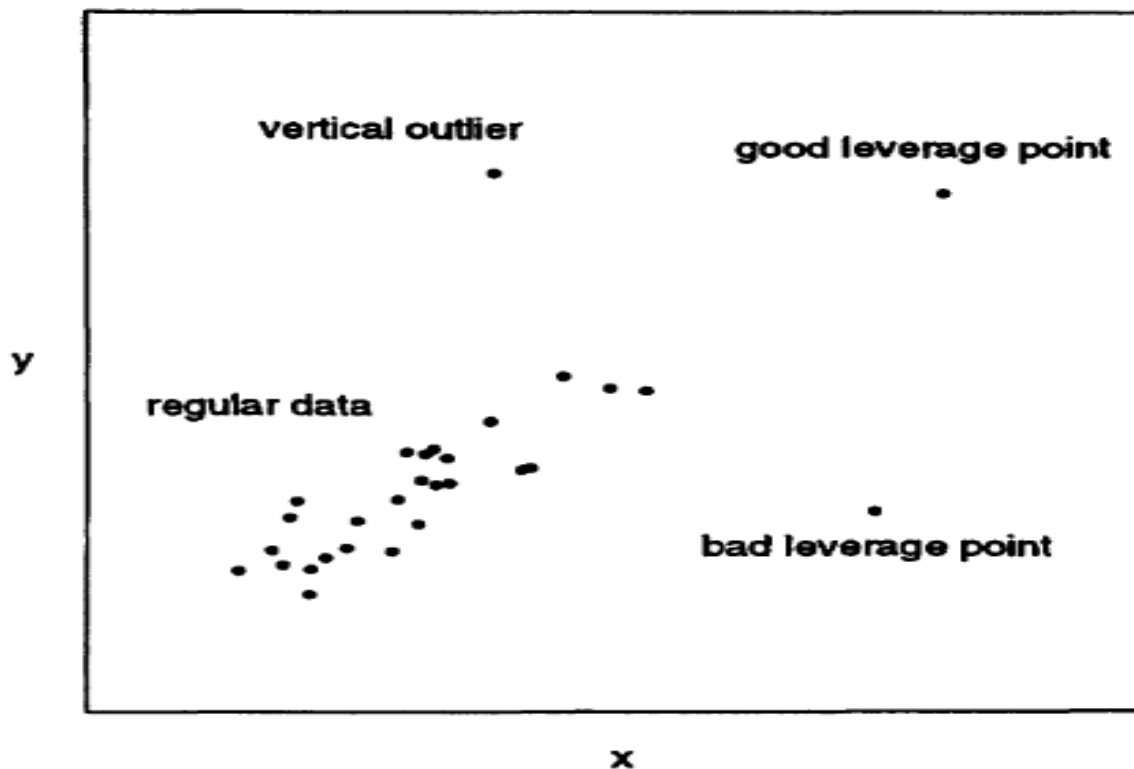
(c) *Bad leverage point*

Menentukan vektor \mathbf{y} berdasarkan model $\mathbf{y} = 5 + 2\mathbf{x} + \mathbf{t}$

dengan \mathbf{t} merupakan vektor komponen galat yang diperoleh dari gabungan penarikan sampel data $\boldsymbol{\varepsilon}^* \sim N(30, 1)$ sebanyak $\frac{1}{2}m \times n$ dengan data $\boldsymbol{\varepsilon} \sim N(0, 1)$ sebanyak $(1 - \frac{1}{2}m) \times n$. Kemudian sebanyak $m \times n$ data pertama pada vektor penjelas (\mathbf{x}) ditambahkan dengan 20.

Vektor \mathbf{y} kemudian akan diregresikan dengan vektor \mathbf{x} baru tersebut pada tahap berikutnya.

Ilustrasi mengenai tampilan ketiga tipe pencilan yang akan bangkitkan dapat dilihat pada Gambar 1 berikut:



Gambar 1: Tipe-tipe pencilan

5. Meregresikan semua gugus data dengan menggunakan MKT, LMS, dan MM dengan rincian langkah-langkah sebagai berikut:

(a) Algoritma regresi metode MKT.

- i. Menyusun vektor data \mathbf{y} dan matriks \mathbf{X} . \mathbf{X} merupakan matriks data peubah penjelas berukuran $n \times (k+1)$ dengan k merupakan jumlah peubah penjelas dan kolom pertama berisi vektor 1.
- ii. Menghitung koefisien penduga parameter β

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

(b) Algoritma regresi metode LMS.

Langkah-langkah pendugaan parameter dengan metode LMS yang terangkum dalam suatu algoritma PROGRESS oleh Rousseeuw dan Hubert (Rousseeuw & Hubert, 1997) pada *package robust R software* adalah sebagai berikut:

- i. Menentukan nilai g

$$g = \left\lceil \frac{n+k+1}{2} \right\rceil \tag{3}$$

- ii. Mengambil himpunan bagian data berukuran g amatan secara acak dari himpunan data berukuran n amatan, sehingga akan terdapat $f = \binom{n}{g}$ himpunan bagian data.
- iii. Melakukan analisis regresi dari setiap himpunan bagian data dengan menggunakan MKT.
- iv. Menghitung nilai M_d (median sisaan kuadrat e_{vd}^2) pada masing-masing himpunan bagian data. Indeks v merupakan indeks untuk banyaknya amatan pada setiap himpunan subset, $v = 1, 2, 3, \dots, g$ dan indeks d adalah banyaknya himpunan bagian yang terbentuk, $d = 1, 2, 3, \dots, \binom{n}{g}$
- v. Menentukan nilai M yaitu minimum M_d berdasarkan hasil tahap (iv).
- vi. Melakukan pendugaan awal simpangan baku sisaan LMS ($\hat{\sigma}_{LMS}$)

$$\hat{\sigma}_{LMS} = 1.4826 \left(1 + \frac{5}{(n-g)} \right) \sqrt{M} \tag{4}$$

- vii. Menghitung bobot penduga LMS (w_{iLMS})

$$w_{iLMS} = \begin{cases} 1, & \left| \frac{e_i}{\hat{\sigma}_{LMS}} \right| \leq 2.5 \\ 0, & \text{lainnya} \end{cases} \tag{5}$$

$w_{i(LMS)}$ merupakan elemen diagonal matriks \mathbf{W}^{LMS} yang memiliki ukuran $n \times n$ dan elemen lainnya bernilai 0.

$$\mathbf{W}^{LMS} = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{pmatrix}$$

- viii. Menduga parameter regresi dengan metode regresi terboboti

$$\widehat{\beta}_{LMS} = \left(\mathbf{X}' \mathbf{W}^{LMS} \mathbf{X} \right)^{-1} \left(\mathbf{X}' \mathbf{W}^{LMS} \mathbf{y} \right) \quad (6)$$

dengan,

\mathbf{y} : vektor data respon berukuran $n \times 1$.

\mathbf{X} : matriks data peubah penjelas berukuran $n \times (k + 1)$ dengan kolom pertama berisi vektor 1 dan k merupakan jumlah peubah penjelas.

$\widehat{\beta}_{LMS}$: vektor koefisien regresi berukuran $(k + 1) \times 1$.

(c) Algoritma penduga-MM

Metode penduga-MM diperoleh melalui dua tahap.

Pertama, menghitung nilai *scale estimate* ($\widehat{\sigma}_s$) dengan metode penduga-S berdasarkan langkah-langkah berikut:

- i. Menduga koefisien regresi pada data menggunakan metode MKT
- ii. Menghitung nilai sisaan $e_i = y_i - \hat{y}_i$
- iii. Menghitung $\widehat{\sigma}_s$

$$\widehat{\sigma}_s = \begin{cases} \frac{\text{median}|e_i - \text{median}(e)|}{0.6745} & ; q = 1 \\ \sqrt{\frac{1}{nK} \sum_{i=1}^n w_i e_i^2} & ; q > 1. \end{cases}$$

dengan $K = 0.199$ dan q merupakan iterasi.

- iv. Menghitung pembobot penduga-S w_{i_s} sebagai berikut:

$$w_{i_s} = \begin{cases} \begin{cases} \left[1 - \left(\frac{u_i}{c} \right)^2 \right]^2, & |u_i| \leq c; q = 1 \\ 0, & |u_i| > c; \end{cases} \\ \frac{\rho(u_i)}{u_i^2} & ; q > 1 \end{cases}$$

Fungsi $\rho(u_i)$ menggunakan fungsi pembobot Tukey adalah:

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

dengan $u_i = \frac{e_i}{\widehat{\sigma}_s}$ dan $c = 4.685$

- v. Menghitung $\widehat{\beta}_S$ menggunakan MKT terboboti dengan bobot w_{i_s} .
- vi. Mengulang tahap ii-v hingga diperoleh nilai $\widehat{\beta}_S$ yang konvergen.
- vii. Setelah didapat nilai $\widehat{\beta}_S$ yang konvergen selanjutnya menghitung $\widehat{\sigma}_S$ pada iterasi terakhir untuk kemudian dijadikan *scale estimate* pada perhitungan penduga-MM berikutnya.

Kedua, menduga parameter regresi dengan melakukan *iteratively reweighted least squares* (IRLS) sebagai berikut:

- i. Menghitung nilai $u_i = \frac{e_i}{\widehat{\sigma}_i}$, nilai $\widehat{\sigma}_S$ diperoleh dari tahap (vii)
- ii. Menghitung nilai pembobotan metode MM (w_{iMM})

$$w_{iMM} = \begin{cases} \left[1 - \left(\frac{u_i}{4.685} \right)^2 \right]^2, & |u_i| \leq 4,685; \\ 0, & |u_i| > 4,685. \end{cases}$$

iii. Menghitung $\hat{\beta}_{MM}$ menggunakan kuadrat terkecil terboboti:

$$\hat{\beta}_{MM} = (\mathbf{X}'\mathbf{W}^{MM}\mathbf{X})^{-1} (\mathbf{X}'\mathbf{W}^{MM}\mathbf{Y}) \quad (7)$$

iv. Mengulang tahap i-iii hingga diperoleh nilai $\hat{\beta}_{MM}$ yang konvergen.

Proses algoritma penduga-MM dilakukan menggunakan bantuan *package robustbase R software*.

6. Melakukan pengecekan model diagnostik terhadap masing-masing metode pada tahap 5 dengan menggunakan uji berikut:

- (a) Asumsi kenormalan sisaan diuji menggunakan uji *Kolmogorov-Smirnov*,
- (b) Asumsi kehomogenan sisaan dengan menggunakan *Breusch-Pagan*,
- (c) Asumsi kebebasan sisaan (nonautokorelasi) diuji menggunakan uji *Runs-test*.

7. Mengulangi langkah 1 sampai 6 sebanyak $r = 1000$ ulangan.

8. Membandingkan rata-rata dari nilai bias pendugaan parameter, RMSE, dan R^2 .

$$\text{rata-rata bias } (\hat{\beta}_j) = \frac{1}{r} \sum_{l=1}^r |\beta_j - \hat{\beta}_{lj}|, \quad j = 0, 1 \quad (8)$$

$$\text{rata-rata RMSE} = \frac{1}{r} \sum_{l=1}^r \sqrt{\left(\frac{\sum_{i=1}^n (y_{il} - \hat{y}_{il})^2}{n - p} \right)} \quad (9)$$

$$\text{rata-rata } R^2 = \frac{1}{r} \sum_{l=1}^r \left(1 - \frac{\sum_{i=1}^n (y_{il} - \hat{y}_{il})^2}{\sum_{i=1}^n (y_{il} - \bar{y}_l)^2} \right) \quad (10)$$

dengan,

n : jumlah observasi

y_{il} : data ke ke- i pada ulangan ke- l

\bar{y}_l : rata-rata data respon ke- l

\hat{y}_{il} : dugaan respon ke- i pada ulangan ke- l

p : jumlah parameter yang diduga

r : banyaknya ulangan

$\hat{\beta}_{lj}$: dugaan parameter ke- j pada ulangan ke- l ,

9. Merangkum hasil nilai evaluasi pada setiap kombinasi ukuran data, tipe pencilan, persentase pencilan, dan metode pendugaan yang digunakan.

Data Aktual

Langkah-langkah analisis data aktual produksi padi tahun 2017 adalah sebagai berikut:

1. Melakukan plot pencarian data penggunaan pupuk organik (ribu ton) dan data jumlah produksi padi (juta ton).
2. Melakukan pendugaan model regresi dengan metode MKT dan menghitung nilai sisaannya.

3. Mengidentifikasi adanya pencilan *vertical outlier* melalui plot nilai *standardized residuals*. Pencilan *vertical outlier* diindikasikan oleh nilai *standardized residuals* lebih dari 2 atau 2.5 relatif terhadap kondisi tebaran siaan.
4. Mengidentifikasi adanya pencilan *good leverage point* melalui plot nilai unsur diagonal matriks *hat*. Data yang memiliki nilai diagonal matriks *hat* lebih besar dari dua kali rata-rata keseluruhan unsur diagonal matriks *hat* merupakan pencilan *good leverage point*.
5. Mengidentifikasi adanya pencilan *bad leverage point*. Data yang teridentifikasi sebagai *vertical outlier* dan juga pencilan *good leverage point* merupakan data pencilan *bad leverage point*.
6. Menduga model regresi dengan metode MKT, LMS, dan MM untuk data yang mengandung pencilan tersebut.
7. Membandingkan nilai RMSE dan R^2 dari metode MKT, LMS, dan MM pada tahap 6.
8. Menentukan dugaan model terbaik, yaitu dugaan model dengan nilai RMSE terkecil dan R^2 tertinggi.

3 Hasil dan Pembahasan

3.1 Hasil Simulasi

Simulasi analisis regresi dengan metode MKT, LMS, dan MM pada data yang mengandung pencilan diperoleh hasil sebagai berikut:

1. Rataan nilai bias pendugaan parameter β_0

(a) *Vertical outlier*

Tabel 1 memperlihatkan rata-rata nilai bias $\hat{\beta}_0$ pada masing-masing persentase pencilan *vertical outlier* dan pada setiap ukuran data. Ketika data tidak mengandung pencilan, rata-rata nilai bias $\hat{\beta}_0$ terkecil berada pada metode MKT, diikuti oleh metode MM, kemudian metode LMS. Namun demikian, ketika data mengandung pencilan, rata-rata nilai bias $\hat{\beta}_0$ metode MKT mengalami peningkatan yang sangat tinggi seiring pertambahan persentase pencilan, sedangkan metode LMS dan MM cenderung dapat mempertahankan rata-rata nilai bias $\hat{\beta}_0$ pada nilai yang kecil. Adanya pencilan *vertical outlier* sangat berpengaruh buruk pada pendugaan parameter β_0 metode MKT.

(b) *Good leverage point*

Rataan nilai bias $\hat{\beta}_0$ dari data-data yang mengandung pencilan *good leverage point* disajikan pada Tabel 2. Tabel 2 menunjukkan bahwa metode MM dan MKT memiliki rata-rata nilai bias $\hat{\beta}_0$ yang cenderung sama dan mendekati nilai 0. Metode LMS memiliki rata-rata nilai bias $\hat{\beta}_0$ terbesar dibandingkan dengan MKT dan MM. Akan tetapi, seiring peningkatan ukuran data dan persentase pencilan, rata-rata bias $\hat{\beta}_0$ metode LMS semakin menurun mendekati nilai 0. Secara keseluruhan, ketiga metode tersebut memiliki kinerja yang baik pada

Tabel 1: Rataan nilai bias $\hat{\beta}_0$ pada berbagai ukuran data dan persentase pencilan *vertical outlier*.

Jumlah data	Metode	Persentase Pencilan (%)					
		0	5	10	15	20	30
50	MKT	0.30	7.16	11.58	13.82	17.10	22.5
	LMS	0.72	0.68	0.67	0.65	0.65	0.62
	MM	0.31	0.32	0.33	0.33	0.34	0.36
200	MKT	0.15	4.45	7.11	9.96	12.7	17.9
	LMS	0.39	0.40	0.38	0.37	0.37	0.36
	MM	0.15	0.15	0.16	0.16	0.17	0.17
1000	MKT	0.06	3.16	6.10	9.09	12.10	18.2
	LMS	0.22	0.21	0.21	0.21	0.21	0.20
	MM	0.07	0.07	0.07	0.07	0.07	0.08

kasus data mengandung pencilan *good leverage point*. Hal tersebut ditunjukkan oleh rataan nilai bias $\hat{\beta}_0$ yang cukup kecil untuk ketiga metode tersebut.

Tabel 2: Rataan nilai bias $\hat{\beta}_0$ pada berbagai ukuran data dan persentase pencilan *good leverage point*.

Jumlah Data	Metode	Persentase Pencilan (%)					
		0	5	10	15	20	30
50	MKT	0.30	0.09	0.08	0.08	0.08	0.08
	LMS	0.71	0.52	0.30	0.23	0.21	0.20
	MM	0.31	0.14	0.09	0.08	0.08	0.09
200	MKT	0.14	0.04	0.04	0.04	0.04	0.04
	LMS	0.40	0.15	0.12	0.11	0.11	0.12
	MM	0.15	0.04	0.04	0.04	0.04	0.04
1000	MKT	0.06	0.02	0.02	0.02	0.02	0.02
	LMS	0.22	0.07	0.06	0.06	0.07	0.07
	MM	0.06	0.02	0.02	0.02	0.02	0.02

(c) *Bad leverage point*

Rataan nilai bias $\hat{\beta}_0$ dari data-data yang mengandung pencilan *bad leverage point* untuk setiap ukuran data disajikan pada Tabel 3. Tabel 3 memperlihatkan bahwa ketika persentase pencilan *bad leverage point* 0%, diperoleh hasil rataan nilai bias $\hat{\beta}_0$ terkecil terletak pada metode MKT, diikuti oleh metode MM dan terakhir adalah metode LMS. Akan tetapi, ketika data mengandung pencilan *bad leverage point*, metode MKT menghasilkan rataan nilai bias $\hat{\beta}_0$ paling besar dibandingkan dengan kedua metode lainnya. Selanjutnya, setiap kenaikan persentase pencilan dalam data, rataan nilai bias $\hat{\beta}_0$ metode MKT dan LMS semakin menurun, sedangkan rataan nilai bias $\hat{\beta}_0$ metode MM mengalami peningkatan. Rataan nilai bias $\hat{\beta}_0$ metode MM lebih besar dibandingkan dengan LMS ketika ukuran data 200 dengan persentase pencilan 30% dan ketika ukuran data 1000 dengan persentase pencilan 15%, 20%, 30%. Secara keseluruhan, semakin bertambahnya ukuran data menyebabkan rataan nilai bias $\hat{\beta}_0$ ketiga metode mengalami penurunan. Peningkatan maupun penurunan rataan nilai

bias $\hat{\beta}_0$ tidak signifikan. Hal tersebut menunjukkan bahwa adanya amatan pencilan *bad leverage point* dalam data tidak memberikan pengaruh yang nyata terhadap rata-rata nilai bias $\hat{\beta}_0$ ketiga metode.

Tabel 3: rata-rata nilai bias $\hat{\beta}_0$ pada berbagai ukuran data dan persentase pencilan *bad leverage point*.

Jumlah Data	Metode	Persentase Pencilan (%)					
		0	5	10	15	20	30
50	MKT	0.12	0.54	0.55	0.55	0.55	0.54
	LMS	0.30	0.29	0.29	0.29	0.29	0.30
	MM	0.13	0.16	0.16	0.17	0.18	0.22
200	MKT	0.06	0.53	0.54	0.54	0.54	0.53
	LMS	0.18	0.18	0.17	0.18	0.18	0.19
	MM	0.06	0.08	0.10	0.13	0.16	0.21
1000	MKT	0.03	0.53	0.54	0.54	0.54	0.53
	LMS	0.10	0.10	0.10	0.10	0.10	0.12
	MM	0.03	0.05	0.10	0.16	0.19	0.21

2. Rataan nilai bias pendugaan parameter β_1

(a) *Vertical outlier*

Rataan nilai bias pendugaan parameter β_1 yang disajikan pada Tabel 4 memperlihatkan bahwa metode MKT memiliki nilai bias terkecil hanya pada saat persentase pencilan 0%. Selanjutnya, ketika data mengandung pencilan maka rata-rata nilai bias pendugaan β_1 MKT selalu mengalami kenaikan dan berada pada nilai bias terbesar dibandingkan kedua metode lainnya. Rataan nilai bias pendugaan β_1 metode LMS dan MM untuk semua ukuran data tetap konsisten pada nilai yang cukup rendah, sehingga metode LMS dan MM baik digunakan untuk pendugaan parameter β_1 pada data yang mengandung *vertical outlier*.

Tabel 4: Rataan nilai bias $\hat{\beta}_1$ pada berbagai ukuran data dan persentase pencilan *vertical outlier*.

Jumlah Data	Metode	Persentase Pencilan (%)					
		0	5	10	15	20	30
50	MKT	0.06	1.36	2.06	2.36	2.78	3.21
	LMS	0.14	0.13	0.13	0.13	0.13	0.12
	MM	0.06	0.06	0.06	0.06	0.07	0.07
200	MKT	0.03	0.72	1.00	1.21	1.37	1.59
	LMS	0.08	0.08	0.07	0.07	0.07	0.07
	MM	0.03	0.03	0.03	0.03	0.03	0.03
1000	MKT	0.01	0.33	0.45	0.54	0.59	0.68
	LMS	0.04	0.04	0.04	0.04	0.04	0.04
	MM	0.01	0.01	0.01	0.01	0.01	0.01

(b) *Good leverage point*

Amatan pencilan tipe *good leverage point* yang berada pada himpunan data

tidak memengaruhi kinerja dari metode MKT dalam pendugaan parameter β_1 . Hal tersebut ditunjukkan dengan rata-rata nilai bias pendugaan parameter β_1 metode MKT yang selalu lebih kecil dari kedua metode lainnya, bahkan seiring pertambahan persentase pencilan ini rata-rata nilai bias pendugaan β_1 MKT semakin menurun mendekati 0. Penurunan rata-rata nilai bias pendugaan β_1 ini juga terjadi pada LMS dan MM. Tabel 5 menunjukkan bahwa ketiga metode memiliki nilai bias $\hat{\beta}_1$ mendekati 0. Adanya pencilan *good leverage point* tidak memberikan pengaruh buruk pada pendugaan β_1 ketiga metode.

Tabel 5: Rataan nilai bias $\hat{\beta}_1$ pada berbagai ukuran data dan persentase pencilan *good leverage point*.

Jumlah Data	Metode	Persentase Pencilan (%)					
		0	5	10	15	20	30
50	MKT	0.06	0.01	0.01	0.01	0.01	0.00
	LMS	0.14	0.09	0.04	0.02	0.01	0.01
	MM	0.06	0.02	0.01	0.01	0.01	0.01
200	MKT	0.03	0.00	0.00	0.00	0.00	0.00
	LMS	0.08	0.02	0.01	0.01	0.01	0.01
	MM	0.03	0.01	0.00	0.00	0.00	0.00
1000	MKT	0.01	0.00	0.00	0.00	0.00	0.00
	LMS	0.04	0.01	0.00	0.00	0.00	0.00
	MM	0.01	0.00	0.00	0.00	0.00	0.00

(c) *Bad leverage point*

Rataan nilai bias $\hat{\beta}_1$ dari data-data yang mengandung pencilan *bad leverage point* pada Tabel 6 menunjukkan bahwa ketika persentase *outlier* 0%, untuk semua ukuran data diperoleh hasil rata-rata nilai bias $\hat{\beta}_1$ terkecil terletak pada metode MKT, diikuti oleh MM, kemudian LMS. Selanjutnya, setiap kenaikan persentase pencilan dalam data, rata-rata nilai bias $\hat{\beta}_1$ ketiga metode semakin meningkat, dengan rata-rata nilai bias tertinggi dimiliki oleh metode MKT untuk semua ukuran data. Metode kekar LMS maupun metode MM dapat mempertahankan rata-rata bias $\hat{\beta}_1$ yang konsisten berada pada nilai mendekati nilai 0 untuk semua ukuran data dan berbagai persentase pencilan. Dengan demikian, metode LMS dan MM lebih kekar terhadap adanya pencilan *bad leverage point* dibandingkan dengan metode MKT.

3. Rataan nilai Koefisien Determinasi

(a) *Vertical outlier*

Rataan nilai R^2 oleh ketiga metode disajikan pada Tabel 7. MKT memiliki rata-rata nilai R^2 tinggi hanya pada saat tidak ada pencilan dalam data. Selanjutnya ketika terdapat pencilan pada data maka rata-rata nilai R^2 oleh MKT mengalami penurunan. Adanya pencilan *vertical outlier* sangat memengaruhi nilai rata-rata R^2 MKT. Rataan nilai R^2 metode LMS untuk semua ukuran data berada pada nilai paling tinggi dibandingkan MKT dan MM yaitu berada pada nilai mendekati 1. Rataan nilai R^2 metode MM juga konsisten berada pada nilai mendekati 1 untuk semua ukuran data.

Tabel 6: Rataan nilai bias $\hat{\beta}_1$ pada berbagai ukuran data dan persentase pencilan *bad leverage point*.

Jumlah Data	Metode	Persentase Pencilan (%)					
		0	5	10	15	20	30
50	MKT	0.12	1.21	1.22	1.23	1.23	1.24
	LMS	0.27	0.27	0.28	0.28	0.29	0.32
	MM	0.12	0.21	0.23	0.26	0.30	0.41
200	MKT	0.05	1.19	1.21	1.23	1.23	1.23
	LMS	0.16	0.16	0.17	0.18	0.19	0.23
	MM	0.06	0.11	0.17	0.25	0.33	0.45
1000	MKT	0.03	1.19	1.22	1.23	1.23	1.92
	LMS	0.09	0.09	0.09	0.10	0.12	0.18
	MM	0.03	0.09	0.22	0.35	0.42	0.47

Tabel 7: Rataan nilai R^2 pada berbagai ukuran data dan persentase pencilan *vertical outlier*.

Jumlah Data	Metode	Persentase Pencilan (%)					
		0	5	10	15	20	30
50	MKT	0.94	0.04	0.03	0.03	0.03	0.02
	LMS	1.00	1.00	1.00	0.99	0.99	0.98
	MM	0.95	0.95	0.95	0.95	0.95	0.94
200	MKT	0.94	0.03	0.02	0.01	0.01	0.01
	LMS	1.00	0.99	0.99	0.99	0.99	0.98
	MM	0.95	0.95	0.95	0.95	0.95	0.94
1000	MKT	0.94	0.02	0.01	0.01	0.01	0.01
	LMS	0.99	0.99	0.99	0.98	0.98	0.97
	MM	0.95	0.95	0.95	0.95	0.95	0.94

(b) *Good leverage point*

Tabel 8 menunjukkan bahwa metode MKT, LMS, dan MM pada data yang mengandung pencilan *good leverage point* memiliki rataan nilai R^2 yang tergolong sangat tinggi yakni di atas 90% untuk semua ukuran data. Adanya pencilan *good leverage point* tidak memperburuk nilai rataan R^2 yang dihasilkan oleh masing-masing metode. Semakin tinggi persentase pencilan ini akan meningkatkan rataan nilai R^2 untuk ketiga metode tersebut. Rataan nilai R^2 metode LMS berada pada posisi tertinggi, sedangkan metode MM cenderung memiliki rataan nilai R^2 yang sama dengan metode MKT.

(c) *Bad leverage point*

Tabel 9 menyajikan rataan nilai R^2 pada berbagai ukuran data yang mengandung pencilan *bad leverage point*. Ketika data mengandung pencilan *bad leverage point* 0%, rataan nilai R^2 ketiga metode untuk semua ukuran data berada pada kisaran nilai lebih besar dari 75%. Akan tetapi, ketika data mulai mengandung pencilan *bad leverage point*, rataan nilai R^2 metode MKT mengalami penurunan hingga kisaran nilai 40%. Sebaliknya, rataan nilai R^2 metode LMS dan MM mengalami peningkatan hingga mencapai nilai mendekati 100%. Rataan nilai R^2 metode LMS berada pada nilai tertinggi

Tabel 8: Rataan nilai R^2 pada berbagai ukuran data dan persentase pencilan *good leverage point*.

Jumlah Data	Metode	Persentase Pencilan (%)					
		0	5	10	15	20	30
50	MKT	0.94	1.00	1.00	1.00	1.00	1.00
	LMS	1.00	1.00	1.00	1.00	1.00	1.00
	MM	0.94	0.99	1.00	1.00	1.00	1.00
200	MKT	0.94	1.00	1.00	1.00	1.00	1.00
	LMS	1.00	1.00	1.00	1.00	1.00	1.00
	MM	0.94	1.00	1.00	1.00	1.00	1.00
1000	MKT	0.94	1.00	1.00	1.00	1.00	1.00
	LMS	0.99	1.00	1.00	1.00	1.00	1.00
	MM	0.94	1.00	1.00	1.00	1.00	1.00

dibandingkan dengan kedua metode lainnya untuk keseluruhan ukuran data. Kehadiran pencilan *bad leverage point* dalam data sangat memengaruhi kinerja metode MKT tetapi tidak dengan metode LMS dan MM.

Tabel 9: Rataan nilai R^2 pada berbagai ukuran data dan persentase pencilan *bad leverage point*.

Jumlah Data	Metode	Persentase Pencilan (%)					
		0	5	10	15	20	30
50	MKT	0.80	0.47	0.46	0.45	0.44	0.40
	LMS	0.85	0.98	0.98	0.99	0.99	0.99
	MM	0.83	0.91	0.93	0.94	0.96	0.98
200	MKT	0.80	0.48	0.47	0.45	0.44	0.40
	LMS	0.82	0.97	0.98	0.99	0.99	0.98
	MM	0.83	0.90	0.94	0.96	0.98	0.98
1000	MKT	0.80	0.48	0.46	0.45	0.44	0.40
	LMS	0.82	0.97	0.98	0.99	0.99	0.99
	MM	0.83	0.91	0.96	0.97	0.98	0.98

4. Rataan nilai RMSE

(a) *Vertical outlier*

Rataan nilai RMSE metode MKT, LMS, dan MM pada Tabel 10 menunjukkan bahwa metode MKT memiliki rataan nilai RMSE kecil mendekati nilai 0 hanya ketika tidak terdapat pencilan pada data. Selanjutnya ketika data terkontaminasi oleh adanya pencilan *vertical outlier* maka nilai rataan RMSE oleh MKT mengalami peningkatan yang sangat tinggi, sedangkan rataan nilai RMSE LMS dan MM berada pada nilai yang tergolong kecil.

(b) *Good leverage point*

Tabel 11 memperlihatkan bahwa adanya pencilan tipe *good leverage point* tidak terlalu berpengaruh buruk pada kinerja masing-masing metode. Rataan nilai RMSE ketiga metode tetap berada pada nilai mendekati 0. Tidak ada

Tabel 10: Rataan nilai RMSE pada berbagai ukuran data dan persentase pencilan *vertical outlier*.

Jumlah Data	Metode	Persentase Pencilan (%)					
		0	5	10	15	20	30
50	MKT	0.50	11.89	18.19	21.05	24.26	27.78
	LMS	0.42	0.44	0.46	0.48	0.49	0.51
	MM	0.49	0.52	0.58	0.62	0.70	0.89
200	MKT	0.50	13.12	18.05	21.48	24.07	27.57
	LMS	0.47	0.48	0.49	0.50	0.50	0.51
	MM	0.50	0.53	0.58	0.63	0.69	0.87
1000	MKT	0.50	13.09	18.02	21.44	24.02	27.51
	LMS	0.48	0.48	0.49	0.50	0.50	0.50
	MM	0.50	0.53	0.57	0.62	0.68	0.86

peningkatan rataan nilai RMSE yang sangat tinggi untuk masing-masing metode. Rataan nilai RMSE untuk metode MM hampir sama dengan rataan nilai RMSE metode MKT, sedangkan rataan nilai RMSE metode LMS berada pada nilai terendah dibandingkan dengan metode MKT maupun MM.

Tabel 11: Rataan nilai RMSE pada berbagai ukuran data dan persentase pencilan *good leverage point*.

Jumlah Data	Metode	Persentase Pencilan (%)					
		0	5	10	15	20	30
50	MKT	0.50	0.50	0.50	0.50	0.50	0.50
	LMS	0.43	0.43	0.43	0.43	0.43	0.43
	MM	0.49	0.49	0.49	0.49	0.49	0.49
200	MKT	0.50	0.50	0.50	0.50	0.50	0.50
	LMS	0.47	0.47	0.47	0.47	0.47	0.47
	MM	0.50	0.50	0.50	0.50	0.50	0.50
1000	MKT	0.50	0.50	0.50	0.50	0.50	0.50
	LMS	0.48	0.48	0.48	0.48	0.48	0.48
	MM	0.50	0.50	0.50	0.50	0.50	0.50

(c) *Bad leverage point*

Rataan nilai RMSE metode MKT, LMS, dan MM untuk setiap ukuran data dan persentase pencilan *bad leverage point* disajikan pada Tabel 12. Ketika persentase pencilan 0%, rataan nilai RMSE ketiga metode berada pada kisaran nilai cukup kecil untuk semua ukuran data. Selanjutnya, seiring meningkatnya persentase pencilan *bad leverage point* dalam data, rataan nilai RMSE masing-masing metode meningkat. Peningkatan tertinggi dialami oleh metode MKT. Peningkatan rataan nilai RMSE metode LMS dan MM tidak nyata, dengan rataan nilai RMSE terkecil adalah rataan nilai RMSE yang dihasilkan oleh metode LMS.

3.2 Kajian Data Aktual

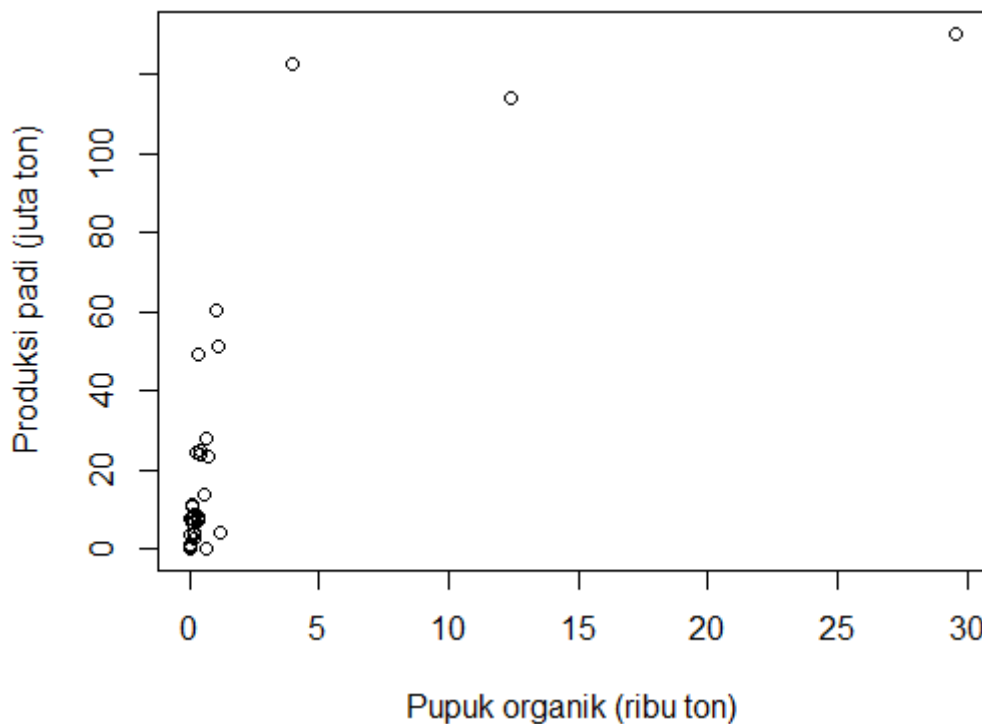
1. Data

Plot tebaran data yang digunakan disajikan pada Gambar 2. Plot tebaran data

Tabel 12: Rataan nilai RMSE pada berbagai ukuran data dan persentase pencilan bad leverage point.

Jumlah Data	Metode	Persentase Pencilan (%)					
		0	5	10	15	20	30
50	MKT	1.00	4.66	5.64	6.45	7.17	9.01
	LMS	0.56	0.67	0.71	0.76	0.81	0.93
	MM	0.86	0.87	0.87	0.88	0.90	1.07
200	MKT	1.00	3.80	5.14	6.20	7.11	8.63
	LMS	0.81	0.86	0.89	0.92	0.94	1.03
	MM	0.87	0.87	0.88	0.91	0.96	1.12
1000	MKT	1.00	3.78	5.12	6.17	7.07	8.59
	LMS	0.89	0.91	0.93	0.94	0.94	1.10
	MM	0.87	0.87	0.90	0.95	1.00	1.14

pada Gambar 2 menunjukkan secara eksplorasi terlihat bahwa terdapat beberapa data yang terletak jauh dari kumpulan data lainnya. Data-data tersebut berpotensi menjadi pencilan. Oleh karena itu, perlu dilihat nilai *standardized residual* untuk mengidentifikasi pencilan *vertical*, sedangkan pencilan *good leverage point* diidentifikasi dengan menggunakan plot nilai diagonal matriks *hat*.

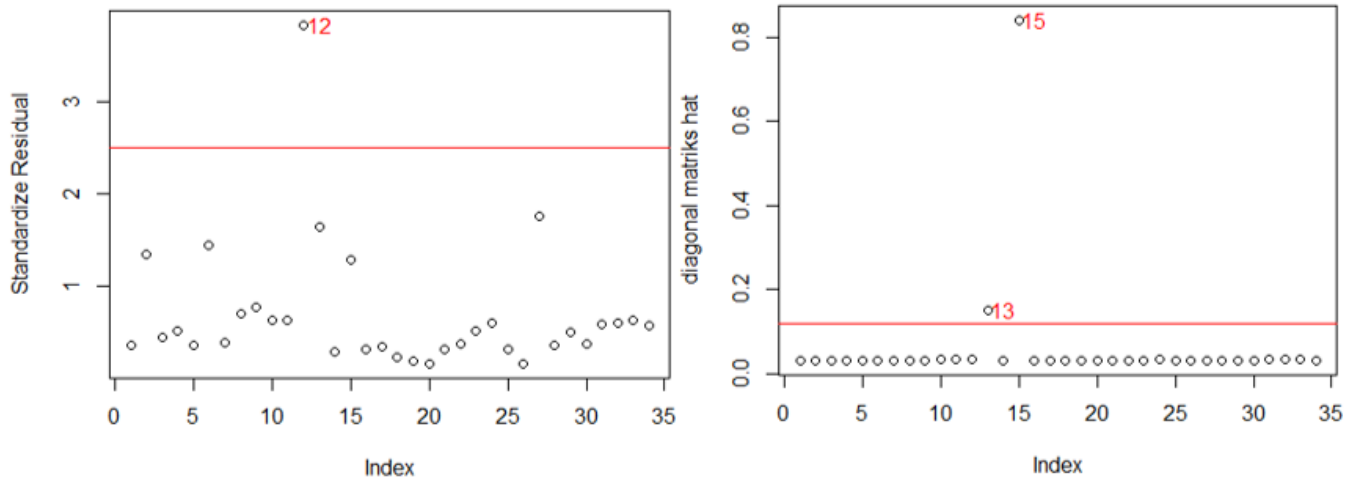


Gambar 2: Plot tebaran data aktual.

2. Pengecekan pencilan

Nilai *standardized residual* berdasarkan plot pada Gambar 3(a) berada pada selang nilai 0 hingga 4. Mayoritas data memiliki nilai *standardized residuals* yang tersebar pada nilai 0-2. Akan tetapi, terdapat 1 amatan yang memiliki nilai di atas 2.5, yaitu

data ke-12 sehingga data ini merupakan data pencilan *vertical outlier*. Di sisi lain, berdasarkan plot tebaran nilai diagonal matriks *hat* pada Gambar 3(b), terdapat 2 data yang memiliki nilai diagonal matriks *hat* lebih besar dari 0.12 (dua kali rata-rata nilai diagonal matriks *hat*), yaitu data ke-13 dan 15. Kedua data tersebut merupakan data pencilan *good leverage point*. Tidak ada data yang teridentifikasi sebagai pencilan *vertical outlier* sekaligus *good leverage point*, sehingga tidak ada data pencilan *bad leverage point* pada data aktual yang digunakan.



Gambar 3: (a) Plot *standardize residual* dan (b) plot diagonal matriks *hat*.

3. Analisis regresi

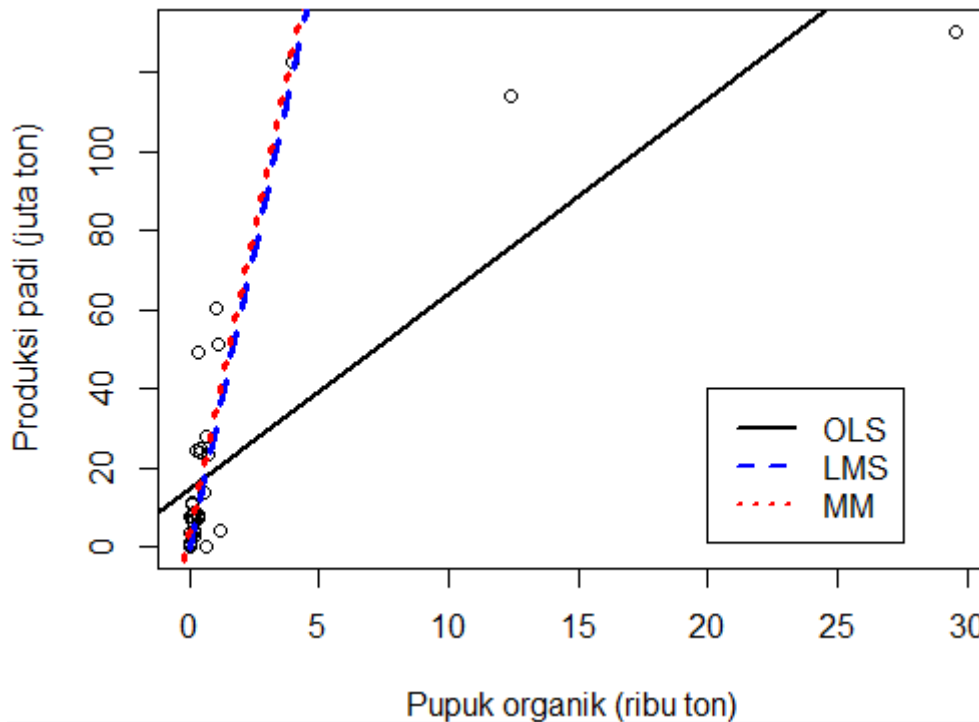
Plot garis persamaan regresi metode MKT, LMS, dan MM disajikan pada Gambar 4. Garis berwarna hitam, biru, dan merah pada Gambar 4 secara berturut-turut merupakan garis persamaan regresi dengan menggunakan metode MKT, LMS, dan MM. Gambar 4 memperlihatkan bahwa garis regresi LMS dan MM cenderung hampir sama, sedangkan garis regresi MKT jauh berbeda dengan LMS dan MM. Garis regresi MKT tertarik oleh adanya amatan yang terpencil. Hal tersebut didukung oleh nilai dugaan parameter ketiga metode. Tabel 13 menyajikan nilai koefisien regresi atau dugaan parameter regresi yang dihasilkan ketiga metode tersebut. Koefisien regresi yang dihasilkan oleh metode LMS dan MM cenderung memiliki nilai yang tidak berbeda jauh, sedangkan koefisien regresi oleh MKT memiliki karakteristik yang berbeda. Nilai dugaan parameter β_0 metode LMS dan MM lebih kecil daripada nilai dugaan parameter β_1 . Sebaliknya, Nilai dugaan parameter β_0 oleh MKT lebih besar dibandingkan nilai dugaan parameter β_1 .

Tabel 13: Koefisien regresi.

Metode	$\hat{\beta}_0$	$\hat{\beta}_1$
MKT	14.62	4.93
LMS	-1.01	30.42
MM	2.53	30.87

4. Pengecekan Asumsi Regresi

Nilai-p uji fomal asumsi regresi disajikan pada Tabel 14. Nilai-p *KS-test* dan *Breusch-Pagan* MKT berturut-turut sebesar 0.0124 dan 0.0195, kurang dari taraf



Gambar 4: Garis regresi metode MKT, LMS, dan MM.

nyata 0.05 sehingga asumsi kenormalan dan kehomogenan ragam sisaan pendugaan menggunakan MKT tidak terpenuhi. Asumsi kebebasan sisaan pendugaan menggunakan metode MKT terpenuhi yang ditunjukkan oleh nilai-p *Runst-test* lebih dari 0.05, yaitu sebesar 0.1278. Nilai-p *KS-test*, *BP-test*, dan *Runs-test* metode kekar LMS dan MM berada pada nilai lebih dari 0.05. Hal tersebut berarti asumsi kenormalan, kehomogenan, dan kebebasan sisaan metode LMS dan MM terpenuhi.

Tabel 14: Nilai-p uji asumsi regresi data aktual.

Metode	KS-test	BP-test	Runs-test
MKT	0.0124	0.0195	0.1278
LMS	0.2791	0.3959	0.4860
MM	0.4421	0.4017	0.1635

5. Kriteria kebaikan model

Nilai RMSE dan R^2 MKT, LMS, dan MM dalam pendugaan model regresi data produksi padi yang mengandung pencilan tersebut telah disajikan pada Tabel 15. Metode yang menghasilkan nilai RMSE terbesar hingga terkecil secara berturut-turut adalah metode MKT sebesar 23.15, metode MM sebesar 6.78, dan metode LMS sebesar 4.44. Nilai R^2 tertinggi dihasilkan oleh metode LMS, yaitu sebesar 98%, kemudian diikuti oleh metode MM yang menghasilkan nilai R^2 sebesar 96%, sedangkan MKT menghasilkan nilai R^2 paling rendah yaitu sebesar 58%.

Metode yang baik digunakan dalam analisis regresi adalah metode yang menghasilkan nilai RMSE kecil dan sebaliknya menghasilkan nilai R^2 terbesar.

Tabel 15 menunjukkan bahwa metode LMS pada penelitian ini menghasilkan nilai RMSE terkecil yaitu sebesar 4.44 dan nilai R^2 tertinggi sebesar 98%. Oleh karena itu, metode LMS merupakan metode terbaik dalam analisis regresi sederhana untuk mengetahui pengaruh data jumlah penggunaan pupuk organik terhadap total produksi padi di provinsi-provinsi Indonesia tahun 2017. Hasil simulasi regresi linear menunjukkan bahwa pada data ukuran kecil, yaitu ukuran 50 amatan dengan kandungan pencilan 5% hingga 10%, metode LMS menghasilkan rata-rata RMSE terkecil dan rata-rata nilai R^2 terbesar. Hal tersebut berarti kesimpulan yang diperoleh dari pendugaan parameter terhadap data aktual produksi padi tahun 2017 selaras dengan hasil pendugaan parameter regresi melalui simulasi. Persamaan garis regresi terbaik yang dibentuk oleh metode LMS adalah sebagai berikut:

$$\widehat{ProduksiPadi}_i = -1.0100 + 30.4200PupukOrganik_i$$

Dugaan β_0 ($\widehat{\beta}_0$) bernilai negatif, yaitu sebesar -1.0100. Interpretasi dari nilai $\widehat{\beta}_0$ tersebut adalah bahwa ketika tidak ada pupuk organik yang digunakan, dugaan rata-rata produksi padi yang dihasilkan akan mengalami penurunan sebesar 1.0100 juta ton. Di sisi lain, nilai $\widehat{\beta}_1$ bernilai positif yaitu sebesar 30.4200. Nilai tersebut menggambarkan bahwa setiap penambahan seribu ton penggunaan jumlah pupuk organik maka produksi padi akan mengalami peningkatan sebesar 30.4200 juta ton.

Tabel 15: Nilai RMSE dan R^2

Metode	RMSE	R^2
MKT	23.1500	0.5800
LMS	4.4400	0.9800
MM	6.7800	0.9600

4 Simpulan

Hasil simulasi menunjukkan bahwa metode kekar LMS dan MM sangat baik digunakan dalam melakukan pendugaan parameter regresi meskipun terdapat hingga 30% pencilan, baik pencilan *vertical outlier*, *good leverage point*, maupun *bad leverage point* pada semua ukuran data. MKT tidak kekar terhadap adanya pencilan *vertical outlier* dan *bad leverage point*, tetapi sangat baik digunakan ketika tipe pencilan dalam data adalah *good leverage point* untuk semua kombinasi ukuran data dan persentase pencilan. Oleh karena itu, apabila tidak ada pencilan lain selain *good leverage point* maka MKT lebih disarankan untuk digunakan, sedangkan LMS dan MM tepat digunakan ketika data mengandung pencilan *vertical outlier* dan atau *bad leverage point*.

Penerapan MKT dan metode regresi kekar LMS serta MM pada data aktual total produksi padi di Indonesia tahun 2017 menghasilkan kesimpulan bahwa metode kekar LMS merupakan metode terbaik dalam pendugaan parameter regresi. Hal tersebut diindikasikan oleh nilai RMSE yang paling rendah, yaitu sebesar 4.44 serta nilai R^2 yang paling tinggi, yaitu mencapai 98%.

Daftar Pustaka

- Almetwally, E. & Almongy, H. (2018). Comparison between m estimation, s estimation, and mm estimation methods of robust estimation with application and simulation. *International Journal of Mathematical Archive*, 9(11): 1–9.
- Atilgan, Y. & Gunay, S. (2011). Least median of squares solution of multiple linear regression models through the origine. *Communication in Statistics—Theory and Methods*, 40: 4125–4137.
- Chen, C., SAS, I., & Cary, N. (2002). Robust regression and outlier detection with the robustreg procedure. *SUGI Proceedings*, (): 265–270.
- D’Urso, P. & Massari, R. (2013). Weighted least squares and least median squares estimation for the fuzzy linear regression analysis. *Metron*, 71: 279–306.
- Montgomery, D., Peck, E., & Vining, G. (2012). *Introduction to Linear Regression Analysis*. New York(USA): Wiley.
- Oktarinanda, A. (2014). Perbandingan efisiensi metode least trimmed square (Its) dan metode least median square (lms) dalam penduga parameter regresi robust. *Jurnal Statistik*, 2(3): 177–180.
- Rousseeuw, P. & Hubert, M. (1997). Recent developments in progress. *L1-Statistical Procedure and Related Topics*, 31: 201–214.