 http://dx.doi.org/10.29313/.v0i0.29248

Perbandingan Estimasi *Least Trimmed Square* (LTS), Estimasi *Scale* (S), dan Estimasi *Method of Moment* (MM) pada Analisis Regresi *Robust*

Nur Fatinah Rahman*, Lisnur Wachidah

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*nurfatiinah@gmail.com, wachidah.lisnur07@gmail.com

Abstract. Multiple linear regression is a statistical method that studies and measures the statistical relationship between two or more variables that contain causality in order to obtain a regression model. One method of estimating the parameters of the multiple linear regression model is the *Ordinary Least Square*. The method must pass several regression assumptions. The residual normality test is one of the regression assumptions that sometimes cannot be fulfilled. The reason is because there are outliers in the data. Therefore, to overcome these problems, a *robust regression* method is needed to overcome outliers. *Robust regression* is a method used when the distribution of the residuals is not normal and or there are some outliers that affect the model. The types of estimation methods found in robust regression are *Least Trimmed Square* (LTS) estimation, *Scale* (S) estimation, and *Method Of Moment* (MM) estimation. The purpose of this study is to obtain the best robust regression method resulting from the comparison of the *LTS* estimation method, *S* estimation, and *MM* estimation based on the smallest *Residual Standard Error* (RSE) value and largest *adjusted R-square* value. The data in this study is agricultural data for the Indonesian country in 2019. This data includes three variables, namely the agricultural product (Y), the agricultural land area (X_1) and the amount of rainfall (X_2). The results of this study indicate that the robust regression estimation of *LTS* has a smaller *RSE* value and a larger *adjusted R-square* value than the *S* estimation method and the *MM* estimation . *LTS* estimation is better in estimating regression parameters than the S estimation method and *MM* estimation.

Keywords: Outlier, Robust regression, LTS-estimation, S-estimation, MM-estimation.

Abstrak. Regresi linear berganda merupakan salah satu metode statistika yang mempelajari dan mengukur hubungan statistik antara dua atau lebih variabel yang mengandung sebab akibat dengan tujuan untuk mendapatkan model regresi. Salah satu metode penduga parameter model regresi linier berganda adalah metode kuadrat terkecil (MKT). Metode tersebut harus memenuhi beberapa asumsi regresi. Uji normalitas *residual* salah satu asumsi regresi yang cukup sering tidak dapat terpenuhi. Alasan asumsi tersebut tidak dapat terpenuhi karena terdapat *outlier* pada data. Oleh karena itu, dalam mengatasi permasalahan tersebut diperlukan metode regresi *robust* untuk mengatasi pencilan. Regresi *robust* adalah metode yang digunakan ketika distribusi dari *residual* tidak normal dan atau adanya beberapa pencilan yang berpengaruh pada model. Beberapa metode estimasi yang terdapat pada regresi robust adalah estimasi *Least Trimmed Square* (LTS), estimasi *Scale* (S), dan estimasi *Method Of Moment* (MM). Adapun tujuan dari penelitian ini untuk menetukan metode regresi *robust* terbaik yang dihasilkan dari perbandingan metode estimasi LTS , estimasi S , dan estimasi MM berdasarkan nilai *Residual Standard Error* (RSE) terkecil dan nilai *adjusted R-square* terbesar. Data yang digunakan dalam penelitian ini adalah data pertanian dan curah hujan di Indonesia tahun 2019. Data ini meliputi tiga variabel yaitu variabel produksi padi (Y), luas panen (X_1) dan curah hujan (X_2). Hasil penelitian ini

menunjukkan bahwa regresi *robust* estimasi LTS memiliki nilai RSE yang lebih kecil dan *adjusted R-square* yang lebih besar dibandingkan metode estimasi S maupun estimasi MM sehingga metode estimasi *Least Trimmed Square* (LTS) lebih baik dalam mengestimasi parameter regresi dibandingkan metode estimasi *Scale* (S) maupun estimasi *Method Of Moment* (MM).

Kata Kunci: *Outlier*, Regresi *robust*, estimasi LTS, estimasi S, estimasi MM.

1. Pendahuluan

Analisis regresi linear berganda merupakan salah satu metode statistika yang mempelajari pola dan mengukur hubungan statistik antara dua atau lebih variabel yang mengandung sebab akibat. Salah satu tujuan analisis regresi adalah mengestimasi koefisien regresi dalam model regresi. Salah satu metode penduga parameter model regresi linier berganda adalah metode kuadrat terkecil (MKT). Penggunaan MKT harus memenuhi beberapa asumsi regresi. Pada kenyataannya, tidak jarang ditemukan hal-hal yang menyebabkan tidak terpenuhinya asumsi regresi tersebut sehingga penggunaan MKT akan memberikan kesimpulan yang bersifat kurang baik. Asumsi yang sering tidak terpenuhi adalah asumsi normalitas residual. Terdapatnya pencilan (*outlier*) menjadi salah satu penyebab tidak terpenuhinya asumsi ini sehingga diperlukan metode lain untuk menangani *outlier*, metode tersebut adalah metode regresi *robust*. Metode ini dapat mengatasi *outlier* dengan mencocokkan model regresi terhadap sebagian besar data tanpa menghapus data outlier (Rousseeuw dan Leroy, 1987). Terdapat beberapa metode estimasi pada regresi robust diantaranya estimasi *Least Trimmed Square* (LTS), *Scale* (S), dan *Method Of Moment* (MM). Ketiga metode estimasi tersebut memiliki nilai *breakdown point* yang tinggi dibandingkan dengan metode estimasi lainnya. Nilai breakdown point ketiga metode estimasi tersebut adalah 50%. Oleh karena itu, berdasarkan kesamaan nilai *breakdown point* dari ketiga estimasi tersebut, maka tujuan pada penelitian untuk melakukan perbandingan dan mencari metode estimasi mana yang lebih baik digunakan dalam mengestimasi data yang mengandung *outlier*. Dalam menentukan metode terbaik, penulis menggunakan nilai *Residual Standard Error* (RSE) dan *adjusted R-square*. Semakin kecil nilai RSE dan semakin besar nilai *adjusted R-square* maka semakin baik metode estimasi.

2. Metodologi

Analisis Regresi Linier Berganda

Analisis regresi linier berganda adalah hubungan secara linier antara dua atau lebih variabel independen dengan variabel dependen. Analisis regresi linier berganda dilakukan dengan cara menetapkan persamaan sebagai berikut :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (1)$$

Keterangan :

Y = Variabel dependen produksi padi (ton)
β_0 = *Intercept*
β_1 = Koefisien regresi pertama
β_2 = Koefisien regresi kedua
X_1 = Variabel independen luas panen (ha)
X_2 = Variabel independen curah hujan (mm)
β_k = Koefisien regresi ($i = 1,2, \dots, k$)
X_k = Variabel independen ($i = 1,2, \dots, k$)
ε = *Error* atau galat

Metode Kuadrat Terkecil (MKT)

Ada beberapa metode untuk menduga parameter regresi, diantaranya adalah metode kuadrat terkecil (MKT=OLS, *ordinary least square*), penduga kemungkinan maksimum (PKM = MLE, *maximum likelihood estimator*), dan lain-lain. Metode kuadrat terkecil pada prinsipnya adalah meminimumkan jumlah kuadrat sisaan (JKS). Sehingga diperoleh nilai estimasi model regresi linear metode kuadrat terkecil (MKT) sebagai berikut :

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{ip} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}x_{ip} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2}x_{ip} + \cdots + \hat{\beta}_p \sum_{i=1}^{n} x_{i^2p} = \sum_{i=1}^{n} x_{ip}y_i \quad (2)$$

Uji Asumsi Regresi

Uji Normalitas Residual

Uji normalitas residual digunakan untuk menguji apakah model regresi variabel independen dan variabel dependen keduanya mempunyai distribusi normal atau tidak (Ghozali, 2009) . Salah satu uji yang dapat digunakan adalah uji *Kolmogorov Smirnov*.

Adapun dalam pengujian ini menggunakan hipotesis sebagai berikut :

Hipotesis :

H_0 : *Residual* berdistribusi normal

H_1 : *Residual* tidak berdistribusi normal

Statistik Uji :

$$D = max|F_0(X_i) - F_n(X_i)| \quad (3)$$

Kriteria Uji :

Apabila nilai $D > D_{tabel}$ atau nilai $P_{value} < \alpha$ maka H_0 ditolak yang artinya asumsi normalitas residual tidak terpenuhi atau data tidak berdistribusi normal.

Uji Heteroskedastisitas Varians

Uji heteroskedastisitas varians adalah untuk melihat apakah terdapat ketidaksamaan varians dari residual satu ke pengamatan yang lain. Salah satu uji statistik yang dapat digunakan adalah uji *Glejser*. Hipotesis dalam pengujian ini adalah (Gujarati, 2003:406) :

Hipotesis :

H_0 : Varians residual homogen atau tidak terjadi heteroskedastisitas

H_1 : Varians residual tidak homogen atau terjadi heteroskedastisitas

Statistik Uji :

$$|\hat{e}_i| = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad (4)$$

Jika terdapat koefesien regresi b yang signifikan secara statistika, maka diindikasi terjadi heteroskedastisitas. Namun jika tidak terdapat koefisien regresi b yang siginifikan secara statistika, maka asumsi homoskedastisitas terpenuhi (Gujarati, 2004:406)

Uji Multikolinieritas

Uji ini bertujuan untuk mengetahui ada tidaknya hubungan antara variabel independen atau variabel bebas..

Pengujian ada tidaknya masalah multikolineritas salah satunya dengan melihat nilai *Variance Inflation Factor* (VIF). Pengujian ini menggunakan hipotesis sebagai berikut :

Hipotesis :

H_0 : Tidak terdapat multikolineritas

H_1 : Terdapat multikolinieritas

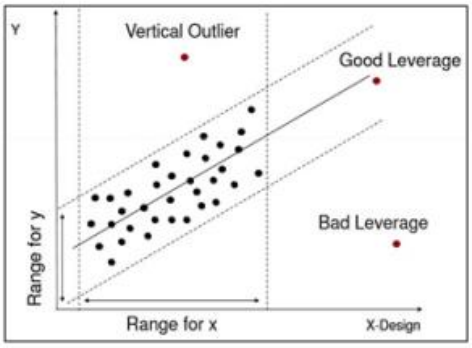
Perhitungan nilai *VIF* dapat dilakukan dengan menggunakan persamaan sebagai berikut (Montgomery, Peck & Vinning, 1982) :

$$VIF = \frac{1}{1-R^2} \quad (5)$$

Masalah multikolinieritas dapat terjadi ketika nilai $VIF \geq 10$ (Ghozali, 2009).

Data Outlier

Data *Outlier* adalah data yang tidak mengikuti pola umum pada model regresi yang dihasilkan atau tidak mengikuti pola data secara keseluruhan. Beberapa tipe data *outlier* (**Gambar 1**.) yang umum ditemukan dalam analisis sebagai berikut (Verardi, 2008) :



Gambar 1. Tipe Data Outlier

Berdasarkan Gambar 1, pada analisis regresi, terdapat 3 tipe *outlier* yang berpengaruh terhadap estimasi kuadrat terkecil yaitu sebagai berikut (Soemartini, 2007):

1. *Outlier* Vertikal (*Vertical Outlier*) merupakan semua pengamatan yang terpencil pada variabel dependen, tetapi tidak terpencil pada variabel independen.
2. *Good Leverage Point* merupakan pengamatan yang terpencil pada variabel independen tetapi terletak dekat dengan garis regresi.
3. *Bad leverage point* merupakan pengamatan yang terpencil pada variabel independen dan terletak jauh dari garis regresi.

Identifikasi Data Outlier

Beberapa metode identifikasi outlier dalam sebuah analisis adalah sebagai berikut:

1. Metode $DFBETAS_{ij}$ (Difference in Beta)

   $DFBETAS_{ij}$ merupakan ukuran pengaruh dengan melihat selisih nilai taksiran koefisien regresi ke-j, β_j dengan nilai taksiran koefisien regresi ke-j saat observasi ke-i dikeluarkan ($\beta_{j(i)}$). Secara matematis, $DFBETAS_{ij}$ dapat didefinisikan seperti persamaan berikut ini (Montgomery & Peck, 1982) :

   $$DFBETAS_{ij} = \frac{\widehat{\beta}_j - \widehat{\beta}_{j(i)}}{\sqrt{s^2_{(i)} c_{ij}}} \qquad (6)$$

   Berdasarkan persamaan (8), pendeteksian outlier menggunakan DFBETAS adalah ketika nilai DFBETAS $> \frac{2}{\sqrt{n}}$ untuk ukuran sampel yang besar.

2. Metode $DFFITS$ (Difference Fitted Value FITS)

   Metode ini menampilkan nilai perubahan dalam nilai yang diprediksi ketika pengamatan ke-i dihapuskan pada data penelitian yang sudah distandarkan (Soemartini, 2007).

   Statistik Uji :

   Rumus perhitungan $DFFITS$ adalah sebagai berikut :

   $$DFFITS = t_i \left(\frac{h_{ii}}{1-h_{ii}}\right)^{\frac{1}{2}} \qquad (7)$$

   Data dikatakan *outlier* apabila nilai $|DfFITS| > 2\sqrt{\frac{p}{n}}$

3. Metode *Cook's Distance*

   Cook's Distance merupakan suatu ukuran untuk mendeteksi besarnya pengaruh adanya outlier terhadap semua estimasi koefisien regresi, yaitu

   $$D_i = \frac{e_i^2 h_{ii}}{pMSE(1-h_{ii}^2)} \qquad (8)$$

   Suatu data disebut *outlier* apabila nilai $D_i > \frac{4}{n}$ (Yaffe, 2002:44)

4. Metode Nilai Pengaruh (*Leverage Values)*

   Metode ini mengukur pengaruh suatu observasi terhadap besarnya estimasi parameter. Hal ini dapat dilihat dari jarak nilai X semua observasi. Nilai leverage untuk linier sederhana dapat ditentukan sebagai berikut (Wijaya, 2009):

$$Leverage\ (h_{ii}) = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{(n-1)s_x^2} \quad (9)$$

5. Metode $R_{Student}$

Metode $R_{Student}$ ini memiliki perhitungan yang hampir sama dengan *studentized residuals*, tetapi varians yang digunakan untuk perhitungan saat observasi ke-i dikeluarkan dari pengamatan (Wijaya, 2009). Hipotesis yang digunakan dalam pengujian ini adalah :

H_0 : Pengamatan ke-i bukan merupakan *outlier*

H_1 : Pengamatan ke-i merupakan *outlier*

Statistik Uji :

Metode $R_{student}$ didefinisikan dengan lambang t_i dengan rumus sebagai berikut (Chatterjee dan Hadi, 1986) :

$$t_i = \frac{e_i}{\sqrt{\hat{\sigma}^2_{(i)}(1-h_{ii})}} \quad (10)$$

Kriteria Uji

H_0 ditolak ketika $|t_i| > t_{\frac{\propto}{2}, n-k-1}$ yang artinya pengamatan ke-i merupakan *outlier* (Montgomery dan Peck, 1982).

Regresi *Robust*

Regresi *robust* merupakan alat yang penting untuk menganalisis data yang terkontaminasi oleh *outlier* dan memberikan hasil yang lebih fleksibel. Regresi robust tetap menggunakan seluruh data, tetapi dengan memberikan bobot yang kecil untuk data pencilan (Soemartini, 2007). Pada regresi robust terdapat beberapa metode estimasi yang dapat digunakan, yakni (1) estimasi LTS, (2) estimasi S, (3) estimasi MM (Chen, 2002).

Estimasi Least Trimmed Square (LTS)

Estimasi *LTS* merupakan metode penduga regresi *robust* untuk meminimumkan jumlah kuadrat *h* residual (fungsi objektif) (Akbar dan Maftukhah, 2007). Persamaan metode untuk penduga LTS (β) dinyatakan dalam bentuk rumus sebagai berikut (Rousseeuw dan Leroy, 1987) :

$$\hat{\beta}_{LTS} = \arg min \sum_{i=1}^{h} e_i^2 \quad (11)$$

dengan :

$$h = \left[\frac{n}{2}\right] + \left[\frac{k+2}{2}\right]; e_i = \left(\hat{Y}_i - x_i\hat{\beta}_0\right) \quad (12)$$

dimana :

Estimasi *Scale* (S)

Estimasi *Scale* (S) akan meminimumkan jumlah kuadrat *error* pada persamaan umum regresi linear. Estimasi *S* dirumuskan persamaan sebagai berikut (Chen, 2002) :

$$\hat{\beta}_s = arg\beta \min S_s \left(n \sum_{i=0}^{n}\left(e_i^2\right) - , e_2, \dots, e_n\right) \quad (13)$$

dengan menentukan nilai estimator skala *robust* ($\hat{\sigma_s}$) yang minimum dan memenuhi :

$$min \sum_{i=1}^{n} \rho\left(\frac{y_i \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}}\right) \quad (14)$$

dengan

$$\hat{\sigma} = \sqrt{\frac{n\sum_{i=0}^{n}(e_i^2) - n\sum_{i=0}^{n}(e_i)^2}{n(n-1)}} \quad (15)$$

Estimasi Method Of Moment (MM)

Estimasi *Method Of Moment* (MM) merupakan salah satu metode regresi *robust* yang dikenalkan oleh Yohai pada tahun 1987. Persamaan dari estimasi *MM* adalah sebagai berikut (Chen, 2002) :

$$\hat{\beta}_{MM} = arg_\beta min \sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}_s}\right) = arg_\beta min \sum_{i=1}^{n} \rho\left(\frac{Y_i - \sum_{j=0}^{k} X_i\beta_j}{\hat{\sigma}_s}\right) \quad (16)$$

Residual Standard Error (RSE)

Residual Standard Error (RSE) merupakan suatu ukuran ketepatan perhitungan dengan mengkuadratkan setiap residual untuk setiap penduga dalam sebuah kumpulan data dan

kemudian diperoleh rata-rata atau nilai tengah jumlah kuadrat tersebut. Persamaan RSE dirumuskan sebagai berikut (Sembiring, 1995) :

$$RSE = \frac{JK_{Res}}{db} = \frac{\sum_{i=1}^{n} \hat{e}_i^2}{(n-k-1)} = \frac{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}{(n-k-1)} \quad (17)$$

Semakin kecil nilai RSE, maka semakin baik kecocokan suatu model dengan data karena nilai prediksi dari *Y* semakin mendeteksi nilai sebenarnya (Sembiring, 1995).

R^2_{adj} (Koefisien Determinasi Terkoreksi)

Koefisien determinasi terkoreksi (R^2_{adj}) berkisar antara $0 \leq R^2_{adj} \leq 1$. Semakin mendekati nilai 1, maka nilai R^2_{adj} maka semakin baik prediksi model regresi. Koefisien determinasi (R^2_{adj}) dirumuskan dalam persamaan (2.31) sebagai berikut (Drapper dan Smith, 1992) :

$$R^2_{adj} = 1 - \frac{(n-1)}{(n-k-1)}(1 - R^2) \quad (18)$$

Data

Data yang digunakan adalah data sekunder dari *website* Badan Pusat Statistik Indonesia. Operasional variabel penelitian ini terdiri atas tiga variabel yaitu dua variabel independen dan satu variabel dependen. Variabel independen (X) dalam penelitian ini adalah luas panen (X_1) dan curah hujan (X_2) di Indonesia tahun 2019 sedangkan variabel dependen (Y) dalam penelitian ini adalah produksi padi di Indonesia tahun 2019. Data dalam penelitian ini disajikan sepada **Tabel 1** sebagai berikut :

Tabel 1. Data Pertanian dan Curah Hujan Indonesia Tahun 2019

Provinsi	Produksi Padi (Y)	Luas Panen (X_1)	Curah Hujan (X_2)
Aceh	1714437,60	310012,46	1437,20
Sumatera Utara	2078901,59	413141,24	1884,30
Sumatera Barat	1482996,01	311671,23	6100,30
…	…	…	…
…	…	…	…
Papua Barat	29943,56	5080,15	2158,30
Papua	235339,51	54131,72	2171,40

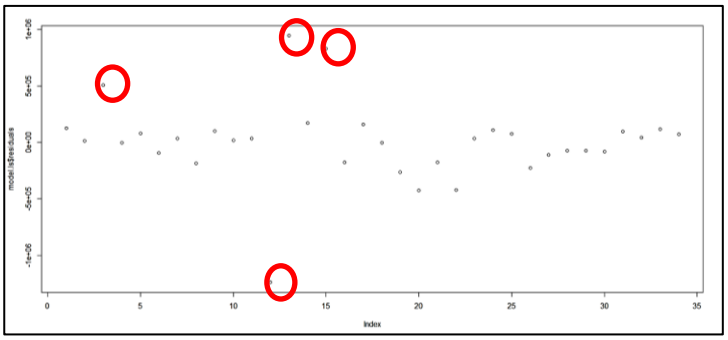
3. Pembahasan dan Diskusi

Metode Kuadrat Terkecil (MKT)

Dengan menggunakan bantuan *software R 4.1.0* diperoleh model persamaan regresi yang terbentuk dari estimasi nilai parameter MKT sebagai berikut :

$$\hat{y} = 173200 + 5.186x_1 - 133.4x_2$$

Diperoleh nilai *Residual Standard Error* (RSE) sebesar 364900 artinya kesalahan dalam memprediksi *Y* sebesar 364900 dan nilai *adjusted* R^2 sebesar 0,9814 artinya 98,14% variasi *Y* dapat dijelaskan oleh X_1 dan X_2, sedangkan sisanya dijelaskan oleh variabel lain.Kemudian dari nilai *residual* yang diperoleh dibuatkan plot sebarannya untuk melihat apakah terdapat *outlier* pada data. Plot *residual* untuk Metode Kuadrat Terkecil (MKT) mayoritas *residual* terletak disekitar 0. Namun, ada beberapa *residual* yang terletak jauh dari 0. Hal ini mengidentifikasi bahwa terdapat *outlier* pada data . Dengan bantuan *software R 4.1.0*, diperoleh plot *residual* MKT pada **Gambar 2** sebagai berikut :



Gambar 2. Plot Residual Metode Kuadrat Terkecil (MKT)

Pada **Gambar 2**, terdapat *residual* yang nilainya paling besar diantara nilai *residual* yang ada pada observasi. Hal ini mengindikasikan adanya *outlier* pada data tersebut. Oleh karena itu, untuk menangani *outlier* digunakan regresi *robust*.

Uji Asumsi Regresi

Uji Normalitas Residual

Hipotesis :

H_0 : Sisaan berdistribusi normal

H_1 : Sisaan tidak berdistribusi normal

Statistik Uji :

Dengan bantuan *software R 4.1.0* diperoleh nilai $p_{\text{-value}}$ sebesar 0,0001038.

Dengan menggunakan tingkat signifikasi α =5% maka H_0 ditolak artinya *residual* atau sisaan tidak berdistribusi normal

Uji Heteroskedastisitas Varians

Hipotesis :

H_0 : Varians residual homogen atau tidak terjadi heteroskedastisitas

H_1 : Varians residual tidak homogen atau terjadi heteroskedastisitas

Statistik Uji :

Dengan menggunakan bantuan *software R 4.1.0* diperoleh nilai $p_{\text{-value}}$ sebesar 0,1001

Dengan menggunakan tingkat signifikasi α =5%, maka H_0 diterima artinya tidak terjadi heteroskedastisitas.

Uji Multikolinieritas

Hipotesis :

H_0 : Tidak terdapat multikolineritas

H_1 : Terdapat multikolinieritas

Statistik Uji :

Dengan menggunakan bantuan *software R 4.1.0* diperoleh nilai VIF untuk variabel independen (X_1 dan X_2) sebesar 1,174481

Karena variabel independen (X_1 dan X_2) menghasilkan nilai $VIF = 1{,}174481 < 10$ maka keputusan H_0 diterima artinya tidak terdapat multikolinearitas dalam penelitian ini.

Pendeteksian *Outlier*

Nilai kriteria ukuran *outlier* dihitung berdasarkan persamaan (7) sampai dengan persamaan (11). Pada kasus ini, menggunakan jumlah pengamatan (n) sebanyak 34 dan banyaknya parameter (k) sebanyak 3 variabel. Jadi, observasi dikatakan sebagai *outlier* apabila kriteria nilainya yaitu $|DfBETAS| > \frac{2}{\sqrt{n}} = \frac{2}{\sqrt{34}} = 0{,}342$, $|DfFITS| > 2\sqrt{\frac{k}{n}} = 2\sqrt{\frac{3}{34}} = 0{,}594$, $Cook's\ Distance\ (D_i) > \frac{4}{n} = \frac{4}{34} = 0{,}117$, $Leverage\ (h_{ii}) > \frac{2 \cdot k}{n} = \frac{2 \cdot 3}{34} = 0{,}176$, $|R_{student}\ (t_i)| > t_{0{,}025;31} = 2{,}040$

Berdasarkan perhitungan nilai kriteria ukuran *outlier* di atas, maka dapat diketahui observasi data yang merupakan *outlier*. Observasi yang terindikasi sebagai *outlier* disajikan pada **Tabel 2** sebagai berikut :

Tabel 2. Observasi Terindikasi Sebagai *Outlier*

No.	*DFBETA S0,i*	*DFBETA S1,i*	*DFBETAS 2,i*	*DFFIT Si*	*Cook's D_i*	*hii*	*R-student*
3	2,190249	0,350	2,820	2,89401	2,430	0,6011	2,357747056
12	2,586347	5,260	2,040	-6,11931	4,680	0,4158	7,252952022
13	0,509889	2,010	0,835	2,22233	1,190	0,2782	3,579417305
15	0,057684	1,610	0,276	1,71031	0,782	0,2530	2,938546339

Berdasarkan **Tabel 2**, diperoleh data yang terindikasi sebagai *outlier* yakni data ke 3, 12, 13 dan 15.

Regresi Robust Estimasi LTS (*Least Trimmed Square*)

Estimasi LTS dihitung menggunakan bantuan *software R 1.4.*0 sehingga diperoleh model persamaan regresi yang terbentuk dari estimasi nilai parameter metode estimasi *Least Trimmed Square* (LTS) adalah:

$$\hat{y} = -11690 + 5{,}001x_1 + 9{,}523x_2$$

Diperoleh nilai *Residual Standard Error* (RSE) sebesar 76910 artinya kesalahan dalam memprediksi *Y* sebesar 76910 dan *adjusted* R^2 sebesar 0,9953 artinya 99,53% variasi *Y* dapat dijelaskan oleh X_1 dan X_2, sedangkan sisanya dijelaskan oleh variabel lain.

Regresi Robust Estimasi *Scale* (S)

Estimasi S dihitung menggunakan bantuan software *R 1.4.0* sehingga diperoleh model persamaan regresi yang terbentuk dari estimasi nilai parameter metode estimasi S (*Scale*) adalah:

$$\hat{y} = -31060 + 5{,}019x_1 + 12{,}38x_2$$

Diperoleh nilai *Residual Standard Error* (RSE) sebesar 78450 artinya kesalahan dalam memprediksi *Y* sebesar 78450 dan *adjusted* R^2 sebesar 0.9994 artinya 99,94% variasi *Y* dapat dijelaskan oleh X_1 dan X_2, sedangkan sisanya dijelaskan oleh variabel lain.

Regresi Robust Estimasi MM (*Method Of Moment*)

Estimasi MM dihitung dengan menggunakan bantuan *software R 1.4.0*, diperoleh model persamaan regresi yang terbentuk dari estimasi nilai parameter metode estimasi MM (*Method of Moment*) adalah:

$$\hat{y} = -30611{,}9812 + 4{,}9683x_1 + -3{,}9703x_2$$

Dengan menggunakan bantuan *software R 4.1.*0, diperoleh nilai *Residual Standard Error* (RSE) sebesar 78440 artinya kesalahan dalam memprediksi *Y* sebesar 78440 dan *adjusted* R^2 sebesar 0,9988 artinya 99,88% variasi *Y* dapat dijelaskan oleh X_1 dan X_2, sedangkan sisanya dijelaskan oleh variabel lain.

Pemilihan Metode Estimasi Terbaik

Nilai *RSE* dan *adjusted* R^2 dari ketiga metode estimasi dikelompokkan pada **Tabel 3** sebagai berikut :

Tabel 3. Residual Standard Error (RSE) dan Adjusted R^2 Pada Tiga Metode Estimasi

Metode Estimasi	*RSE*	*Adjusted R^2*
MKT/Ordinary Least Square (OLS)	364900	0,9814
Least Trimmed Square (LTS)	76910	0,9956
Scale (S)	78450	0,9994
Methode Of Moment (MM)	78440	0,9988

Berdasarkan **Tabel 3**, dapat dilihat nilai *adjusted* R^2 dari ketiga metode memiliki nilai memiliki nilai berkisar antara 0,9814-0,9988 atau 98.14% – 99.94% artinya persamaan yang

dihasilkan ketiga metode ini mempunyai kemampuan menjelaskan variasi Y yang baik karena berdasarkan kriteria nilai *adjusted* R^2 , apabila nilai *adjusted* R^2 mendekati 1 maka semakin baik prediksi model regresi. Namun, jika ditinjau dari nilai *Residual Standard Error (RSE)* maka estimasi LTS (*Least Trimmed Square*) menjadi metode yang memiliki nilai *RSE* paling kecil sedangkan estimasi MKT/*Ordinary Least Square* (OLS) memiliki nilai *RSE* paling besar jika dibandingkan dengan metode estimasi lainnya. Oleh karena itu, estimasi LTS (*Least Trimmed Square*) merupakan metode terbaik untuk digunakan dalam mengestimasi parameter regresi untuk kasus data produksi padi dan curah hujan di Indonesia tahun 2019..

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan untuk kasus data pertanian dan curah hujan di Indonesia tahun 2019, maka dapat disimpulkan bahwa dari perbandingan estimasi *Least Trimmed Square* (LTS), estimasi *Scale* (S) dan estimasi *Method Of Moment* (MM) berdasarkan nilai *Residual Standard Error* (RSE) terkecil dan *adjusted* R^2 terbesar diperoleh metode estimasi regresi robust terbaik yaitu metode estimasi *Least Trimmed Square* (LTS).

Acknowledge

Ucapan terima kasih saya sampaikan kepada semua pihak yang berperan dalam penyusunan artikel ini, khususnya untuk Program Studi Statistika Universitas Islam Bandung dan Ibu Dr.Lisnur Wachidah., Dra., M.Si selaku dosen pembimbing yang telah memberikan dukungan kepada saya, sehingga saya dapat menyelesaikan artikel ilmiah ini.

Daftar Pustaka

[1] Badan Pusat Statistik Indonesia. (2020). Statistik Indonesia 2020.

[2] Chen, C. (2002). *Robust Regression and Outlier Detection with ROBUSTREG Procedure*. SAS Institute Inc. (Online). (https://pdfs.semanticscholar.org/ccb3/3dfc93f60dd-b9f488533b8d85081c550a7d8.pd, diakses 20 Januari 2021)

[3] Dehon, C., Gassner, M., & Verardi, V. (2008). *A New Hausmann Type Test to Detect the Presence of Influential Outliers* (No. 2008_006). ULB Universite Libre de Bruxelles.

[4] Fox, J. & Weisberg, S. (2010). *Robust Regression in R . Apendix to An R and S-Plus Companion to Applied Regression, Second Edition*. (Online). (https://soc-serv.socsci.mcmaster.ca/jfox/Books/Companion/appendix/Appendix-Robust-Regression.pdf, diakses 20 Januari 2021)

[5] Maharani, I.F., Satyahadewi, N., & Kusnandar, D. (2014). Metode *Ordinary Least Squares* dan *Least Trimmed Squares* Dalam Mengestimasi Parameter Regresi Ketika Terdapat *Outlier*. Buletin Ilmiah Mat, Stat, dan Terapannya (Bimaster), Vol. 03, No. 3, 163-168.

[6] Nurcahyadi, H. (2010). Analisis regresi pada data *outlier* dengan menggunakan *Least Trimmed Square* (LTS) dan MM-Estimasi.

[7] Olive, D.J. (2005). *Applied Robust Statistics*. Carbondale: Southern Illinois University.

[8] Rousseeuw, P.J., & Leroy, A.M. (1987). *Robust Regression and Outlier Detection*. New York: John Wiley and Sons.

[9] Yohai, Victor J. (1987). *High Breakdown Point and High Efficiency Robust Estimates For Regression. The Annals of Statistics*, 642-656.

[10] Sembiring, R.K. (1995). Analisis Regresi. Penerbit ITB. Bandung.

[11] Soemartini. (2007). Pencilan (*Outlier*). Bandung: Universitas Padjajaran.

[12] Utama Muhammad Bangkit Riksa, Hajarisman Nusar. (2021). *Metode Pemilihan Variabel pada Model Regresi Poisson Menggunakan Metode Nordberg*. Jurnal Riset Statistika, 1(1), 35-42.