Jurnal Matematika Thales (JMT): 2021 ISSN: 2715-1891
Vol. 03 No. 01

ESTIMASI *ROBUST* PADA MODEL REGRESI UNTUK MENANGANI OUTLIER DAN HETEROSKEDASTISITAS

(ROBUST ESTIMATION IN REGRESSION MODEL FOR HANDLING OUTLIER ANDHETEROSCEDASTICITY)

ARLINDA AMALIA DEWAYANTI*, HERNI UTAMI

Abstract. Regression analysis is a method in statistics to determine the relationship between the dependent variable and independent variables. The method used in the estimation of parameters in the model is Ordinary Least Square (OLS). This method is very sensitive to deviations assumptions on the data. The assumption is often not met is assuming heteroscedasticity. One cause of non-fulfillment of this assumption because there are outlier in the data. Therefore, another method used to handle the data outlier.The solve of this case is the robust regression method using estimates Least Trimmed Square (LTS) and Least Median Square (LMS). This paper will be discussed handling outlier and heteroscedasticity by comparing both the robust estimation by OLS seen from the residual standard error, standard error, and the regression coefficient. The results obtained from the Central Java province GRDP data in 2016-2017, show that data containing the direction y outlier, the method of estimating at least the median square is better to use compared to other methods.
Keywords: LMS estimation, LTS estimation, Heteroskidastity, OLS, Outlier, Robust Regression

Abstrak. Regresi semiparametrik *spline* merupakan model regresi yang mengombinasikan komponen parametrik dan komponen nonparametrik dalam satu model dimana komponen nonparametriknya didekatkan dengan regresi *spline*.Metode estimasi yang umumnya digunakan untuk mengestimasi model regresi semiparametrik *spline* adalah metode kuadrat terkecil (l*east square*). Namun estimasi yang dikonstruksikan dengan metode tersebut sensitif terhadap *outlier* sehingga menyebabkan estimasi nilai parameter menjadi bias dan interpretasi hasil menjadi tidak akurat. Dalam mengatasi hal tersebut, *outlier* tidak dapat dihilangkan begitu saja karena *outlier* dapat mengandung informasi penting yang tidak dapat diberikan oleh pengamatan lain. Oleh karena itu, dibutuhkan suatu metode estimasi yang kokoh terhadap outlier, yaitu metode *robust*. Metode robust yang digunakan dalam penelitian ini adalah metode estimasi M. Metode estimasi M mengestimasi parameter dengan cara meminimumkan fungsi objektif dari residual. Hasil penelitian menunjukkan bahwa metode estimasi M menghasilkan nilai GCV (*Generalized Cross Validation*) yang lebih kecil dibandingkan nilai GCV yang diperoleh dari metode kuadrat terkecil. Dengan demikian, estimator parameter yang dihasilkan oleh metode estimasi M lebih baik dibandingkan metode kuadrat terkecil.
Kata-kata kunci: Estimasi M, *Outlier*, Regresi Semiparametrik *Spline*, *Robust*

1. PENDAHULUAN

Analisis regresi adalah salah satu alat dalam analisis statistik yang memanfaatkan hubungan antara dua variabel atau lebih. Tujuannya adalah membuat suatu perkiraan variabel dependen dengan menggunakan variabel independen (Qudratullah) [6]. Salah satu metode dalam mengestimasi parameter-parameter pada model regresi linear adalah Ordinary Least Square (OLS). OLS adalah salah satu metode estimasi parameter pada regresi yang dilakukan dengan meminimumkan jumlah kuadrat error. Didalam estimasi OLS terdapat sifat BLUE (Best Linear Unbiased Estimator), syaratnya asumsi klasik harus terpenuhi. Asumsi klasik tersebut yaitu asumsi normalitas, asumsi homoskedastisitas, asumsi tidak autokorelasi, dan asumsi tidak multikolinearitas. Pada berbagai kasus, tidak jarang ditemukan kondisi dimana asumsi-asumsi tersebut tidak terpenuhi. Jika asumsi tidak terpenuhi akan mengakibatkan hasil estimasi parameter pada OLS kurang baik. Pada paper ini akan membahas kasus pelanggaran mengenai asumsi homoskedastisitas atau heteroskedatisitas. Hal ini disebabkan adanya outlier pada data pengamatan.

Outlier adalah pengamatan yang jauh dari pusat data dan berpengaruh besar terhadap koefisien regresi. Keberadaan data outlier akan mengganggu dalam proses analisis data. Menurut Soemartini [10] ada tiga jenis outlier yaitu outlier pada variabel dependen atau pada arah y (outlier vertikal), outlier pada variabel independen atau pada arah x (good leverage point), dan outlier pada arah x dan y (bad leverage point). Adanya outlier dapat menyebabkan residual yang besar. Oleh karena itu diperlukan metode lain untuk menangani adanya outlier yaitu regresi robust.

Outlier pada umumnya memberikan masalah, terutama dalam pemenuhan asumsi normalitas maupun homoskedastisitas, akan tetapi outlier bukan berarti data yang tidak berguna. Outlier kaya akan informasi yang penting sehingga dalam setiap analisa, outlier biasanya menjadi salah satu yang diperhatikan dan tidak dihilangkan. Salahsatu cara untuk menganalisa outlier adalah dengan memanfaatkan error pada data. Error sendiri informasi yang tidak terserap dalam model. Ketika suatu data individu teridentifikasi sebagai outlier, maka error dari data tersebut akan berbeda nyata secara signifikan dengan error dari individu lainnya dalam data set yang sama.

Pelanggaran pada asumsi ini mengakibatkan penduga OLS yang diperoleh tetap memenuhi persyaratan tak bias, tetapi variansi yang diperoleh menjadi tidak efisien. Variansi yang diperoleh cenderung membesar sehingga tidak lagi merupakan variansi minimum. Hal tersebut mengharuskan model harus diperbaiki supaya pengaruh dari heteroskedastisitas hilang. Secara statistik permasalahan heteroskedastisitas tersebut akan menganggu model yang akan diestimasi. Cook & Weisberg [3] memberikan tes skor untuk mendeteksi heteroskedastisitas dengan transformasi. Dalam penanggulangan heteroskedastisitas, transformasi sering digunakan, tetapi bukti untuk transformasi kadang-kadang sangat tergantung pada satu atau beberapa pengamatan. Beberapa penulis menunjukkan bahwa transformasi bisa sangat sensitif terhadap outlier. Selain itu, kasus pemodelan variansi itu sendiri sangat menarik atau transformasi sederhana tidak memadai untuk mengoreksi ketidaksetaraan perbedaan.

Salah satu metode yang menjadi alternatif untuk menyelesaikan kasus ini adalah dengan menggunakan metode robust. Suatu estimasi yang resistance adalah relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data (Mashitah, dkk) [5]. Terdapat berbagai macam metode robust diantaranya estimasi M (Maximum Likelihood Type), estimasi S (Scale), estimasi MM (Method Of Moment), estimasi LTS (Least Trimmed Square) dan estimasi LMS (Least Median Square).

Estimasi LMS merupakan metode estimasi memiliki nilai effisiensi yang tinggi. Pada estimasi LMS dilakukan dengan meminimumkan median kuadrat error. Pada estimasi LTS merupakan metode yang mempunyai nilai breakdown point yang tinggi yaitu hampir 50%. Pada estimasi LTS pertama-tama menghitung h, banyak data yang menjadikan estimasi robust, dengan sebelumnya menyusun residual kuadrat dari yang terkecil sampai dengan yang terbesar (Rousseeuw) [8].

2. RUMUSAN MASALAH

Berdasarkan uraian latar belakang masalah tersebut, maka permasalahan yang dapat dirumuskan dari penelitian ini adalah sebagai berikut:

(1) Bagaimana menangani outlier dalam kasus heteroskedastisitas?
(2) Bagaimana langkah-langkah untuk mendapatkan model regresi pada kasus yang mengandung outlier dan heterokedatisitas dengan menggunakan estimasi LTS dan LMS?
(3) Bagaimana perbandingan kedua estimator tersebut dalam mengestimasi data yang mengandung outlier pada kasus heteroskedastisitas?

3. OUTLIER

Outlier menurut Sembiring [9] adalah pengamatan yang jauh dari pusat data yang mungkin berpengaruh besar terhadap koefesien regresi. Sehingga dapat disimpulkan outlier adalah suatu data pengamatan yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat data. Pada analisis regresi terdapat tiga tipe outlier yang berpengaruh terhadap estimasi OLS. Menurut Soemartini [10] ada tiga jenis outlier yaitu:

(1) Outlier Vertikal (Vertical Outlier)
Outlier vertikal adalah semua pengamatan terpencil pada variabel dependen, tetapi tidak terpencil dalam variabel independen. Keberadaan vertical outlier berpengaruh terhadap estimasi Least Square (LS).
(2) Good Leverage Point
Good leverage point adalah pengamatan terpencil pada variabel independen tetapi terletak dekat dengan garis regresi. Keberadaan outlier jenis ini tidak berpengaruh terhadap estimasi Least Square (LS), tetapi berpengaruh terhadap inferensi statistik karena meningkatkan estimasi standar error.

(3) Bad Leverage Point
Bad leverage point adalah pengamatan yang terpencil pada variabel independen dan terletak jauh dari garis regresi. Keberadaan outlier ini berpengaruh pada intersep maupun slope dari persamaan regresi.

Menurut Soemartini [10] | $R - studentized$ | dapat digunakan untuk identifikasi outlier dengan melihat arah y. Metode ini memiliki perhitungan hampir sama dengan studentized residuals, tetapi variansi yang digunakan dalam perhitungan R-student saat observasi ke-i dikeluarkan dari pengamatan. Jika $| R - student | > 2$ dapat dikatakan data tersebut outlier.

4. HETEROSKEDASTISITAS

Analisis regresi adalah suatu metode yang berguna untuk menentkan hubungan suatu variabel dependen dengan satu atau lebih variabel independen. Bentuk umum analisis regresi sebagai berikut:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{4.1}$$

dengan $\boldsymbol{Y}$: vektor respon (n×1), $\boldsymbol{X}$: Matriks variabel prediktor berukuran (n×(p+1)), $\boldsymbol{\beta}$: vektor parameter (n × 1), dan $\boldsymbol{\epsilon}$: vektor error yang berdistribusi normal.

Estimator $\hat{\boldsymbol{\beta}}_{\text{ols}}$ pada persamaan (4.1), dapat diperoleh sebagai berikut:

$$\hat{\boldsymbol{\beta}}_{\text{ols}} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}. \tag{4.2}$$

Estimator OLS (Ordinary Least Square) tersebut akan bersifat BLUE jika semua asumsi dari variabel error terpenuhi, termasuk varabel error homoskedastisitas yang secara simbolis dituliskan sebagai berikut:

$$E\left(\boldsymbol{\epsilon}_i^2\right) = \boldsymbol{\sigma}^2.$$

dengan $\boldsymbol{\epsilon_i} = \boldsymbol{Y_i} - \boldsymbol{X_i}\boldsymbol{\beta}$ dan $\boldsymbol{\epsilon}_i^2$ adalah kuadrat error, $\forall_i = 1, 2, \ldots, n$.

Pelanggaran dalam asumsi homoskedastisitas sering disebut heteroskedastisitas. Heteroskedastisitas merupakan variansi setiap residualnya tidak sama. Deteksi adanya heteroskedastisitas dapat dilakukan secara grafis dengan melihat apakah terdapat pola non acak dari plot residual. Selain itu dapat dilakukan perhitungan secara statistik dengan menggunakan uji Breusch-Pagan. Apabila terjadi heteroskedatisitas, estimator OLS tidak bersifat BLUE (Best Linear Unbiased Estimator), tetapi hanya bersifat LUE.

Model persamaan regresi yang mengandung heteroskedastisitas dapat ditunjukkan dengan mengasumsikan didalam OLS bahwa $E\left(\boldsymbol{\epsilon}_i\right) = 0$ dan $E\left(\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}\right) = \Phi$, didapatkan taksiran variansi β sebagai berikut:

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{\Phi}\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}. \tag{4.3}$$

dengan matriks variansi kovariansi residual ($\boldsymbol{\Phi}$). Jika persamaan (4.3) memuat unsur heteroskedastisitas, maka variansi residual pada regresi tersebut tidak konstan karena tergantung nilai variabel bebas (X). Apabila nilai matriks variansi kovariansi residual ($\boldsymbol{\Phi}$) diketahui, maka dapat dipergunakan dalam penanganan heteroskedastisitas. Akan tetapi, jarang nilai variansi residual diketahui dalam kasus heteroskedastisitas, sehingga

harus didapatkan estimasi variansi residual tersebut terlebih dahulu agar bisa dilakukan estimasi parameter regresi yang memuat unsur heteroskedastisitas secara eksplisit.

Apabila diasumsikan $\boldsymbol{\Phi} = \boldsymbol{\sigma}^2\boldsymbol{\Psi}$ dengan $\boldsymbol{\Phi}$ adalah matriks yang memuat variansi residual sebagai unsur diagonal utamanya, dan Ψ merupakan unsur hetersokedastisitas maka persamaan diatas menjadi :

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{\Psi}\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}.$$

dengan Ψ merupakan unsur hetersokedastisitas dan bukan matriks diagonal. Sehingga diketahui bahwa untuk mendapatkan nilai variansi residual (Φ) harus didapatkan dari data yang diketahui memuat unsur heteroskedastisitas.

5. UKURAN ROBUST

Pada estimasi robust terdapat ukuran untuk menyatakan seberapa baik suatu estimasi diantaranya breakdown point, fungsi objektif, dan fungsi influence. Salah satu ukuran tersebut yang sering digunakan adalah breakdown point. Breakdown point merupakan proporsi minimal dari banyaknya outlier dibandingkan dengan seluruh data pengamatan. Diasumsikan bahwa pada sebuah sampel Z (berdstribusi normal dengan ukuran sampel n) dan T merupakan estimasi regresi. Nilai breakdown point dari sebuah estimator $T = T(Z)$ dapat didefinisikan sebagai berikut:

$$T(Z) = \hat{\beta} \quad \text{dengan} \quad \hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}.$$

dengan mempertimbangkan kemungkinan sampel z^* yang mengganti sebarang m titik data asli dan supremum adalah semua kemungkinan dari z^*. Sehingga bias$(m; T, Z)$ yang merupakan maksimum bias dapat dijabarkan sebagai berikut:

$$\text{bias}(m; T, Z) = \sup_{Z^*} \|T(Z^*) - T(Z)\|.$$

Jika bias$(m; T, Z)$ tak terbatas maka m outlier dapat memiliki efek yang besar untuk T yang dinyatakan dalam estimator breakdown. Oleh karena itu, breakdown point dapat dinyatakan sebagai berikut :

$$\epsilon^*(T, Z) = \min\left\{\frac{m}{n}; \text{bias}(m; T, Z) \text{ adalah tak terbatas }\right\}. \tag{5.1}$$

Berdasarkan persamaan (5.1) dapat disimpulkan bahwa breakdown point merupakan pecahan terkecil dari kontaminasi yang menyebabkan estimator T untuk mengambil sebarang nilai-nilai yang jauh dari $T(Z)$.

Di dalam estimasi regresi robust dikenal berbagai macam fungsi diantaranya fungsi objektif dan fungsi influence. Fungsi objektif digunakan untuk representasi pembobot dari residual atau $\rho(u)$. Fungsi influence ($\psi(u)$) digunakan untuk mengukur pengaruh dari sebuah data terhadap estimasi parameter. Fungsi influence secara matematis didefinisikan sebagai berikut:

$$\psi(u) = \frac{\partial \rho(u)}{\partial u}.$$

dengan $\rho(u)$ adalah representasi pembobot dari residual (fungsi objektif).

6. ESTIMASI LTS

Salah satu metode estimasi robust dalam menduga parameter model regresi pada data yang mengandung outlier adalah estimasi LTS. Estimasi LTS mempunyai nilai high breakdown point dan diperkenalkan oleh Rousseeuw pada tahun 1984. Estimasi LTS merupakan suatu metode pendugaan parameter regresi robust untuk meminimumkan jumlah kuadrat h residual (fungsi objektif). Menurut Chen [2] rumus pada estimasi LTS sebagai berikut:

$$\hat{\boldsymbol{\beta}}_{LTS} = \arg\min_{\beta} Q_{LTS}.$$

dengan $Q_{LTS} = \sum_{i=1}^{h} \epsilon_i^2, h = \left[\frac{n}{2}\right] + \left[\frac{(k+2)}{2}\right], \epsilon_i^2 = \left(\hat{y}_i - X_i\hat{\boldsymbol{\beta}}\right)^2$ merupakan kuadrat residual untuk pengamatan ke- i yang diurutkan dari terkecil ke terbesar, n banyaknya pengamatan dan k adalah banyaknya parameter, sehingga diperoleh estimasi parameter $\hat{\beta}$ model regresi LTS sebagai berikut:

$$\hat{\boldsymbol{\beta}}_{LTS} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}.$$

Prosedur estimasi dengan menggunakan estimasi LTS atau algoritma FAST-LTS, C-steps, dan FWLS diuraikan sebagai berikut:

(1) Hitung estimasi dari $\boldsymbol{\beta}$ dengan metode OLS, dinotasikan $\hat{\boldsymbol{\beta}}_{\text{ols}}$.
(2) Hitung jumlah nilai $h_1 = \left[\frac{n+k+2}{2}\right]$ dengan pembulatan nilai keatas, kemudian ditentukan jumlah n kuadrat residual $\epsilon_i^2 = \left(\hat{Y}_i - X_i^T\hat{\boldsymbol{\beta}}\right)^2$, dan mhitung jumlah pengamatan dengan nilai ϵ_i^2 dari yang terkecil.
(3) Lakukan estimasi parameter $\hat{\boldsymbol{\beta}}_{LTS}^{(1)}$ dengan meminimalkan $\sum_{i=1}^{h_1} \epsilon_i^2$.
(4) Hitung jumlah nilai $h_2 = \left[\frac{h_1+k+2}{2}\right]$, kemudian ditentukan jumlah n kuadrat residual $\epsilon_i^2 = \left(\hat{Y}_i - X_i^T\hat{\boldsymbol{\beta}}\right)^2$, dan hitung jumlah pengamatan dengan nilai ϵ_i^2 dari yang terkecil.
(5) Lakukan C-steps yaitu tahap 3 sampai 5 untuk mendapatkan fungsi objektif h yang kecil dan konvergen $\left(\left\|\hat{\beta}_{LTS}^{m} - \hat{\beta}_{LTS}^{m+1}\right\| < \epsilon\right)$.

7. ESTIMASI LMS

Prinsip dasar pada metode estimasi robust dengan penduga Least Median Squares (LMS) adalah mencocokkan sebagian besar data, setelah outlier teridentifikasi sebagai titik yang tidak berhubungan dengan data. Jika pada OLS yang perlu dilakukan adalah meminimumkan kuadrat error $\left(\sum_{i=1}^{n} \epsilon_i^2\right)$ maka pada LMS yang perlu dilakukan adalah meminimumkan median kuadrat error yaitu:

$$M_J = \min\left\{\text{med } \epsilon_i^2\right\} = \min\left\{M_1, M_2, \ldots, M_s\right\}. \tag{7.1}$$

dengan ϵ_i^2 adalah kuadrat error hasil taksiran dengan OLS.

Menurut Rousseeuw [8] untuk mendapatkan nilai M_1, dicari himpunan bagian dari matriks $\mathbf{X}$ sejumlah h_i pengamatan, yaitu:

$$h_i = h_1 = \left[\frac{n}{2}\right] + \left[\frac{p+1}{2}\right]. \tag{7.2}$$

dimana n banyaknya data, dan p banyaknya parameter.

Adapun algoritma penduga parameter regresi estimasi robust dengan metode LMS secara teoritis sebagai berikut:

(1) Hitung nteraksi ke-0 atau $h_0 = n$ dengan n adalah banyaknya data.
(2) Bentuk nilai M_j seperti pada persamaan (7.1).
(3) Hitung bobot w_{ii}, menurut Rousseeuw[8] jika nilai $r = 2.5$, sehingga didapatkan persamaan :

$$w_{ii} = \begin{cases} 1, & \text{jika } \left|\frac{\epsilon_i}{\hat{\sigma}}\right| \leq 2.5 \\ 0, & \text{lainnya} \end{cases}.$$

(4) Bentuk matriks $\mathbf{W}_{ij}$ seperti pada persamaan:

$$\mathbf{W}_{ij} = \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1n} \\ W_{21} & W_{22} & \ddots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n1} & W_{n2} & \cdots & W_{nn} \end{bmatrix}$$

(5) Hitung penduga parameter $\hat{\boldsymbol{\beta}}_{LMS}^{(0)}$ seperti pada persamaan:

$$\hat{\boldsymbol{\beta}}_{LMS}^{(0)} = \left(\boldsymbol{X^T W X}\right)^{-1}\left(\boldsymbol{X^T W X}\right).$$

(6) Cari himpunan bagian data dari matriks $\mathbf{X}$ sejumlah h_i pengamatan, yaitu: $h_i = h_0 = \left[\frac{n}{2}\right] + \left[\frac{p+1}{2}\right]$ dengan n banyaknya data, dan p banyaknya parameter.
(7) Lakukan langkah 6 sampai iterasi berakhir pada iterasi ke- s yaitu saat $h_s = h_{s+1}$ dan lakukan langkah 2 sampai 5 sehingga mendapatkan nilai konvergen $\left(\left\|\hat{\beta}_{LMS}^{s} - \hat{\beta}_{LMS}^{s+1}\right\| < \epsilon\right)$.

8. STUDI KASUS

Pada penelitian kali ini akan dicri model terbaik unutk memprediksi PDRB di provinsi Jawa Tengah tahun 2016-2017. Variabel independen yang akan digunakan diantaranya:

(1) X_1 = Jumlah Tenaga Kerja.
(2) X_2 = Upah Minimum Kabupaten.
(3) X_3 = Pendapatan Asli Daerah.
(4) X_4 = Indeks Pembangnan Manusia.

Pemodelan dilakukan pada saat data teridentifikasi mengandung heteroskedastisitas sehingga metode OLS tidak dapat digunakan ntuk memodelkan parameter. Pemilihan model terbaik dilakukan dengan membandingkan nilai Residual Standard Error dan dipilih nilai yang terkecil.

8.1. **Uji Heteroskedastisitas.** Uji Heteroskedastisitas bertujuan untuk menguji apakah dalam model regresi terjadi ketidaksamaan variansi dari error satu pengamatan ke pengamatan yang lain. Jika variansi dari error satu pengamatan ke pengamatan lain tetap, maka disebut homokedastisitas. Jika berbeda maka disebut heteroskedastisitas. Model regresi yang baik adalah tidak mengandung unsur heteroskedastisitas (Baltagi [1]). Situasi heterokedastisitas akan menyebabkan penaksiran koefisien regresi menjadi tidak efisien dan bias.

Pengujian heteroskedastisitas dengan menggunakan uji Breusch-Pagan. Hipotesis yang digunakan sebagai dugaan awal adalah sebagai berikut:

$$H_0 = \text{ Tidak ada heteroskedastisitas}$$
$$H_1 = \text{ Terdapat heteroskedastisitas}$$

Keputusan H_0 dilakukan ketika p-value $> \alpha : 5\%$ yang artinya model dikatakan bebas masalah heteroskedastisitas. Dari perhitungan data, didapatkan nilai p-value sebesar 0.02668 maka H_0 ditolak. Sehingga, dapat disimpulkan bahwa terdapat unsur heteroskedastisitas.

8.2. **PENDETEKSIAN OUTLIER.** Sebelum menghitung nilai suatu estimasi dari parameter $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$, dan $\hat{\beta}_4$ akan diperiksa terlebih dahulu keberadaan outlier dalam data. Pendeteksian outlier dalam data dilihat berdasarkan jenis outlier, diantaranya outlier pada arah x, outlier pada arah y, dan outlier pada arah x dan y (influence). Metode yang digunakan untuk mendeteksi outlier arah influence adalah DFBETAS $_{ji}$ dan DFFITS DF $_i$-student digunakan untuk melihat outlier pada arah y, dan leverage untuk melihat outlier pada arah x. Data diolah dengan menggunakan software R.

Pada kasus ini, nilai cut-off untuk masing-masing metode ditentukan berdasarkan jumlah sampel (n) yaitu 70 dan banyaknya parameter (k) adalah 4 . Oleh sebab itu, observasi dikatakan outlier jika:

(1) $\mid R-$ student $\mid > 2$
(2) Leverage $(h_{ii}) > \frac{2k}{n} = \frac{2.4}{70} = 0.11428$
(3) $\mid$ DFFITS $_i \Big| > 2\sqrt{\frac{k}{n}} = 2\sqrt{\frac{4}{70}} = 0.47809$

(4) $\left|\text{ DFBETAS } S_{ji}\right| > \frac{2}{\sqrt{n}} = \frac{2}{\sqrt{70}} = 0.23904$

Berdasarkan kriteria tersebut, dapat diketahui observasi yang merupakan outlier. Terdapat data observasi yang outlier pada arah x yaitu data observasi ke-32, 33, 56, 67 dan ke-68. Dimana nilai kedua observasi tersebut sebagai berikut:

(1) Pada observasi ke-32 yaitu Leverage $(h_{ii})_{32} = 0.1413 > 0.11428$
(2) Pada observasi ke-33 yaitu Leverage $(h_{ii})_{33} = 0.3716 > 0.11428$
(3) Pada observasi ke-56 yaitu Leverage $(h_{ii})_{56} = 0.1793 > 0.11428$
(4) Pada observasi ke-67 yaitu Leverage $(h_{ii})_{67} = 0.1477 > 0.11428$
(5) Pada observasi ke-68 yaitu Leverage $(h_{ii})_{68} = 0.4960 > 0.11428$

Data yang outlier pada arah y adalah data observasi $1, 19, 36, 54$, dan 56 . Hal ini dikarenakan pada pendeteksian outlier menggunakan R-student diperoleh hasil bahwa nilai $|$ R-student $|$ lebih dari 2. Berikut ini merupakan hasil nilai $|$ R-student $|$ tersebut:

(1) Pada observasi ke-1, nilai $|$ R - student $|= 3.8144 > 2$.
(2) Pada observasi ke-19, nilai $|$ R - student $|= 4.21857 > 2$.
(3) Pada observasi ke-36, nilai $|$ R-student $|= 3.07281 > 2$.
(4) Pada observasi ke-54, nilai $|$ R - student $|= 4.22429 > 2$.
(5) Pada observasi ke-56, nilai $|$ R - student $|= 2.14053 > 2$.

Pada kasus ini, outlier yang digunakan adalah outlier pada arah y. Sehingga, data observasi yang dikatakan merupakan outlier berdasarkan kriteria tersebut adalah data observasi data observasi 1, 19,36, 54, dan 56 . Sehingga karena adanya outlier mengakibatkan heteroskedastisitas.

8.3. **Pemodelan Parameter.** Estimasi untuk mencari nilai parameter pada data yang mengandung outlier dilakukan dengan beberapa metode, yaitu metode estimasi OLS, metode regresi robust estimasi LTS, dan metode regresi robust estimasi LMS. Hasil dari nilai estimasi parameter pada metode OLS tersebut adalah sebagai berikut:

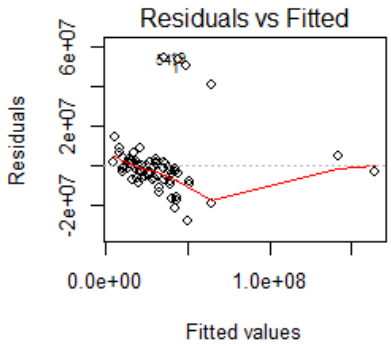
Parameter	Nilai Estimasi	Standar Error
(intercept)	-124600000	46050000
x_1	45.33	15.05
x_2	39.55	13.91
x_3	0.08198	0.1830
x_4	768.5	628.6

TABEL 1. Hasil Estimasi Parameter Menggunakan OLS Pada Data Mengandung Outlier

Model persamaan regresi yang terbentuk dari estimasi nilai parameter metode OLS adalah:

$$\hat{Y} = -124600000 + 45.33X_1 + 39.55X_2 + 0.08198X_3 + 768.5X_4.$$

Estimasi pada metode ini menghasilkan Residual Standard Error (RSE) sebesar 14730000. Plot residual pada metode OLS mayoritas terletak disekitar 0. Namun, ada beberapa residual yang terletak jauh dari 0 . Hal ini teridentifikasi bahwa data terdapat outlier. Berikut ini merupakan plot residual metode OLS:



Gambar 1. Plot Residual Metode OLS Mengandung Outlier

Pada gambar 1 dapat dilihat bahwa terdapat residual yang nilainya paling besar diantara nilai residual yang pada observasi lain. Hal ini teridentifikasi adanya outlier sehingga terdapat metode estimasi lain yang dapat digunakan menangani outlier yaitu regresi robust.

Regresi robust yang pertama digunakan adalah metode estimasi LTS. Berikut ini merupakan hasil estimasi nilai parameter menggunakan metode estimasi LTS:

Parameter	Nilai Estimasi	Standar Error
(intercept)	-60470000	11030000
x_1	45.38	3.847
x_2	-8.566	3.863
x_3	371.2	0.008168
x_4	952.1	149.2

Tabel 2. Hasil Estimasi Parameter Menggunakan LTS Pada Data Mengandung Outlier

Model persamaan regresi yang terbentuk dari estimasi nilai parameter metode estimasi LTS adalah:

$$\hat{Y} = -60470000 + 45.38X_1 - 8.566X_2 + 371.2X_3 + 952.1X_4.$$

Estimasi pada metode ini menghasilkan residual standard error sebesar 3317000. Plot residual pada metode LTS memiliki hasil yang hampir sama dengan metode OLS

yaitu tedapat residual observasi yang terletak jauh dari 0 . Oleh sebab itu metode ini, digunakan untuk mengatasi hal tersebut tanpa perlu mengeluarkan observasi yang diindikasi sebagai outlier.

Regresi robust selanjutnya adalah estimasi LMS. Berikut ini merupakan hasil estimasi nilai parameter menggunakan metode estimasi LMS:

Parameter	Nilai Estimasi	Standar Error
(intercept)	3585000	3403794
x_1	23.27	22895.61
x_2	1.159	19237.23
x_3	0.03364	2014223
x_4	-46.42	528.3125

Tabel 3. Hasil Estimasi Parameter Menggunakan LMS Pada Data Mengandung Outlier

Model persamaan regresi yang terbentuk dari estimasi nilai parameter metode estimasi LMS adalah:

$$\hat{Y} = -60470000 + 45.38X_1 - 8.566X_2 + 371.2X_3 + 952.1X_4.$$

Estimasi pada metode ini menghasilkan residual standard error sebesar 2604409. Plot residualpada metode LMS memiliki hasil yang hampir sama dengan metode OLS dan metode LMS yaitu tedapat residual observasi yang terletak jauh dari 0 . Oleh sebab itu metode ini, digunakan untuk mengatasi hal tersebut tanpa perlu mengeluarkan observasi yang diindikasi sebagai outlier.

Berdasarkan hasil estimasi parameter menggunakan ketiga estimator tersebut diperoleh residual standard error secara keseluruhan seperti berikut:

Metode	Residual Standard Error
OLS	14730000
Estimasi LTS	3317000
Estimasi LMS	2604409

Tabel 4. Hasil Estimasi Parameter Menggunakan LMS Pada Data Mengandung Outlier

Pada tabel 4 dapat dilihat bahwa metode estimasi LTS dan estimasi LMS lebih baik digunakan dari pada OLS jika data mengandung outlier. Metode estimasi LMS lebih baik digunakan dari pada metode estimasi LTS. Hal ini disebabkan karena data mengandung outlier pada arah y. Oleh sebab itu, motode yang cocok digunakan untuk mengatasi outlier adalah regresi robust dikatakan baik apabila menggunakan estimasi LTS dan estimasi LMS. Namun, ketika terdapat data outlier pada arah y metode yang cocok digunakan adalah metode estimasi LMS.

9. PENUTUP

Berdasarkan hasil analisis dan pembahasan maka dapat diperoleh kesimpulan sebagai berikut:

(1) Estimasi parameter menggunakan metode estimasi LTS dilakukan dengan cara meminimumkan jumlah kuadrat h residual (fungsi objektif) dan menggunakn FWLS. Pada estimasi parameter menggunkan metode estimasi LTS dilakukan dengan cara meminimumkan median kuadrat residual

(2) Estimasi parameter untuk $\hat{\beta}$ pada model regresi robust dengan estimasi LTS, diperoleh sebagai berikut:

$$\hat{\boldsymbol{\beta}}_{LTS} = (\boldsymbol{X^T X})^{-1}(x^T \boldsymbol{X})$$

Estimasi parameter untuk $\hat{\beta}$ pada model regresi robust dengan estimasi LMS, diperoleh sebagai berikut:

$$\hat{\boldsymbol{\beta}}_{LMS} = (\boldsymbol{X^T W X})^{-1}(x^T \boldsymbol{W X})$$

(3) Pada kasus data penelitian PDRB di provinsi Jawa Tengah tahun 2016-2017, yang teridentifikasi adanya heteroskedastisitas dan outlier pada arah y, metode yang paling baik digunakan adalah metode estimasi LMS pada regresi robust dibandingkan dengan OLS dan metode estimasi LTS. Perbandingan dilakukan dengan melihat nilai residual standard error.

Referensi

[1] Baltagi, B.H, *Econometrics* (4th ed), Verlag Berlin Heidelberg, Springer, 2008

[2] Chen, C., *Robust Regression and Outlier Detection with the ROBUSTREG Procedure*, 27, pp.265-270, 2002

[3] Cook, R.D. dan Weisberg, S., *Diagnostics for Heteroscedasticity in Regression*. Biometrika, 70, pp. 110, 1983

[4] Draper, N. dan Smith, H., *Analisis Regresi Terapan Edisi Kedua*, Gramedia Pustaka Utama, Jakarta, 1992

[5] Mashitah, Wibowo, A. dan Indriani, D., Metode Robust Regression on Ordered Statistics (ROS) pada Data Tersensor Kiri dengan Outlier. *Jurnal Biometrika dan Kependudukan*, 02, pp. 148157, 2013.

[6] Qudratullah, M.F., *Analisis Regresi Terapan Teori Contoh Kasus dan Aplikasi dengan SPSS*, ANDI, Yogyakarta, 2013

[7] Riani, M., Atkinson, A.C., dan Torti, F., Robust Methods For Heteroskedastic Regression, *Computational Statistics and Data Analysis*, 104, pp. 209222, 2016.

[8] Rousseeuw, P.J., *Robust Regression and Outlier Detection*, Wiley and Sons, New York, 1987.

[9] Sembiring, Analisis Regresi Edisi Kedua,Institute Teknologi Bandung, Bandung, 2003.

[10] Soemartini, *Pencilan (Outlier)*, Universitas Padjajaran, Bandung, 2007.

Arlinda Amalia Dewayanti* (Penulis Korespondensi)
Departemen Matematika, Fakultas MIPA, Universitas Gadjah Mada, Indonesia
arlinda.amalia.d@mail.ugm.ac.id

Herni Utami
Departemen Matematika, Fakultas MIPA, Universitas Gadjah Mada, Indonesia
herni utami@ugm.ac.id