



PERBANDINGAN METODE *ROBUST LEAST MEDIAN OF SQUARE (LMS)* DAN PENDUGA *S* UNTUK MENANGANI *OUTLIER* PADA REGRESI LINIER BERGANDA

Laeli Sidik Febrianto[✉], Nur Karomah Dwidayati, Putriaji Hendikawati

Jurusan Matematika, FMIPA, Universitas Negeri Semarang, Indonesia
Gedung D7 Lt.1, Kampus Sekaran Gunungpati, Semarang, 50229

Info Artikel

Sejarah Artikel:
Diterima : Mei 2016
Disetujui : Agustus 2016
Dipublikasikan : Mei 2018

Keywords:
Outlier, Robust Method,
LMS, S Estimation

Abstrak

Permasalahan yang dikaji dalam penelitian ini adalah menentukan metode terbaik dalam mengatasi permasalahan *outlier* menggunakan metode regresi *robust LMS* dan Penduga *S*. Penelitian ini menggunakan data Anggaran Pendapatan dan Belanja Daerah (APBD) kabupaten/kota di Pulau Jawa tahun 2010 dengan variabel bebas meliputi Pendapatan Asli Daerah (X_1), Dana Bagi Hasil (X_2), Dana Alokasi Umum (X_3), Luas Wilayah (X_4), dan variabel terikat yaitu Belanja Modal (Y). Analisis dimulai dengan uji asumsi normalitas, linieritas, keberartian simultan, keberartian parsial, multikolinearitas, heteroskedastisitas, dan autokorelasi. Model regresi yang dapat diterima yaitu regresi data transformasi logaritma natural dari data APBD dengan variabel bebas meliputi Pendapatan Asli Daerah ($\log X_1$) dan Dana Bagi Hasil ($\log X_2$), serta variabel terikat yaitu Belanja Modal ($\log Y$). Pendeteksian *outlier* menggunakan metode *boxplot* dan *Cook's Distance* menunjukkan bahwa terdapat *outlier*, sehingga dilakukan pendugaan parameter regresi *robust* dengan metode *LMS* dan Penduga *S*. Metode *LMS* menghasilkan nilai AIC sebesar 25,54423 dan SIC sebesar 27,76414, sedangkan dengan metode Penduga *S* menghasilkan nilai AIC sebesar 40,22523 dan SIC sebesar 43,72099. Berdasarkan hasil penelitian dan pembahasan disimpulkan bahwa *LMS* merupakan metode terbaik, karena *LMS* memiliki nilai AIC dan SIC yang lebih kecil dibandingkan dengan Penduga *S*.

Abstract

Issues examined in this study is to determine the best method to solve outlier problems using robust regression method *LMS* and *S Estimation*. This study uses data Budget (APBD) districts / cities in Java in 2010 with the independent variables include revenue (X_1), DBH (X_2), General Allocation Fund (X_3), Total Region (X_4) and the dependent variable is the Capital Expenditure (Y). The analysis begins with the assumption of normality, linearity, significance simultaneous, partial significance, multicollinearity, heteroscedasticity, and autocorrelation. Regression models were acceptable ie data regression natural logarithm transformation of budget data with independent variables include revenue ($\log X_1$) and DBH ($\log X_2$), as well as the dependent variable is the Capital Expenditure ($\log Y$). The detection of outliers using *boxplot* and *Cook's Distance* shows that there are outliers, so that a robust regression parameter estimation with the *LMS* method and *S Estimation*. *LMS* method produces a value of AIC 27.76414 and SIC 25.54423, while the *S Estimation* method produces a value of AIC 43.72099 and SIC 40.22523. Based on the results and discussion can be concluded that the *LMS* is the best, since the *LMS* method has a value of AIC and SIC are smaller than the *S Estimation*.

How to Cite

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

PENDAHULUAN

Analisis Regresi Linier Berganda digunakan untuk mengukur pengaruh antara lebih dari satu variabel prediktor (variabel bebas) terhadap variabel terikat. *Ordinary Least Square* (OLS) (Draper dan Smith, 1992) merupakan salah satu metode yang sering digunakan untuk mendapatkan nilai-nilai penduga parameter dalam pemodelan regresi. Penggunaan OLS memerlukan beberapa asumsi klasik yang harus dipenuhi oleh komponen residual atau galat dalam model yang dihasilkan. Beberapa asumsi itu antara lain: (1) Residual mengikuti distribusi normal, (2) varians dari residual adalah konstan dan homoskedastisitas, (3) tidak ada autokorelasi, (4) tidak ada multikolinearitas di antara variabel bebas. Pada berbagai kasus tidak jarang ditemui hal-hal yang menyebabkan tidak terpenuhinya asumsi klasik. Salah satu penyebabnya adalah adanya pencilan (*outlier*) dalam data amatan.

Data *outlier* (Makkulau et al., 2010) adalah data pengamatan yang berada jauh (ekstrim) dari pengamatan-pengamatan lainnya. *Outlier* yang disebabkan oleh data terkontaminasi dapat dihapuskan dari data penelitian atau jika memungkinkan dapat dilakukan sampling ulang. Jika setelah dilakukan beberapa sampling ulang namun data *outlier* tetap muncul maka data tersebut tidak dapat dihapuskan dari data penelitian, karena analisis data yang dihasilkan akan tidak mencerminkan populasi yang diteliti.

Menurut Sembiring (Paludi, 2009), *outlier* adalah pengamatan yang jauh dari pusat data yang mungkin berpengaruh besar terhadap koefisien regresi. Keberadaan data *outlier* akan mengganggu dalam proses analisis data. Kaitannya dalam analisis regresi, *outlier* dapat menyebabkan hal-hal berikut: (1) Residual yang besar dari model yang terbentuk atau $E(e) \neq 0$. (2) Varians pada data tersebut menjadi lebih besar. (3) Taksiran interval memiliki rentang yang lebar.

Pendeteksian *outlier* dapat dilakukan dengan beberapa metode diantaranya dengan metode *Boxplot* dan metode *Cook's Distance*. Metode *Boxplot* merupakan metode yang menggunakan nilai kuartil dan jangkauan untuk mendeteksi *outlier*, sehingga pada metode ini dapat mengetahui adanya *outlier* untuk masing-masing variabel. Sedangkan menggunakan metode *Cook's Distance* dapat mengetahui adanya *outlier* secara simultan pada variabel bebas.

Saat ada asumsi yang tidak terpenuhi, maka penggunaan metode OLS akan memberikan kesimpulan yang bersifat kurang baik atau nilai penduga parameternya bersifat bias sehingga berakibat interpretasi hasil yang diperoleh menjadi tidak valid (Nurchayadi, 2010). Untuk mengatasinya diperlukan metode lain supaya analisis data dengan adanya data *outlier* tetap tahan (*robust*) terhadap asumsi yang diterapkan pada analisis datanya. Metode tersebut dikenal dengan metode *robust*.

Regresi *robust* diperkenalkan oleh Andrews (1972), yaitu metode regresi yang digunakan ketika distribusi dari residual tidak normal atau adanya beberapa *outlier* yang berpengaruh pada model (Wijayanti, 2015). Menurut Chen (2014) regresi *robust* terdiri dari 5 metode penduga, yaitu estimasi *robust* M, estimasi *robust* least median of square (LMS), estimasi *robust* least trimmed square (LTS), estimasi *robust* S dan estimasi *robust* MM.

Penggunaan metode LMS data *outlier* yang ada tidak dibuang begitu saja, tetapi diproses dan dieliminasi melalui sebuah iterasi. Metode ini mempunyai keuntungan untuk mengurangi pengaruh dari residual terhadap keakuratan koefisien regresi. Menggunakan median dari kuadrat residual, penduga yang dihasilkan akan lebih kekar dalam menghadapi *outlier*.

Pada penelitian Oktarinanda (2014) mengenai perbandingan efisiensi metode LTS dan metode LMS dalam estimasi parameter regresi *robust*. Perbandingan keakuratan model menggunakan koefisien determinasi dan RMSE diperoleh kesimpulan bahwa data yang digunakan dalam penelitian lebih sesuai menggunakan penduga LMS dalam menduga parameter regresi.

Penduga S bertujuan untuk memperoleh penduga dengan nilai simpangan baku terkecil. Hasil penelitian Permana (2014) menunjukkan metode Penduga S merupakan metode yang lebih baik dibandingkan metode *Least Trimmed Square* (LTS) untuk menangani *outlier* pada regresi karena memiliki nilai Mean Square Error (MSE) lebih kecil.

Penelitian ini mempunyai tujuan: (1) Memperoleh hasil estimasi regresi linier berganda metode *robust* LMS dan Penduga S. (2) Memperoleh metode regresi *robust* terbaik di antara metode LMS dan metode Penduga S.

METODE

Masalah yang diangkat dalam penelitian ini adalah memperoleh metode regresi *robust* terbaik di antara metode LMS dan metode Penduga S. Selain itu penelitian ini menjawab suatu masalah dengan cara mengkaji sejumlah literatur (bahan pustaka) secara mendalam, penelitian pustaka yang digunakan adalah jurnal, skripsi, prosiding, dan buku yang berkaitan dengan analisis regresi *robust* metode LMS dan Penduga S.

Data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data rekap Anggaran Pendapatan dan Belanja Daerah (APBD) kabupaten/kota di Pulau Jawa tahun 2010 (Solikhah, 2014). Oleh karena itu, akan dibahas model regresi belanja modal (Y) berdasarkan pendapatan asli daerah (X_1), dana bagi hasil (X_2), dana alokasi umum (X_3), dan luas wilayah (X_4). *Software* yang digunakan dalam penelitian ini yaitu SPSS16 dan SAS9.1.

Langkah-langkah penyelesaian masalah yaitu: (1) menyiapkan data; (2) menguji asumsi normalitas, linieritas, keberartian simultan, keberartian parsial, multikolinearitas, heteroskedastisitas, dan autokorelasi sampai diperoleh model regresi yang BLUE, apabila belum memenuhi semua asumsi tersebut maka dilakukan transformasi data dan/atau eliminasi beberapa variabel; (3) membuat regresi awal dengan metode OLS; (4) mendeteksi adanya *outlier* menggunakan metode *boxplot* dan *Cook's Distance*, apabila dalam pendeteksian *outlier* tidak diperoleh *outlier* dalam data maka penelitian berhenti, apabila diperoleh *outlier* maka dilanjutkan ke tahap penyelesaian berikutnya; (5) pendugaan parameter regresi *robust* dengan metode LMS; (6) pendugaan parameter regresi *robust* dengan metode Penduga S; (7) membandingkan nilai AIC dan

SIC dari masing-masing metode untuk memperoleh metode terbaik.

HASIL DAN PEMBAHASAN

HASIL

Tahap pertama setelah mempersiapkan data adalah uji asumsi regresi linier berganda. Pengujian asumsi regresi linier berganda dilakukan untuk mengetahui apakah model regresi yang diperoleh merupakan model regresi yang menghasilkan estimator BLUE.

1. Uji Asumsi Regresi pada Data APBD

Pengujian asumsi regresi linier berganda awal dilakukan terhadap data rekap APBD di Pulau Jawa pada tahun 2010.

1.1. Uji Normalitas

Uji normalitas dimaksudkan untuk mendeteksi apakah nilai residual berdistribusi normal atau tidak, menggunakan uji Kolmogorov-Smirnov (K-S). Berdasarkan tabel 1 diperoleh nilai $\hat{\alpha}$ (Asymp. Sig. (2-tailed)) sebesar 0,858. Jelas $\hat{\alpha} < \alpha$ yang berarti terima H_0 . Dengan demikian artinya residual berdistribusi normal.

1.2. Uji Linieritas

Pengujian linieritas menggunakan uji LM-Test dengan meregresikan semua variabel bebas yang telah dikuadratkan dengan nilai residual regresi data APBD. Berdasarkan tabel 2 nilai R^2 sebesar 0,037 dengan jumlah pengamatan sebanyak 90, maka $X_{hitung}^2 = (90 \times 0,037) = 3,33$. Nilai X_{tabel}^2 dengan df:(90;0,05) yaitu sebesar 113,15. Karena nilai $X_{hitung}^2 < X_{tabel}^2$ maka tolak H_0 . Dengan demikian artinya ada hubungan linier antara variabel X_1, X_2, X_3, X_4 terhadap variabel Y . Jadi pada data APBD model persamaan regresi linier.

Tabel 1 Hasil uji Kolmogorov-Smirnov data APBD

		Unstandardized Residual
N		90
Normal Parameters ^a	Mean	0
	Std. Deviation	502,517
	Most Extreme Differences	
	Absolute	0,064
	Positive	0,049
	Negative	-64
Kolmogorov-Smirnov Z		0,605
Asymp. Sig. (2-tailed)		0,858

Tabel 2 Model Summary LM-Test data APBD

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0,193	0,037	-0,008	50450,821

1.3. Uji Keberartian Simultan

Berdasarkan tabel3 nilai $\hat{\alpha}$ sebesar 0,000. Karena nilai $\hat{\alpha} < \alpha$ maka tolak H_0 , artinya tidak semua $\beta_i = 0$ ($i = 1,2,3,4$). Dengan demikian variabel bebas X_1, X_2, X_3, X_4 data APBD secara simultan (bersama-sama) berpengaruh atau mampu menjelaskan perubahan nilai variabel terikat Y .

1.4. Uji Keberartian Parsial

Berdasarkan hasil yang diperoleh pada tabel 4 diketahui bahwa $\hat{\alpha} < \alpha$ untuk variabel bebas X_1, X_2, X_4 dan $\hat{\alpha} > \alpha$ untuk variabel bebas X_3 . Hal ini berarti variabel bebas X_1, X_2, X_4 secara parsial berpengaruh terhadap variabel terikat Y , sedangkan variabel bebas X_3 secara parsial tidak berpengaruh terhadap variabel terikat Y . Dengan demikian model regresi linier berganda yang dapat diterima adalah regresi X_1, X_2, X_4 terhadap Y . Tahapan selanjutnya adalah pengujian multikolinearitas, heteroskedastisitas, dan autokorelasi.

1.5. Uji Multikolinearitas

Berdasarkan tabel 5 nilai VIF variabel $X_1 < 10$, nilai VIF variabel $X_2 < 10$, nilai VIF variabel

$X_3 < 10$ artinya terima H_0 . Jadi tidak terdapat masalah multikolinearitas pada regresi linier berganda data APBD.

1.6. Uji Heteroskedastisitas

Pengujian heteroskedastisitas dilakukan dengan Uji Glejser yaitu dengan meregresikan variabel bebas X_1, X_2 , dan X_4 terhadap nilai mutlak residualnya. Berdasarkan tabel 6 diperoleh nilai $\hat{\alpha} < \alpha$ pada variabel bebas X_2 dan X_4 , dan nilai $\hat{\alpha} > \alpha$ untuk variabel bebas X_1 . Karena nilai $\hat{\alpha} > \alpha$ untuk variabel bebas X_1 maka tolak H_0 . Dengan demikian bahwa pada model regresi data APBD terdapat masalah heteroskedastisitas.

1.7. Uji Autokorelasi

Pengujian autokorelasi dilakukan dengan Uji Durbin Watson (DW). Berdasarkan tabel 7 dengan $\alpha = 0,05$, $k=5$, dan $n=90$ diperoleh nilai DL sebesar 1,542 dan DU sebesar 1,776. Karena $dL < DW < dU$ maka belum dapat diambil kesimpulan (ragu-ragu) apakah didalam model regresi terdapat autokorelasi atau tidak terdapat autokorelasi.

Tabel 3 ANOVA regresi data APBD

Model	Sum of Squares	df	Mean Square	F	Sig.	
1	Regression	1,27E+12	4	3,17E+11	119,77	0,000 ^a
	Residual	2,25E+11	85	2,64E+09		
	Total	1,49E+12	89			

Tabel 4 Nilai $\hat{\alpha}$ uji keberartian parsial data APBD

Data	Variabel bebas	$\hat{\alpha}$
APBD	X_1	0,000
	X_2	0,001
	X_3	0,427
	X_4	0,011

Tabel 5 Nilai VIF variabel bebas pada data APBD

Data	Variabel bebas	VIF
APBD	X_1	4,140
	X_2	4,121
	X_4	1,033

Tabel 6 Nilai $\hat{\alpha}$ pada Uji Glejser data APBD

Data	Variabel bebas	$\hat{\alpha}$
APBD	X_1	0,000
	X_2	0,298
	X_4	0,200

Tabel 7 Model *Summary* regresi data APBD dengan 3 (tiga) variabel bebas

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	0,921	0,848	0,843	51311,88482	1,769

Berdasarkan uji asumsi regresi pada data APBD diperoleh bahwa pada regresi data APBD terdapat masalah heteroskedastisitas dan belum dapat diambil kesimpulan mengenai masalah autokorelasi. Sehingga perlu diatasi dengan mentransformasikan data APBD dengan transformasi logaritma natural (log). Selanjutnya dilakukan uji asumsi regresi pada data logAPBD untuk mengetahui apakah semua asumsi regresi sudah tidak ada yang dilanggar agar model regresi linier berganda bersifat BLUE.

2. Uji Asumsi Regresi pada Data logAPBD

Pengujian asumsi regresi linier berganda selanjutnya dilakukan terhadap data logAPBD.

2.1. Uji Normalitas

Uji normalitas dimaksudkan untuk mendeteksi apakah nilai residual berdistribusi normal atau tidak, menggunakan uji Kolmogorov-Smirnov (K-S). Berdasarkan tabel 8 diperoleh nilai $\hat{\alpha}$ (Asymp. Sig. (2-tailed)) sebesar 0,835. Jelas nilai $\hat{\alpha} > \alpha$ yang berarti terima H_0 . Dengan demikian artinya residual berdistribusi normal.

2.2. Uji Linieritas

Pengujian linieritas menggunakan uji LM-Test dengan meregresikan semua variabel bebas yang telah dikuadratkan dengan nilai residual regresi data logAPBD. Berdasarkan tabel 9 nilai R^2 sebesar 0,000 dengan jumlah pengamatan sebanyak 90, maka $X_{hitung}^2 =$

$(90 \times 0,000) = 0,000$. Nilai X_{tabel}^2 dengan df:(90;0,05) yaitu sebesar 113,15. Karena nilai $X_{hitung}^2 < X_{tabel}^2$ maka tolak H_0 . Dengan demikian artinya ada hubungan linier antara variabel $\log X_1$, $\log X_2$, $\log X_4$ terhadap variabel $\log Y$. Jadi pada data logAPBD model persamaan regresi linier.

2.3. Uji Keberartian Simultan

Berdasarkan tabel 10 nilai $\hat{\alpha}$ sebesar 0,000. Karena nilai $\hat{\alpha} < \alpha$ maka tolak H_0 , artinya tidak semua $\beta_i = 0$ ($i = 1,2,4$). Dengan demikian variabel bebas $\log X_1$, $\log X_2$, $\log X_4$ data logAPBD secara simultan (bersama-sama) berpengaruh atau mampu menjelaskan perubahan nilai variabel terikat $\log Y$.

2.4. Uji Keberartian Parsial

Berdasarkan hasil yang diperoleh pada tabel 11 diketahui bahwa $\hat{\alpha} < \alpha$ untuk variabel bebas $\log X_1$, $\log X_2$, dan $\hat{\alpha} > \alpha$ untuk variabel bebas $\log X_4$. Hal ini berarti variabel bebas $\log X_1$, $\log X_2$, secara parsial berpengaruh terhadap variabel terikat $\log Y$, sedangkan variabel bebas $\log X_4$ secara parsial tidak berpengaruh terhadap variabel terikat $\log Y$. Dengan demikian model regresi linier berganda yang dapat diterima adalah regresi $\log X_1$, $\log X_2$ terhadap $\log Y$. Tahapan selanjutnya adalah pengujian multikolinearitas, heteroskedastisitas, dan autokorelasi.

Tabel 8 Hasil uji Kolmogorov-Smirnov data logAPBD

		Unstandardized Residual
N		90
Normal Parameters ^a	Mean	0
	Std. Deviation	0,34946653
Most Extreme Differences	Absolute	0,065
	Positive	0,03
	Negative	-0,065
Kolmogorov-Smirnov Z		0,621
Asymp. Sig. (2-tailed)		0,835

Tabel 9 Model Summary LM-Test data logAPBD

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0,021	0,000	-0,034	0,35543319

Tabel 10 ANOVA regresi data logAPBD

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	21,011	3	7,004	55,415	0.000
Residual	10,869	86	0,126		
Total	31,880	89			

Tabel 11 Nilai $\hat{\alpha}$ uji keberartian parsial data logAPBD

Data	Variabel bebas	$\hat{\alpha}$
logAPBD	$\log X_1$	0,019
	$\log X_2$	0,000
	$\log X_4$	0,505

2.5. Uji Multikolinearitas

Berdasarkan tabel 12 nilai VIF variabel $\log X_1 < 10$ dan nilai VIF variabel $\log X_2 < 10$ yang berarti terima H_0 . Jadi tidak terdapat masalah multikolinearitas pada regresi linier berganda data logAPBD.

2.6. Uji Heteroskedastisitas

Pengujian heteroskedastisitas dilakukan dengan Uji Glejser yaitu dengan meregresikan variabel bebas $\log X_1$ dan $\log X_2$ terhadap nilai mutlak residualnya. Berdasarkan tabel 13 diperoleh nilai $\hat{\alpha} < \alpha$ untuk masing-masing

variabel bebas $\log X_1$ dan $\log X_2$. Karena $\hat{\alpha} < \alpha$ untuk semua variabel bebas maka terima H_0 . Dengan demikian bahwa pada model regresi data logAPBD tidak terdapat masalah heteroskedastisitas.

2.7. Uji Autokorelasi

Pengujian autokorelasi dilakukan dengan Uji DW. Berdasarkan tabel 14 dengan $\alpha = 0,05$, $k=3$, dan $n=90$ diperoleh nilai DL sebesar 1,589 dan DU sebesar 1,726. Karena $dU < DW < 4 - dU$ maka dalam model regresi tidak terdapat masalah autokorelasi.

Tabel 12 Nilai VIF variabel bebas pada data logAPBD

Data	Variabel bebas	VIF
logAPBD	$\log X_1$	2,549
	$\log X_2$	2,549

Tabel 13 Nilai $\hat{\alpha}$ pada Uji Glejser data logAPBD

Data	Variabel bebas	$\hat{\alpha}$
logAPBD	$\log X_1$	0,825
	$\log X_2$	0,756

Tabel 14 Model *Summary* regresi data logAPBD dengan 2 (dua) variabel bebas

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	0,811	0,657	0,649	0,35438	1,819

Berdasarkan uji asumsi regresi data logAPBD diketahui bahwa residual berdistribusi normal, persamaan regresi linier, secara simultan seluruh variabel bebas $\log X_1$, $\log X_2$, dan $\log X_4$ berpengaruh terhadap variabel terikat $\log Y$, secara parsial variabel $\log X_1$ dan $\log X_2$ berpengaruh terhadap $\log Y$, tidak terdapat masalah multikolinearitas, tidak terdapat masalah heteroskedastisitas, dan tidak terdapat masalah autokorelasi. Oleh sebab itu, untuk keperluan penelitian selanjutnya menggunakan data logAPBD dengan variabel bebas hanya $\log X_1$ dan $\log X_2$.

3. Estimasi Regresi Linier Berganda dengan OLS

Model regresi linier berganda pada data logAPBD dengan metode OLS diperoleh dengan bantuan program SAS 9.1.

Tabel 15 Koefisien regresi linier berganda pada data logAPBD

Model	Koefisien
Konstanta	3,520
$\log X_1$	0,215
$\log X_2$	0,500

Berdasarkan Tabel 15 model regresi linier berganda pada data logAPBD dengan metode OLS dituliskan seperti persamaan:

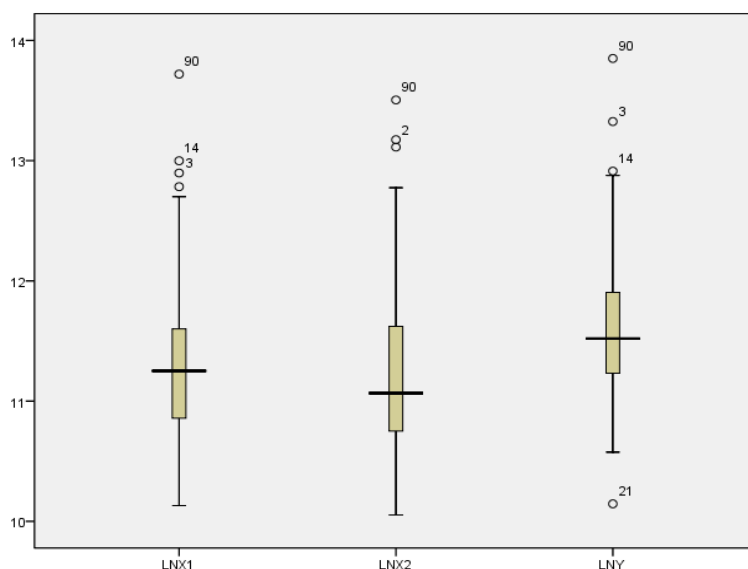
$$\log \hat{Y} = 3,520 + 0,215 \log X_1 + 0,500 \log X_2$$

4. Pendeteksian *Outlier*

Pendeteksian *outlier* dilakukan untuk mengetahui apakah di dalam data yang digunakan untuk regresi terdapat pencilan atau tidak. Pendeteksian *outlier* menggunakan metode *boxplot* untuk mendeteksi *outlier* pada masing-masing data. Sedangkan, pendeteksian *outlier* menggunakan metode *Cook's Distance* untuk mendeteksi *outlier* secara simultan.

4.1. Metode *Boxplot*

Pendeteksian *outlier* menggunakan metode *boxplot* dilakukan dengan bantuan program SPSS16. Berdasarkan gambar 1 dapat diketahui bahwa secara individu terdapat *outlier* pada semua variabel data logAPBD. Pada variabel $\log X_1$ *outlier* terdapat pada baris data ke-3, ke-14, dan ke-90. Pada variabel $\log X_2$ *outlier* terdapat pada baris data ke-2 dan ke-90. Pada variabel $\log Y$ *outlier* terdapat pada baris data ke-3, ke-14, ke-21 dan ke-90.



Gambar 4.1 Boxplot data logAPBD

4.2. Metode Cook's Distance

Pendeteksian outlier menggunakan metode Cook's Distance dilakukan dengan bantuan program SPSS16. Data yang mempunyai nilai $D_i > \frac{4}{90}$ disebut outlier. Berdasarkan tabel 16 dapat diketahui bahwa terdapat outlier pada data logAPBD. Outlier terletak pada baris data ke-3, ke-21, ke-22, ke-52, ke-55, ke-89, dan ke-90.

5. Pendugaan Parameter dengan Metode LMS

Pendugaan parameter regresi robust pada data logAPBD dengan metode LMS dilakukan dengan bantuan program SAS9.1.

Tabel 17 Koefisien regresi robust pada data logAPBD dengan metode LMS

Model	Koefisien
Konstanta	4,946
$\log X_1$	0,072
$\log X_2$	0,520

Berdasarkan Tabel 17 model regresi linier berganda pada data logAPBD dengan metode LMS dituliskan seperti persamaan:

$$\log \hat{Y}_{LMS} = 4,946 + 0,072 \log X_1 + 0,520 \log X_2 .$$

6. Pendugaan Parameter dengan Metode Penduga S

Pendugaan parameter regresi robust pada data logAPBD dengan metode Penduga S dilakukan dengan bantuan program SAS9.1.

Tabel 18 Koefisien regresi robust pada data logAPBD dengan metode Penduga S

Model	Koefisien
Konstanta	4,717
$\log X_1$	0,119
$\log X_2$	0,492

Berdasarkan Tabel 18 model regresi linier berganda pada data logAPBD dengan metode Penduga S dituliskan seperti persamaan:

$$\log \hat{Y}_S = 4,717 + 0,119 \log X_1 + 0,492 \log X_2 .$$

Tabel 16 Hasil pendeteksian *outlier* dengan metode *Cook's Distance*

Data ke-	Nilai D_i	Hasil Deteksi
1	0,001996	bukan <i>outlier</i>
2	0,002735	bukan <i>outlier</i>
3	0,102059	<i>outlier</i>
4	0,004320	bukan <i>outlier</i>
5	0,000137	bukan <i>outlier</i>
6	0,000128	bukan <i>outlier</i>
7	0,018370	bukan <i>outlier</i>
8	0,005168	bukan <i>outlier</i>
9	0,006478	bukan <i>outlier</i>
10	0,000001	bukan <i>outlier</i>
11	0,002266	bukan <i>outlier</i>
12	0,018916	bukan <i>outlier</i>
13	0,000171	bukan <i>outlier</i>
14	0,000757	bukan <i>outlier</i>
15	0,000890	bukan <i>outlier</i>
16	0,021524	bukan <i>outlier</i>
17	0,000544	bukan <i>outlier</i>
18	0,000649	bukan <i>outlier</i>
19	0,021030	bukan <i>outlier</i>
20	0,004649	bukan <i>outlier</i>
21	0,060574	<i>outlier</i>
22	0,045474	<i>outlier</i>
23	0,002427	bukan <i>outlier</i>
24	0,004186	bukan <i>outlier</i>
25	0,009061	bukan <i>outlier</i>
26	0,000010	bukan <i>outlier</i>
27	0,008894	bukan <i>outlier</i>
28	0,003812	bukan <i>outlier</i>
29	0,019696	bukan <i>outlier</i>
30	0,001299	bukan <i>outlier</i>
31	0,000786	bukan <i>outlier</i>
32	0,011133	bukan <i>outlier</i>
33	0,000038	bukan <i>outlier</i>
34	0,000189	bukan <i>outlier</i>
35	0,026663	bukan <i>outlier</i>
36	0,002726	bukan <i>outlier</i>
37	0,002345	bukan <i>outlier</i>
38	0,002243	bukan <i>outlier</i>
39	0,001537	bukan <i>outlier</i>

Lanjutan tabel 16 Hasil pendeteksian *outlier* dengan metode *Cook's Distance*

Data ke-	Nilai D_i	Hasil Deteksi
40	0,009743	bukan <i>outlier</i>
42	0,019237	bukan <i>outlier</i>
43	0,002960	bukan <i>outlier</i>
44	0,000137	bukan <i>outlier</i>
45	0,024290	bukan <i>outlier</i>
46	0,012288	bukan <i>outlier</i>
47	0,006598	bukan <i>outlier</i>
48	0,000001	bukan <i>outlier</i>
49	0,005495	bukan <i>outlier</i>
50	0,005736	bukan <i>outlier</i>
51	0,010365	bukan <i>outlier</i>
52	0,107752	<i>outlier</i>
53	0,001155	bukan <i>outlier</i>
54	0,014950	bukan <i>outlier</i>
55	0,125798	<i>outlier</i>
56	0,008276	bukan <i>outlier</i>
57	0,001809	bukan <i>outlier</i>
58	0,000019	bukan <i>outlier</i>
59	0,003464	bukan <i>outlier</i>
60	0,000571	bukan <i>outlier</i>
61	0,014009	bukan <i>outlier</i>
62	0,002260	bukan <i>outlier</i>
63	0,027405	bukan <i>outlier</i>
64	0,000260	bukan <i>outlier</i>
65	0,000931	bukan <i>outlier</i>
66	0,001187	bukan <i>outlier</i>
67	0,000545	bukan <i>outlier</i>
68	0,000987	bukan <i>outlier</i>
69	0,005955	bukan <i>outlier</i>
70	0,002469	bukan <i>outlier</i>
71	0,003903	bukan <i>outlier</i>
72	0,013429	bukan <i>outlier</i>
73	0,003008	bukan <i>outlier</i>
74	0,009621	bukan <i>outlier</i>
75	0,001208	bukan <i>outlier</i>
76	0,001251	bukan <i>outlier</i>
77	0,004860	bukan <i>outlier</i>
78	0,003088	bukan <i>outlier</i>
79	0,018464	bukan <i>outlier</i>
80	0,014385	bukan <i>outlier</i>
81	0,002911	bukan <i>outlier</i>

Lanjutan tabel 16 Hasil pendeteksian outlier dengan metode Cook's Distance

Data ke-	Nilai D_i	Hasil Deteksi
82	0,003563	bukan outlier
83	0,005273	bukan outlier
84	0,001204	bukan outlier
85	0,000774	bukan outlier
86	0,007621	bukan outlier
87	0,002925	bukan outlier
88	0,032697	bukan outlier
89	0,045083	outlier
90	0,251683	outlier

7. Nilai AIC dan SIC Estimasi Regresi yang Diperoleh dengan Metode LMS dan Metode Penduga S

Perhitungan nilai AIC dan SIC estimasi regresi yang diperoleh dengan metode LMS dan metode Penduga S dilakukan secara manual dengan rumus (Fathurahman, 2009):

$$AIC = 2,718 \frac{2k}{n} \sum_{i=1}^n e_i^2$$

$$SIC = \frac{k}{n} \frac{\sum_{i=1}^n e_i^2}{n}$$

Tabel 19 Hasil perhitungan nilai AIC dan SIC

Nilai	Metode LMS	Metode Penduga S
AIC	25,54423	40,22523
SIC	27,76414	43,72099

Berdasarkan tabel 19 diketahui bahwa model regresi dengan metode LMS menghasilkan nilai AIC sebesar 25,54423 dan SIC sebesar 27,76414. Model regresi dengan metode Penduga S menghasilkan nilai AIC sebesar 40,22523 dan SIC sebesar 43,72099. Dengan demikian nilai AIC dan SIC model regresi dengan metode LMS lebih kecil daripada model regresi dengan metode Penduga S.

PEMBAHASAN

Pengujian asumsi regresi linier berganda pertama dilakukan pada data rekap APBD kabupaten/kota di Pulau Jawa (Data APBD). Berdasarkan uji normalitas dan linieritas model regresi data APBD diperoleh hasil bahwa residual berdistribusi normal dan model regresi linier. Uji keberartian simultan menunjukan bahwa secara bersama-sama variabel Pendapatan Asli Daerah (PAD), Dana Bagi Hasil (DBH), Dana Alokasi Umum (DAU), dan Luas Wilayah (LW) berpengaruh

signifikan terhadap variabel Belanja Modal (BM). Berbeda dengan uji keberartian parsial menunjukan bahwa secara parsial variabel PAD, DBH, dan LW berpengaruh signifikan mempengaruhi variabel BM, sedangkan variabel DAU secara parsial tidak mempengaruhi BM. Dengan demikian variabel DAU harus dieliminasi dari model regresi. Jika dilihat berdasarkan koefisien regresi sebelum variabel DAU dieliminasi dari model bernilai -0,032 yang berarti secara statistik variabel DAU memberikan kontribusi yang sangat kecil terhadap variabel BM. Hal ini tidak sejalan dengan kajian teori bahwa variabel DAU mempengaruhi variabel BM. Variabel DAU menjadi tidak berpengaruh terhadap BM karena ada indikasi bahwa DAU lebih banyak digunakan untuk membiayai belanja pegawai dan belanja rutin. Dana Perimbangan yang ditransfer oleh Pemerintah Pusat ke Daerah lebih banyak pada pos DAU, namun penggunaan DAU ini lebih banyak digunakan untuk menutup Belanja-belanja Operasional daripada Belanja Modal, sehingga DAU menjadi tidak berpengaruh terhadap Belanja Modal. Dengan demikian untuk analisis selanjutnya dilakukan pada model regresi variabel PAD, DBH, dan LW terhadap variabel BM. Model regresi tersebut tidak terdapat masalah multikolinearitas namun terjadi masalah heteroskedastisitas, sedangkan asumsi autokorelasi belum dapat diambil kesimpulan. Oleh karena terdapat masalah heteroskedastisitas, maka berdasarkan kajian teori bahwa pada data APBD perlu dilakukan transformasi logaritma natural (log).

Pengujian asumsi regresi linier berganda selanjutnya dilakukan pada hasil transformasi log data APBD (data logAPBD). Berdasarkan uji normalitas dan linieritas model regresi data logAPBD diperoleh hasil bahwa residual berdistribusi normal dan model regresi linier. Uji keberartian simultan menunjukan

bahwa secara bersama-sama variabel PAD, DBH, dan LW berpengaruh signifikan terhadap variabel BM. Berbeda dengan uji keberartian parsial menunjukkan bahwa secara parsial variabel PAD dan DBH berpengaruh signifikan mempengaruhi variabel BM, sedangkan variabel LW secara parsial tidak mempengaruhi BM. Dengan demikian variabel BM harus dieliminasi dari model regresi. Jika dilihat berdasarkan koefisien regresi sebelum variabel LW dieliminasi dari model bernilai 0,019. Artinya, secara statistik variabel LW memberikan kontribusi yang sangat kecil terhadap variabel BM. Hal ini tidak sejalan dengan kajian teori bahwa variabel LW mempengaruhi variabel BM secara parsial. Hal ini disebabkan karena semakin luas wilayah suatu daerah menyebabkan kebutuhan jumlah pegawai yang cukup besar, sehingga kebutuhan Belanja Pegawai untuk kegiatan operasional menjadi cukup besar dibandingkan untuk Belanja Modal. Dengan demikian untuk analisis selanjutnya dilakukan pada model regresi variabel PAD dan DBH terhadap variabel BM. Model regresi tersebut tidak terdapat masalah multikolinearitas, heteroskedastisitas, dan autokorelasi. Dengan demikian, sesuai dengan kajian teori bahwa model regresi linier berganda variabel PAD dan DBH terhadap variabel BM bersifat BLUE

Estimasi regresi linier berganda dengan OLS dilakukan pada data logAPBD. Model regresi linier berganda data logAPBD yaitu $\log \hat{Y} = 3,520 + 0,215 \log X_1 + 0,500 \log X_2$. Interpretasi model tersebut sebagai berikut.

1. Konstanta sebesar 3,520, artinya jika variabel PAD dan DBH bernilai nol maka BM bernilai 3,520. Hal ini karena BM dipengaruhi oleh faktor variabel bebas selain pada model.
2. Koefisien $\log X_1$ sebesar 0,215, artinya semakin bertambah PAD sebesar satu satuan maka semakin bertambah BM sebesar 0,215 satuan.
3. Koefisien $\log X_2$ sebesar 0,500, artinya semakin bertambah DBH sebesar satu satuan maka semakin bertambah BM sebesar 0,500 satuan.

Pendeteksian *outlier* dilakukan untuk mendeteksi adanya *outlier* pada data logAPBD. Pendeteksian *outlier* dilakukan dengan metode *boxplot* dan *Cook's Distance*.

Berdasarkan metode *boxplot* dan *Cook's Distance* memberikan hasil bahwa pada data logAPBD terdapat *outlier* sehingga perlu dilakukan pendugaan parameter dengan metode LMS dan Penduga S.

Pendugaan parameter menggunakan metode LMS pada data logAPBD menghasilkan model regresi linier berganda yaitu

$$\log \hat{Y}_{LMS} = 4,946 + 0,072 \log X_1 + 0,520 \log X_2.$$

Interpretasi dari model tersebut sebagai berikut.

1. Konstanta sebesar 4,946, artinya jika variabel PAD dan DBH bernilai nol maka BM bernilai 4,946. Hal ini karena BM dipengaruhi oleh faktor variabel bebas selain pada model.
2. Koefisien $\log X_1$ sebesar 0,072, artinya semakin bertambah PAD sebesar satu satuan maka semakin bertambah BM sebesar 0,072 satuan.
3. Koefisien $\log X_2$ sebesar 0,520, artinya semakin bertambah DBH sebesar satu satuan maka semakin bertambah BM sebesar 0,520 satuan.

Pendugaan parameter menggunakan metode Penduga S pada data logAPBD menghasilkan model regresi linier berganda yaitu

$$\log \hat{Y}_S = 4,717 + 0,119 \log X_1 + 0,492 \log X_2.$$

Interpretasi dari model tersebut sebagai berikut.

1. Konstanta sebesar 4,717, artinya jika variabel PAD dan DBH bernilai nol maka BM bernilai 4,717. Hal ini karena BM dipengaruhi oleh faktor variabel bebas selain pada model.
2. Koefisien $\log X_1$ sebesar 0,119, artinya semakin bertambah PAD sebesar satu satuan maka semakin bertambah BM sebesar 0,119 satuan.
3. Koefisien $\log X_2$ sebesar 0,492, artinya semakin bertambah DBH sebesar satu satuan maka semakin bertambah BM sebesar 0,492 satuan.

Berdasarkan hasil penelitian nilai AIC dan SIC model regresi dengan metode LMS lebih kecil daripada nilai AIC dan SIC model regresi dengan metode Penduga S. Dengan demikian, bahwa metode LMS merupakan metode *robust* terbaik diantara LMS dan Penduga S. Metode LMS memiliki kelebihan yaitu meminimumkan kuadrat median pada data regresi sedangkan pada metode penduga S tidak demikian. Berdasarkan teori, hal ini salah satu yang menyebabkan variansi parameter duga dari metode LMS lebih kecil. Sebagai akibat selanjutnya, metode LMS menghasilkan MSE yang lebih kecil sehingga berdampak pada nilai AIC dan SIC yang lebih kecil pula.

PENUTUP

Berdasarkan hasil penelitian dan pembahasan, maka dapat disimpulkan bahwa hasil estimasi regresi linier berganda metode *robust* LMS yaitu $\log \hat{Y}_{LMS} = 4,94 + 0,072 \log X_1 + 0,52 \log X_2$ dan hasil estimasi regresi linier berganda metode *robust* Penduga S yaitu $\log \hat{Y}_S = 4,717 + 0,119 \log X_1 + 0,492 \log X_2$ dengan X_1 adalah Pendapatan Asli Daerah (PAD), X_2 adalah Dana Bagi Hasil (DBH), dan Y adalah Belanja Modal (BM). LMS merupakan metode regresi *robust* terbaik dibandingkan metode Penduga S.

DAFTAR PUSTAKA

- Chen, C. 2014. Robust Regression and Outlier Detection with the Robustreg Procedure. *Proceedings International. America: SAS Institute Inc.* Tersedia di <http://www2.sas.com/proceedings/ugi27/p265-27.pdf> [diakses 5-4-2015].
- Draper, N. R., & Smith, H. 1992. *Analisis Regresi Terapan*. Jakarta: Gramedia Pustaka Utama.
- Fathurahman, M. 2009. Pemilihan Model Regresi Terbaik Menggunakan Metode Akaike's Information Criterion dan Schwarz Information Criterion. *Jurnal Informatika Mulawarman*, 4(3): 37-39. Tersedia di <https://informatikamulawarman.files.wordpress.com> [diakses 1-12-2015].
- Makkulau, et al. 2010. Pendeteksian Outlier dan Penentuan Faktor-Faktor yang Mempengaruhi Produksi Gula dan Tetes Tebu dengan Metode Likelihood Displacement Statistic-Lagrange. *Jurnal Teknik Industri*, 12(2): 95-100. Tersedia di <http://ced.petra.ac.id/index.php> [diakses 1-4-2015].
- Nurchayadi, H. 2010. *Analisis Regresi pada Data Outlier dengan Menggunakan Least Trimmed Square (LTS) dan MM-Estimasi*. Skripsi. Jakarta: Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah. Tersedia di <http://repository.uinjkt.ac.id> [diakses 10-5-2015].
- Oktarinanda, A. 2014. Perbandingan Efisiensi Metode Least Trimmed Square (Lts) dan Metode Least Median Square (LMS) dalam Estimasi Parameter Regresi Robust. *Jurnal Statistik*, 2(3): 177-180. Tersedia di <http://statistik.studentjournal.ub.ac.id/> [diakses 1-4-2015].
- Paludi, S. 2009. Identifikasi dan Pengaruh Keberadaan Data Pencilon (Outlier). *Majalah Panorama Nasional*, Januari-Juni. Hlm. 56-62. Tersedia di http://stein.ac.id/e-journal/pn_6/PN_6.pdf [diakses 2-4-2015].
- Permana, A.T. 2014. Perbandingan Metode Least Trimmed Square (LTS) dan Penduga-S sebagai Metode Pendugaan Parameter Regresi Robust. *Jurnal FMIPA Universitas Brawijaya*, 2(2): 125-128. Tersedia di <http://statistik.studentjournal.ub.ac.id/> [diakses 1-4-2015].
- Rousseeuw, P.J. 1984. Least Median of Squares Regression. *Journal of the American Statistical Association*, 79(388): 871-880. Tersedia di <http://www.cse.yorku.ca> [diakses 3-4-2015].
- Sembiring, R. K. 2003. *Analisis Regresi* (2th ed.). Bandung: Institut Teknologi Bandung.
- Solikhah, I. 2014. *Analisis Belanja Modal pada Pemerintah Kabupaten/Kota di Jawa*. Skripsi. Semarang: FE Universitas Negeri Semarang.
- Wijayanti, L.U. 2015. *Analisis Perbandingan Regresi Robust Estimasi-M Huber dan Estimasi-S dalam Mengatasi Outlier*. Skripsi. Yogyakarta: FMIPA Universitas Negeri Yogyakarta. Tersedia di <http://eprints.uny.ac.id/17923/> [diakses 4-4-2015].