

REGRESI ROBUST UNTUK MENGATASI DATA PENCILAN

Faula Arina[†]

Jurusan Teknik Industri Universitas Sultan Ageng Tirtayasa Cilegon

Jl. Jend. Sudirman Km. 3 Cilegon, Banten 42435

Faula.arina71@gmail.com

ABSTRAK

Analisis regresi merupakan metode analisis dalam statistika yang paling banyak diaplikasikan di berbagai bidang baik ekonomi maupun sosial. Metode Kuadrat Terkecil (MKT) merupakan metode yang sering digunakan untuk mengestimasi parameter model regresi. Pencilan adalah data yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat data. Jika terdapat pencilan, MKT tidak akurat untuk mengestimasi parameter. Untuk mengatasi masalah ini, salah satu metode yang digunakan adalah metode regresi *robust*. *Least Trimmed Square* (LTS) merupakan salah satu estimator regresi *robust* terhadap pencilan. Tujuan dari penelitian ini adalah membandingkan tingkat efektifitas metode LTS dan MKT pada data yang mengandung pencilan berdasarkan nilai Standard Error (SE). Hasil penelitian menunjukkan bahwa metode LTS lebih baik dalam mengestimasi parameter pada data yang mengandung pencilan.

Kata Kunci: LTS, MKT, Regresi Robust, Standard Error

I. PENDAHULUAN

Regresi merupakan suatu metode statistika yang digunakan untuk menyelidiki pola hubungan antara variabel dependen dan variabel independen. Secara umum regresi linier terdiri dari regresi linier sederhana dan regresi linier berganda (Draper dan Smith, 1996).

Metode Kuadrat Terkecil (MKT) merupakan metode yang sering digunakan untuk mengestimasi parameter model regresi dan mempunyai asumsi-asumsi yang harus dipenuhi diantaranya kenormalan galat e_i . Namun metode ini mempunyai kelemahan ketika datanya ada pencilan. Pencilan adalah data yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat data. Akibat dari adanya pencilan galat e_i tidak lagi berdistribusi normal atau variansi dari galatnya tidak lagi homogen. Pencilan dapat dideteksi dengan menggunakan *leverage*, nilai *discrepancy*, nilai *influence*. Pencilan atau sisaan besar adalah nilai sisaan yang lebih besar atau lebih kecil dari 3 kali galat baku. Bila terdapat pencilan maka model regresi dengan pendekatan MKT akan sangat sensitif terjadinya bias. Walaupun metode MKT menjamin adanya BLUE (*Best Linier unbiased Estimator*) $E(\hat{\beta}) = \beta$. Permasalahannya pada sifat BLUE yaitu ragam atau variansi yang tidak konstan yang disebabkan adanya data pencilan, maka MKT yang diperoleh menjadi tidak efisien (Ryan, 1997). Bila ada pencilan yang berpengaruh terhadap model maka model yang dihasilkan MKT akan menjadi tidak sesuai. Untuk mengatasi adanya pencilan dalam analisis regresi digunakan metode regresi kekar (*robust regression*) (Huber, 1981). MKT bukan merupakan prosedur regresi yang *robust* terhadap adanya pencilan, sehingga estimasinya menjadi tidak sesuai (Ryan, 1997). Jika terdapat pencilan, MKT tidak akurat untuk mengestimasi parameter. Untuk mengatasi masalah ini, salah satu metode yang digunakan adalah metode regresi *robust*.

Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi dari residual tidak normal dan atau mengandung beberapa pencilan yang berpengaruh pada model (Chen, 2002). *Least Trimmed Square* (LTS) merupakan salah satu estimator regresi *robust* terhadap pencilan dan memiliki *breakdown point* yang tinggi dikenalkan oleh Rousseeuw (1984). Metode LTS mengestimasi parameter regresi dengan meminimumkan jumlah kuadrat residual yang terurut dari terkecil hingga terbesar dari data yang sudah terpotong (*trimmed*). Sedangkan metode MKT estimasi parameter dengan meminimumkan jumlah kuadrat residual dari keseluruhan data. Zaman et al (2001) telah melakukan penelitian mengenai perbandingan MKT dan LTS.. Dalam penelitian tersebut MKT menghasilkan standard error (SE) yang lebih besar dibanding LTS. LTS mengestimasi parameter model dengan meminimumkan kuadrat residual untuk pengamatan sebanyak h dimana $h < n$. Nilai h akan membangun *breakdown point* yang besar sebanding dengan 50% (Rousseeuw dan Leroy, 1987). Tujuan dari penelitian ini adalah membandingkan tingkat efektifitas metode LTS dan MKT pada data yang mengandung pencilan. Tingkat efisiensi dari kedua metode dibandingkan berdasarkan Standard Error (SE) dari nilai estimasi parameter yang dihasilkan.

II. METODE PENELITIAN

2.1 Data

Data penelitian ini merupakan data dari Nurcahya (2009) berupa data perusahaan asuransi pensiun dari 18 cabang perusahaan dengan x = pendapatan premi dan y = Premi cadangan.

Tabel 1 Data Penelitian

No	1	2	3	4	5	6
----	---	---	---	---	---	---

[†] Corresponding Author

X	10.4	15.6	16.2	17.9	37.8	46.9
Y	272.2	212.9	120.7	163.6	226.1	622.9

No	7	8	9	10	11	12
X	52.4	52.9	71	73.9	16.3	77
Y	1353.2	363.6	951.7	307.2	588.4	952.5

No	13	14	15	16	17	18
X	131.2	151	206.1	314.7	470.8	1406.3
Y	1157.3	2105.6	3581.4	3404.7	4095.3	2802.7

2.2 Metode

Metode penelitian ini adalah

1. Mengestimasi koefisien regresi linier sederhana menggunakan MKT. Dalam model regresi, persamaan umum yang digunakan adalah

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

2. Pengujian asumsi klasik analisis regresi linier
3. Mendeteksi adanya pencilan dengan metode leverage (h_{ii}) dan *externally studentized residuals*
4. Membuat regresi robust dengan penduga LTS.

$$\hat{\theta}_{LTS} = \arg \min_{\theta} E_{LTS}^2(\theta) \text{ dengan } E_{LTS}^2 = \sum_{i=1}^h e_{(i=n)}^2$$

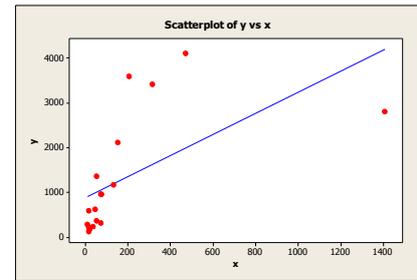
Tahap algoritma LTS yaitu

- a. Mengestimasi parameter dengan MKT
 - b. Menghitung kuadrat galat $e^2_{(i)}$ kemudian diurutkan dari nilai terkecil hingga terbesar, $e^2_{(1)} < e^2_{(2)} < \dots < e^2_{(n)}$
 - c. Menghitung h , $[n/2] + 1 \leq h \leq [\frac{3n+p+1}{4}]$ dan breakdown value = $(([n/2] - p + 2)/n)$
 - d. Menghitung $E_{LTS}^2 = \sum_{i=1}^h e^2_{(i=n)}$
 - e. Melakukan estimasi parameter $b_{baru(i)}$ dari $h_{baru(i)}$ pengamatan
 - f. Menentukan kuadrat residual $e^2_{(i)}$ dari $h_{baru(i)}$ pengamatan
 - g. Menghitung $E_{LTS}^2(baru)$
 - h. Melakukan C-steps yaitu tahap f sampai h untuk mendapatkan fungsi obyektif h yang terkecil dan konvergen ke 0.
- Pada penelitian ini menggunakan ROBUSTREG procedure di SAS algoritmanya menggunakan FAST-LTS (Rousseeuw dan Van, 2000).
5. Membandingkan nilai koefisien determinasi (R^2) dari MKT dan LTS

III. HASIL DAN PEMBAHASAN

3.1 Estimasi Parameter regresi linier sederhana dengan MKT

Berdasarkan Tabel 1 hasil estimasi parameter dengan MKT diperoleh $\hat{y} = 878 + 2.36 x$ dengan nilai $R^2 = 34.68\%$. Yang artinya hanya 34.68 % pendapatan premi yang mempengaruhi premi cadangan.

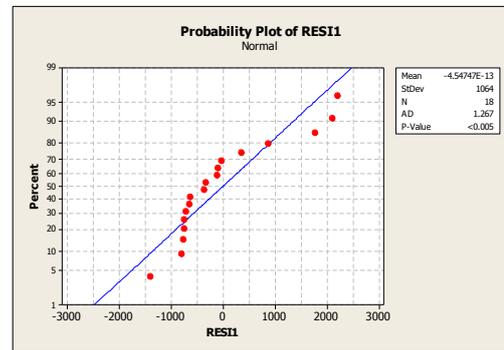


Gambar 1 Plot data penelitian

Tabel 2 Hasil Olah Program SAS menggunakan MKT

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	878.31987	294.80023	2.98	0.0089
x	1	2.35836	0.80606	2.93	0.0099

Dari Pengujian asumsi klasik analisis regresi linier diperoleh pelanggaran asumsi MKT yaitu galat yang tidak normal, dapat dilihat pada Gambar (2).



Gambar 2 Plot galat yang tidak normal

Asumsi ini dilanggar karena diduga adanya data pencilan. Maka dilakukan deteksi pencilan. Pendeteksian pencilan dengan metode leverage (h_{ii}) $> \frac{2(p+1)}{n} = 0.222$ dan *externally studentized residuals* (TRES) $> t$ table $(\alpha/2, n-2) = 2.120$ diperoleh pencilan :

Tabel 3 Data Pencilan dari Data penelitian

Data	Pengamatan	h_{ii}	TRESS	Keputusan
1	15	0.056045	2.3599	Pencilan
2	18	0.873729	-7.7012	Pencilan

3.2 Estimasi Parameter Menggunakan Metode Least Trimmed Square

Dalam mengestimasi parameter regresi menggunakan LTS pada data yang mengandung pencilan dilakukan langkah-langkah sebagai berikut:

- a. Mengestimasi parameter dengan MKT diperoleh $\hat{y} = 878 + 2.36 x$
 - b. Menghitung kuadrat galat $e^2_{(i)}$ dan Menghitung h
- Pada iterasi pertama digunakan model regresi awal untuk menghitung kuadrat galat dan dilakukan

perhitungan terhadap h pengamatan, maka diperoleh h = 14, maka akan diurutkan nilai kuadrat galat dari yang terkecil ke yang terbesar dari data 1– 14.

c. Menghitung $E_{LTS}^2 = \sum_{i=1}^h e_{(i=n)}^2 = 4756738$

d. Melakukan estimasi parameter b baru (i) dari h baru (i) pengamatan diperoleh $\hat{y} = 75 + 10.8 x$ sehingga dibuat tabelnya pada Tabel 4

Tabel 4 Estimasi Parameter b baru (i) dari h baru (i)

Iterasi	H	b0	b1	E_{LTS}^2
0	18	878	2.36	
1	14	75	10.8	4756738

Pada penelitian ini menggunakan ROBUSTREG procedure di SAS dan algoritma nya menggunakan FAST-LTS diperoleh nilai estimasi parameter pada data asuransi yang mengandung pencilan yaitu $\hat{y} = 175.74 + 9.0528 x$. Dengan nilai *breakdown point* 0.2778 dan nilai $R^2 = 80.72 \%$.

Tabel 5 Hasil Olah Program SAS menggunakan LTS

Parameter	DF	Estimate	Standard Error	95% Confidence Limits	Chi-Square	Pr > ChiSq
Intercept	1	175.7387	113.2478	-46.2230 397.7004	2.41	0.1207
x	1	9.0528	0.7277	7.6265 10.4790	154.76	<.0001
Scale	0	353.6529				

3.3. Perbandingan Estimasi Parameter dengan MKT dan LTS berdasarkan nilai koefisien Determinasi

Berdasarkan Tabel 6 menunjukkan nilai SE metode MKT lebih besar dari LTS. Hal ini menunjukkan estimasi dengan LTS lebih baik untuk data yang mengandung pencilan.

Tabel 6 Hasil Estimasi Koefisien Regresi dan Koefisien Determinasi

Koefisien	Metode	
	MKT	LTS
Bo	878	175.7387
SE bo	294.80023	113.2478
b1	2.36	9.0528
SE b1	0.80606	0.7277

IV. KESIMPULAN

Berdasarkan pengolahan data maka dapat disimpulkan sebagai berikut :

1. Regresi dengan menggunakan pendugaan MKT sangat rentan terpengaruh dengan adanya pencilan
2. Pendugaan LTS dapat mengatasi adanya pencilan dalam regresi.
3. Berdasarkan nilai Standar error (SE) diperoleh bahwa metode LTS lebih efektif dari pada metode MKT. Hal ini menunjukkan metode LTS lebih baik dalam mengestimasi parameter pada data yang mengandung pencilan.

DAFTAR PUSTAKA

Chen, C .2002. Robust Regression and Outlier detection with The ROBUSTREG Procedure.. SUGI paper 265-267. Cary, NC: SAS Institute Inc.

Draper dan Smith.1996. *Applied Regression Analysis 2nd edition* New York :John Wiley & Sons. Chapman and Hall.

Nurcahyadi .2009. Analisis Regresi pada Data Outlier dengan Menggunakan Least Trimmed Square (LTS) dan MM-Estimasi. Skripsi. Program Studi Matematika UIN Syarif Hidayatullah Jakarta.

Rousseeuw, P.J. 1984. Least Median Squares Regression. *Journal of the American Statistical Association*. Vol. 79. Number 388.

Rousseeuw dan Leroy, 1987. *Robust Regression and Outlier Detection*. Canada : John Wiley & Sons, Inc.

Rousseeuw dan Van driessen, K.2000. An Algorithm for positive Breakdown Regression based on Concentration Steps. In *Data Analysis*. New York : Springer-Verlag

Ryan, T.P. 1997. *Modern Regression Methods*. Canada: John Wiley & Sons, Inc.

Zaman, A., Rousseeuw, PJ and Orhan, M. 2001, Econometric Application of High Breakdown Robust Regression Techniques. *Economic Letters*. 1-8.