KERAGAAN REGRESI LS, LAD, DAN MLAD PADA DATA *DELIVERY TIME*

(*The Performance of LS, LAD, and MLAD Regression on Delivery Time Data*)

Setyono[1], IM Sumertajaya[2], A Kurnia[2], dan AA Mattjik[2]

[1]Jurusan Agroteknologi, Fakultas Pertanian, Universitas Djuanda Bogor
[2]Jurusan Statistika, Fakultas MIPA, Institut Pertanian Bogor.
Korespondensi: Setyono 081318949696, E-mail: setyono@unida.ac.id

Abstrak

Pendugaan koefisien regresi berbasis optimasi sisaan yang dikenal adalah dengan cara meminimumkan jumlah kuadrat sisaan (LS) dan meminimumkan jumlah sisaan mutlak (LAD). Pendugaan dengan cara meminimumkan maksimum sisaan mutlak (MLAD) belum dikembangkan. Tujuan penelitian ini adalah untuk mengetahui apakah program linier dapat digunakan untuk mendapatkan penduga koefisien regresi yang meminimumkan maksimum sisaan mutlak dan membandingkan hasilnya dengan hasil pendugaan menggunakan metode LS dan LAD. Data yang digunakan adalah data *Delivery Time* yang biasa digunakan untuk uji coba metode regresi. Hasil penelitian menunjukkan bahwa program linier dapat digunakan untuk mendapatkan penduga koefisien regresi yang meminimumkan maksimum sisaan mutlak, pada data *Delivery Time* regresi LAD paling baik menurut kriteria validasi silang, sedangkan regresi LS paling stabil menurut semua kriteria. Dalam metode MLAD dimungkinkan diperoleh subset pengamatan yang menghasilkan penduga koefisien regresi yang sama besar dengan penduga koefisien regresi dari keseluruhan pengamatan.

Kata kunci : MLAD, program linier, regresi, sisan mutlak, validasi silang

Abstract

Estimation of regression coefficients based on residuals optimization which are commonly known, are by minimizing the residual sum of squares (LS) and by minimizing the amount of absolute residual (LAD). Estimation by minimizing the maximum absolute residual (MLAD) has not been developed. The purpose of this study was to determine whether the linear programming can be used to estimate regression coefficients that minimize the maximum absolute residual and compare its results with the results of the LS and LAD. The data used was the Delivery Time data that commonly used for regression method testing. The results showed that linear programming can be used to estimate regression coefficients that minimize the maximum absolute residual, the LAD regression is the best for cross-validation criteria, whereas LS regression is the most stable according to all of criterias. In MLAD regression it is possible to obtain a subset of observations which its regression coefficient is the same as the regression coefficient resulted by overall observations.

Key words: MLAD, linear programming, regression, absolute residual, cross validation

Setyono, IM Sumertajaya, A Kurnia, dan AA Mattjik, 2015. Keragaan Regresi LS, LAD, dan MLAD Pada Data *Delivery Time*. *Jurnal Agronida* 2 (1): 18 – 26.

PENDAHULUAN

Latar Belakang

Regresi merupakan salah satu analisis statistika yang sering digunakan dalam penelitian. Dalam Statistika, regresi termasuk salah satu bahasan dalam model linier. Teori model linier klasik pada prinsipnya adalah teori untuk pendugaan bersyarat, yaitu $\mu(Y|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$. Regresi seperti itu dikenal dengan model rataan atau dengan istilah lebih umum adalah model pemusatan. Pendugaan koefisien regresi pada model pemusatan pada umumnya dilakukan berdasarkan optimasi sisaan. Cara yang sudah biasa dikenal adalah dengan meminimumkan jumlah kuadrat sisaan (*least square* disingkat LS) dan berikutnya dengan meminimumkan jumlah sisaan mutlak (*least absolute deviation* disingkat LAD). Sementara itu optimasi sisaan dengan cara meminimumkan maksimum sisaan mutlak (*minimum largest absolute deviation* disingkat MLAD) sudah dirintis oleh Rudolf *et al* (1999), tetapi belum dikembangkan dalam paket program komputer untuk statistika.

Metode LS sudah diimplementasikan pada semua paket program komputer untuk statistika. Metode LS dapat diimplementasikan paling awal karena penduga koefisien regresi dan sifat-sifatnya dapat dipeoleh secara analitik. Oleh sebab itu pembelajaran regresi dan model linier menggunakan metode ini. Hasil metode LS sering dijadikan nilai awal bagi pendugaan koefisien regresi yang membutuhkan solusi secara iteratif, misalnya regresi t-Student (Setyono *et al*. 1996).

Metode LAD dapat dikerjakan dengan beberapa cara, antara lain dengan regresi median dan regresi terbobot iteratif *(iteratively weighted least squares* disingkat IWLS). Metode LAD sudah diimplementasikan pada paket program komputer yang menyediakan regresi kuantil. Metode LAD relatif kekar (*robust*) terhadap pencilan, karena memberi bobot yang besarnya berbanding terbalik dengan besarnya sisaan. Penduga LAD untuk ukuran pemusatan tidak khas dan telah dibuktikan oleh Hao dan Naiman (2007) bahwa salah satu penduganya adalah median. Oleh sebab itu regresi LAD dapat dikerjakan dengan regresi median atau regresi kuantil-0.5. Regresi kuantil ini mulai diperkenalkan oleh Koenker dan Bassett (1978) dan dikembangkan oleh Koenker dan Hallock (2001)

Regresi MLAD digunakan ketika yang dikehendaki adalah model yang tidak pernah memiliki sisaan yang besar, atau maksimum sisaannya dibuat sekecil mungkin. Hal ini penting karena pada masalah yang menyangkut kepentingan publik, terjadinya kasus dengan simpangan yang besar menjadi sorotan meskipun kasus lain aman. Ilustrasi sederhana misalnya pada setelan roda, baik terlalu oleng ke kiri maupun terlalu oleng ke kanan di suatu titik dapat menjadi masalah meskipun hampir sepanjang putaran yang lain relatif di tengah. Pada kasus seperti ilustrasi itu regresi LS dan regresi LAD kurang tepat, karena meskipun pada hampir semua titik memiliki sisaan kecil, tidak menutup kemungkinan terjadinya sisaan besar di suatu titik.

Data *Stack Loss* dan data *Delivery Time* adalah dua data yang sering digunakan oleh Statistikawan untuk uji coba metode regresi. Metode regresi yang pernah diterapkan kedua data tersebut antara lain pemodelan normal, Huber, Ramsay, Andrews, dan pemodelan t (Setyono *et al* 1996). Keragaan metode LS, LAD, dan MLAD pada data *Stack Loss* pernah dikaji oleh Setyono *et al* (2014) yang menyertakan hasil simulasi menggunakan sebaran normal. Keragaan MLAD pada data *Delivery Time* belum dikaji secara detail. Oleh sebab itu perlu dikaji keragaaan metode LS, LAD, dan MLAD dalam menduga koefisien regresi pada *Delivery Time*, termasuk beberapa karakteristiknya.

Tujuan

Penelitian ini memiliki 3 tujuan, yaitu:

1. Pada pendugaan parameter yang berbasis pada optimasi sisaan, metode yang digunakan juga merupakan kriteria kebaikannya. Sebagai contoh kalau kriteria yang digunakan adalah jumlah kuadrat

sisaan maka metode yang terbaik adalah metode kuadrat terkecil. Tujuan pertama dari kajian ini adalah memeriksa apakah pada data *Delivery Time* metode MLAD selain terbaik dari segi maksimum sisaan mutlaknya, juga cukup baik dari segi jumlah kuadrat sisaan dan jumlah sisaan mutlaknya.
2. Salah satu kebaikan suatu metode regresi dinilai berdasarkan nilai validasi silang. Tujuan kedua dari kajian ini adalah mengetahui metode mana yang terbaik untuk data *Delivery Time* menurut kriteria validasi silang.
3. Dalam metode MLAD yang dikerjakan melalui program linier, pengamatan (*case*) berperan sebagai kendala (*constraint*) dalam program linier. Tujuan berikutnya adalah mengetahui subset pengamatan dari data *Delivery Time* yang menjadi penentu besarnya koefisien regresi MLAD.

MATERI DAN METODE

Data

Data yang akan digunakan adalah data *Delivery Time*, yang terdiri atas 25 pengamatan dari 3 peubah, yaitu *delivery time* (Y), *the number of cases of product stocked* (X_1), dan *the distance walked by the route driver* (X_2). Data *Delivery Time* dipilih untuk simulasi dengan pertimbangan data ini sering digunakan untuk contoh metode regresi, antara lain Rousseeuw dan Leroy (1987), Golberg dan Cho (2010), dan Montgomery *et al*. (2012). Metode regresi yang pernah diaplikasikan antara lain pemodelan normal, pemodelan normal yang disertai diagnosis terhadap pencilan, metode Huber, metode Andrews, metode Ramsay, metode Hampel, dan pemodelan t (Setyono *et al* 1996). Data *Delivery Time* disajikan pada Tabel 1.

Tabel 1. Data *Delivery Time*

No	Y	X_1	X_2	No	Y	X_1	X_2
1	16.68	7	560	14	19.75	6	462
2	11.50	3	220	15	24.00	9	448
3	12.03	3	340	16	29.00	10	776
4	14.88	4	80	17	15.35	6	200
5	13.75	6	150	18	19.00	7	132
6	18.11	7	330	19	9.50	3	36
7	8.00	2	110	20	35.10	17	770
8	17.83	7	210	21	17.90	10	140
9	79.24	30	1460	22	52.32	26	810
10	21.50	5	605	23	18.75	9	450
11	40.33	16	688	24	19.83	8	635
12	21.00	10	215	25	10.75	4	150
13	13.50	4	255				

Regresi MLAD

Regresi MLAD dilaksanakan dengan program linier dengan panduan sebagai berikut. Misalkan y_i adalah respon pengamatan ke-i, x_i' adalah vektor kovariat pengamatan ke-i, b adalah vektor koefisien regresi, dan z≥0 adalah batas atas (*upper boundary*) sisaan mutlak sehingga

$$0 \le |y_i - \mathbf{x}_i'\mathbf{b}| \le z; \text{untuk semua i}$$

Untuk setiap pengamatan ke-i perlu diperhatikan dua kasus, yaitu ketika sisaan positif

$$0 \le y_i - \mathbf{x}_i'\mathbf{b} \le z \Leftrightarrow \mathbf{x}_i'\mathbf{b} + z \ge y_i$$

dan ketika sisaan negatif

$$-z \le y_i - \mathbf{x}_i'\mathbf{b} \le 0 \Leftrightarrow \mathbf{x}_i'\mathbf{b} - z \le y_i$$

Dengan demikian pada regresi MLAD ini nilai z diminimumkan dengan kendala

$$\mathbf{x}_i'\mathbf{b} - z \le y_i \text{ dan } \mathbf{x}_i'\mathbf{b} + z \ge y_i$$

Panduan untuk pemrograman linier dapat merujuk pada McCarl dan Spreen (1997) atau Winston dan Goldberg (2004), sedangkan untuk mewujudkannya dalam bahasa R dapat merujuk pada Rizzo (2008).

Keragaan MLAD, LAD, dan LS

Kebaikan suatu metode pada satu set data salah satunya menggunakan kriteria validasi silang. Pada kajian ini validasi silang dilakukan dengan langkah-langkah berikut:
1. Dimulai dari i=1
2. Pengamatan ke-i dibuang
3. Dilakukan pendugaan koefisien regresi tanpa menyertakan pengamatan ke-i
4. Dilakukan pendugaan terhadap nilai respon pengamatan ke-i berdasarkan model langkah ketiga
5. Dihitung selisih antara nilai repon pengamatan ke-i yang dibuang pada

langkah kedua dengan hasil pendugaan respon pengamatan ke-i pada langkah keempat, lalu dicatat sebagai e_i

6. Lakukan langkah 2 sampai 5 untuk i=2, 3, …, n

Setelah itu dilakukan pengukuran validasi silang (CV) dengan tiga cara, yaitu:

$$\text{Rata} - \text{rata kuadrat galat prediksi} = CV(1) = \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

$$\text{Rata} - \text{rata galat mutlak prediksi} = CV(2) = \frac{1}{n}\sum_{i=1}^{n} |e_i|$$

$$\text{Maksimum galat mutlak prediksi} = CV(3) = max(|e_i|)$$

Pendugaan Galat Baku M

Pada regresi LS galat baku koefisien regresi dapat diturunkan secara matematis menjadi sebuah rumus atau bentuk tertutup (*close form*). Pada regresi MLAD koefisien regresinya saja tidak dapat dinyatakan dalam bentuk tertutup, apa lagi galat bakunya. Untuk keperluan inferensia dibutuhkan pendugaan galat baku bagi koefisien regresi berdasarkan satu set data. Salah satu cara mencari penduga galat baku adalah melalui *bootstrap*. *Bootstrap* adalah mengambil contoh dengan pemulihan berulang-ulang. *Bootstrap* dilakukan dengan dua cara, yaitu *bootstrap* pengamatan dan *bootstrap* sisaan (Givens dan Hoeting 2005). *Bootstrap* pengamatan berarti menganggap nilai pasangan pengamatan (x,y) adalah contoh acak dari populasi pasangan pengamatan (x,y). *Bootstrap* terhadap pengamatan pernah dilakukan oleh Setyono *et al* (1996). Sementara itu *bootstrap* sisaan berarti menganggap matriks rancangan bersifat tetap, sedangkan galat bersifat acak. *Bootstrap* terhadap sisaan pernah dilakukan oleh Zhu dan Jing (2010).

Pada pendugaan galat baku bagi koefisien regresi lebih tepat menggunakan *bootstrap* sisaan. Untuk itu diasumsikan bahwa sebaran sisaan e_i mewakili sebaran galat ε_i, sehingga dapat dilakukan *bootstrap* berdasarkan e_i sebanyak n. Langkah detailnya sebagai berikut:

1) Dilakukan regresi terhadap data yang akan dianalisis, sehingga diperoleh koefisien regresi **b** dan sisaan e.
2) Dihitung $\hat{\mathbf{u}} = \mathbf{Xb}$
3) Diambil contoh nilai d_i sebanyak n dengan pemulihan dari e_i hasil langkah 1
4) Dihitung nilai $z_i = \hat{u}_i + d_i$
5) Diregresikan **z** terhadap **X**, sehingga diperoleh koefisien regresi **a**
6) Dilakukan pengulangan 5000 kali terhadap langkah 3-5
7) Simpangan baku dari **a** dianggap sebagai galat baku bagi **b**

HASIL DAN PEMBAHASAN

Komputasi Regresi Linier Sederhana

Misalkan gugus data berpasangan (x,y) yang akan diregresikan adalah {(2,2), (4,3), (6,5), (8,7), (10,11)}. Regresi linier sederhana $y=b_0+b_1x$ menggunakan metode MLAD dilakukan dengan fungsi obyektif meminimumkan z dengan kendala seperti disajikan pada Tabel 2.

Tabel 2 Matriks kendala pada program linier regresi MLAD

	b_0	b_1	z	Tanda	Y
Tujuan	0	0	1		
Kendala	1	2	1	>=	2
	1	4	1	>=	3
	1	6	1	>=	5
	1	8	1	>=	7
	1	10	1	>=	11
	1	2	-1	<=	2
	1	4	-1	<=	3
	1	6	-1	<=	5
	1	8	-1	<=	7
	1	10	-1	<=	11
Hasil	-1.125	1.125	0.875		

Persamaan garis regresi yang dihasilkan adalah y= -1.125+1.125x dengan maksimum sisaan mutlak (z) sebesar 0.875. Nilai z ini paling kecil dibandingkan metode lain. Sebagai contoh, kalau digunakan regresi kuadrat terkecil diperoleh persamaan garis regresi y=-1+1.1x dengan nilai z sebesar 1.0, kalau digunakan regresi median diperoleh persamaan garis regresi y= -1+1x dengan nilai z sebesar 2.0. Perbandingan nilai sisaan dari regresi LS, LAD, dan MLAD untuk data tersebut disajikan pada Tabel 3.

Tabel 3 Perbandingan b_0, b_1, dan z pada LS, LAD, dan MLAD

Statistik	LS	LAD	MLAD
b_0	-1.000	-1.000	-1.125
b_1	1.100	1.000	1.125
e_1	0.800	1.000	0.875
e_2	-0.400	0.000	-0.375
e_3	-0.600	0.000	-0.625
e_4	-0.800	0.000	-0.875
e_5	1.000	2.000	0.875
z	1.000	2.000	0.875

Pendugaan Koefisien Regresi Data *Delivery Time*

Analisis regresi pada data *Delivery Time* sudah pernah dilakukan menggunakan beberapa metode, yaitu LS, Huber, Ramsay, Andrews, Hampel, dan t_5 (Setyono *et al* 1996), dan kali ini juga digunakan LAD dan MLAD. Nilai dugaan koefisien regresi berikut nilai maksimum sisaan mutlak (MSM), rata-rata sisaan mutlak (RSM), dan rata-rata kuadrat sisaan (RKS) dari beberapa metode untuk data *Delivery Time* disajikan pada Tabel 4.

Tabel 4. Nilai MSM, RSM, dan RKS dari beberapa metode

Metode	b0	b1	b2	MSM	RSM	RKS
LS	2.34	1.62	0.01	7.42	2.28	9.35
LAD	3.66	1.43	0.01	11.94	2.12	11.20
MLAD	0.53	1.86	0.01	5.98	2.85	11.11
Huber	3.37	1.53	0.01	15.37	2.54	15.28
Ramsay	3.80	1.49	0.01	16.14	2.50	16.05
Andrews	4.65	1.46	0.01	16.19	2.37	15.55
Hampel	4.62	1.47	0.01	15.92	2.36	15.17
t(v=5)	2.35	1.56	0.01	15.49	2.96	17.49

Tampak bahwa pada data *Delivery Time*, penduga MLAD yang diperoleh melalui pemrograman linier selalu menghasilkan maksimum sisaan mutlak yang paling kecil. Regresi LS menghasilkan rata-rata kuadrat sisaan paling kecil, sedangkan regresi LAD menghasilkan rata-rata sisaan mutlak paling kecil. Dengan demikian program linier yang dibuat sudah berhasil mendapatkan koefisien regresi yang meminimumkan maksimum sisaan mutlak.

Berdasarkan kedekatan dugaan koefisien regresi yang dihasilkan, sejumlah metode di atas dapat dikelompokkan ke dalam tiga kelompok. Kelompok pertama adalah metode LS, kelompok kedua adalah metode MLAD, dan kelompok ketiga adalah metode LAD, Huber, Ramsay, Andrews, Hampel, dan t. Kelompok ketiga dikenal sebagai metode kekar (*robust*) yang tidak mudah terpengaruh pencilan. Karakteristiknya adalah memberi bobot besar kepada sisaan kecil dan memberi bobot kecil kepada sisaan yang besar. Metode LAD dapat diselesaikan melalui regresi terbobot iteratif yang memberi bobot besar kepada sisaan kecil dan memberi bobot kecil kepada sisaan yang besar, sehingga layak masuk kategori tersebut.

Pendugaan Galat Baku

Pada Tabel 5 disajikan galat baku bagi koefisien regresi yang diperoleh melalui *bootstrap*. Dari tabel tersebut tampak bahwa galat baku terkecil diraih oleh metode LAD, sementara itu galat baku pada metode LS dan MLAD relatif berdekatan. Regresi LAD kurang mempedulikan sisaan besar, sehingga variasi koefisien regresi yang diperoleh dari setiap ulangan *bootstrap* tidak besar. Namun galat baku melalui *bootstrap* ini tergantung pengambilan contoh pada saat *bootstrap*, dan semakin konvergen ketika ulangan *bootstrap* semakin besar.

Tabel 5 Rata-rata dan galat baku koefisien regresi data *Delivery Time* berdasarkan *bootstrap*

Metode	b_0	b_1	b_2
MLAD	0.5068	1.8790	0.0115
	(1.3290)	(0.2294)	(0.0045)
LAD	3.8729	1.4284	0.0142
	(0.7337)	(0.1252)	(0.0025)
LS	2.3151	1.6082	0.0146
	(1.0284)	(0.1687)	(0.0035)

Validasi Silang

Kebaikan suatu metode dapat dievaluasi berdasarkan kemampuannya memprediksi, salah satunya menggunakan kriteria validasi silang. Validasi silang ini dilakukan dengan membuang sebuah pengamatan, melakukan pendugaan koefisien regresi berdasarkan pengamatan yang tersisa, kemudian menduga nilai respon pada pengamatan yang dibuang dan menghitung selisihnya (galatnya). Validasi silang pada data *Delivery Time* mula-mula dilakukan terhadap pengamatan pertama, sedangkan pengamatan lainnya sebagai *trained*. Kemudian dilakukan validasi silang terhadap pengamatan kedua, sedangkan pengamatan lainnya sebagai *trained*. Begitu seterusnya sampai pengamatan terakhir. Berdasarkan nilai galat prediksi pengamatan pertama sampai terakhir diperoleh rata-rata kuadrat galat prediksi (CV1), rata-rata galat mutlak prediksi (CV2), dan maksimum galat mutlak prediksi (CV3).

Hasil rekapitulasi nilai validasi silang penduga LS, LAD, dan MLAD untuk data *Delivery Time* disajikan pada

Tabel 6.

Tabel 6. Validasi silang regresi MLAD, LAD, dan LS

Jenis Validasi Silang	MLAD	LAD	LS
Rata-rata Kuadrat Galat Prediksi (CV1)	17.8714	13.3456	18.3616
Rata-rata Galat Mutlak Prediksi (CV2)	3.2592	2.5361	2.8780
Maksimum Galat Mutlak Prediksi (CV3)	12.5443	12.2503	14.7889

Tampak bahwa maksimum galat mutlak prediksi paling kecil diraih oleh regresi LAD diikuti oleh MLAD dan terakhir LS. Sementara itu rata-rata galat mutlak prediksi terkecil diraih oleh regresi LAD diikuti oleh LS dan MLAD, sedangkan kuadrat galat prediksi terkecil diraih oleh regresi LAD, kemudian diikuti oleh MLAD dan selanjutya LS. Dengan demikian secara umum metode regresi terbaik menurut kriteria validasi silang untuk data *Delivery Time* adalah regresi LAD.

Data *Delivery Time* dikenal sebagai data yang "tidak disukai" oleh regresi LS, sehingga banyak regresi kekar yang mencoba data tersebut sebagai alternatif bagi regresi LS. Kalaupun regresi LS diterapkan pada data ini, biasanya disertai diagnosis terhadap pencilan. Regresi LAD termasuk regresi kekar, atau paling tidak, lebih kekar dari pada LS dan MLAD. Oleh sebab itu wajar jika galat prediksi yang dihasilkan relatif lebih baik.

Pend ugaan Koefisien Regresi Berbasis Subset Pengamatan

Kenyataan bahwa regresi MLAD hanya memperhatikan kendala yang berbeda dan membatasi lebih ketat membuka peluang diperolehnya subset pengamatan yang menghasilkan koefisien regresi sama dengan yang dihasilkan oleh keseluruhan pengamatan. Ketika metode MLAD digunakan untuk menduga ukuran pemusatan (regresi model intersep) berdasarkan data contoh berukuran n, terdapat subset pengamatan berukuran dua yang menghasilkan nilai dugaan yang sama dengan nilai dugaan dari n pengamatan, yaitu ketika subset yang terpilih adalah y[1] dan y[n].
Pada kajian sebelumnya sudah diketahui bahwa koefisien regresi MLAD untuk data *Delivery Time* adalah b_0=0.528, b_1=1.862, b_2=0.012, dan maksimum sisaan mutlak=5.98. Kalau diambil contoh berukuran 4 tanpa pemulihan akan diperoleh $\binom{25}{4} = 12650$ kemungkinan contoh. Sebagian hasil sampling disajikan pada

Tabel 7.

Tabel 7 Nilai koefisien regresi hasil sampling data *Delivery Time*

No Contoh	Nomor Pengamatan				b_0	b_1	b_2
1	1	2	3	4	7.993	1.678	-0.005
2	1	2	3	5	7.758	0.881	0.005
...	...	...	...	...	...	...	...
6090	4	9	19	25	-4.319	5.695	-0.059
6091	4	9	20	21	0.000	2.074	0.008
6092	**4**	**9**	**20**	**22**	**0.528**	**1.862**	**0.012**
...	...	...	...	...	...	...	...
12649	21	23	24	25	5.419	1.100	0.008
12650	22	23	24	25	2.030	1.764	0.004

Tampak bahwa pada pengambilan contoh ke-6092, ketika pengamatan yang terpilih adalah nomor 4, 9, 20, dan 22 diperoleh koefisien regresi sebesar koefisien regresi yang diperoleh pada keseluruhan data. Kalau pada empat pasangan pengamatan tersebut ditambahkan satu pasangan pengamatan lagi, hasilnya tetap (Tabel 8). Dengan demikian empat pasangan pengamatan tersebut menjadi penentu koefisien regresi berbasis MLAD.

Tabel 8 Hasil regresi MLAD pada subset pengamatan berukuran 5 yang mengandung pengamatan nomor 4, 9, 20, dan 22

Nomor Pengamatan					b_0	b_1	b_2	z
4	9	20	22	1	0.528	1.862	0.012	5.98
4	9	20	22	2	0.528	1.862	0.012	5.98
4	9	20	22	3	0.528	1.862	0.012	5.98
4	9	20	22	4	0.528	1.862	0.012	5.98
4	9	20	22	5	0.528	1.862	0.012	5.98
4	9	20	22	6	0.528	1.862	0.012	5.98
4	9	20	22	7	0.528	1.862	0.012	5.98
4	9	20	22	8	0.528	1.862	0.012	5.98
4	9	20	22	9	0.528	1.862	0.012	5.98
4	9	20	22	10	0.528	1.862	0.012	5.98
4	9	20	22	11	0.528	1.862	0.012	5.98
4	9	20	22	12	0.528	1.862	0.012	5.98
4	9	20	22	13	0.528	1.862	0.012	5.98
4	9	20	22	14	0.528	1.862	0.012	5.98
4	9	20	22	15	0.528	1.862	0.012	5.98
4	9	20	22	16	0.528	1.862	0.012	5.98
4	9	20	22	17	0.528	1.862	0.012	5.98
4	9	20	22	18	0.528	1.862	0.012	5.98
4	9	20	22	19	0.528	1.862	0.012	5.98
4	9	20	22	20	0.528	1.862	0.012	5.98
4	9	20	22	21	0.528	1.862	0.012	5.98
4	9	20	22	22	0.528	1.862	0.012	5.98
4	9	20	22	23	0.528	1.862	0.012	5.98
4	9	20	22	24	0.528	1.862	0.012	5.98
4	9	20	22	25	0.528	1.862	0.012	5.98

Penjelasan secara induksi lengkap - bahwa pada regresi dengan p paramater dapat diperoleh subset berukuran p+1 pasangan pengamatan yang menghasilkan koefisien regresi sama dengan koefisien regresi yang dihasilkan oleh keseluruhan data - sulit dilakukan, tetapi hal itu dapat dijelaskan secara matematis melalui cara lain. Program linier terdiri atas beberapa pernyataan berupa pertidaksamaan yang berperan sebagai kendala, dan sebuah pernyataan yang berperan sebagai fungsi tujuan. Program linier akan mengeliminasi kendala-kendala yang kalah ketat sehingga yang tersisa hanyalah kendala yang tidak ada yang mengungguli keketatannya. Jadi ketika ada kendala $2x+3y>5$ dan $2x+3y>7$ maka $2x+3y>5$ dieliminasi karena dapat diwakili oleh $2x+3y>7$. Nilai fungsi tujuan akan khas ketika kendala yang tersisa sebanyak peubah dan tidak khas ketika kendala yang tersisa kurang dari banyaknya peubah.

Pada saat pendugaan regresi dengan satu parameter, pada prinsipnya melakukan optimasi dengan program linier yang melibatkan dua peubah, yaitu k dan z. Pada program linier dengan kendala dua peubah pada prinsipnya hanya dibutuhkan dua kendala agar optimasi dapat dilakukan. Kalau tersedia banyak kendala tentu dapat diperoleh dua buah kendala yang membuat kendala lain kalah ketat dari dua kendala tersebut. Oleh sebab itu ketika menduga ukuran pemusatan menggunakan metode MLAD, dapat diperoleh contoh dua pengamatan yang menghasilkan penduga sama dengan yang diperoleh dari keseluruhan data.

Pada pendugaan koefisien regresi linier sederhana menggunakan metode MLAD pada prinsipnya melakukan optimasi program linier yang melibatkan tiga peubah, yaitu b_0, b_1, dan z. Sesuai dengan prinsip program linier meskipun tersedia banyak kendala sebenarnya ada tiga kendala yang menentukan atau memberi batasan paling ketat. Oleh sebab itu pada regresi liner sederhana dapat diperoleh subset tiga pengamatan yang menghasilkan koefisien regresi sama dengan koefisien regresi yang diperoleh dari keseluruhan pengamatan.

Dari penjelasan di atas dapat dimengerti bahwa pada regresi MLAD dengan p parameter dapat diperoleh subset pengamatan berukuran p+1 yang menghasilkan koefisien regresi sama besar dengan koefisien regresi yang dihasilkan oleh keseluruhan pengamatan. Sifat ini merupakan hal yang penting karena dimungkinkan mendapatkan data contoh yang menghasilkan statistik sama dengan parameter populasi.

KESIMPULAN

Optimasi sisaan dengan cara meminimumkan maksimum sisaan mutlak dapat dikembangkan menjadi metode untuk pendugaan koefisien regresi menggunakan program linier. Regresi LAD merupakan metode terbaik untuk data *Delivery Time* menurut kriteria validasi silang. Metode LS merupakan metode paling stabil pada semua kriteria optimasi sisaan. Pada analisis data *Delivery Time* yang terdiri atas 25 pengamatan, terdapat pengamatan yang menjadi penentu koefisien regresi MLAD, yaitu pengamatan nomor 4, 9, 20, dan 22.

DAFTAR PUSTAKA

Givens GH, Hoeting JA. 2005. *Computational Statistics*. New Jersey: John Wiley & Sons.

Golberg, M.A and H.A Cho. 2010. *Introduction to Regression Analysis*. Southampton: WIT Press,.

Hao L, Naiman DQ. 2007. *Quantile Regression*. California: Sage Publications, Inc.

Koenker R, Bassett G. 1978. Regression quantiles. *Econometrica* 46 (1): 33-50.

Koenker R, Hallock KF. 2001. Quantile regression. *Journal of Economic Perspectives* 15 (4): 143–156.

McCarl BA, Spreen TH. 1997. *Applied Mathematical Programming Using Algebraic Systems*. Copyright Bruce A. McCarl and Thomas H. Spreen

Montgomery DC, Peck EA, Vining GG. 2012. *Introduction to Linear Regression Analysis*. Fifth Edition. New York: John Wiley and Sons.

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

Rudolf M, Wolter H, Zimmermann H. 1999. A linear model for tracking error minimization. *Journal of Banking & Finance* 23 (1999) 85-103

Setyono, Notodiputro K, Aunuddin, Mattjik AA. 1996. Pemodelan statistika atas dasar sebaran t student. *Forum Statistika dan Komputasi* Vol 1 No 2: 10-16

Setyono, Sumertajaya IM, Kurnia A, Mattjik AA. 2014. The performance of LS, LAD, and MLAD regression on the stack loss data. *Proc. ICCS-13*, Bogor, Indonesia December 18-21, 2014, Vol. 27, pp. 41-54

Winston WL, Goldberg JB. 2004. *Operations Research Applications and Algorithms*. Belmont: Brooks/Cole—Thomson Learning.

Zhu J, Jing P. 2010. The analysis of bootstrap method in linear regression effect. *Journal of Mathematics Research*.Vol. 2, No. 4; November 2010: 64-69.