

PEMBANDINGAN BEBERAPA METODE ANALISIS REGRESI KOMPONEN UTAMA ROBUST

NOTIRAGAYU

Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung
Email : notiragayu@unila.ac.id

ABSTRACT

Principal Component Regression (PCR) is one of the widely used statistical techniques for regression analysis with colinearity and a robust technique on PCR when data contains outlier is urgently needed. In this paper we consider combination between robust principal component analysis (PCA): Minimum Covariance Determinant (MCD), Minimum Volume Ellipsoid (MVE) and robust regression methods: Least Trimmed Square (LTS), Least Median Square (LMS).

We compare resistance level of MCD-LMS, MCD-LTS, MVE-LMS and MVE-LTS through the bias and the mean square error on some samples size and outlier's presentation.

The result shows that the MVE-LTS and MVE-LMS perform better than MCD-LTS and MCD-LMS.

Keywords: *Robust Regression, colinearity, MCD, MVE, LTS, LMS.*

PENDAHULUAN

Salah satu masalah yang sering muncul dalam analisis regresi berganda adalah masalah multikolinieritas, yaitu adanya hubungan linier (korelasi) antara dua variabel bebas atau lebih dalam suatu persamaan regresi. Hal ini akan mengakibatkan dugaan koefisien variabel tidak tunggal melainkan tidak terhingga banyaknya sehingga tidak mungkin lagi untuk menduganya (Myers, 1990). Hal ini disebabkan matriks $X^T X$ menjadi singular sehingga persamaan normalnya tidak lagi mempunyai jawaban yang tunggal.

Regresi komponen utama (RKU) merupakan salah satu metode yang digunakan untuk mengatasi masalah multikolinieritas. Metode ini mengatasi multikolinieritas dengan cara membentuk komponen-komponen utama yang saling bebas dari variabel bebasnya. Selanjutnya komponen-komponen utama yang terbentuk diregresikan dengan peubah respon. Dalam analisis komponen utama klasik, perhitungannya didasarkan pada matriks kovarian sampel (S) biasa, yaitu

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ji} - \bar{x}_j)(x_{ki} - \bar{x}_k) \quad , j = 1, 2, \dots, p \quad , k = 1, 2, \dots, p$$
 sebagai elemen matriks S . Matriks

kovarian ini akan optimal jika data berasal dari suatu distribusi normal multivariat, tetapi sangat sensitif terhadap adanya pencilan (*outlier*), terutama jika data mengandung pencilan yang ekstrim yang mengakibatkan distribusi data menjadi sangat menjulur (*heavy tailed distribution*), pada kasus seperti ini S akan kehilangan efisiensinya (Oja, 2002). Untuk mengatasi masalah pencilan diperlukan suatu metode penduga yang tegar terhadap pencilan yang disebut sebagai metode *robust*. Metode *robust* bagi S yang cukup populer adalah metode *Minimum Covariance Determinant* (MCD) dan metode *Minimum Volume ellipsoid* (MVE) yang diperkenalkan oleh Rousseeuw pada tahun 1985 (Hubert, Rousseeuw, Van Aelst, 2004). Penduga MCD bagi S diperoleh dari subhimpunan data berukuran h yang memiliki nilai determinan matriks kovarian terkecil. Sedangkan penduga MVE bagi S diperoleh dari subhimpunan data berukuran h yang memiliki volume elipsoid terkecil.

Selanjutnya hasil komponen-komponen utama robust yang terbentuk diregresikan dengan peubah respon menggunakan metode OLS (*Ordinary Least Square*). Metode OLS dikenal sebagai metode penduga terbaik dalam analisis regresi, namun metode ini sangat peka terhadap adanya penyimpangan asumsi pada data. Jika data tidak memenuhi salah satu asumsi regresi maka penduga OLS tidak lagi efisien. Salah satu asumsi penting dalam analisis regresi yang berkaitan dengan inferensia model adalah asumsi sebaran normal (normalitas). Asumsi normalitas seringkali dilanggar saat data mengandung pencilan (*outlier*). Jika terdapat pencilan dalam data, maka bentuk sebaran data tidak lagi simetrik tetapi cenderung menjulur ke arah pencilan sehingga melanggar asumsi normalitas. Kadang untuk mengatasi hal ini seorang peneliti melakukan transformasi terhadap data dengan maksud agar asumsi dapat terpenuhi, namun seringkali transformasi yang dilakukan terhadap data tidak dapat memperkecil nilai *leverage* pencilan yang pada akhirnya membiaskan pendugaan. Dalam kasus seperti ini, analisis regresi *robust* merupakan satu-satunya metode yang paling layak untuk digunakan.

Sampai saat ini berbagai metode *robust* untuk analisis regresi terus berkembang dan digunakan dalam berbagai bidang, diantaranya adalah *Least Median Square* (LMS) dan *Least Trimmed Square* (LTS). Metode LMS menduga koefisien regresi dari data yang mengandung pencilan dengan meminimumkan median dari kuadrat galatnya ($\min \text{median}(e_i)^2$), sedangkan metode LTS dengan melakukan analisis regresi kuadrat terkecil ($\min \Sigma(e_i)^2$) terhadap sebaran data yang sudah terpankas (*trimmed*) atau sebaran terwinsorkan (*winsorized distribution*).

Dalam tulisan ini kami membandingkan tingkat ketegaran (resistensi) metode MCD-LMS, MCD-LTS, MVE-LMS dan MVE-LTS sebagai metode RKU robust dengan menggunakan nilai bias dan MSE pada beberapa ukuran sampel dan prosentase pencilan.

Metode *Least Median Square* (LMS)

Metode LMS menduga koefisien regresi dengan meminimumkan median dari kuadrat galatnya, yaitu :

$$\hat{\beta} = \arg \min_{\beta} \text{median}(e_i^2) = \arg \min_{\beta} \text{median}(y_i - \hat{y}_i)^2, \quad i = 1, 2, 3, \dots, n$$

Solusi persamaan di atas dari dugaan LMS juga dapat diperoleh dengan menggunakan pemrograman linier (Chen, 2002)

Metode *Least Trimmed Square* (LTS)

Metode LTS menduga koefisien regresi dengan menggunakan metode OLS terhadap subhimpunan data berukuran h , yaitu

$$\hat{\beta} = \arg \min_{\beta} \left(\sum_{i=1}^h e_i^2 \right) = \arg \min_{\beta} \left(\sum_{i=1}^h (y_i - \hat{y}_i)^2 \right), \quad \frac{(3n + p + 1)}{4} \leq h \leq n.$$

Solusi $\hat{\beta}$ pada persamaan di atas dapat diperoleh dengan menggunakan turunan (diferensial) seperti pada penyelesaian pada metode OLS, hanya pada LTS persamaan tersebut dihitung pada subhimpunan data terbaik (disimbolkan dengan H) yang berukuran h . Penentuan subhimpunan H terbaik dilakukan dengan menggunakan algoritma resampling dari seluruh kemungkinan subhimpunan yang didapat dibentuk yaitu sebanyak $\binom{n}{h}$. Algoritma resampling secara lengkap dapat dilihat dalam Nisa (2006). Subhimpunan H yang diperoleh merupakan sebaran data yang sudah terpankas (*trimmed distribution*) (Chen, 2002).

Metode Minimum Covariance Determinant (MCD)

Misalkan $\mathbf{X} = \{x_1, \dots, x_n\}$ merupakan suatu sampel dari n pengamatan dalam \mathbf{R}^k dan ambil $h \approx n/2$, cari subhimpunan \mathbf{J}^* berukuran h sedemikian sehingga :

$$\mathbf{J}^* = \min_{J \subset \{1, \dots, n\}, |J|=h} \det \hat{S}_J$$

Dimana \hat{S}_J adalah matriks kovarians berdasarkan pada pengamatan x_i dengan $i \in J$.

Penduga MCD diberikan oleh :

$$\hat{\mu} = \bar{x}_{J^*} = \frac{1}{h} \sum_{i \in J^*} x_i$$

$$\hat{S} = \hat{S}_{J^*} = \frac{1}{h} \sum_{i \in J^*} (x_i - \hat{\mu})' (x_i - \hat{\mu})$$

(Critchley, Schyns & Haesbroeck, 2003).

Metode Minimum Volume Ellipsoid (MVE)

Pada prinsipnya metode ini adalah mencari ellipsoid dengan volume paling minimum yang melingkupi suatu subhimpunan dari minimal h pengamatan.

Penduga MVE didefinisikan sebagai pasangan $(\bar{\mathbf{X}}, \mathbf{S})$ di mana $\bar{\mathbf{X}}$ adalah vektor- p dan \mathbf{S} adalah matriks semi-definit positif $p \times p$ sedemikian sehingga determinan \mathbf{S} diminimalkan untuk memenuhi,

$$\#\{i \mid (x_i - \bar{\mathbf{X}})' \mathbf{S}^{-1} (x_i - \bar{\mathbf{X}}) \leq c^2\} \geq h$$

dengan n adalah jumlah pengamatan, p adalah jumlah variabel, $h = [(n+p+1)/2]$, c adalah konstanta dan X adalah himpunan data pengamatan

(Rousseeuw and Van Zomeren, 1990,1991).

Berdasarkan definisi di atas maka mencari penduga MVE sebenarnya sulit dilakukan dalam prakteknya. Misalkan untuk sampel berukuran $n = 30$ dan dimensi data $p = 3$, maka $h = (30+3+1)/2 = 17$, maka ada total $(30!/13! 17!) = 119.759.850$ *halfset* yang berpotensi menjadi basis bagi penduga MVE. Karena perhitungannya yang sulit, maka Rousseeuw dan Leroy (1987) dalam Jensen *et.al* (2004) menggunakan sebuah pendekatan untuk mencari penduga MVE dengan algoritma resampling (*resampling algorithm*).

Regresi Komponen Utama

Tahap pertama pada prosedur regresi komponen utama yaitu menghitung komponen utama yang merupakan kombinasi linear dari beberapa peubah X , dan tahap kedua adalah peubah tak-bebas diregresikan pada komponen utama dalam sebuah model regresi linear.

Bentuk persamaan regresi dalam bentuk peubah asli X dapat ditulis sebagai :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

dengan: Y = peubah tak-bebas
 X_h = peubah bebas ke- i yang dispesifikasikan sejak awal, $h = 1, 2, \dots, p$.
 β_0 = konstanta
 β_h = koefisien regresi dari peubah bebas ke- h , $h = 1, 2, \dots, p$.

Peubah baru (K) sebagai komponen utama adalah hasil transformasi dari peubah asal (X) yang modelnya dalam bentuk matriks adalah $K = A X$, dan komponen ke- j ditulis

$$K = a_{1j} X_1 + a_{2j} X_2 + \dots + a_{pj} X_p, \text{ atau}$$

$$K = \underline{a}_j' \underline{x}$$

dimana vektor pembobot \underline{a}_j' diperoleh dengan memaksimalkan keragaman komponen utama ke- j , yaitu

$$S_{y_j}^2 = \underline{a}_j' S \underline{a}_j$$

dengan kendala $\underline{a}_j' \underline{a}_j = 1$ serta $\underline{a}_h' \underline{a}_j = 0$, untuk $h \neq j$. Vektor pembobot \underline{a}_j' diperoleh dari matriks peragam Σ yang diduga dengan matriks S , yaitu

$$S = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

Misalkan diberikan notasi K_1, K_2, \dots, K_m sebagai banyaknya komponen utama dan Y sebagai peubah tak-bebas, maka model regresi komponen utama dapat ditulis sebagai

$$Y = w_0 + w_1 K_1 + w_2 K_2 + \dots + w_m K_m + \varepsilon$$

dengan: Y = peubah tak-bebas.
 K_j = peubah bebas komponen utama yang merupakan kombinasi linear dari semua peubah baku Z ($j = 1, 2, \dots, m$).
 w_0 = konstanta.
 w_j = parameter model regresi atau koefisien regresi, ($j = 1, 2, \dots, m$).
 ε = galat.

METODE PENELITIAN

Metode penelitian yang digunakan adalah studi pustaka terhadap buku-buku dan jurnal-jurnal yang berkaitan dengan penelitian ini dan melakukan simulasi data dengan membuat program makro simulasinya menggunakan perangkat lunak SAS/IML versi 8.

Pada simulasi data ini banyaknya sampel yang digunakan secara keseluruhan sebanyak 2000 sampel, yaitu untuk kombinasi 4 ukuran sampel ($n = 20, 60, 100, \text{ dan } 200$) dan 5 jenis prosentase pencilan (5%, 10%, 15%, 20%, 25%) dengan pengulangan 100 kali. Berikut tahapan dalam simulasi:

1. Membangkitkan satu kelompok data berukuran $n=20$ dan $p=5$ variabel dari sebaran normal multivariat, yaitu $X \sim MN(\mu, \Sigma)$. Data yang diperoleh dimisalkan $X^{(0)}$.
2. Membangkitkan vektor galat (e) dari sebaran normal $(0,1)$. Selanjutnya vektor $y^{(0)}$ dibangkitkan dengan berdasarkan kombinasi linier x_i ditambah galat yaitu : $y^{(0)} = a^t X^{(0)} + e$. Vektor a ditetapkan sebagai vektor satuan.

- Melakukan analisis regresi komponen utama terhadap data $X^{(0)}$ dan $y^{(0)}$. Vektor koefisien regresi yang diperoleh disimbolkan dengan $\hat{\beta}^{(0)}$
- Membangkitkan matriks pencilan ($X_{out}^{(s)}$) berukuran $n \times p$. Matriks ($X_{out}^{(s)}$) ini elemennya adalah nol kecuali beberapa elemen yang dijadikan pencilan. Elemen pencilan dibangkitkan dari sebaran normal $(10, (0,05)^2)$. Banyaknya pencilan yang diberikan adalah 5% dari data.
- Menambahkan matriks $X_{out}^{(s)}$ pada data $X^{(0)}$ sehingga diperoleh matriks $X^{(s)}$ yang sudah terkontaminasi :

$$X^{(s)} = X^{(0)} + X_{out}^{(s)}$$

- Definisikan $y^{(s)}$ sebagai kombinasi linier berikut : $y^{(s)} = a^t X^{(s)} + e$.
- Menghitung matriks kovarian *robust* dengan metode MCD dan MVE dari $X^{(s)}$.
- Melakukan analisis komponen utama berdasarkan matriks kovarian MCD dan MVE, kemudian menghitung skor komponen utama berdasarkan hasil yang diperoleh dari metode MCD dan MVE.
- Meregresikan skor komponen yang diperoleh metode MCD dan MVE terhadap peubah respon $y^{(s)}$ dengan menggunakan metode LMS dan LTS, vektor koefisien regresi yang diperoleh disimbolkan dengan $\hat{\beta}^{(s)}$
- Ulangi langkah 4 sampai 9 sebanyak 100 kali.
- Menghitung nilai bias dan MSE dari $\hat{\beta}$ yang dihasilkan oleh masing-masing metode dengan rumus sebagai berikut :

$$\text{Bias}(\beta_k) = \left(\frac{1}{100} \sum_{s=1}^{100} \hat{\beta}_k^{(s)} \right) - \hat{\beta}_k^{(0)} \quad , k = 1, 2, \dots, 5$$

$$\text{MSE}(\beta_k) = \frac{1}{100} \sum_{s=1}^{100} (\hat{\beta}_k^{(0)} - \hat{\beta}_k^{(s)})^2 \quad , k = 1, 2, \dots, 5$$

- Menghitung rata-rata bias dan MSE dengan rumus sebagai berikut :

$$\text{Bias}(\beta) = \frac{1}{5} \sum_{k=1}^5 |\text{bias}(\beta_k)| \quad \text{dan} \quad \text{MSE}(\beta) = \frac{1}{5} \sum_{k=1}^5 \text{MSE}(\beta_k)$$

- Ulangi langkah 4 sampai 12 untuk jumlah pencilan sebanyak 10%, 15%, 20%, dan 25%.
- Ulangi langkah 1 sampai 13 untuk $n = 60, 100$ dan 200 .

☛ Membandingkan bias dan MSE dari seluruh metode.

☛ Menyimpulkan konsistensi seluruh metode

HASIL DAN PEMBAHASAN

Hasil simulasi data dengan menggunakan 100 sampel memperlihatkan bahwa nilai dugaan keempat metode yaitu LTS-MCD, LTS-MVE, LMS-MCD dan LMS-MVE cukup baik dan tidak terlalu

terpengaruh oleh pencilan, hal ini dapat dilihat dari nilai bias dan MSE yang dihasilkan keempat metode relative kecil dan tidak terpengaruh oleh pencilan. Nilai bias dan MSE dihitung dari nilai dugaan koefisien regresi yang dihasilkan pada 100 sampel. Rata-rata nilai bias dan MSE untuk seluruh koefisien regresi disajikan dalam Tabel 1 dan Tabel 2. Selanjutnya untuk mempermudah analisis data bias dan MSE disajikan dalam bentuk grafik pada Gambar 1 sampai dengan Gambar 8.

Pada tabel terlihat nilai bias dan MSE untuk semua metode tidak konsisten, yaitu tidak menunjukkan trend meningkat seiring dengan meningkatnya prosentase pencilan yang diberikan pada data. Penulis beranggapan bahwa ini disebabkan oleh jumlah replikasi yang kurang besar. Jika jumlah replikasi ditambah, maka nilai yang dihasilkan akan konvergen menuju suatu angka yang konsisten. Namun demikian, pada tabel 1 dan tabel 2 secara keseluruhan dapat disimpulkan bahwa perbedaan yang terjadi antara keempat metode lebih disebabkan oleh metode komponen utama MCD dan MVE, sedangkan metode regresi LTS dan LMS hanya menyumbang sedikit perbedaan. Dan secara keseluruhan pula dapat disimpulkan bahwa metode LMS-MVE memberikan performa yang lebih baik dari yang lainnya.

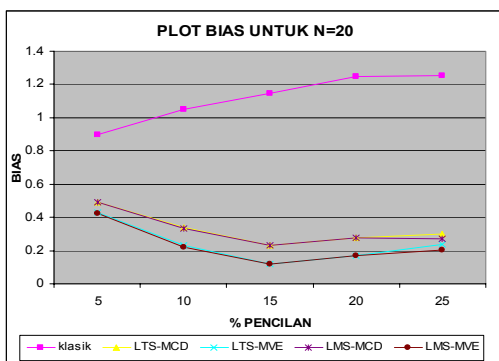
Tabel 1. Nilai bias pada setiap prosentase pencilan (p) dan ukuran sampel (n)

N	% PENCILAN	BIAS				
		KLASIK	LTSMCD	LTSMVE	LMSMCD	LMSMVE
20	5	0.8962123	0.4896378	0.4288657	0.4918324	0.4205673
	10	1.050048	0.3408935	0.2290004	0.3307189	0.2189367
	15	1.1440371	0.233542	0.1191178	0.2307746	0.1157403
	20	1.2470711	0.2784146	0.1682937	0.2773144	0.1686776
	25	1.2527912	0.2983096	0.2347079	0.2719516	0.2028341
60	5	0.9818016	0.0768792	0.0760318	0.1056201	0.1328001
	10	0.8653206	0.1753474	0.1856846	0.1526462	0.1746675
	15	0.9529399	0.2251624	0.2007516	0.2005548	0.1768927
	20	0.9508991	0.2709492	0.2467314	0.2753392	0.2528783
	25	1.1550009	0.1288936	0.1103816	0.0965153	0.0671091
100	5	1.0830519	0.2141613	0.0725615	0.2672859	0.1314448
	10	1.2454994	0.2268605	0.1713791	0.1907021	0.1187346
	15	1.2251826	0.1943125	0.1646563	0.2341912	0.1320659
	20	1.1445903	0.2437153	0.1403637	0.2327924	0.1101854
	25	1.1817469	0.1197517	0.0738313	0.0839394	0.0520445
200	5	1.1826909	0.278073	0.1383636	0.3080679	0.1798922
	10	1.0938423	0.1262051	0.0861206	0.1135608	0.0743638
	15	1.0379701	0.1289789	0.149424	0.123144	0.1402877
	20	1.1149175	0.0336784	0.0625588	0.0336802	0.0624259
	25	1.1256372	0.0889862	0.1012589	0.0872424	0.0993631

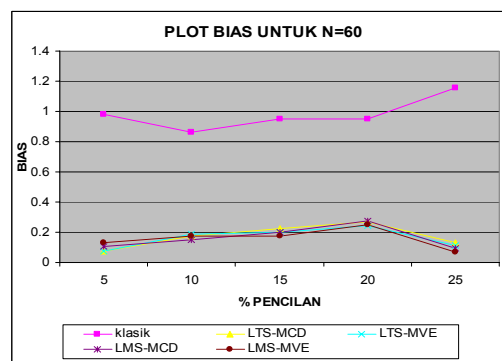
Tabel 2. Nilai MSE pada setiap prosentase pencilan (p) dan ukuran sampel (n)

N	% PENCILAN	MSE				
		KLASIK	LTSMCD	LTSMVE	LMSMCD	LMSMVE
20	5	1.280195	0.6413125	0.582443	0.6588761	0.5960618
	10	1.432132	0.223244	0.1107012	0.2174835	0.1078777
	15	1.7659621	0.2087474	0.1274898	0.2109524	0.1369414
	20	1.9342399	0.2285233	0.0776635	0.2340347	0.0835575
	25	2.0366276	0.1644556	0.1076799	0.1400718	0.0901589
60	5	1.3699782	0.0280914	0.0188757	0.0358768	0.0283957
	10	1.2975954	0.0640583	0.0667754	0.0728068	0.075934
	15	1.1419907	0.0782736	0.0688611	0.0713057	0.0600509

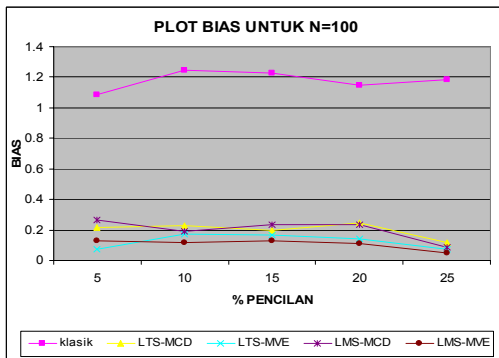
N	% PENCILAN	MSE				
		KLASIK	LTSMCD	LTSMVE	LMSMCD	LMSMVE
	20	1.1829789	0.0567057	0.0420832	0.0575384	0.0427541
	25	1.2976603	0.0321107	0.0312206	0.0417693	0.0305549
	5	1.4946043	0.11457	0.0417702	0.1726054	0.1066541
100	10	1.9960957	0.0785748	0.0676011	0.0647055	0.0416382
	15	2.0130277	0.0991786	0.0917606	0.1136128	0.1144806
	20	1.6416902	0.0753283	0.0341653	0.0694896	0.0254188
	25	1.7153301	0.0344776	0.0205594	0.0280945	0.0157713
	5	1.6887151	0.1747177	0.0761387	0.1894758	0.0876513
200	10	1.5215067	0.0799186	0.0550372	0.0738152	0.0521656
	15	1.3125195	0.0521765	0.0449973	0.0502045	0.043085
	20	1.5270669	0.0252164	0.0273619	0.0252639	0.0266743
	25	1.552197	0.0054432	0.0127547	0.0052384	0.0126671
	5					



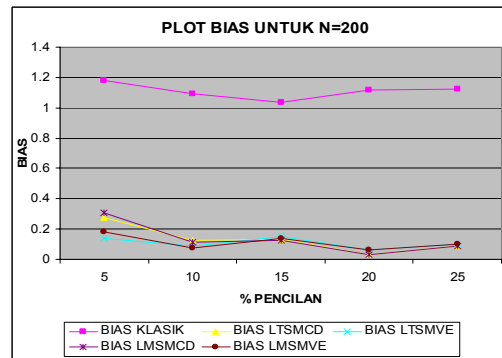
Gambar 1. Plot bias pada n=20



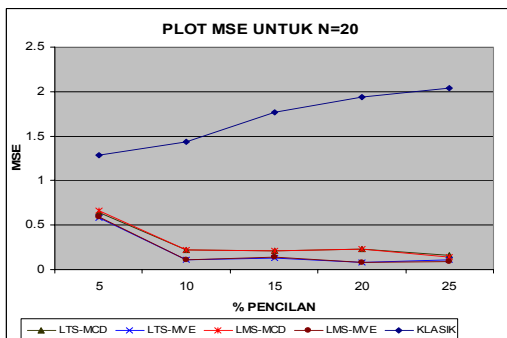
Gambar 2. Plot bias pada n=60



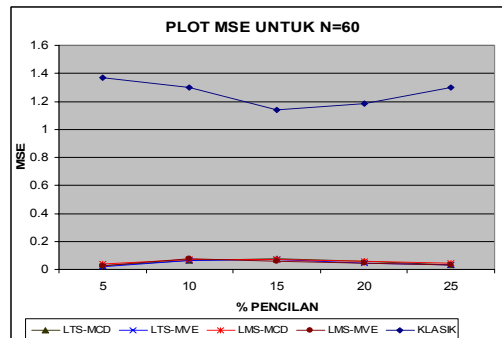
Gambar 3. Plot bias pada n=100



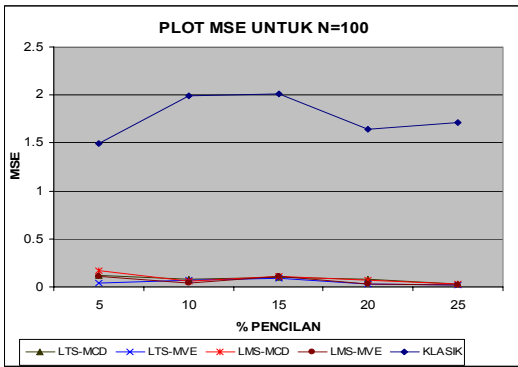
Gambar 4. Plot bias pada n=200



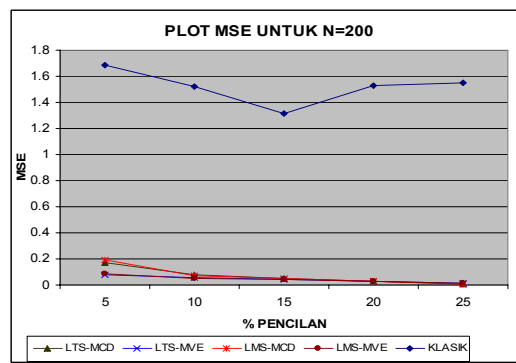
Gambar 5. Plot MSE pada n=20



Gambar 6. Plot MSE pada n=60

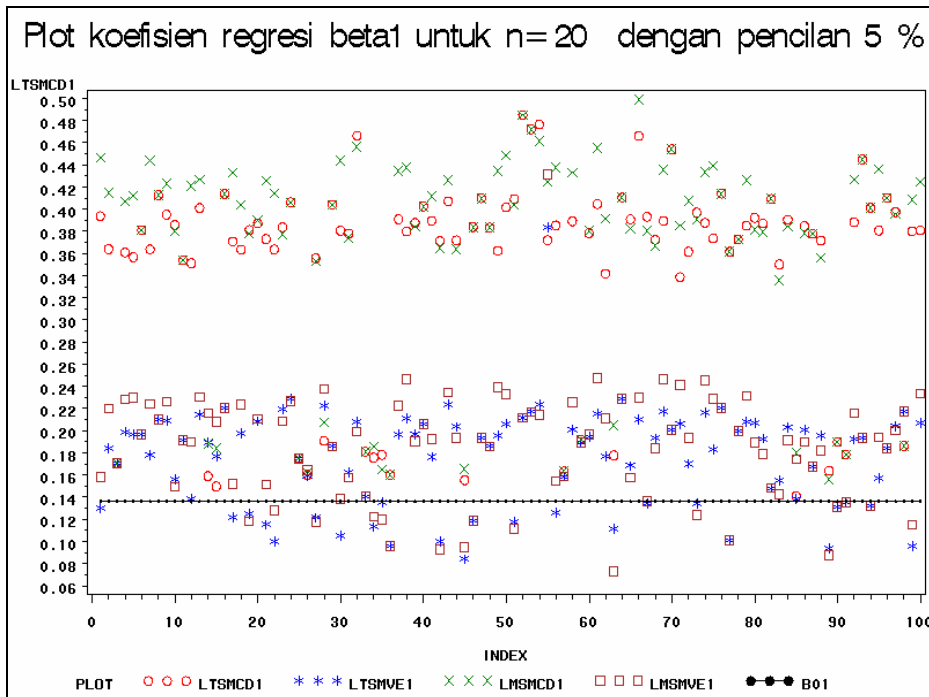


Gambar 7. Plot MSE pada n=100



Gambar 8. Plot MSE pada n=200

Pada Gambar 1-4 dapat dilihat grafik nilai bias dari metode klasik, LTS-MCD, LTS-MVE, LMS-MCD, LMS-MVE. Pada keempat gambar tersebut terlihat bahwa bias yang dihasilkan oleh keempat metode robust relative kecil dibandingkan dengan metode klasik pada setiap prosentase pencilan dan ukuran sampel. Demikian pula nilai MSE pada Gambar 5 - 8. Metode robust LTS-MVE dan LMS-MVE memberikan hasil yang hampir sama baiknya, dan keduanya lebih baik dari metode LTS-MCD maupun LMS-MCD. Untuk lebih jelasnya berikut ini diperlihatkan plot dugaan koefisien regresi dari keempat metode robust untuk ukuran sampel $n=20$ pada setiap prosentase pencilan.



Gambar 9. Plot koefisien regresi pada n=20 dan dengan 5% pencilan

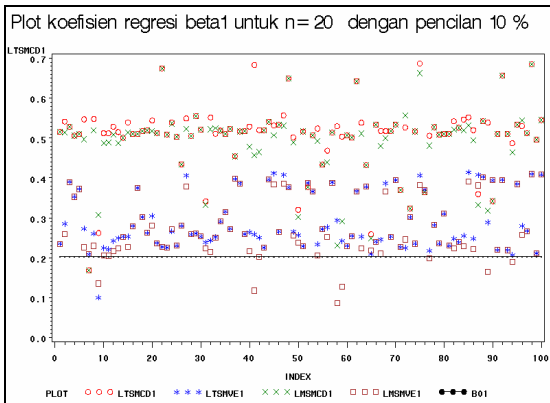
Pada gambar 9 di atas, dugaan koefisien regresi dari data awal (yaitu data tanpa diberi kontaminasi pencilan) dibuat menjadi garis lurus. Pada gambar di atas terlihat bahwa dugaan koefisien regresi komponen utama yang dihasilkan metode LTS-MVE dan LMS-MVE (selanjutnya disebut sebagai RKU MVE) lebih mendekati nilai yang sebenarnya dibandingkan dengan metode LTS-MCD dan LMS-MCD (selanjutnya disebut sebagai RKU MCD). Sedangkan keragaman dari dugaan kedua kelompok menunjukkan sebaran yang hampir sama.

Gambar 10 di bawah merupakan plot koefisien regresi komponen utama untuk $n=20$ dengan pencilan 10%. Pada gambar tersebut terlihat plot dugaannya memiliki pola yang hampir sama dengan plot untuk pencilan 5% di atas. Sedangkan untuk plot dugaan pada pencilan 15% (Gambar 11) memperlihatkan bahwa sebaran titik-titik dugaan metode RKU-MCD lebih mendekati garis

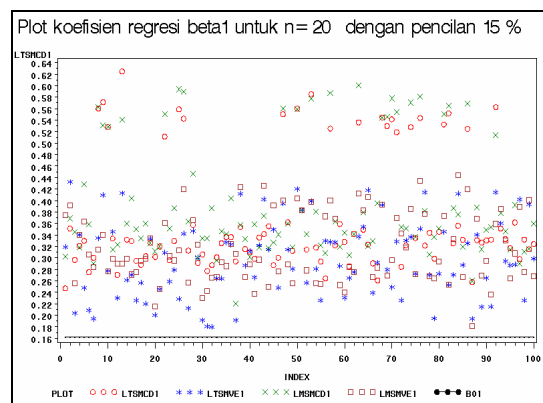
lurus, namun metode ini memiliki keragaman yang lebih besar dibanding RKM-MVE sehingga metode RKM-MVE lebih efisien dibandingkan dengan metode RKM-MCD.

Untuk pencilan 20% memiliki pola yang hampir sama dengan pencilan 10%, sedangkan pencilan 25% memiliki pola yang hampir sama dengan pencilan 15%.

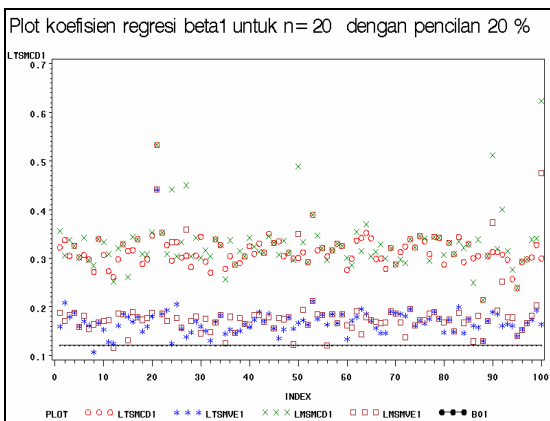
Grafik plot dugaan koefisien regresi komponen utama untuk ukuran sampel lainnya cenderung memiliki pola yang hampir sama dengan ukuran sampel 20 dan tidak diperlihatkan disini.



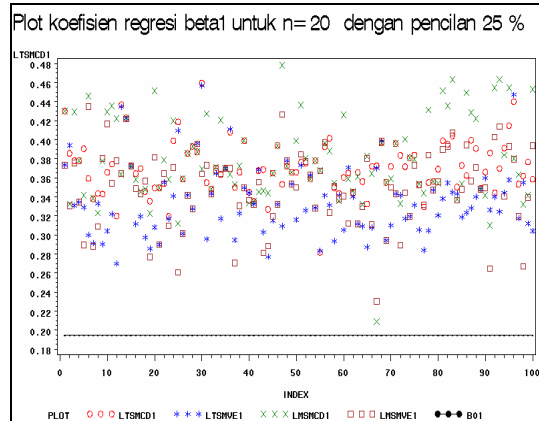
Gambar 10. Plot koefisien regresi pada n=20 dan dengan 10% pencilan



Gambar 11. Plot koefisien regresi pada n=20 dan dengan 15% pencilan



Gambar 12. Plot koefisien regresi pada n=20 dan dengan 20% pencilan



Gambar 11. Plot koefisien regresi pada n=20 dan dengan 25% pencilan

SIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan bahwa secara keseluruhan untuk pencilan 5% sampai 25% menunjukkan bahwa metode LTS-MVE dan LMS-MVE memberikan hasil yang lebih baik dibandingkan dengan metode LTS-MCD dan LMS-MCD.

Untuk penelitian selanjutnya disarankan agar menambah jumlah replikasi dalam simulasi sehingga akan didapat hasil yang lebih akurat.

DAFTAR PUSTAKA

Chen, Colin (2002). Robust Regression and Outlier Detection with the ROBUSTREG procedure. SUGI Paper 265-27. SAS Institute: Cary, NC. SAS OnLineDoc. SAS Institute, Cary, NC: IML Robust Regression, <http://v8doc.sas.com/sashtml>

- Critchley F., Schyns M. & Haesbroeck G. (2003). Smooth Optimization for the MCD estimator. *International Conference On Robust Statistics*, Wed 16 July, Belgium.
- Croux C. & Ruiz-Gazen A.(2005). High Breakdown Estimators for Principal Components: the Projection-Pursuit Approach Revisited. *Computational Statistics*. Physica-Verlag, Heidelberg,
- Croux. C , P. Filzmoser, G. Pison & P.J. Rousseeuw. (2004) Fitting Multiplicative Models by Robust Alternating Regressions. Technical Report No. 350
- Filzmoser P. (2005). Identification of Multivariate Outliers: A Performance Study. *Austrian Journal Of Statistics*. Volume 34, No.2, hal 127-138
- Filzmoser P. (2004). A Multivariate Outlier Detection Method. Dalam : Aivazian S., Filzmoser P. & Kharin Y. (editors). *Proceedings Of The Seventh International On Computer Data Analysis and Modelling* : 18-22.
- Hubert, M., Rousseeuw, P.J., Van Aelst, S. (2004), Robustness, *Encyclopedia of Actuarial Sciences*, edited by Sundt, B. and Teugels, J., Wiley, New York, pp. 1515-1529.
- Myers, R.H. 1990. *Classical and Modern Regression with Applications*. PWS-KENT Publishing Company, Boston
- Nisa, K. (2006). Analisis Regresi Robust Menggunakan Metode Least Trimmed Square untuk Data Mengandung Pencilan. *Jurnal Ilmiah MIPA*. Vol IX, No. 2.
- Notiragayu & Nisa, K. (2008). Analisis Regresi Komponen Utama Robust Untuk Data Mengandung Pencilan. *Jurnal Sains MIPA*. Vol. 14, NO. 1.
- Oja H. (2002). Robust And Nonparametric Multivariate Methods. Department of Mathematic and Statistics University of Jyväskylä. Finland
- Olive D. J. (2004). A resistant estimator of multivariate location and dispersion. *Computational Statistical Data Analysis*. Vol 46 : 99-102.
- Olive, David J. .(2005). Two Simple Resistant Regression Estimators. Southern Illinois University.
- Rousseeuw, R. J. & A. M. Leroy. (1987). Robust Regression and Outlier Detection. New York: Wiley.
- Rousseeuw, Peter J.,) (1984). "Least Median of Squares Regression," *Journal of the American Statistical Association*, vol 79 (388), 871-880.
- Rousseeuw P.J. & Van Driessen K. (1999). A fast algorithm for the minimum covariance determinant estimator, *Technometrics*, 41:212-223.
- Rousseeuw P.J. & van Zomeren B.C. (1990). Unmasking multivariate outliers and leverage points, *Journal of the American Statistical Association*, 85:633-639.
- Rousseeuw, P.J. and van Zomeren, B.C. (1991). Robust Distance: Simulations and Cutoff Values. *The IMA Volumes in Mathematics and its Application*. 34: 195- 203.