

ANALISIS KETEGARAN REGRESI ROBUST TERHADAP LETAK PENCILAN: STUDI PERBANDINGAN

NETTI HERAWATI, KHOIRIN NISA DAN ERI SETIAWAN

Abstract. *The effect of outlier numbers and location on regression coefficient is studied by comparing the estimated regression coefficient and Mean Square Error (MSE) resulted by M-estimator, MM-estimator, and by Ordinary least square (OLS). In order to show the robustness of robust regression methods we used four sample sizes: $n = 20, 60, 100$ and 200 which contaminated by 5%, 10%, 15%, 20%, 25% and 30% outliers located in the center and end of the data. The result shows that MM-Estimator is the most robust method compare to M-estimator and OLS.*

1. PENDAHULUAN

Untuk menyelidiki atau membangun model hubungan antara dua variabel kuantitatif atau lebih analisis regresi adalah salah satu metode statistika yang sering digunakan. Jika X merupakan variabel independen and Y variabel dependen, fungsi relasi dari keduanya dapat ditulis:

$$Y = f(X)$$

Secara umum model regresi linier dapat ditulis:

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I)$$

Received 02-12-2010, Accepted 29-12-2010.

2000 Mathematics Subject Classification: 46N30, 62J05

Key words and Phrases: Ordinary least square, M-Estimator, MM-Estimator

dengan:

$Y_{(n \times 1)}$ adalah vektor variabel tak bebas

$X_{(n \times k)}$ adalah matriks variabel bebas

$\beta_{(k \times 1)}$ adalah vektor parameter yang harus diduga

$\varepsilon_{(n \times 1)}$ adalah vektor galat

Pedugaan parameter regresi pada model di atas umum dilakukan dengan menggunakan metode kuadrat terkecil (MKT) yang merupakan metode penduga terbaik karena bersifat tak bias dan memiliki ragam minimum (efisien) di antara semua penduga tak bias linier. Metode MKT dapat digunakan bila asumsi yang mendasarinya terpenuhi. Jika salah satu asumsi tidak terpenuhi maka penduga MKT dapat menjadi bias dan tidak lagi efisien [9].

Salah satu asumsi yang harus dipenuhi adalah asumsi normalitas. Asumsi ini sangat berkaitan dengan inferensia model. Umumnya asumsi normalitas tidak akan terpenuhi bila data mengandung pencilan (outlier). Pencilan akan membuat bentuk sebaran data menjadi tidak simetrik dan lebih menjulur ke arah pencilan. Untuk mengatasi masalah tersebut data dapat di transformasi. Tetapi pada umumnya peneliti kurang menyukai cara tersebut. Salah satu metode alternatif yang bisa digunakan adalah metode regresi robust. Metode ini dapat langsung diterapkan pada data yang tidak memenuhi asumsi normalitas atau data yang mengandung pencilan tanpa harus mentransformasi data terlebih dahulu [6].

Dalam tulisan ini kami akan memperlihatkan ketegaran metode penduga-M dan penduga-MM terhadap data pencilan dan membandingkannya dengan metode MKT dengan melihat kuadrat tengah galat (KTG) dari dugaan koefisien regresi yang dihasilkan masing-masing metode.

2. PENDUGA-M DAN PENDUGA-MM

Penduga-M (M-estimator) diperkenalkan oleh Huber pada tahun 1964. Penduga-M menggunakan pendekatan yang sederhana antara komputasi dan teoritis. Prinsip dasar penduga-M adalah meminimumkan fungsi objektif:

$$\sum_{i=1}^n \rho(e_i^*) = \sum_{i=1}^n \rho(e_i/\hat{\sigma}) = \sum_{i=1}^n \rho((y_i - x_i b)/\hat{\sigma}) \quad (1)$$

dengan:

e_i^* : skala / derajat residual ke-i;

$\rho(e_i)$: fungsi simetris dari residual atau fungsi yang memberikan kontribusi pada masing-masing residual pada fungsi objektif;

$\hat{\sigma} = 1, 5$ med $|e_i|$, $i = 1, 2, \dots, n$.

Fox [3] menyatakan bahwa sebuah ρ yang baik harus mempunyai sifat-sifat:

1. $\rho(e) \geq 0$
2. $\rho(0) = 0$
3. $\rho(e) = \rho(-e)$
4. $\rho(e_i) \geq \rho(e_{i'})$ untuk $|e_i| > |e_{i'}|$

Contohnya, untuk penduga kuadrat terkecil $\rho(e_i) = e_i^2$. Jika $\psi = \rho$ adalah turunan dari ρ . Maka untuk meminimumkan persamaan (1)

$$\sum_{i=1}^n \psi((y_i - x_i b)/\hat{\sigma}) x_i = 0 \quad (2)$$

$\psi(\cdot)$ merupakan fungsi influence yang digunakan untuk memperoleh bobot (weight). Dengan fungsi pembobot $w_i = \frac{\psi(e_i^*)}{(e_i^*)^2}$, maka persamaan (2) menjadi:

$$\sum_{i=1}^n w_i ((y_i - x_i b)/\hat{\sigma}) x_i = 0 \quad (3)$$

Persamaan (3) jika dibuat kedalam bentuk matriks menjadi:

$$X^T W X b = X^T W y \quad (4)$$

Persamaan (4) meminimumkan $\sum_{i=1}^n w_i (y_i - \hat{y}_i)^2$, persamaan ini disebut juga kuadrat terkecil terboboti (weighted least square). Weighted least

square tersebut dapat digunakan sebagai alat untuk menghitung penduga-M. Sehingga parameter penduga menjadi:

$$b = (X^T W X)^{-1} X^T W y$$

Pembobotan tersebut bergantung residual dan koefisien dugaan. Untuk mendapatkan estimasi parameter parameter diperlukan solusi iterasi yang disebut IRLS (Iteratively reweighted least squares). Metode IRLS terdiri dari tahap-tahap berikut:

1. Taksir vektor awal menduga b_0 , dari b_0 didapatkan residual $e_{i,0}$.
2. Berdasarkan residual awal, hitung $\hat{\sigma}_0$ dan bobot awal $(w_{i,0})$, $w_{i,0} = \psi(e_{i,0}^*) / (e_{i,0}^*)$.
3. Dengan menggunakan Weighted Least Square didapatkan penduga parameter robust yang baru, $b_{R,0} = (X^T W_0 X)^{-1} X^T W_0 y$ dimana W_0 adalah matriks diagonal dari bobot dengan elemen diagonal ke- i adalah $w_{i,0}$.
4. Dugaan parameter $b_{R,0}$ dari tahap 3 digunakan menjadi b_0 dalam tahap 1, lalu pilih juga residual, $\hat{\sigma}$, dan bobot yang baru. Ulangi kembali tahap 3.

Tahapan diatas dilanjutkan sampai diperoleh penaksir parameter yang konvergen.

Penduga MM diperkenalkan oleh Yohai [10], yaitu sebuah metode yang secara simultan mempunyai dua sifat, yaitu penduga yang bersifat breakdown point tinggi dan efisiensi tinggi, atau dengan kata lain penduga-MM bertujuan menghasilkan sebuah penduga yang breakdown point tinggi serta mempertahankan efisiensinya, dimana breakdown point dan efisiensi merupakan sifat terpenting dalam penduga robust. Untuk memperoleh sifat penduga yang memiliki breakdown point tinggi dan efisiensi tinggi, metode penduga-MM menggabungkan 2 metode resgresi robust dalam tahap perhitungannya, yaitu metode robust yang memiliki breakdown point tinggi sebagai penduga awal, dan metode robust yang memiliki efisiensi tinggi pada proses perhitungan iteratifnya.

Prosedur penduga-MM dapat dibagi dalam tiga tahap sebagai berikut:

1. Menghitung nilai penduga awal dengan metode Least Trimmed Square (LTS) yang memiliki nilai breakdown-point tinggi. Perhitungan penduga LTS sama dengan metode MKT, yaitu meminimumkan jumlah

kuadrat galat, namun metode LTS menggunakan subhimpunan data terbaik berukuran h [4].

2. Menghitung parameter skala (scale parameter) $\hat{\sigma}$ menggunakan penduga M berdasarkan galat dari penduga awal dengan rumus:

$$\hat{\sigma}_{(m+1)} = \text{median} \left\{ \frac{|e|}{0,6745} \right\} = \text{median} \left\{ \frac{|y_i - x\hat{\beta}_{LTS}^m|}{0,6745} \right\};$$

$i = 1, 2, \dots, n; m = \text{iterasi } 0, 1, \dots, m$

3. Menghitung penduga akhir menggunakan algoritma penduga M. Prinsip dasar penduga-M adalah meminimumkan fungsi objektif:

$$\sum_{i=1}^n \rho \left(\frac{e_i}{\hat{\sigma}} \right)$$

dengan $e_{(i)}^2 = (y_i - \hat{y}_i)^2 = (y_i - x_i\beta)^2$

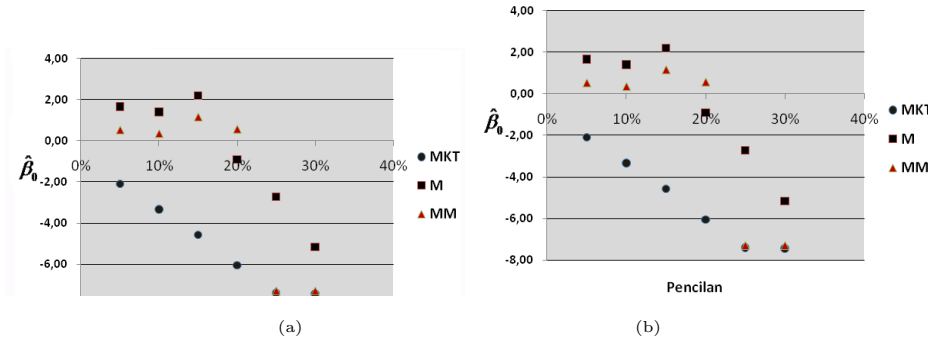
4. Jika ψ turunan dari ρ , maka fungsi objektif diatas akan menjadi:

$$\sum_{i=1}^n \psi \left(\frac{e_i}{\hat{\sigma}} \right) x_i = 0$$

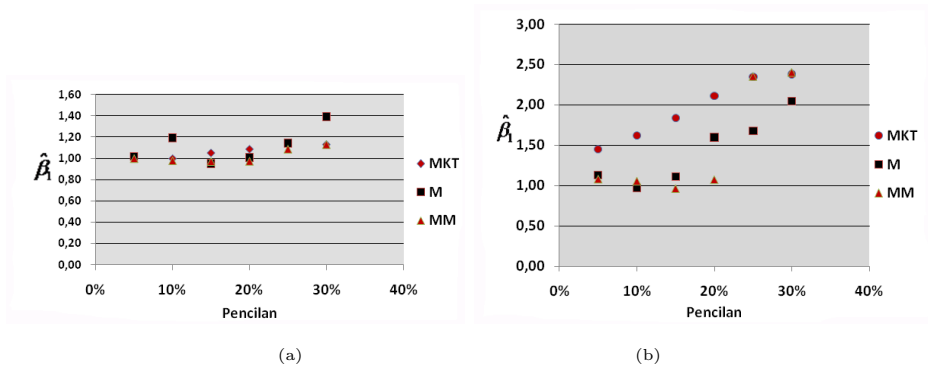
3. METODE PENELITIAN

Dalam penelitian ini kami meneliti dua kasus letak pencilan, yaitu pencilan terletak di tengah garis regresi dan pencilan terletak di ujung atas garis regresi. Simulasi data menggunakan bantuan perangkat lunak SAS versi 9. Prosedur atau tahapan-tahapan simulasi sebagai berikut:

1. Membangkitkan vektor galat ($e^{(0)}$) dari sebaran normal baku ($N(0, 1)$) berukuran $n \times 1$ dengan $n=20, 60, 100$ dan 200 .
2. Menetapkan elemen vektor x dari bilangan bulat mulai dari 1 sampai 20 (untuk $n = 60$ maka elemen vektor x ditetapkan sebagai bilangan bulat dari 1 sampai 60, untuk $n = 100$ vektor x ditetapkan sebagai bilangan buat 1 sampai 100, dan untuk $n = 200$ vektor x ditetapkan sebagai bilangan bulat dari 1 sampai 200).

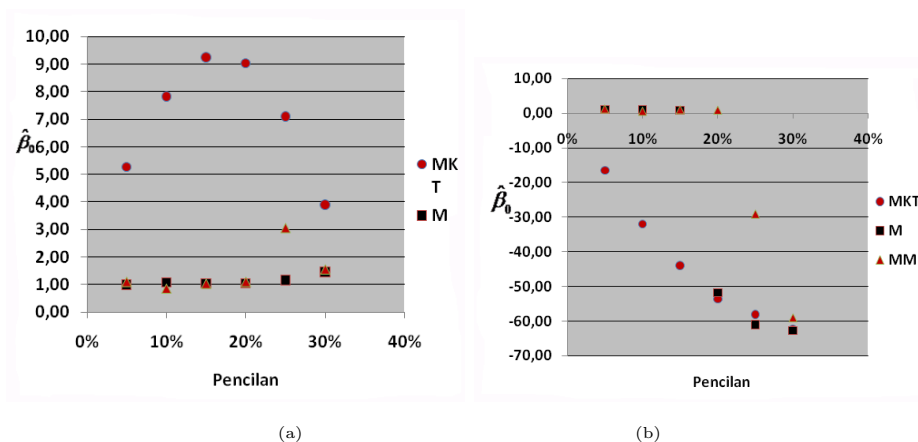


Gambar 1: Plot sebaran dugaan koefisien $\hat{\beta}_0$ letak pencilan (a) di tengah dan (b) di ujung pada $n = 20$

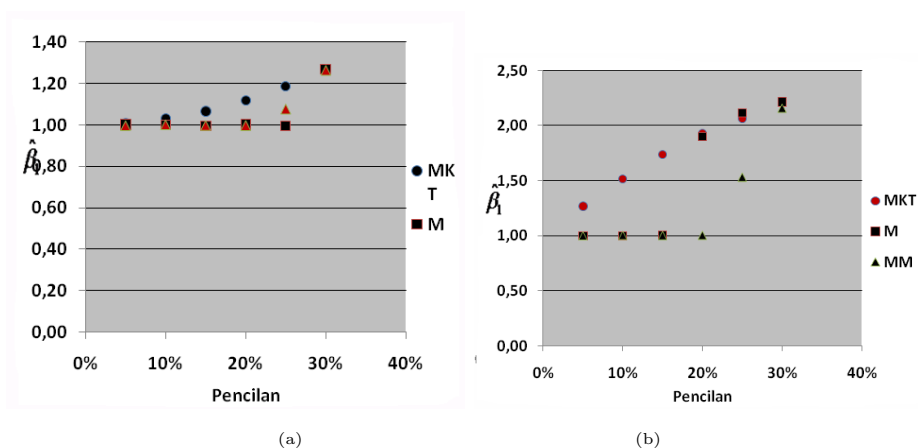


Gambar 2: Plot sebaran dugaan koefisien $\hat{\beta}_1$ letak pencilan (a) di tengah dan (b) di ujung pada $n = 20$

3. Membangkitkan vektor pencilan ($out^{(s)}$) berukuran $n \times 1$ yang elemennya adalah nol kecuali beberapa elemen yang dijadikan pencilan. Pencilan yang diberikan sebanyak 5%, 10%, 15%, 20%, 25% dan 30% dari ukuran data (n).
4. Menambahkan vektor $out^{(s)}$ pada vektor $e^{(0)}$ sehingga diperoleh vektor e yang sudah terkontaminasi pencilan, yaitu: $e^{(s)} = e^{(0)} + out^{(s)}$
5. Dengan menetapkan $\beta_0 = 1$ dan $\beta_1 = 1$, bangkitkan vektor y yaitu $y = 1 + x + e^{(s)}$. Gabungkan x dan y sebagai himpunan data sampel.
6. Menduga koefisien regresi berdasarkan model $y = \beta x + \varepsilon$ terhadap x dan y menggunakan metode MKT, penduga-M, dan penduga-MM. Simpan nilai $\hat{\beta}$ yang diperoleh dari masing-masing metode.



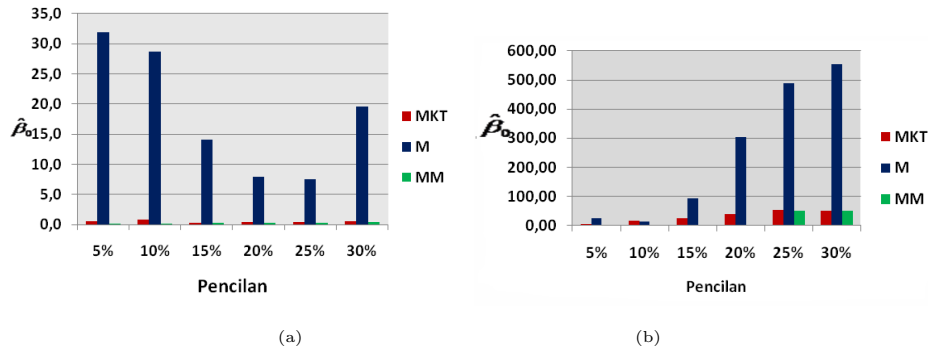
Gambar 3: Plot sebaran dugaan koefisien $\hat{\beta}_0$ letak pencilan (a) di tengah dan (b) di ujung pada $n = 200$



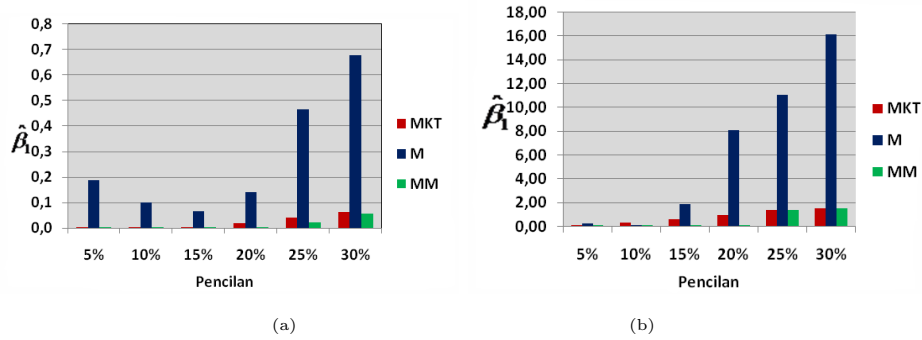
Gambar 4: Plot sebaran dugaan koefisien $\hat{\beta}_1$ letak pencilan (a) di tengah dan (b) di ujung pada $n = 200$

7. Ulangi langkah 1 sampai 6 sebanyak 10 kali.
8. Menghitung nilai KTG dari β masing-masing metode dengan rumus sebagai berikut:

$$KTG(\beta_0) = \frac{1}{m} \sum_{i=1}^m (1 - \hat{\beta}_{0i})^2, \quad m = 10$$



Gambar 5: Nilai KTG $\hat{\beta}_0$ letak pencilan (a) di tengah dan (b) di ujung pada $n = 20$

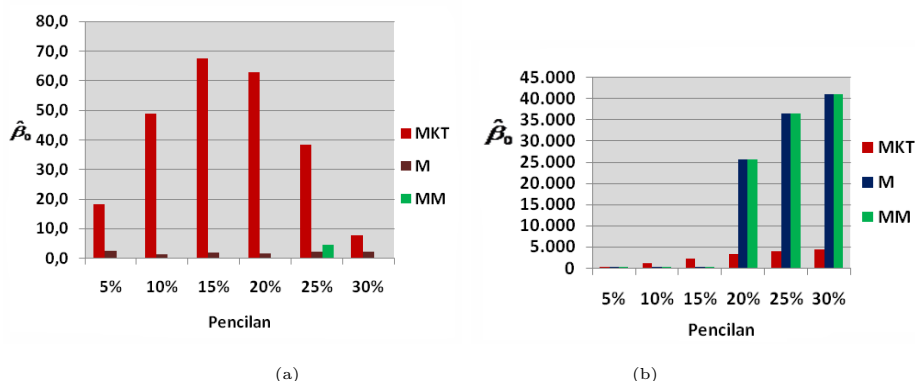


Gambar 6: Nilai KTG $\hat{\beta}_1$ letak pencilan (a) di tengah dan (b) di ujung pada $n = 20$

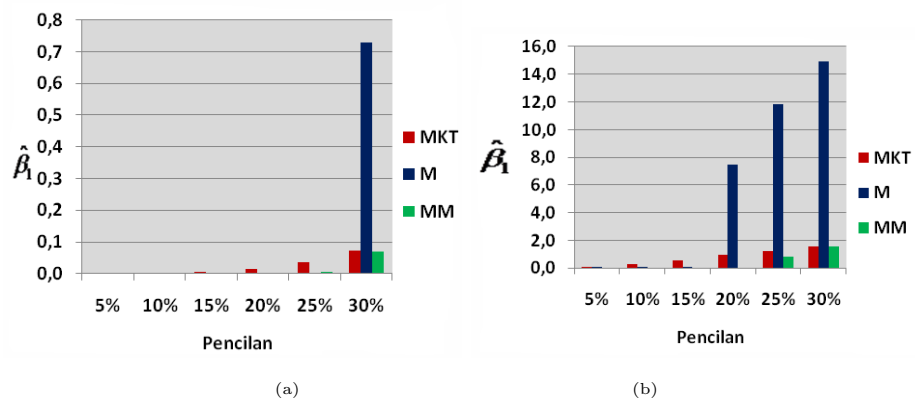
$$KTG(\beta_1) = \frac{1}{m} \sum_{i=1}^m (1 - \hat{\beta}_{1i})^2, \quad m = 10$$

4. HASIL DAN PEMBAHASAN

Nilai dugaan koefisien regresi metode MKT, Penduga-M, dan Penduga-MM untuk ukuran sampel 20, 60, 100, dan 200 untuk prosentase pencilan 5, 10, 15, 20, 25, dan 30 pada letak pencilan di tengah dan di ujung disajikan pada Tabel 1. Pada letak pencilan di tengah, nilai dugaan $\hat{\beta}_0$ dari ketiga metode pendugaan sangat dipengaruhi oleh jumlah pencilan, sedangkan pada letak pencilan di ujung metode nilai $\hat{\beta}_0$ dan $\hat{\beta}_1$ MKT saja yang



Gambar 7: Nilai KTG $\hat{\beta}_0$ letak pencilan (a) di tengah dan (b) di ujung pada $n = 200$



Gambar 8: Nilai KTG $\hat{\beta}_1$ letak pencilan (a) di tengah dan (b) di ujung pada $n = 200$

sangat terpengaruh oleh pencilan. Hal ini juga dapat dilihat lebih jelas pada Gambar 1.a- 4.b. yang merupakan contoh plot koefisien regresi untuk $n = 20$ dan $n = 200$ pada pencilan di tengah dan di ujung. Pada Gambar-gambar tersebut terlihat bahwa nilai $\hat{\beta}_0$ metode MKT bergeser jauh dari angka 1, sedangkan nilai MKT bergeser sedikit dari angka 1.

Nilai dugaan koefisien regresi Penduga-MM untuk $n = 20, 60, 100,$ dan 200 mendekati nilai awal $\beta_0 = \beta_1 = 1$ pada pencilan 5-20 persen (Tabel 1). Hal yang sama ditunjukkan pada Gambar 1.a-4.b. Ini menunjukkan bahwa Penduga-MM merupakan penduga yang paling baik dalam menduga koefisien regresi dibandingkan MKT dan Penduga-M. Nilai dugaan MKT

sangat terpengaruh oleh pencilan dengan semakin menyimpangnya nilai dugaan koefisien regresi terhadap nilai parameter yang sebenarnya yaitu $\beta_0 = \beta_1 = 1$. Pada pencilan 25% dan 30% ketiga penduga memberikan nilai dugaan yang kurang baik dengan nilai bias yang cukup besar. Hal ini menunjukkan bahwa ketiga penduga ini sangat terpengaruh oleh pencilan.

Nilai-nilai KTG koefisien regresi kedua metode dari hasil pengulangan sebanyak 10 kali pada setiap letak pencilan disajikan pada Tabel 2. Nilai KTG dihitung untuk mengetahui ketepatan kestabilan nilai dugaan. Bila nilai KTG kecil berarti nilai dugaan lebih baik. Pada Tabel 2 juga terlihat bahwa metode MM memberikan hasil yang sangat baik sampai dengan pencilan 20% untuk letak pencilan di tengah dan di ujung. Hasil yang lebih jelas terlihat pada Gambar 5.a-8.b. Nilai KTG Penduga-MM yang jauh lebih kecil dibandingkan kedua penduga yang lain menunjukkan bahwa dari ketiga metode yang dicobakan metode Penduga-MM merupakan metode robust yang sangat bisa diandalkan bila pencilan berjumlah sampai dengan 20% baik pada letak pencilan di tengah ataupun letak pencilan di ujung.

5. KESIMPULAN

Metode penduga- MM merupakan metode robust yang lebih baik dibandingkan metode MKT dan metode robust Penduga-M pada jumlah pencilan sampai dengan 20% dari jumlah data dalam menduga nilai koefisien regresi linier baik pada letak pencilan di tengah maupun pada letak pencilan di ujung.

Daftar Pustaka

- [1] Chen, C. 2002. Robust Regression and Outlier Detection with the ROBUSTREG procedure. Statistics and Data Analysis. *SAS Institute., Inc. Cary, NC:IML Robust Regression*, [<http://v8doc.sas.com/sashtml>] diakses 23-10-2009.
- [2] Croux, C & Ruiz-Gazen, A. 2005. High breakdown estimators for principal components : the projection-pursuit approach revisited. *Journal of Multivariate Analysis*. Vol 95. pp. 206-226.
- [3] Fox, J. 2002. *Robust Regression*. Appendix To An R And S-Plus Companion To Applied Regression.
- [4] Nisa, K. 2006. Analisis Regresi Robust menggunakan Metode Least Trimmed Square. *Jurnal Ilmiah MIPA*. Vol IX, No. 2 : 93-100.

- [5] Molina I., Pea D., & Prez B. 2009. Robust Estimation In Linear Regression Models With Fixed Effects. *Statistics And Econometrics Series*. Working Paper 09-88 (27).
- [6] Ryan, T. P. 1997. *Modern Regression Methods*. A Wiley-Interscience Publication, New York.
- [7] Ullah, I. & Qadir, M.F. 2006. Insha's Redescending M-estimator for Robust Regression: A Comparative Study. *Pakistan journal of statistics and operation research*, Vol.II No.2 2006 pp135-144
- [8] Van Aelst, S. & Willems, G. 2005. Multivariate Regression S-Estimators For Robust Estimation And Inference. *Statistica Sinica* Vol 15, pp. 981-1001
- [9] Yaffee, R. A. 2002. *Robust Regression Analysis: Some Popular Statistical Package Options*. Statistics, Social Science, and Mapping Group.
- [10] Yohai, V. J. 1987. High Breakdown Point and High Efficiency Robust Estimates For Regression. *Annals Of Statistics*. 15, 642-656. [<http://www.stat.ualberta.ca/wiens/stat578/papers/Yohai.pdf>.] 23-10-2009.

NETTI HERAWATI: PS Matematika, Jurusan Matematika, FMIPA Universitas Lampung, Bandar Lampung 35148, Indonesia
E-mail: netti@unila.ac.id

KHOIRIN NISA: PS Matematika, Jurusan Matematika, FMIPA Universitas Lampung, Bandar Lampung 35148, Indonesia
E-mail: nisa@unila.ac.id

ERI SETIAWAN: PS Matematika, Jurusan Matematika, FMIPA Universitas Lampung, Bandar Lampung 35148, Indonesia
E-mail: nisa@unila.ac.id