

# METODE PENDUGAAN MATRIKS RAGAM-PERAGAM DALAM ANALISIS REGRESI KOMPONEN UTAMA (RKU) (*Variance-Covariance Matrix Estimation Method for Principal Component Regression*)

Itasia Dina S<sup>1</sup>, Dian Kusumaningrum, Yani Suryani  
Departemen Statistika FMIPA Institut Pertanian Bogor

E-mail : <sup>1</sup>[itasia@smkwikrama.net](mailto:itasia@smkwikrama.net)

## Abstrak

*Regresi komponen utama (RKU) merupakan salah satu analisis regresi yang menggunakan komponen utama untuk mengatasi adanya multikolinearitas pada regresi berganda. Metode kemungkinan maksimum (MLE) biasanya digunakan untuk menduga matrik ragam-peragam pada analisis regresi komponen utama. Namun, metode pendugaan ini sangat sensitif terhadap adanya data pencilan multivariat. Oleh karena itu, salah satu cara untuk mengatasi masalah ini adalah dengan menggunakan metode minimum covariance determinant (MCD) dalam menduga matriks ragam-peragamnya. Penelitian ini menggunakan metode MLE dan MCD untuk menduga matriks ragam-peragam pada analisis regresi komponen utama. Sedangkan parameter regresinya diduga oleh metode kuadrat terkecil (MKT). Sementara itu, untuk pemilihan jumlah komponen utama digunakan kriteria 80% proporsi keragaman dari data contoh. Hasil penelitian ini menunjukkan bahwa dampak adanya pencilan multivariat pada analisis regresi komponen utama yang matriks ragam-peragamnya diduga oleh metode MCD akan menghasilkan nilai rata-rata akar ciri pertama yang tetap stabil pada komponen utama pertama (KUI), walaupun rasio pencilan multivariat dengan banyaknya data terus bertambah. Saat rasio pencilan multivariat dengan banyaknya data sebesar 5%, metode pendugaan parameter regresi komponen utama dengan MKT-MLE dan MKT-MCD menunjukkan hasil yang sama baik karena kedua metode ini cenderung menghasilkan nilai bias dan mean squared error (MSE) yang relatif sama kecil. Namun, pada saat rasio pencilan multivariat dengan banyaknya data lebih besar dari 5% (10%,15%,20%), metode MKT-MCD menunjukkan hasil yang lebih baik dibandingkan metode MKT-MLE dalam menduga parameter regresi komponen utama. Hal ini terjadi karena metode MKT-MCD cenderung menghasilkan nilai bias dan MSE yang lebih kecil dibandingkan MKT-MLE.*

*Kata kunci : regresi komponen utama, metode kemungkinan maksimum, minimum covariance determinant*

## PENDAHULUAN

Salah satu masalah yang sering muncul dalam analisis regresi linear berganda adalah adanya korelasi yang kuat antar peubah bebas (multikolinearitas). Hal ini menyebabkan metode kuadrat terkecil (MKT) menghasilkan penduga yang tidak efisien karena matriks yang dibangun untuk menduga parameter regresi yaitu  $X'X$  akan memiliki kondisi buruk (*ill-conditioned*) atau *singular* yang pada akhirnya menyebabkan penduga ragam bagi parameter regresi menjadi lebih besar dari seharusnya (Myers 1989).

Salah satu metode untuk mengatasi adanya multikolinearitas dalam analisis regresi berganda

adalah regresi komponen utama (RKU). RKU merupakan salah satu analisis regresi yang menggunakan komponen utama sebagai peubah bebasnya. Komponen utama ini merupakan kombinasi linear dari peubah asal yang bersifat saling bebas dan dihasilkan dari penguraian matriks ragam-peragam.

Metode kemungkinan maksimum (*maximum likelihood estimation* selanjutnya disingkat MLE) biasanya digunakan untuk menduga matriks ragam-peragam pada RKU. Metode MLE pada prinsipnya adalah memaksimumkan fungsi peluang bersama dari data contoh yang kita miliki (Nasoeition & Rambe 1984). Penduga vektor rata-rata dan matriks

ragam-peragam bagi contoh acak tersebut dengan menggunakan metode pendugaan MLE diperoleh

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = n^{-1} \mathbf{X}' \mathbf{1}$$

$$\begin{aligned} \Sigma &= (n-1)^{-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})' \\ &= (n-1)^{-1} \{ \mathbf{X}' \mathbf{X} - n^{-1} \mathbf{X}' \mathbf{1} \mathbf{1}' \mathbf{X} \} \end{aligned}$$

$\mathbf{X}$  adalah matriks berukuran  $n \times p$  dan  $\mathbf{1}$  adalah vektor berukuran  $n \times 1$  yang elemen-elemennya adalah 1.

Metode pendugaan MLE sangat sensitif terhadap adanya data pencilan multivariat. Data pencilan multivariat diidentifikasi sebagai pengamatan yang memiliki jarak Mahalanobis kekar yang besar secara statistik. Oleh karena itu, metode *minimum covariance determinant* (MCD) merupakan salah satu metode pendugaan matriks ragam-peragam yang digunakan untuk mengatasi masalah ini. MCD diperkenalkan oleh Rousseeuw pada tahun 1985. Tujuan dari metode pendugaan MCD adalah mencari himpunan bagian sebanyak  $h$  elemen yang matriks ragam-peragamnya memiliki determinan terkecil (Rousseeuw 1999).

Pada prinsipnya metode MCD adalah mencari himpunan bagian yang anggotanya sebanyak  $h$  elemen dari matriks  $\mathbf{X}$  dengan  $h$  merupakan bilangan bulat terkecil dari  $(n+p+1)/2$ . Misalkan himpunan bagian itu adalah  $\mathbf{X}_h$ , maka terdapat sebanyak  $C_h^n$  kombinasi yang harus ditemukan untuk mendapatkan dugaan vektor rata-rata dan matriks ragam-peragam. Untuk  $n$  kecil, pendugaan MCD mudah dan relatif lebih cepat untuk ditemukan. Tetapi, jika  $n$  besar maka banyak sekali kombinasi subhimpunan yang harus ditemukan untuk mendapatkan pendugaan MCD. Untuk mengatasi keterbatasan ini digunakan pendekatan FAST-MCD dengan algoritma *C-step* yang dikembangkan oleh Rousseeuw & Van Driessen (1999).

Misalkan terdapat  $\mathbf{X}_p = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p]$  merupakan himpunan data sejumlah  $n$  pengamatan dari  $p$  peubah. Misalkan  $H_1 \subset \{1, 2, \dots, n\}$  dengan  $|H_1| = h$ , maka hitung

$$T_1 = \frac{1}{h} \sum_{i \in H_1} x_i$$

$$\mathbf{C}_1 = \frac{1}{h} \sum_{i \in H_1} (x_i - T_1)(x_i - T_1)'$$

jika  $\det(\mathbf{C}_1) \neq 0$  definisikan jarak relatif  $d_i$  yaitu

$$d_i(i) = \sqrt{(x_i - T_1)' \mathbf{C}_1^{-1} (x_i - T_1)}$$

dengan  $i=1, \dots, n$ . Selanjutnya ambil  $H_2$  demikian sehingga  $\{d_1(i); i \in H_2\} = \{(d_1)_{1:n}, \dots, (d_h)_{h:n}\}$  dengan  $(d_1)_{1:n} \leq (d_1)_{2:n} \leq \dots \leq (d_1)_{h:n}$  menyatakan urutan jarak. Hitung nilai  $T_2$  dan  $\mathbf{C}_2$  berdasarkan himpunan  $H_2$ . Maka  $\det(\mathbf{C}_2) \leq \det(\mathbf{C}_1)$  dan akan sama jika dan hanya jika  $T_1 = T_2$  dan  $\mathbf{C}_1 = \mathbf{C}_2$ .

Penjelasan di atas mensyaratkan  $\det(\mathbf{C}_1) \neq 0$ , karena jika  $\det(\mathbf{C}_1) = 0$  maka nilai objektif minimum untuk mendapatkan determinan terkecil telah ditemukan. Selain itu, jika  $\det(\mathbf{C}_1) > 0$ , penggunaan formulasi di atas akan menghasilkan  $\mathbf{C}_2$  yang  $\det(\mathbf{C}_2) \leq \det(\mathbf{C}_1)$ . Dalam FAST-MCD akan

digunakan algoritma *C-step* dengan  $\mathbf{C}$  disebut *concentration* (pemusatan). Pemusatannya dilakukan pada  $h$  amatan agar menghasilkan jarak relatif terkecil dan  $\mathbf{C}_2$  dipusatkan agar memiliki determinan yang lebih kecil dibandingkan  $\mathbf{C}_1$ . Adapun algoritma dari *C-step* sebagai berikut :

1. Hitung jarak relatif  $d_{old}(i)$  untuk  $i=1, 2, \dots, n$
2. Urutkan jarak relatif hasil permutasi dari  $\pi$  dengan  $d_{old}(\pi(1)) \leq d_{old}(\pi(2)) \leq \dots \leq d_{old}(\pi(n))$ .
3. Tentukan  $H_{new} := \{ \pi(1), \pi(2), \dots, \pi(h) \}$ .
4. Hitung  $T_{new}$  dan  $\mathbf{C}_{new}$ .

pengulangan algoritma *C-step* akan menghasilkan sejumlah proses iterasi. Proses iterasi akan berhenti, jika  $\det(\mathbf{C}_2) = 0$  atau  $\det(\mathbf{C}_2) = \det(\mathbf{C}_1)$ . Jika kondisi di atas belum terpenuhi, maka proses iterasi akan terus berlangsung hingga menghasilkan sejumlah  $h$  amatan yang memiliki nilai determinan terkecil dan konvergen ( $T_{full}$ ,  $\mathbf{C}_{full}$ ). Untuk mendapatkan konsistensi ketika data berasal dari sebaran peubah ganda, maka hitung

$$\begin{aligned} T_{MCD} &= T_{full} \\ \mathbf{C}_{MCD} &= \frac{\text{med}_i d_{(T_{full}, \mathbf{C}_{full})}^{(i)}}{\chi_{p,0.05}^2} \mathbf{C}_{full} \end{aligned}$$

Selanjutnya hasil akhir dari pendugaan FAST-MCD adalah melalui pembobot. Pendugaan terboboti diperoleh dengan cara sebagai berikut

$$\begin{aligned} t_{MCD} &= \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \\ \mathbf{C}_{MCD} &= \frac{\sum_{i=1}^n w_i (x_i - t_{MCD})(x_i - t_{MCD})'}{\sum_{i=1}^n w_i - 1} \end{aligned}$$

$$\text{dengan } w_i = \begin{cases} 1 & \text{jika } d_{(T_{MCD}, \mathbf{C}_{MCD})}^{(i)} \leq \chi_{p,0.975}^2 \\ 0 & \text{lainnya} \end{cases}$$

Penelitian ini, menggunakan metode MLE dan MCD untuk menduga matriks ragam-peragam dalam analisis regresi komponen utama. Sedangkan parameter regresinya diduga dengan Metode Kuadrat Terkecil (MKT). Metode MKT-MLE didefinisikan sebagai metode RKU yang matriks ragam-peragamnya diduga dengan metode MLE dan parameter regresinya diduga dengan metode MKT. Sedangkan MKT-MCD didefinisikan sebagai metode RKU yang matriks ragam-peragamnya diduga dengan metode MCD dan pendugaan parameter regresinya diduga dengan metode MKT.

Data yang digunakan dalam penelitian ini merupakan data simulasi. Dimana menurut Huber *et al* (2005) untuk mendapatkan  $n$  data contoh yang terkontaminasi oleh data pencilan multivariat dapat dilakukan dengan cara membangkitkan sejumlah  $n(1 - \delta)$  dari sebaran normal peubah ganda dengan parameter  $(\boldsymbol{\mu}, \Sigma)$ , sedangkan  $n\delta$  dibangkitkan dari sebaran normal peubah ganda dengan parameter  $(\tilde{\boldsymbol{\mu}}, \tilde{\Sigma})$ .  $\delta$  adalah rasio antara banyaknya pencilan multivariat dengan banyaknya data.

$$(1 - \delta)N_p(\boldsymbol{\mu}, \Sigma) + \delta N_p(\tilde{\boldsymbol{\mu}}, \tilde{\Sigma})$$

Skema algoritma simulasi ini dapat dilihat pada Lampiran 1.

**HASIL DAN PEMBAHASAN**

**Karakteristik Data Bangkitan**

Data populasi  $X^{(0)}$  dibangkitkan dengan vektor rata-rata  $\mu = [0\ 0\ 0\ 0\ 0]'$  sedangkan data pencilan multivariat dibangkitkan dengan vektor rata-rata  $\bar{\mu} = [0\ 0\ 0\ 15\ 0]'$ . Sedangkan besarnya korelasi antar peubah dapat dilihat pada Lampiran 2.

Analisis pendugaan parameter regresi komponen utama dilakukan pada ukuran data contoh  $n=20$  dan  $n=100$ . Ukuran contoh  $n=20$  dipilih sebagai representasi ukuran contoh kecil sedangkan  $n=100$  dipilih sebagai representasi ukuran contoh besar. Sedangkan rasio antara banyaknya pencilan multivariat dengan banyaknya data ( $\delta$ ) yang dicobakan adalah 5%, 10%, 15% dan 20%, serta ulangan dilakukan sebanyak 100 kali. Adapun hasil simulasi yang dilakukan sebagai berikut.

**Perbedaan Akar Ciri**

Rata-rata nilai akar ciri pertama pada komponen utama pertama yang dihasilkan oleh metode MLE dan MCD pada saat  $n=20$  dan  $n=100$  serta rasio antara banyaknya pencilan multivariat dengan banyaknya data ( $\delta$ ) sebanyak 5%, 10%, 15% dan 20% adalah sebagai berikut:

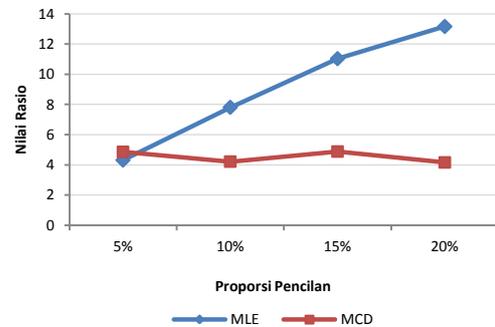
Tabel 1. Rata-rata akar ciri pertama pada komponen utama pertama saat  $n=20$  dan  $n=100$

n	$\delta$	METODE	
		MLE	MCD
n=20	5%	12.10047	3.281058
	10%	21.84225	3.878747
	15%	31.20774	3.098879
	20%	38.92775	3.560835
n=100	5%	11.94905	3.034016
	10%	21.56266	3.223575
	15%	30.06375	3.521161
	20%	37.38446	3.281375

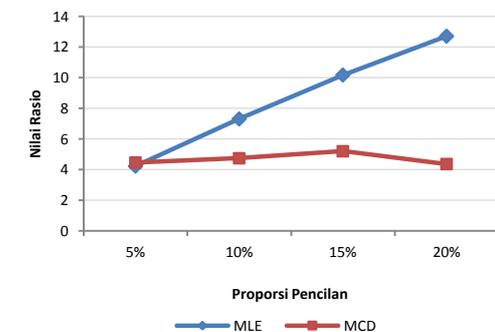
Tabel 1 memperlihatkan bahwa dengan bertambahnya rasio antara banyaknya pencilan multivariat dengan banyaknya data, RKU yang matriks ragam-peragamnya diduga dengan MLE akan menghasilkan rata-rata akar ciri pertama yang lebih besar dibandingkan metode MCD terutama pada komponen utama pertama. Akibatnya, komponen utama pertama pada penduga MLE akan didominasi oleh amatan pencilan, dan tidak mencakup keragaman dari data pada umumnya (Huber *et al* 2005). Sedangkan metode MCD menghasilkan nilai rata-rata akar ciri pertama yang cenderung stabil, karena rata-rata akar ciri pertama yang diduga oleh metode ini diperoleh dari penguraian matriks ragam-peragam yang kekar terhadap adanya pencilan multivariat. Adapun nilai-nilai akar ciri dari setiap komponen utama yang dihasilkan oleh metode MLE dan MCD saat

$n=20$  dan  $n=100$  dapat dilihat pada Lampiran 3 dan Lampiran 4.

Gambar di bawah ini menggambarkan rasio antara nilai akar ciri pertama pada komponen utama pertama dengan nilai akar ciri kedua pada komponen utama kedua saat  $n=20$  dan  $n=100$ .



Gambar 1. Rasio antara akar ciri pertama dan akar ciri kedua saat  $n=20$



Gambar 2. Rasio antara akar ciri pertama dan akar ciri kedua saat  $n=100$

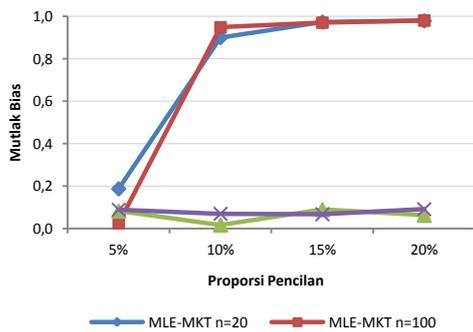
Gambar 1 dan Gambar 2 menunjukkan bahwa metode MLE menghasilkan rasio akar ciri yang besar dengan semakin bertambahnya rasio antara banyaknya pencilan multivariat dengan banyaknya data baik pada saat  $n=20$  dan  $n=100$ . Sedangkan rasio akar ciri yang dihasilkan oleh metode MCD cenderung konstan atau stabil. Perbedaan nilai rata-rata akar ciri pertama pada metode MLE dan MCD dapat menyebabkan perbedaan jumlah komponen utama yang terpilih.

Menurut Johnson (1998) salah satu kriteria penentuan banyaknya jumlah komponen utama yang digunakan adalah dengan mengambil sejumlah komponen utama yang mampu menjelaskan 80% total keragaman dari data contoh. Saat peubah penjelas yang digunakan sebanyak lima, metode MLE memiliki kemungkinan hanya menggunakan satu komponen utama saja untuk menjelaskan 80% total keragaman dari data contoh. Sedangkan metode MCD akan memiliki kemungkinan untuk menggunakan lebih besar atau sama dengan satu komponen utama. Dalam penelitian ini, saat rasio antara banyaknya pencilan multivariat dengan banyaknya data lebih besar dari 5%, metode MCD

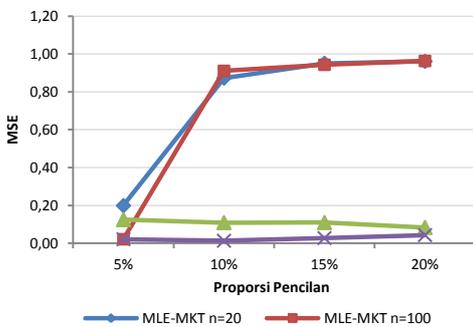
menggunakan satu hingga dua komponen utama saja untuk menjelaskan 80% total keragaman dari data contoh. Walaupun dalam analisis regresi diharapkan hanya sedikit saja komponen utama yang digunakan untuk menjelaskan keragaman dari data contoh, tetapi nilai bias dan *mean squared error* (MSE) yang dihasilkan MKT-MLE lebih besar dibandingkan MKT-MCD.

**Bias dan Mean Squared Error (MSE)**

Nilai bias dan MSE yang diperoleh dengan metode MKT-MLE dan MKT-MCD, pada saat  $n=20$  dan  $n=100$  dapat dilihat pada Lampiran 5 dan Lampiran 6.



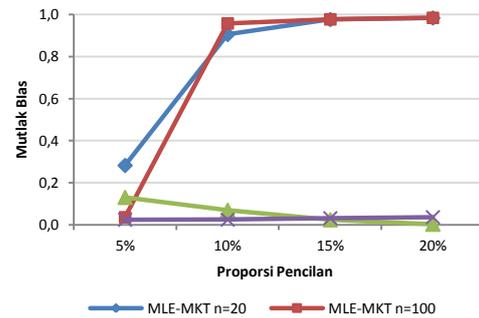
Gambar 3. Perbandingan nilai bias  $\beta_1$  saat  $n=20$  dan  $n=100$  untuk MKT-MLE dan MKT-MCD



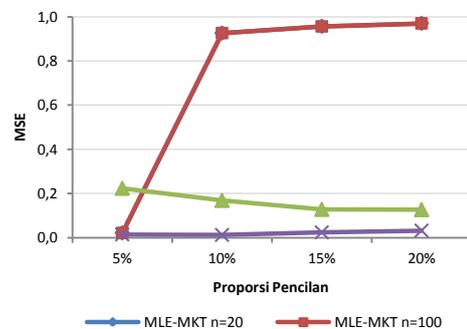
Gambar 4. Perbandingan nilai MSE  $\beta_1$  saat  $n=20$  dan  $n=100$  untuk MKT-MLE dan MKT-MCD

Gambar 1 dan Gambar 2 memperlihatkan bahwa saat rasio antara banyaknya pencilan multivariat dengan banyaknya data meningkat lebih dari 5% nilai bias dan MSE  $\beta_1$  yang dihasilkan oleh metode MKT-MCD akan tetap lebih kecil dibandingkan MKT-MLE baik pada saat ukuran contoh  $n=20$  maupun  $n=100$ . Hal ini dikarenakan metode MCD sebagai metode pendugaan matriks ragam-peragam mampu meminimalisasi adanya pengaruh data pencilan multivariat, sehingga saat bertambahnya rasio antara banyaknya pencilan multivariat terhadap banyaknya data, metode ini akan tetap kekar. Penambahan besarnya ukuran contoh dari

$n=20$  menjadi  $n=100$  dapat memperkecil nilai MSE pada metode MKT-MCD.

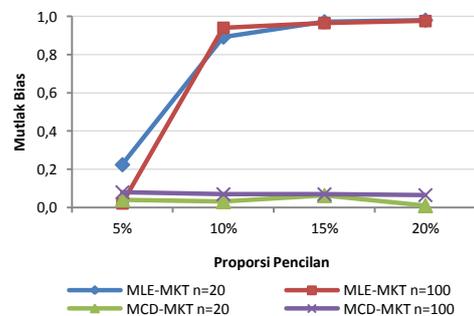


Gambar 5. Perbandingan nilai bias  $\beta_2$  saat  $n=20$  dan  $n=100$  untuk MKT-MLE dan MKT-MCD

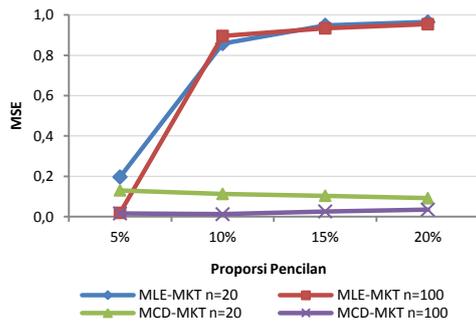


Gambar 6. Perbandingan nilai MSE  $\beta_2$  saat  $n=20$  dan  $n=100$  untuk MKT-MLE dan MKT-MCD

Sama halnya pada pendugaan  $\beta_1$ , Gambar 3 dan Gambar 4 memperlihatkan bahwa nilai bias dan MSE pada pendugaan  $\beta_2$  yang dihasilkan oleh metode MKT-MCD akan lebih kecil dibandingkan MKT-MLE, baik saat ukuran contoh  $n=20$  maupun  $n=100$ . Penambahan banyaknya contoh dari  $n=20$  menjadi  $n=100$  mampu memperkecil nilai MSE pada metode MKT-MCD.

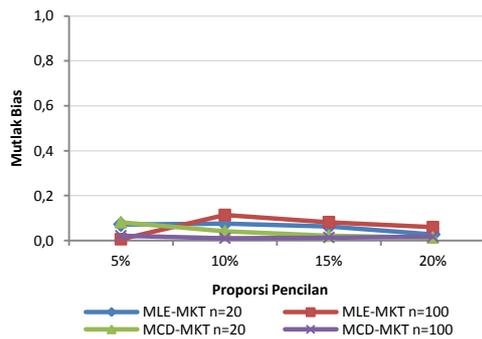


Gambar 7. Perbandingan nilai bias  $\beta_3$  saat  $n=20$  dan  $n=100$  untuk MKT-MLE dan MKT-MCD

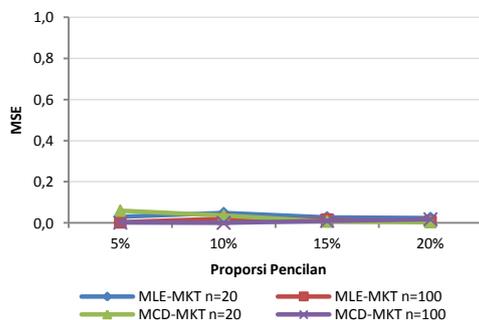


Gambar 8 Perbandingan nilai MSE  $\beta_3$  saat  $n=20$  dan  $n=100$  untuk MKT-MCD dan MKT-MCD

Gambar 7 dan Gambar 8 memperlihatkan kondisi yang relatif sama seperti pendugaan  $\beta_1$  dan  $\beta_2$ . Nilai bias dan MSE dari pendugaan  $\beta_3$  yang dihasilkan oleh metode MKT-MCD akan lebih kecil dibandingkan MKT-MLE baik saat ukuran contoh  $n=20$  maupun  $n=100$ . Pada metode MKT-MCD, penambahan banyaknya contoh  $n=20$  menjadi  $n=100$  mampu menghasilkan nilai MSE yang relatif lebih kecil.



Gambar 9 Perbandingan nilai MSE  $\beta_4$  saat  $n=20$  dan  $n=100$  untuk MKT-MLE dan MKT-MCD



Gambar 10. Perbandingan nilai MSE  $\beta_4$  saat  $n=20$  dan  $n=100$  untuk MKT-MLE dan MKT-MCD

Gambar 9 dan Gambar 10 menunjukkan bahwa bias dan MSE pada pendugaan  $\beta_4$  yang dihasilkan dari kedua metode pendugaan parameter regresi

komponen utama memiliki performa yang hampir sama. Namun, jika dilihat pada Lampiran 5 dan Lampiran 6 untuk pendugaan  $\beta_4$ , metode MKT-MCD tetap menghasilkan bias dan MSE yang lebih kecil dibandingkan metode MKT-MLE.

Nilai bias dan MSE yang dihasilkan metode MKT-MCD tetap kecil dan hampir sama dengan pendugaan  $\beta$  lainnya. Dengan kata lain, metode ini tetap kekar saat bertambahnya rasio antara banyaknya pencilan multivariat dengan banyaknya data. Sedangkan MKT-MLE menunjukkan hal sebaliknya yaitu nilai bias dan MSE yang dihasilkan pada pendugaan  $\beta_4$  berbeda dan cenderung relatif lebih kecil dibandingkan pendugaan  $\beta$  lainnya. Hal ini dikarenakan, pada simulasi data pencilan multivariat dibangkitkan dengan cara menggantikan nilai elemen rata-rata pada  $X_4$  dengan suatu nilai tertentu yang lebih besar dari sebelumnya, sedangkan nilai elemen rata-rata pada  $X$  lainnya tetap. Sehingga saat  $n$  data contoh terkontaminasi oleh pencilan multivariat, karakteristik data pada  $X_4$  akan berbeda dengan data-data peubah penjelas lainnya. Perbedaan yang terjadi pada  $X_4$  adalah  $X_4$  akan memiliki rata-rata dan ragam yang lebih besar dibandingkan sebelumnya. Hal ini akan berpengaruh pada pendugaan akar ciri maupun vektor ciri.

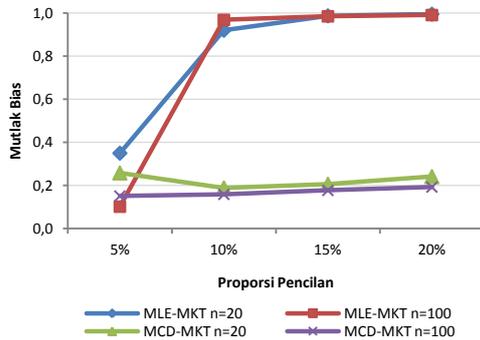
Peubah penjelas yang lebih dominan berpengaruh terhadap komponen utama pertama dapat dilihat pada vektor ciri pertama. Adanya  $n$  data contoh yang terkontaminasi oleh pencilan multivariat, akan mengakibatkan nilai elemen vektor ciri pada  $X_4$  memiliki nilai yang jauh lebih besar dibandingkan elemen vektor ciri pada  $X$  lainnya. Sehingga komponen utama pertama akan didominasi oleh  $X_4$ .

Transformasi peubah baru ke peubah  $X$  atau peubah asal dilakukan untuk mengetahui dugaan parameter regresi pada model awal. Pengaruh  $X_4$  yang dominan pada komponen utama pertama, akan menyebabkan nilai dugaan parameter regresi yang dihasilkannya mendekati nilai parameter sesungguhnya. Sehingga dengan kondisi seperti ini nilai bias dan MSE yang dihasilkan untuk menduga parameter  $X_4$  relatif lebih kecil dibandingkan  $X$  lainnya.

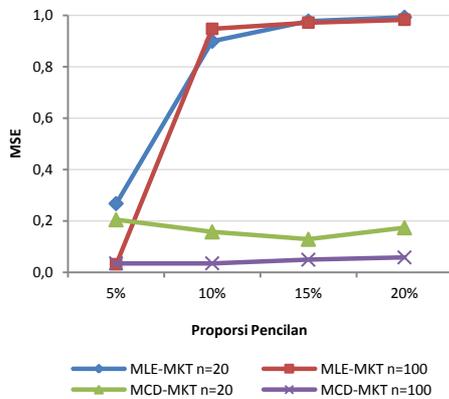
Gambar 11 dan Gambar 12 menunjukkan nilai bias dan MSE yang dihasilkan pada pendugaan  $\beta_5$ . Nilai bias dan MSE yang dihasilkannya cenderung memiliki performa yang relatif sama seperti pada pendugaan  $\beta_1$ ,  $\beta_2$ , dan  $\beta_3$ . Metode MKT-MCD menghasilkan nilai bias dan MSE yang relatif lebih kecil dibandingkan MKT-MLE saat ukuran contoh  $n=20$  maupun  $n=100$ . Penambahan banyaknya contoh  $n=20$  menjadi  $n=100$  pada metode MKT-MCD mampu memperkecil nilai MSE.

Secara umum pada saat  $n=20$  dan  $n=100$  nilai bias dan MSE yang dihasilkan oleh metode MKT-MCD cenderung lebih kecil dibandingkan metode

MKT-MLE dalam menduga parameter regresi komponen utama (RKU).



Gambar 11. Perbandingan nilai MSE  $\beta_5$  saat  $n=20$  dan  $n=100$  untuk MLE- MKT dan MKT-MCD



Gambar 12. Perbandingan nilai MSE  $\beta_5$  saat  $n=20$  dan  $n=100$  untuk MKT-MLE dan MKT-MCD

### KESIMPULAN

Hasil penelitian ini menunjukkan bahwa dampak adanya pencilan multivariat pada analisis regresi komponen utama yang matriks ragam-peragamnya diduga oleh metode MCD akan menghasilkan nilai rata-rata akar ciri pertama yang tetap stabil pada Komponen Utama Pertama ( $KU_1$ ), walaupun rasio pencilan multivariat dengan banyaknya data terus bertambah.

Metode MKT-MCD akan lebih baik dalam menduga parameter regresi komponen utama apabila rasio pencilan multivariat dengan banyaknya data lebih besar dari 5% (10%, 15% dan 20%) karena metode MKT-MCD cenderung menghasilkan nilai bias dan Mean Squared Error (MSE) yang relatif kecil. Sedangkan jika rasio pencilan multivariat dengan banyaknya data kurang dari 5%, metode MKT-MCD dan MKT-MLE akan menghasilkan performa yang sama baik.

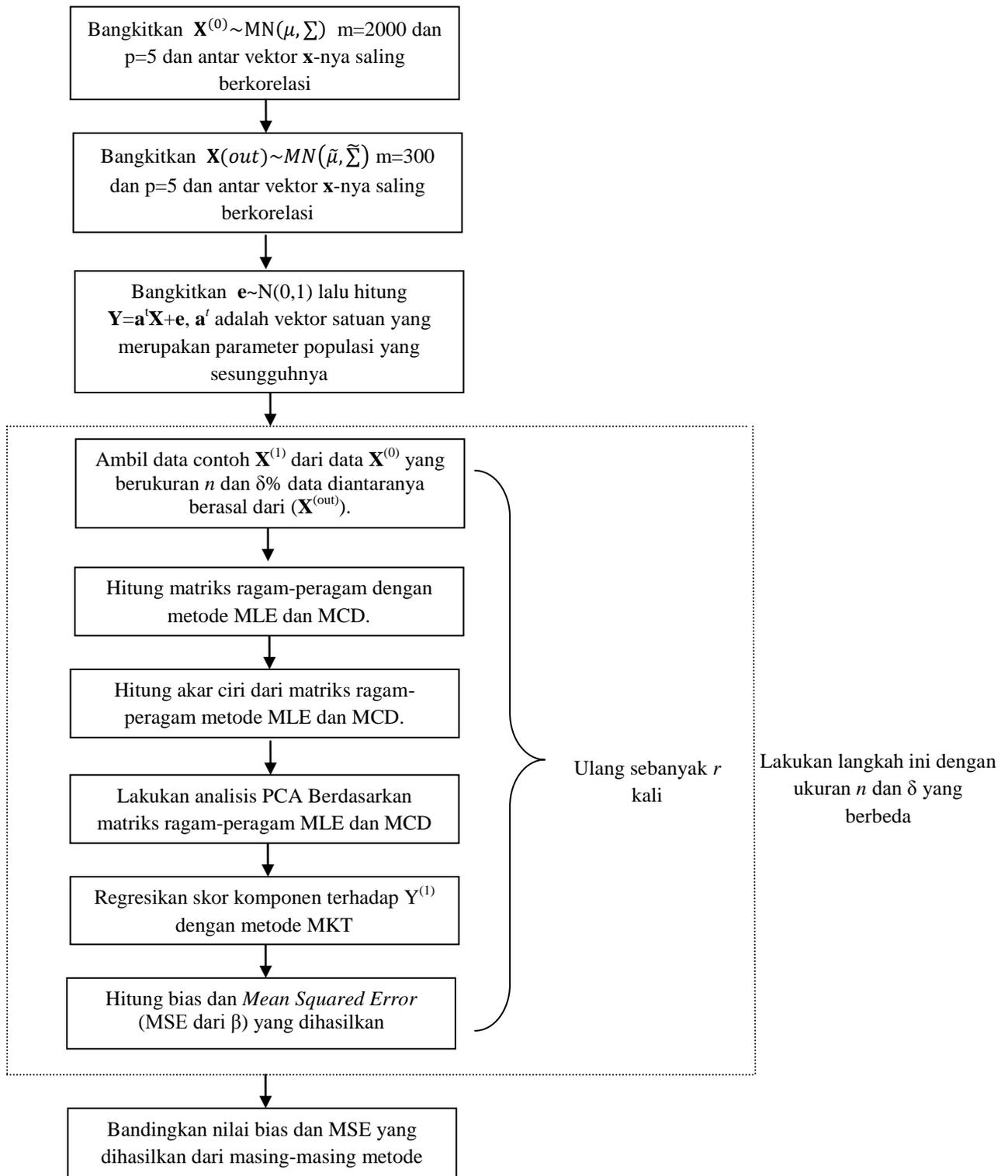
### SARAN

Perlu pengkajian mengenai kombinasi adanya pengaruh pencilan dalam peubah respon maupun peubah penjelas terhadap dugaan parameter regresi. Pengkajian dalam penggunaan matrik ragam-peragam dalam analisis peubah ganda seperti analisis biplot, analisis gerombol maupun analisis diskriminan cukup menarik untuk dikaji.

### DAFTAR PUSTAKA

- Atkinson A & Riani M. 2000. *Robust Diagnostic Regression Analysis*. New York: Springer.
- Huber M, Rousseeuw PJ, & Branden KV. 2005. ROBPCA: A New Approach to Robust Principal Component Analysis. *Technometrics* **47**, 64-79.
- Johnson RA & Wichern DW. 1998. *Applied Multivariate Statistics Analysis*, Fourt edition. London: Prentice-Hall.
- Myers RM. 1989. *Clasical and Modern Regression with Application*, Second Edition. Boston: PWS-KENT.
- Nasoetion AH & Rambe A. 1984. *Teori Statistika untuk Ilmu-Ilmu Kuantitatif*, Edisi kedua. Jakarta: Bhratara Karya Aksara
- Notiragayu. 2008. *Pembandingan Beberapa Metode Analisis Regresi Komponen Utama Robust*. [Makalah Seminar Hasil Penelitian & Pengabdian Kepada Masyarakat]. Bandar Lampung: Universitas Lampung.
- Rocke DM & Woodruff DL. 1996. Identification of Outliers in Multivariate Data. *Journal of the American Statistical Association* **91**, 1047-1061.
- Rousseeuw P & Driessen V. 1999. A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics* **41**, 212-223.
- Smith IL. 2002. *A Tutorial on Principal Components Analysis*. [http://www.cs.otago.ac.nz/cosc453/...tutorial/principal\\_components.pdf](http://www.cs.otago.ac.nz/cosc453/...tutorial/principal_components.pdf). [16 Juli 2009]
- Sumantri B. 1992. *Analisis Regresi Terapan*. Terjemahan dari Draper NR & Smith H. *Applied Regression Analysis*, 2<sup>nd</sup> Edition. Jakarta: PT Gramedia Pustaka Umum.
- Suryana. 2008. *Analisis Deskriminan Robust dengan Menggunakan Penaksiran Minimum Covariance Determinant dan Minimum Weight Covariance Determinant* [Tesis]. Surabaya: Program Pascasarjana, Institut Teknologi Sepuluh Nopember.

Lampiran 1. Skema algoritma simulasi



Lampiran 2. Nilai korelasi antar peubah penjelas

	X1	X2	X3	X4
X2	0.704 (0.000)			
X3	0.817 (0.000)	0.510 (0.000)		
X4	0.718 (0.000)	0.601 (0.000)	0.806 (0.000)	
X5	0.496 (0.000)	0.703 (0.000)	0.587 (0.000)	0.289 (0.000)

Lampiran 3. Nilai akar ciri yang dihasilkan tiap komponen pada saat n=20

Metode	Proporsi Pencilan	Komponen				
		1	2	3	4	5
MLE	5%	12.10047	2.807062	0.609824	0.285303	0.059930
	10%	21.84225	2.830267	0.587560	0.277922	0.060839
	15%	31.20774	2.824154	0.606702	0.285119	0.063123
	20%	38.92775	2.954491	0.603386	0.299287	0.062724
MKT	5%	3.281058	0.675507	0.248653	0.073936	0.003683
	10%	3.878747	0.794977	0.28009	0.081623	0.005329
	15%	3.098879	0.694016	0.265429	0.084094	0.003225
	20%	3.560835	0.684439	0.281912	0.089412	0.003507

Lampiran 4 Nilai akar ciri yang dihasilkan tiap komponen pada saat n=100

Metode	Proporsi Pencilan	Komponen				
		1	2	3	4	5
MLE	5%	11.94905	2.798448	0.616247	0.396761	0.078465
	10%	21.56266	2.958848	0.640975	0.373427	0.081136
	15%	30.06375	2.949049	0.607169	0.374834	0.077176
	20%	37.38446	2.940875	0.613416	0.383644	0.07744
MKT	5%	3.034016	0.720279	0.365457	0.201945	0.006846
	10%	3.223575	0.77513	0.357138	0.200696	0.007519
	15%	3.521161	0.742126	0.367753	0.207052	0.008302
	20%	3.281375	0.751586	0.386627	0.212405	0.007468

Lampiran 5. Nilai bias yang dihasilkan saat  $n=20$  dan  $n=100$

Koefisien Regresi	Proporsi pencilan	n=20		n=100	
		MLE-MKT	MCD-MKT	MLE-MKT	MCD-MKT
$\beta_1$	5%	0.1871686	0.0815523	0.025521	0.088549
	10%	0.9003334	0.0165237	0.948797	0.069279
	15%	0.9739487	0.0890508	0.970855	0.075758
	20%	0.9792076	0.0619884	0.981059	0.090947
$\beta_2$	5%	0.2828108	0.1308402	0.036431	0.024461
	10%	0.9054480	0.0696591	0.958005	0.025063
	15%	0.9766903	0.0243586	0.977445	0.032022
	20%	0.9838950	0.0031605	0.984505	0.036899
$\beta_3$	5%	0.2243521	0.0402294	0.021607	0.080218
	10%	0.8926047	0.0310761	0.941240	0.069998
	15%	0.9729174	0.0628239	0.965870	0.069634
	20%	0.9810937	0.0100834	0.976624	0.065130
$\beta_4$	5%	0.0720078	0.0813621	0.005676	0.021229
	10%	0.0750331	0.0408637	0.113869	0.010196
	15%	0.0625046	0.0213564	0.082006	0.013659
	20%	0.0267614	0.0130492	0.059495	0.017417
$\beta_5$	5%	0.3499703	0.2581955	0.101562	0.151005
	10%	0.9208591	0.1887437	0.969283	0.158529
	15%	0.9877058	0.2058749	0.985708	0.177100
	20%	0.9954906	0.2409160	0.991496	0.192337

Lampiran 6 Nilai *Mean Squared Error* (MSE) yang dihasilkan saat  $n=20$  dan  $n=100$

Koefisien Regresi	Proporsi pencilan	n=20		n=100	
		MLE-MKT	MCD-MKT	MLE-MKT	MCD-MKT
$\beta_1$	5%	0.198453	0.1247659	0.0209341	0.0210521
	10%	0.872634	0.1089853	0.9107728	0.0129883
	15%	0.950779	0.1093314	0.9430642	0.0272208
	20%	0.961175	0.083614	0.9629186	0.0430572
$\beta_2$	5%	0.225625	0.2231951	0.0211983	0.0142203
	10%	0.876144	0.1682665	0.9257220	0.0121820
	15%	0.956875	0.1278128	0.9559315	0.0239750
	20%	0.970193	0.1266318	0.9696739	0.0319691
$\beta_3$	5%	0.196976	0.1305964	0.0191969	0.0163857
	10%	0.858003	0.1129991	0.8957300	0.0124236
	15%	0.948491	0.1039837	0.9334149	0.0261068
	20%	0.964842	0.091797	0.9541786	0.0350019
$\beta_4$	5%	0.03016	0.0600579	0.0033192	0.0016478
	10%	0.047947	0.0360393	0.0206214	0.0006136
	15%	0.026775	0.006834	0.0166542	0.0103657
	20%	0.024004	0.0044697	0.0075620	0.0181516
$\beta_5$	5%	0.267375	0.2044164	0.0312169	0.0347551
	10%	0.8994	0.1570353	0.9479321	0.0348812
	15%	0.978067	0.1283877	0.9721482	0.0489915
	20%	0.992943	0.1732776	0.9833994	0.0578358