PERBANDINGAN METODE KEKAR *BIWEIGHT MIDCOVARIANCE* DAN *MINIMUM COVARIANCE DETERMINANT* DALAM ANALISIS KORELASI KANONIK

Freza Riana, Erfiani, Aji Hamim Wigena
Universitas Ibn Khaldun Bogor
Jln. K.H Sholeh Iskandar Km. 2 Bogor
zarianafre@gmail.com, erfiani_ipb@yahoo.com, ajiwigena@ymail.com

Abstract-Canonical Correlation Analysis (CCA) is a multivariate linear used to identify and quantify associations between two sets of random variables. Its standard computation is based on sample covariance matrices, which are however very sensitive to outlying observations. The robust methods are needed. There are two robust methods, i.e robust Biweight Midcovariance (BICOV) and Minimum Covariance Determinant (MCD) methods. The objective of this research is to compare the performance of both methods based on mean square error. The data simulations are generated from various conditions. The variation data consists of the proportion of outliers, and the kind of outliers: shift, scale, and radial outlier. The performance of robust BICOV method in CCA is the best compared to MCD and Classic.

I. PENDAHULUAN

Analisis korelasi kanonik (AKK) merupakan suatu metode peubah ganda untuk mengidentifikasi dan mengukur hubungan antara dua gugus peubah. AKK berdasarkan pada matriks peragam [1] dan membentuk suatu kombinasi linier dari setiap gugus peubah sedemikian sehingga korelasi di antara kedua gugus peubah tersebut menjadi maksimum [2]. Namun matriks peragam pada AKK sangat sensitif terhadap pengamatan pencilan [3]. AKK dapat diaplikasikan dalam berbagai bidang seperti pemasaran, transportasi, kedokteran, meteorologi, perbankan, pertanian, pendidikan, dan perekonomian.

Ada dua jenis pengamatan pencilan. Pertama, *shift outlier* yaitu pengamatan pencilan dari sebaran rata-ratanya berbeda dengan sebaran dasarnya (tanpa pencilan) tetapi peragamnya sama [4]. Kedua, *scale outlier* yaitu pengamatan pencilan yang peragamnya berbeda tetapi kedua sebarannya sama. Hubert dan Van Driessen menggabungkan keduanya yaitu pengamatan pencilan yang mucul dari sebaran dengan rata-rata dan peragam berbeda, yang dikenal sebagai *radial outlier* [5]. Pencilan tersebut dapat mengakibatkan sebaran data menjadi tidak normal, sehingga matriks peragamnya tidak efisien dan sifat penduganya menjadi berbias [6].

Salah satu pendekatan untuk mengatasi pengamatan pencilan yaitu dengan menggunakan metode kekar [7].

Beberapa metode kekar yang dikembangkan dalam AKK, diantaranya *Minimum Covariance Determinant*, *Projection Pursuit*, *Alternating Regression*, *Sign Test*, dan *Biweight Midcovariance*. Dehon membangkitkan data simulasi dengan proporsi pencilan 10% pada kondisi *scale outlier*, untuk membandingkan metode *Minimum Covariance Determinant* (MCD), *Projection Pursuit*, *Alternating Regression*, dan *Sign Test*. Hasil simulasinya menunjukkan bahwa MCD memberikan nilai *Mean Square Error* (MSE) paling minimum [1]. Cannon dan Hsieh menggunakan metode *Biweight Midcovariance* (BICOV) yang dikembangkan oleh Wilcox pada tahun 1997, untuk mengatasi pencilan pada peramalan curah hujan [7].

Pada penelitian ini akan dibandingkan kinerja metode BICOV dan MCD dalam AKK melalui data simulasi dengan berbagai kondisi pencilan dan proporsi pencilan, dan selanjutnya Menerapkan AKK dengan metode kekar yang terbaik untuk kasus data struktur ekonomi dan kesejahteraan rakyat yang didapat dari data Survey Ekonomi Nasional (SUSENAS) dan Survey Penduduk Antar Sensus (SUPAS) tahun 1995.

II. TINJAUAN PUSTAKA

a) Analisis Korelasi Kanonik

Analisis korelasi kanonik (AKK) yang diperkenalkan oleh Hotelling pada tahun 1936, bertujuan untuk mengidentifikasi dan menghitung hubungan linier antara dua gugus peubah. Perhitungan AKK berfokus pada korelasi antara kombinasi linier dari dua gugus peubah. Ide utama dari AKK adalah mencari pasangan dari kombinasi linier yang memiliki korelasi terbesar. Pasangan kombinasi linier ini disebut peubah kanonik dan korelasinya disebut korelasi kanonik (Johnson dan Wichern 2002)

$$Corr\,(U,V) = \rho = \frac{\boldsymbol{\alpha}\Sigma_{XY}\boldsymbol{\beta}}{\sqrt{\boldsymbol{\alpha}'\Sigma_{XX}\boldsymbol{\alpha}}\sqrt{\boldsymbol{\beta}'\Sigma_{YY}\boldsymbol{\beta}}}$$

Nilai koefisien korelasi kanonik berada pada kisaran $-1 \leq \rho \leq +1$ dan kuadrat korelasi kanonik merupakan proporsi keragaman peubah kanonik *U* yang dapat dijelaskan oleh peubah kanonik *V* (Johnson dan Wichern 2002).

b) Metode Kekar

Perhitungan AKK berdasarkan matriks peragam klasik sangat sensitif terhadap pencilan [3], sehingga diperlukan metode kekar untuk mengatasi pencilan. Beberapa metode kekar telah dikembangkan seperti *Biweight Midcovariance* dan *Minimum Covariance*

Determinat. Kedua matriks peragam yang dihasilkan dari kedua metode tersebut menjadi alternatif sebagai pengganti matriks peragam klasik.

c) *Biweight Midcovariance*

Korelasi Pearson merupakan hubungan antara dua peubah yang bisa dipengaruhi oleh keberadaan suatu pengamatan pencilan [8]. *Biweight midcorrelation* merupakan alternatif sebagai pengganti dari korelasi Pearson. *Biweight* berasal dari pembobot Tukey's *bisquare* yaitu:

$$\psi(x) = \begin{cases} x(1-x^2)^2, & \text{jika } |x| < 1 \\ 0, & \text{jika } |x| \geq 1 \end{cases}$$

Misalkan n adalah jumlah pengamatan, M_x adalah median dari x, dan M_y adalah median dari y, sehingga pembobot untuk E_i dan F_i adalah:

$$E_i = \frac{x_i - M_x}{9MAD_x}$$

$$F_i = \frac{y_i - M_y}{9MAD_y}$$

dengan MAD_x adalah median dari $|x_i - M_x|$ dan MAD_y adalah median dari $|y_i - M_y|$.

Peragam *Biweight Midcovariance* dari x dan y:

$$\Sigma_{bxy} = \frac{n \sum a_i(x_i - M_x)(1-E_i^2)^2 b_i(y_i - M_y)(1-F_i^2)^2}{(\sum a_i(1-E_i^2)(1-5E_i^2))\,(\sum b_i(1-V_i^2)(1-5F_i^2))}$$

dengan:

$$a_i = 1 \text{ jika } -1 \leq E_i \leq 1, \text{ selainnya}$$

$$a_i = 0$$

$$b_i = 1 \text{ jika } -1 \leq F_i \leq 1, \text{ selainnya}$$

$$b_i = 0$$

Peragam *Biweight Midvariance* dari x:

$$\Sigma_{bxx} = \frac{n \sum a_i(x_i - M_x)^2(1-E_i^2)^4}{(\sum a_i(1-E_i^2)(1-5F_i^2))^2}$$

Peragam *Biweight Midvariance* dari y:

$$\Sigma_{byy} = \frac{n \sum b_i(y_i - M_y)^2(1-F_i^2)^4}{(\sum b_i(1-F_i^2)(1-5F_i^2))^2}$$

Matriks peragam BICOV sebagai berikut:

$$\Sigma_{BICOV} = \begin{bmatrix} \Sigma_{bxx} & \Sigma_{bxy} \\ \Sigma_{byx} & \Sigma_{byy} \end{bmatrix} \quad (3)$$

Sehingga didapatkan *Biweight Midcorrelation* adalah:

$$r_{BICOV} = \frac{\Sigma_{bxy}}{\sqrt{\Sigma_{bxx}\Sigma_{byy}}}$$

Nilai korelasi pada *Biweight Midcorrelation* sama dengan nilai korelasi Pearson yaitu berada pada kisaran -1 $\leq r_{BICOV} \leq$ +1.

Matriks peragam BICOV juga dapat digunakan dalam AKK yaitu dengan menggantikan matriks peragam klasik, sehingga didapatkan nilai korelasi kanonik sebagai berikut:

$$\rho_{BICOV} = \frac{\boldsymbol{\alpha}\Sigma_{bxy}\boldsymbol{\beta}}{\sqrt{\boldsymbol{\alpha}'\Sigma_{bxx}\alpha}\sqrt{\boldsymbol{\beta}'\Sigma_{byy}\boldsymbol{\beta}}} \quad (4)$$

d) *Minimum Covariance Determinant*

Minimum Covariance Determinant (MCD) diperkenalkan oleh Rousseeuw pada tahun 1985. Metode MCD bertujuan mencari submatriks $\boldsymbol{H}$ yang berisi unsur-unsur matriks sejumlah h elemen yang matriks peragamnya memiliki determinan terkecil [9]. Pada prinsipnya metode MCD adalah mencari submatriks $\boldsymbol{H}$ berukuran $h \times p$ yang dipilih secara acak sejumlah h elemen dari matriks $\boldsymbol{X}$ berukuran $n \times p$, dengan h merupakan bilangan bulat terkecil dari $(n+p+1)/2$. Kemungkinan banyaknya submatriks $\boldsymbol{H}$ yang dapat dipilih secara acak dari matriks $\boldsymbol{X}$ yaitu sebanyak kombinasi h dari n yang berbeda, C^h_n.

Submatriks $\boldsymbol{H}$ digunakan untuk memperoleh dugaan vektor rataan dan matriks peragam. Jika n kecil, maka penduga MCD relatif mudah dan cepat untuk diperoleh, tetapi jika n besar, maka perlu waktu lama dan banyak sekali kombinasi submatriks yang harus diperoleh untuk mendapatkan penduga MCD. Keterbatasan tersebut dapat diatasi dengan pendekatan FAST-MCD dengan algoritma *C-Step* yang dikembangkan oleh Rousseeuw dan Vandriessen (1999).

Misalkan $\boldsymbol{H}_{(1)} \subset \boldsymbol{X}$, dengan $\boldsymbol{H}_{(1)}$ merupakan submatriks berukuran $h \times p$ dari matriks $\boldsymbol{X}$ berukuran $n \times p$. Hitung vektor rataan dan matriks peragam:

$$\bar{\boldsymbol{t}}_{mcd(1)(p\times 1)} = \frac{1}{h}\sum_{i\in \boldsymbol{H}_1}\sum_{j=1}^{p} x_{ij}$$

$$\Sigma_{mcd(1)(p\times p)} = \frac{1}{h}\sum_{i\in \boldsymbol{H}_1}\sum_{j=1}^{p}(x_{ij}-\bar{\boldsymbol{t}}_{1j})(x_{ij}-\bar{\boldsymbol{t}}_{1j})' \qquad (5)$$

Jika $\det(\Sigma_{mcd(1)}) \neq 0$, maka definisikan jarak relatif, yaitu:

$$d_{(1)}(i) = \sqrt{(x_{ij}-\bar{\boldsymbol{t}}_{1j})'\Sigma^{-1}_{mcd(1)}(x_{ij}-\bar{\boldsymbol{t}}_{1j})}$$, dengan i=1,2,..n.

Urutkan jarak untuk setiap pengamatan, dengan $(d_{(1)})_{1:n} \le (d_{(1)})_{2:n} \le \cdots \le (d_{(1)})_{n:n}$ Selanjutnya, sejumlah h pengamatan yang menghasilkan jarak terkecil menjadi unsur matriks $\boldsymbol{H}_{(2)}$ sedemikian sehingga $\{d_{(1)}(i); i \in \boldsymbol{H}_{(2)}\} \coloneqq \{(d_{(1)})_{1:n}, \dots, (d_{(1)})_{h:n}\}$.

Kemudian, hitung $\bar{\boldsymbol{t}}_{(2)}$ dan $\Sigma_{mcd(2)}$ berdasarkan matriks $\boldsymbol{H}_{(2)}$, dengan $\det(\Sigma_{mcd(2)}) \le \det(\Sigma_{mcd(1)})$.

Penjelasan di atas mensyaratkan $\det(\Sigma_{mcd(1)}) \neq 0$, karena jika $\det(\Sigma_{mcd(1)}) = 0$ maka nilai objektif minimum untuk mendapatkan determinan terkecil telah ditemukan. Selain itu, jika $\det(\Sigma_{mcd(1)}) > 0$, penggunaan formulasi di atas akan menghasilkan $\Sigma_{mcd(2)}$ yang $\det(\Sigma_{mcd(2)}) \le \det(\Sigma_{mcd(1)})$. Dalam FAST-MCD akan digunakan algoritma *C-Step*, ada pun algoritma dari *C-Step* sebagai berikut:

1. Tetapkan $\boldsymbol{H}_{(\text{old})}$, hitung $\bar{\boldsymbol{t}}_{mcd(\text{old})}$ dan $\sum_{mcd(\text{old})}$.
2. Hitung jarak relatif $\sum_{mcd(\text{old})}(i)$ untuk i=1, 2, …, n
3. Urutkan jarak relatif hasil permutasi dari π dengan $d_{(\text{old})}(\pi(1)) \leq d_{(\text{old})}(\pi(2)) \leq \cdots \leq d_{(\text{old})}(\pi(n))$.
4. Tentukan $\boldsymbol{H}_{(\text{new})} := \{\pi(1), \pi(2), \dots, \pi(h)\}$.
5. Hitung $\bar{\boldsymbol{t}}_{mcd(\text{new})}$ dan $\sum_{mcd\ \text{new})}$.
6. Pengulangan algoritma *C-Step* akan menghasilkan sejumlah proses iterasi. Proses iterasi akan berhenti, jika $\det(\sum_{mcd(2)}) = 0$ atau $\det(\sum_{mcd(2)}) = \det(\sum_{mcd(1)})$.
7. Jika kondisi di atas belum terpenuhi, maka proses iterasi akan terus berlangsung hingga menghasilkan sejumlah h amatan yang memiliki nilai determinan terkecil dan konvergen.

$\sum_{mcd}$ merupakan matriks peragam dengan determinan terminan terkecil. Matriks peragam MCD tersebut dapat dituliskan sebagai berikut:

$$\textstyle\sum_{mcd} = \begin{bmatrix} \sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{bmatrix} \qquad (5)$$

Matriks pergam MCD dapat digunakan sebagai pengganti matriks peragam klasik dalam AKK. Sehingga didapatkan nilai korelasi kanonik dari:

$$\rho_{MCD} = \frac{\boldsymbol{\alpha}\sum_{12}\boldsymbol{\beta}}{\sqrt{\boldsymbol{\alpha}'\sum_{11}\alpha}\sqrt{\boldsymbol{\beta}'\sum_{22}\boldsymbol{\beta}}} \qquad (6)$$

e) Pencilan

Pencilan merupakan suatu pengamatan yang menyimpang cukup jauh dari pengamatan lainnya sehingga menimbulkan kecurigaan bahwa pengamatan tersebut berasal dari sebaran data yang berbeda (Hawkins 1997). Berdasarkan pengaruh pengamatan pencilan terhadap data, pencilan dapat dibedakan menjadi tiga jenis. Pencilan pertama yaitu *shift outlier*, merupakan pengamatan pencilan yang berasal dari sebaran yang berbeda dengan sebaran dasarnya (tanpa pencilan) tetapi peragamnya sama. *Shift outlier* mampu menggeser vektor rataan sehingga pusat data menjadi berubah. Pada data menyebar normal, pergeseran vektor rataan bisa melalui penambahan setiap vektor rataan dengan v satuan. Data dengan kondisi *shift outlier* dapat dinyatakan dengan persamaan:

$$\pi_j \sim (1-\varepsilon)N_p(\boldsymbol{\mu}_j, \textstyle\sum_j) + \varepsilon N_p(\boldsymbol{\mu}_j^*, \sum_j);\ j = 1, \dots, n \qquad (7)$$

ε menyatakan proporsi pencilan dalam data dan $\boldsymbol{\mu}_j^*$ menyatakan vektor rataan yang berfungsi sebagai *shift outlier*.

Jenis kedua, *scale outlier*, yaitu pengamatan pencilan yang berasal dari peragam yang berbeda tetapi sebarannya sama. *Scale outlier* mampu merubah bentuk ellipsoid. Data dengan kondisi *scale outlier* dapat dinyatakan dengan persamaan:

$$\pi_j \sim (1-\varepsilon) N_p(\boldsymbol{\mu}_j, \textstyle\sum_j) + \varepsilon N_p j(\boldsymbol{\mu}_j, K \textstyle\sum_j); j = 1, \dots, n \qquad (8)$$

$K\sum_j$ menyatakan matriks peragam yang berfungsi sebagai *scale outlier*.

Jenis ketiga merupakan gabungan dua jenis pencilan yaitu *shift outlier* dan *scale outlier*, yang disebut dengan *radial outlier* (Hubert dan Van Driessen 2004). *Radial outlier* mampu menggeser pusat ellipsoid dan merubah bentuk ellipsoid. Data dengan kondisi *radial outlier* dinyatakan dengan persamaan:

$$\pi_j \sim (1-\varepsilon) N_p(\boldsymbol{\mu}_j, \textstyle\sum_j) + \varepsilon N_p(\boldsymbol{\mu}_j^*, K \textstyle\sum_j); j = 1, \dots, n \qquad (9)$$

f) Pendeteksian Pencilan

Identifikasi pencilan pada peubah ganda umumnya didasarkan pada jarak Mahalanobis (Suryana 2008). Johnson dan Wichern (1998) menyatakan bahwa pengamatan ke-i diidefinisikan sebagai pencilan jika jaraknya lebih besar dari nilai khi-kuadrat pada sejumlah p peubah. Perhitungan tersebut sebagai berikut:

$$d_{MD}^2 = (x_i - \boldsymbol{\mu})' \textstyle\sum^{-1} (x_i - \boldsymbol{\mu}) > \chi^2_{P(1-\alpha)}$$

$\boldsymbol{\mu}$ dan $\sum$ merupakan vektor rataan dan matriks peragam.

Rousseeuw dan Von Zomeren (1990) menjelaskan bahwa penggunaan jarak Mahalanobis untuk pendeteksian pencilan pada peubah ganda menjadi tidak maksimal jika terdapat lebih dari satu pengamatan pencilan, karena adanya pengaruh *masking* dan *swamping*. *Masking* terjadi pada saat pengamatan pencilan tidak terdeteksi karena adanya pengamatan pencilan yang berdekatan, sedangkan *swamping* terjadi pada saat pengamatan bukan pencilan teridentifikasi sebagai pengamatan pencilan.

Salah satu metode kekar yang dikembangkan untuk mengatasi pencilan dengan jumlah lebih dari satu pengamatan pada kasus peubah ganda yaitu jarak Mahalanobis kekar MCD (Hubert *et al*. 2007).

Suatu pengamatan ke-i didefinisikan sebagai pencilan jika:

$$d^2_{RD} = (x_{ij} - \boldsymbol{t}_{1j})'\Sigma^{-1}_{mcd(1)}(x_{ij} - \boldsymbol{t}_{1j}) > \chi^2_{p(1-\alpha)} \quad (10)$$

dengan $\boldsymbol{t}$ dan $\sum_{mcd}$ merupakan vektor rataan dan matriks peragam dari sebagian data $\boldsymbol{X}$ yang mempunyai determinan matriks peragam terkecil.

III. METODOLOGI

Langkah-langkah analisis data yang dilakukan berkaitan dengan tujuan penelitian dilakukan melalui tahapan sebagai berikut:

1. Perbandingan metode BICOV dan MCD

a) Membangkitkan data populasi

Membangkitkan data populasi untuk gugus X berukuran $N \times p$ dan gugus Y berukuran $N \times q$, dengan $p = 2$, $q = 3$ dan $N = 5000$. Data populasi gugus tersebut dibangkitkan dengan sebaran normal ganda $N_p(\boldsymbol{\mu}, \Sigma)$, dengan $\boldsymbol{\mu} = [0,0,0,0,0]$ dan matriks peragamnya

$$\Sigma = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$$

b) Membangkitkan data contoh tanpa pencilan

Membangkitkan data contoh mengikuti sebaran seperti data populasi, dengan jumlah pengamatan n_c=50 dan 100. Data contoh dibangkitkan sebanyak M = 500 kali.

c) Membangkitkan data dengan pencilan

Simulasi untuk data pencilan didapatkan dengan mengubah data contoh sejumlah proporsi pencilan (ε) dengan berbagai jenis kondisi pencilan. Kondisi berbagai jenis pencilan dibangkitkan pada gugus X*Y dan gugus X*Y*, sebagai berikut:

i. Pada kondisi ini pengamatan dalam bentuk *shift outlier* dengan rata-rata dan matriks peragam mengikuti persamaan (7). Masing-masing parameter diberikan nilai:

1) $\boldsymbol{\mu}_1 = [0,0,0,0,0]$

$\boldsymbol{\mu}_1^* = \boldsymbol{\mu}_1 + (v_1, \dots, v_1)$, dengan $v_1 = +10$

$$\Sigma_1 = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$$

2) $\boldsymbol{\mu}_2 = [0,0,0,0,0]$

$\boldsymbol{\mu}_2^* = \boldsymbol{\mu}_2 + (v_2, \dots, v_2)$, dengan $v_2 = +20$

$$\Sigma^2 = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$$

ii. Pada kondisi ini pengamatan dalam bentuk *scale outlier* dengan rata-rata dan matriks peragam mengikuti persamaan (8). Masing-masing parameter diberikan nilai:

1) $\boldsymbol{\mu}_1 = [0,0,0,0,0]$

$$\Sigma_1 = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$$

$\Sigma_1^* = K\,\Sigma_1 = 100\Sigma_1$.

2) $\boldsymbol{\mu}_2 = [0,0,0,0,0]$

$$\Sigma_2 = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$$

$\Sigma_2^* = K\,\Sigma_2 = 144\Sigma_2$.

iii. Pada kondisi ini pengamatan dalam bentuk *radial outlier* dengan rata-rata dan matriks peragam mengikuti persamaan (9). Masing-masing parameter diberikan nilai:

1) $\boldsymbol{\mu}_1 = [0,0,0,0,0]$
$\boldsymbol{\mu}_1^* = \boldsymbol{\mu}_1 + (v_1, \dots, v_1)$, dengan $v_1 = +10$

$$\Sigma_1 = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$$

$\Sigma_1^* = K\,\Sigma_1 = 100\Sigma_1$.

2) $\boldsymbol{\mu}_2 = [0,0,0,0,0]$
$\boldsymbol{\mu}_2^* = \boldsymbol{\mu}_2 + (v_2, \dots, v_2)$, dengan $v_2 = +20$

$$\Sigma_2 = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$$

$\Sigma_2^* = K\,\Sigma_2 = 144\Sigma_2$.

iv. Menentukan matriks peragam
Menghitung matriks peragam dengan metode klasik, BICOV, MCD untuk gugus populasi, gugus XY, gugus X*Y dan gugus X*Y*.

v. Menentukan nilai korelasi kanonikMenghitung nilai korelasi kanonik klasik untuk gugus populasi, gugus XY, gugus X*Y dan gugus X*Y*.

vi. Menghitung nilai MSE untuk metode klasik, BICOV, dan MCD.

$$\text{MSE}(\hat{\rho}) = \frac{1}{M}\sum_{m=1}^{M}\left(\varphi(\hat{\rho})^m - \varphi(\rho)\right)^2$$

dengan:

m: $1,2,\dots,M$, $M = 500$

$(\hat{\rho})^m$ adalan nilai korelasi contoh bangkitan ke-*m*

ρ adalah nilai korelasi populasi

$\varphi(\rho)$ didapatkan dari $\tanh^{-1}(\rho)$ (Dehon *et al*. 2000)

vii. Mengevaluasi kinerja metode klasik, BICOV dan MCD

Membandingkan nilai MSE dari ketiga metode tersebut. Metode tersebut dikatakan terbaik atau paling kekar apabila memberikan nilai MSE paling minimum (Dehon *et al*. 2000).

IV. PEMBAHASAN

Kondisi *Shift Outlier*

Keseluruhan hasil simulasi pada kondisi *shift outlier* menunjukkan bahwa metode BICOV mampu meminimumkan nilai MSE dengan pola grafik yang konsisten mulai dari gugus data tanpa pencilan sampai dengan proporsi pencilan 12 %, baik untuk gugus X*Y maupun gugus X*Y*.

Jenis Pencilan	**Gugus Peubah**	**Metode**	**Proporsi Pencilan**						
			0%	**2%**	**4%**	**6%**	**8%**	**10%**	**12%**
Shift Outlier (+10)	Gugus X*Y	MLE	0.02	0.13	0.19	0.27	0.28	0.29	0.29
		BICOV	0.02	0.02	0.02	0.02	0.02	0.02	0.02
		MCD	0.08	0.08	0.07	0.07	0.06	0.05	0.05
	GugusX*Y*	MLE	0.02	0.30	0.47	0.42	0.28	0.19	0.43
		BICOV	0.02	0.02	0.02	0.02	0.02	0.02	0.02
		MCD	0.08	0.07	0.05	0.05	0.05	0.13	0.04
Shift Outlier (+20)	Gugus X*Y	MLE	0.02	0.23	0.27	0.29	0.30	0.30	0.31
		BICOV	0.02	0.02	0.02	0.02	0.02	0.02	0.02
		MCD	0.07	0.06	0.06	0.05	0.05	0.04	0.04
	GugusX*Y*	MLE	0.02	0.54	0.61	0.32	0.43	0.26	0.34
		BICOV	0.02	0.02	0.02	0.02	0.02	0.02	0.02
		MCD	0.07	0.05	0.05	0.04	0.04	0.12	0.05

Jenis Pencilan	**Gugus Peubah**	**Metode**	**Prporsi Pencilan**						
			0%	**2%**	**4%**	**6%**	**8%**	**10%**	**12%**
Shift Outlier (+10)	Gugus X*Y	MLE	0.01	0.15	0.23	0.28	0.31	0.33	0.34
		BICOV	0.01	0.01	0.01	0.01	0.01	0.01	0.02
		MCD	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	GugusX*Y*	MLE	0.01	0.34	0.53	0.64	0.39	0.39	0.45
		BICOV	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		MCD	0.02	0.02	0.02	0.02	0.02	0.02	0.26

Shift Outlier	Gugus X*Y	MLE	0.01	0.33	0.39	0.42	0.44	0.45	0.45
(+20)		BICOV	0.01	0.01	0.01	0.01	0.01	0.01	0.02
		MCD	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	GugusX*Y*	MLE	0.01	0.63	0.71	0.65	0.44	0.43	0.48
		BICOV	0.01	0.01	0.01	0.01	0.01	0.02	0.02
		MCD	0.02	0.02	0.02	0.02	0.02	0.02	0.32

Kondisi *Scale Outlier*

Pada jumlah pengamatan N_c=50 dan N_c=100 dengan kondisi pencilan *scale outlier* menunjukkan bahwa metode klasik yang paling buruk dengan nilai MSE yang paling maksimum dan pola grafik yang berubah-ubah. Sedangkan metode MCD merupakan kinerja metode kekar yang lebih baik dibanding metode klasik. Namun dibandingkan MCD, metode BICOV merupakan metode paling kekar, dengan memberikan nilai MSE paling minimum dan pola grafik yang konsisten untuk setiap proporsi pencilan.

Jenis Pencilan	Gugus Peubah	Metode	Proporsi Pencilan						
			0%	2%	4%	6%	8%	10%	12%
Scale Outlier	Gugus X*Y	MLE	0.02	0.08	0.12	0.18	0.21	0.24	0.26
K=100		BICOV	0.02	0.02	0.02	0.02	0.03	0.03	0.03
		MCD	0.08	0.08	0.08	0.07	0.07	0.07	0.07
	GugusX*Y*	MLE	0.02	0.16	0.19	0.16	0.15	0.12	0.10
		BICOV	0.02	0.02	0.02	0.02	0.02	0.02	0.03
		MCD	0.08	0.07	0.08	0.03	0.07	0.07	0.07
Scale Outlier	Gugus X*Y	MLE	0.02	0.09	0.16	0.20	0.24	0.27	0.28
K=144		BICOV	0.02	0.02	0.02	0.02	0.03	0.03	0.04
		MCD	0.09	0.09	0.09	0.08	0.07	0.07	0.07
	GugusX*Y*	MLE	0.02	0.25	0.27	0.28	0.21	0.17	0.13
		BICOV	0.02	0.02	0.02	0.02	0.02	0.02	0.03
		MCD	0.09	0.09	0.09	0.08	0.07	0.07	0.07

Jenis Pencilan	Gugus Peubah	Metode	Prporsi Pencilan						
			0%	2%	4%	6%	8%	10%	12%
Scale Outlier	Gugus X*Y	MLE	0.01	0.11	0.18	0.24	0.29	0.31	0.33
K=100		BICOV	0.01	0.01	0.01	0.01	0.01	0.02	0.02
		MCD	0.03	0.03	0.01	0.03	0.03	0.03	0.03
	GugusX*Y*	MLE	0.01	0.10	0.06	0.06	0.07	0.08	0.09
		BICOV	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		MCD	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Scale Outlier	Gugus X*Y	MLE	0.01	0.13	0.22	0.28	0.33	0.36	0.37
K=144		BICOV	0.01	0.01	0.01	0.01	0.01	0.02	0.02
		MCD	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	GugusX*Y*	MLE	0.01	0.12	0.09	0.07	0.07	0.08	0.10
		BICOV	0.01	0.01	0.01	0.01	0.01	0.01	0.02
		MCD	0.02	0.02	0.02	0.02	0.02	0.02	0.02

Kondisi *Radial Outlier*

Pada kasus gugus data dengan kondisi *radial outlier*, tidak satupun hasil simulasi data menunjukkan metode MCD lebih baik daripada metode BICOV. Sedangkan metode klasik merupakan metode paling buruk di antara metode lainnya. Keseluruhan hasil simulasi menunjukkan bahwa metode BICOV memberikan nilai MSE paling minimum untuk setiap proporsi pencilan 12% untuk gugus X*Y dan gugus X*Y*.

Jenis Pencilan	Gugus Peubah	Metode	Proporsi Pencilan						
			0%	2%	4%	6%	8%	10%	12%
Radial Oulier	Gugus X*Y	MLE	0.02	0.17	0.27	0.03	0.33	0.36	0.38
(+10,*K*=100)		BICOV	0.02	0.02	0.03	0.04	0.03	0.04	0.04
		MCD	0.08	0.08	0.03	0.06	0.06	0.06	0.06
	GugusX*Y*	MLE	0.02	0.50	0.62	0.54	0.48	0.34	0.25
		BICOV	0.02	0.02	0.02	0.02	0.03	0.03	0.03
		MCD	0.08	0.08	0.07	0.07	0.06	0.06	0.06
Radial Oulier	Gugus X*Y	MLE	0.02	0.21	0.29	0.31	0.32	0.32	0.32
(+20,*K*=144)		BICOV	0.02	0.02	0.02	0.02	0.02	0.02	0.03

		MCD	0.08	0.07	0.07	0.07	0.07	0.06	0.06
	GugusX*Y*	MLE	0.02	1.60	1.80	1.72	1.27	1.02	0.82
		BICOV	0.02	0.02	0.02	0.02	0.02	0.02	0.02
		MCD	0.08	0.07	0.07	0.07	0.07	0.06	0.06

Jenis Pencilan	**Gugus Peubah**	**Metode**	**Prporsi Pencilan**						
			0%	**2%**	**4%**	**6%**	**8%**	**10%**	**12%**
Radial Oulier (+10,*K*=100)	Gugus X*Y	MLE	0.01	0.20	0.30	0.35	0.38	0.39	0.40
		BICOV	0.01	0.01	0.01	0.01	0.02	0.02	0.02
		MCD	0.03	0.03	0.02	0.02	0.02	0.02	0.02
	GugusX*Y*	MLE	0.01	0.32	0.29	0.21	0.14	0.09	0.06
		BICOV	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		MCD	0.03	0.03	0.02	0.02	0.02	0.02	0.02
Radial Oulier (+20,*K*=144)	Gugus X*Y	MLE	0.01	0.34	0.40	0.42	0.43	0.44	0.44
		BICOV	0.01	0.01	0.01	0.01	0.01	0.01	0.02
		MCD	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	GugusX*Y*	MLE	0.01	1.09	0.92	0.66	0.45	0.35	0.27
		BICOV	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		MCD	0.02	0.02	0.02	0.02	0.02	0.02	0.02

V. PENUTUP

Biweight Midcovariance merupakan metode kekar terbaik yang memberikan nilai *mean square error* paling minimum dibandingkan metode klasik dan metode *Minimum Covariance Determinant* dalam Analisis korelasi kanonik dengan berbagai kondisi pencilan (*shift outlier*, *scale outlier*, *radial outlier)* dan berbagai proporsi pencilan (2%, 4%, 6%, 8%, 10%, 12%).

VI. DAFTAR PUSTAKA

[1] Dehon C, Filzmoser P, Croux C. 2000. Robust Methods for Canonical Correlation Analysis. http://www.statistik.tuwien.ac.at/public/filz/papers/namur00. pdf.

[2] Johnson RA, Winchern DW. 2002. Applied Multivariate Statistical Analysis. Fourth Edition. New Jersey: Prentice-Hall International inc.

[3] Romanazzi M. 1992. Influence in Canonical Correlation Analysis.

Psychometrika. 57:237-259. http://www.springerlink.com/content/l24p13843114jr65/.

[4] Hawkins DM, McLachlan GJ. 1997. High Breakdown Linear Discriminant Anaysis. *Journal of the American Statistical Association*, 437:136 –143.

[5] Hubert M, Van Driessen K. 2004. Fast and Robust Discriminant Analysis, *Computational Statistics and Data Analysis*. 45: 301-320. ftp://adrem.ua.ac.be/pub/preprints/02/Fasrob02.pdf.

[6] Yohai VJ. 2006. A Fast Algorithm for S-regression Estimates. *Journal of Computational and Graphical Statistics*. 15:414-427. http://www.stat.ualberta.ca/~wiens/stat578/papers/Salibian-Barrera%20&%20Yohai.pdf.

[7] Cannon AJ, Hsieh WW. 2008. Robust Nonlinear Canonical Correlation Analysis: Application to Seasonal Climate Forecasting. http://www.nonlin-processes-geophys.net/15/221/2008/npg-15-221-2008.pdf.

[8] Rancher AC. 2002. Methods of Multivariate Analysis. Second Edition. John Wiley & Sons. New York.

[9] Wilcox RR. 2004. Introduction to Robust Estimation and Hypotesis Testing. Second Edition. Academic press.

[10] Rousseeuw PJ, Van Driessen K. 1999. A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*. 3:212-223. ftp://ftp.win.ua.ac.be/pub/preprints/99/Fasalg99.pdf.