Forum Statistika dan Komputasi : *Indonesian Journal of Statistics*
ISSN : 0853-8115
Vol. 18 No.1, April 2013, p: 11-20
available online at:
journal.ipb.ac.id/index.php/statistika

PENDEKATAN KUADRAT TERKECIL PARSIAL KEKAR UNTUK PENANGANAN PENCILAN PADA DATA KALIBRASI

(*Partial Least Squares Robust Regression Approach to Handle Outliers in Calibration Data*)

Enny Keristiana Sinaga[1], Anik Djuraidah[2], Aji Hamim Wigena[3]
Departemen Statistika IPB
E-mail: [1]christy_math02@yahoo.com, [2]anikdjuraidah@gmail.com, [3]ajiwigena@ymail.com

Abstract

The serious problems in the calibration of multivariate estimation are multicollinearity and outliers. Partial Least Squares (PLS) is one of the statistical method used in chemometrics, to handle high or perfect multicollinearity in independent variables. Straightforward Implementation Partial Least Squares (SIMPLS) is the extension of PLS regression proposed by De Jong (1993). The SIMPLS algorithm is based on the empirical cross-variance matrix between the independent variables and the regressors. This method does not resistant toward outlier observations. Robust PLS method is used to handle the multicollinearity and outliers in the data sets. This method can be classified in two groups, there are iteratively reweighting technique and robustication of covariance matrix. Partial Regression-M (PRM) method is one of the robust PLS methods used the idea of iteratively reweighting technique that proposed by Serneels et al. (2005). Robust SIMPLS (RSIMPLS) method is one of the robust PLS methods used the idea of robustication of covariance that proposed by Huber and Branden (2003). A modified RSIMPLS used M estimator with the Huber weight function called RSIMPLS-M was proposed by Ismah (2010). These two methods (RSIMPLS-M and PRM) are applied to Fish data (Naes 1985) to know their performances. The research results indicated that the values of R^2 and RMSEP of RSIMPLS-M are higher than those of PRM method. Whereas based on the confidence interval estimation of the regression coefficients by jackknife method, estimation of PRM is narrower than that RSIMPLS-M method. Therefore RSIMPLS-M method is better than PRM method for prediction, whereas PRM method is better than RSIMPLS-M method for estimation.

Keywords : Partial least squares regression robust (PLSRR), partial robust M-regression (PRM), straightforward implementation partial least squares robust (RSIMPLS)

PENDAHULUAN

Analisis regresi merupakan bagian dalam analisis yang digunakan untuk memodelkan hubungan peubah tak bebas ($\boldsymbol{Y}$) dengan satu atau beberapa peubah bebas ($\boldsymbol{X}$). Bila dalam analisisnya melibatkan dua atau lebih peubah bebas, maka analisis yang digunakan adalah analisis regresi berganda. Permasalahan dalam analisis regresi berganda adalah terjadinya korelasi yang tinggi antar dua atau lebih peubah bebas (multikolinieritas) dan pencilan (*outliers*). Adanya masalah multikolinieritas mengakibatkan ketidakstabilan pendugaan parameter (Myers 1990). Sedangkan pencilan dapat menyebabkan sisaan yang besar dari model yang diperoleh, ragam pada data menjadi lebih besar dan dugaan selang kepercayaan memiliki rentang yang lebar. Menurut Liebmann *et al*. (2009), ada dua tipe pengamatan pencilan yakni pencilan sisaan dan pengamatan berpengaruh (*leverage point*). Pencilan sisaan adalah pengamatan pencilan yang mempunyai sisaan baku yang besar, sedangkan pengamatan berpengaruh merupakan pencilan berganda (*multivariate outliers*) yang terdapat dalam ruang peubah bebas.

Pada bidang kemometrik, sering dijumpai data dengan kondisi adanya multikolinieritas antar peubah bebas dan masalah pencilan. Masalah multikolinieritas dapat diatasi dengan beberapa metode, antara lain metode Regresi Komponen Utama (*Principal Component Regression*/PCR), metode Regresi Gulud (*Ridge Regression*), dan metode Regresi Kuadrat Terkecil Parsial (*Partial Least Squares*/PLS). Menurut Wigena dan Aunuddin (1997), PLS memberikan hasil yang lebih baik daripada metode lainnya dalam mengatasi masalah multikolinieritas. Salah satu algoritma pendugaan PLS adalah SIMPLS (*Straightforward Implementation Partial Least Square*) yang dapat diterapkan dalam mengatasi multikolinieritas, tetapi tidak resisten terhadap pencilan. Pengembangan dari metode PLS dengan menggunakan algoritma SIMPLS sebagai alternatif untuk mengatasi pencilan adalah dengan menggunakan regresi kekar. Secara umum, PLS kekar dapat dibagi ke dalam dua kelompok, yaitu menggunakan teknik memboboti kembali secara iteratif dan penggunaan matriks peragam kekar (Turkmen 2008). Salah satu metode yang termasuk kelompok pertama adalah Regresi-M Kekar Parsial (*Partial Robust Regression*-M/PRM) yang dikenalkan oleh Serneels *et al*. (2005) dan salah satu metode yang termasuk kelompok kedua adalah SIMPLS kekar (*Robust* SIMPLS/ RSIMPLS) yang dikenalkan oleh Huber dan Branden (2003).

Dalam regresi kekar terdapat beberapa metode pendugaan parameternya, diantaranya adalah penduga-M (*M-estimators*). Penduga-M merupakan generalisasi dari metode kemungkinan maksimum, yaitu penduga yang meminimumkan fungsi objektif tertentu dalam data (Huber 1981). Fungsi pembobot dari penduga-M telah banyak tersedia, diantaranya adalah fungsi pembobot Huber dan fungsi pembobot Fair.

Ismah (2010) telah melakukan kajian mengenai RSIMPLS dan RSIMPLS-M. RSIMPLS-M merupakan pendekatan PLS kekar dengan memodifikasi pembobot RSIMPLS berdasarkan penduga-M dengan fungsi pembobot Huber. Hasil analisis yang diperoleh oleh Ismah (2010) menunjukkan bahwa RSIMPLS-M lebih baik daripada RSIMPLS untuk satu peubah tak bebas dan jumlah pencilan yang dideteksi RSIMPLS-M lebih kecil atau sama dengan jumlah pencilan yang dideteksi RSIMPLS. Metode PRM yang dikaji oleh Serneels *et al*. (2005) merupakan pendekatan PLS kekar dengan fungsi pembobot Fair. Bobot yang digunakan pada pendugaan PRM agar resisten terhadap pencilan sisaan dan pengamatan berpengaruh adalah dengan mengalikan bobot pencilan sisaan dan bobot untuk setiap pengamatan berpengaruh.

Dalam makalah ini akan dilakukan kajian mengenai metode RSIMPLS-M dan metode PRM untuk penanganan pencilan pada data kalibrasi. Fungsi pembobot yang digunakan pada metode PRM yaitu fungsi pembobot Fair dan Huber. Bobot yang dibutuhkan pada pendugaan PRM agar resisten terhadap pencilan sisaan dan pengamatan berpengaruh ditentukan dengan dua cara yakni, (1) mengalikan bobot pencilan sisaan dengan bobot untuk setiap pengamatan berpengaruh dan (2) mengambil nilai minimum dari bobot pencilan sisaan dan bobot untuk setiap pengamatan berpengaruh. Untuk menilai baik tidaknya hasil dugaan pada kedua metode (PRM dan RSIMPLS-M) akan dilakukan pendugaan selang kepercayaan parameter dengan menggunakan pendekatan nonparametrik yaitu dengan metode *Jackknife.*

LANDASAN TEORI

Regresi Kuadrat Terkecil Parsial

PLS adalah salah satu metode di kemometrik, dengan proses pembentukan model melalui struktur keragaman peubah bebas ($\boldsymbol{X}$) dan struktur keragaman peubah tak bebas ($\boldsymbol{Y}$) yang dilakukan secara iterasi. Pemodelan dilakukan tanpa asumsi sebaran (Wigena dan Aunuddin, 1997).

Salah satu algoritma pendugaan PLS adalah SIMPLS. Algoritma SIMPLS ini dikemukakan oleh De Jong pada tahun 1993 (Norliza 2006). Algoritma ini didasarkan pada matriks peragam empiris antara peubah tak bebas dan peubah bebas pada regresi linier.

Metode SIMPLS mengasumsikan peubah-peubah $\boldsymbol{X}$ dan $\boldsymbol{Y}$ dihubungkan dalam model bilinier seperti berikut ini :

$$\boldsymbol{x}_i = \bar{x} + \mathbf{P}\mathbf{t}_i + \mathrm{g}_i \qquad (1)$$

$$\boldsymbol{y}_i = \bar{y} + \mathbf{A}'\mathbf{t}_i + \mathrm{f}_i \qquad (2)$$

$\bar{x}$ dan $\bar{y}$ merupakan rata-rata dari peubah $\boldsymbol{X}$ dan $\boldsymbol{Y}$. $\mathbf{t}_i$ adalah skor berdimensi *k*, dengan $k < p$ dan $i = 1,\ldots,n$. $\mathbf{P}$ adalah matriks loading $\boldsymbol{X}$ berdimensi *p*x*k*, sedangkan sisaan dalam model ini dinotasikan dengan g_i dan f_i. Matriks $\mathbf{A}$ direpresentasikan sebagai matriks koefisien regresi (2) berdimensi $q \times k$. Berdasarkan struktur model bilinier (1) dan (2), algoritma SIMPLS adalah sebagai berikut (Norliza 2006) :

1. Data terpusat peubah $\widetilde{\boldsymbol{X}}_{n,p}$ dan $\widetilde{\boldsymbol{Y}}_{n,q}$
$$\tilde{x}_i = x_i - \bar{x}$$
$$\tilde{y}_i = y_i - \bar{y}$$
2. Untuk setiap $a = 1,2,\ldots,k$ vektor bobot SIMPLS $\mathbf{r}_a$ dan $\mathbf{q}_a$, didefinisikan sebagai vektor yang memaksimumkan
$$\mathrm{cov}\big(\widetilde{\boldsymbol{Y}}_{n,q}\mathbf{q}_a, \widetilde{\boldsymbol{X}}_{n,p}\mathbf{r}_a\big) = \mathbf{q}_a' \frac{\widetilde{\boldsymbol{X}}'_{p,n}\widetilde{\boldsymbol{Y}}_{n,q}}{n-1}\mathbf{r}_a = \mathbf{q}_a'\boldsymbol{S}_{xy}\mathbf{r}_a,$$
$\boldsymbol{S}_{yx}' = \boldsymbol{S}_{xy} = \frac{\widetilde{\boldsymbol{X}}'_{p,n}\widetilde{\boldsymbol{Y}}_{n,q}}{n-1}$ adalah matriks peragam antara peubah $\boldsymbol{X}$ dan $\boldsymbol{Y}$, dengan normalisasi $\mathbf{r}_a$ dan $\mathbf{q}_a$ $(\|\mathbf{r}_a\| = \|\mathbf{q}_a\| = 1)$ terdapat restriksi bahwa komponen $\widetilde{\mathbf{X}}\mathbf{r}_a$ tidak berkorelasi (ortogonal) agar diperoleh solusi lebih dari satu dan menghindari multikolinieritas antara peubah-peubah bebas.
3. Hitung skor SIMPLS:
$$\widetilde{\boldsymbol{T}}_{n,k} = (\mathbf{t}_1, \cdots, \mathbf{t}_n)' = \widetilde{\boldsymbol{X}}_{n,p}\boldsymbol{R}_{p,k}$$
dengan $\boldsymbol{R}_{p,k} = (\mathbf{r}_1, \cdots, \mathbf{r}_k)$
skor pertama SIMPLS yaitu : $\mathbf{t}_1 = \tilde{\boldsymbol{x}}_1'\mathbf{r}_1$
4. Periksa restriksi:
$$\mathbf{r}_j'\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}\mathbf{r}_a = \sum_{i=1}^{n} \tilde{t}_{ij}\tilde{t}_{ia} = 0, \quad a > j$$
komponen $\widetilde{\boldsymbol{X}}\mathbf{r}_j$ ortogonal agar diperoleh solusi lebih dari satu dan menghindari multikolinieritas antara peubah-peubah bebas.

5. Hitung loading $\boldsymbol{X}$ yaitu $\mathbf{p}_j$ yang menggambarkan hubungan linier antara peubah $\boldsymbol{X}$ dan komponen $\widetilde{\boldsymbol{X}}\mathbf{r}_j$ ke-j.
$$\mathbf{p}_j = \left(\mathbf{r}_j'\boldsymbol{S}_x\mathbf{r}_j\right)^{-1}\boldsymbol{S}_x\mathbf{r}_j$$
$\boldsymbol{S}_X$ merupakan matriks ragam peragam dari peubah $\boldsymbol{X}$ dan $j = 1,\ldots k$.
6. Langkah 5 terpenuhi jika $\mathbf{p}_j'\mathbf{r}_a = 0$ untuk $a > j$.
7. Hitung sebuah basis ortonormal $\{\mathbf{v}_1,\ldots,\mathbf{v}_{a-1}\}$ terhadap loading-x $\{\mathbf{p}_1,\ldots,\mathbf{p}_{a-1}\}$ untuk $2 \le a \le k$
Basis $\mathbf{v}_1 = \mathbf{p}_1$
$$\mathbf{v}_i = \mathbf{p}_i - \frac{\mathbf{v}_1'\mathbf{p}_i}{\mathbf{v}_1'\mathbf{v}_1}\mathbf{v}_1 - \cdots - \frac{\mathbf{v}_{i-1}'\mathbf{p}_i}{\mathbf{v}_{i-1}'\mathbf{v}_{i-1}}\mathbf{v}_{i-1}$$
8. Hitung matriks peragam silang $\boldsymbol{S}_{XY}^{(a)}$
$\boldsymbol{S}_{xy}^{(a)} = \boldsymbol{S}_{xy}^{(a-1)} - \mathbf{v}_a\left(\mathbf{v}_a'\boldsymbol{S}_{xy}^{(a-1)}\right)$ dan $\boldsymbol{S}_{xy}^{(1)} = \boldsymbol{S}_{xy}$, untuk $2 \le a \le k$
9. Tentukan vektor bobot SIMPLS $\mathbf{r}_a$ dan $\mathbf{q}_a$, untuk $2 \le a \le k$
Vektor-vektor bobot SIMPLS yang pertama, $\mathbf{q}_1$ adalah vektor ciri dari $\boldsymbol{S}_{yx}\boldsymbol{S}_{xy}$ dan $\mathbf{r}_1$ adalah vektor ciri dari $\boldsymbol{S}_{xy}\boldsymbol{S}_{yx}$. Sedangkan sepasang vektor bobot SIMPLS $(\mathbf{r}_a,\mathbf{q}_a)$ dengan $2 \le a \le k$ adalah vektor ciri $\mathrm{S}_{yx}^{(a)}\mathrm{S}_{xy}^{(a)}$ dan $\mathrm{S}_{xy}^{a}\mathrm{S}_{yx}^{a}$.
10. Hitung skor SIMPLS berikutnya untuk $2 \le a \le k$
$$\boldsymbol{T}_a = \widetilde{\boldsymbol{X}}_{n,p}\mathbf{r}_a$$
11. Ulangi langkah 3 untuk $2 \le a \le k$
12. Regresikan skor SIMPLS dengan peubah tak bebas. Model regresi secara matematis diberikan seperti berikut :
$$\boldsymbol{y}_i = \boldsymbol{\alpha}_i + \mathbf{A}_{q,k}'\mathbf{t}_i + \mathrm{f}_i$$
13. Hitung pendugaan algoritma SIMPLS
$$\widehat{\boldsymbol{A}}_{k,q} = (\boldsymbol{S}_t)^{-1}\boldsymbol{S}_{ty} = \left(\boldsymbol{R}_{k,p}'\boldsymbol{S}_x\boldsymbol{R}_{p,k}\right)^{-1}\boldsymbol{R}_{k,p}'\boldsymbol{S}_{xy}$$
$$\widehat{\boldsymbol{\alpha}}_0 = \overline{\boldsymbol{y}}$$
dimana $\boldsymbol{S}_y$ dan $\boldsymbol{S}_t$ merupakan matriks peragam peubah y dan t
14. Hitung koefisien regresi SIMPLS terhadap peubah asli (penduga paremeter untuk regresi linier $\boldsymbol{y}_i = \boldsymbol{\beta}_0 + {}_q\boldsymbol{B}'_p\,\boldsymbol{x}_i + \mathrm{e}_i$)
$$\widehat{\boldsymbol{B}}_{ktp} = \boldsymbol{R}_{p,k}\widehat{\boldsymbol{A}}_{k,q}$$
$$\widehat{\boldsymbol{\beta}}_0 = \overline{\boldsymbol{y}} - \widehat{\boldsymbol{B}}_{ktp}'\overline{\boldsymbol{x}}$$

Penentuan jumlah komponen k (peubah laten) menggunakan kriteria PRESS (*Prediction Sum of Squares*). Jumlah komponen ditentukan dari komponen k yang memiliki nilai PRESS minimum (Geladi dan Kowalski 1986).

Regresi Penduga-M

Model regresi linier berganda yang melibatkan p peubah bebas adalah
$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad (3)$$
dengan asumsi bahwa galat $\boldsymbol{\varepsilon}$ menyebar normal dengan $\mathrm{E}(\boldsymbol{\varepsilon}) = 0$ dan $\mathrm{Var}(\boldsymbol{\varepsilon}) = \mathbf{I}\sigma^2$ untuk $i = 1, 2, \ldots, n$. Salah satu metode yang digunakan untuk menduga parameter regresi dalam analisis regresi berganda adalah Metode Kuadrat Terkecil (*Ordinary Least* Square/OLS). Konsep dasar dari OLS adalah menduga parameter regresi dengan meminimumkan kuadrat sisaan
$$\hat{\beta}_{KT} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta})^2 \qquad (4)$$
Akan tetapi, jika galat menyebar tidak normal, sekalipun sebarannya mirip dengan normal namun memiliki ekor lebih panjang, maka OLS tidak tepat digunakan untuk menduga parameter regresi. Masalah ini, dapat diatasi dengan menggunakan metode pendugaan yang bersifat kekar yaitu penduga-M. Penduga-M diperoleh dengan mengganti fungsi kuadrat dalam (4) dengan fungsi kerugian (*loss function*) ρ sebagai berikut :
$$\hat{\beta}_{M} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \rho(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta}) \qquad (5)$$
dengan fungsi ρ simetrik dan merupakan fungsi tidak turun. Dengan mengambil $\rho(u) = u^2$, untuk u sembarang fungsi, maka kriteria meminimumkan akan sama dengan persamaan (4), sehingga penduga OLS tampak sebagai kasus khusus. Untuk mengurangi pengaruh sisaan yang besar dipilih fungsi ρ tertentu sehingga menghasilkan penduga yang lebih kekar dari OLS. Misalkan $r_i = \boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta}$ merupakan sisaan model dan w_y adalah bobot dari sisaan yang didefinisikan sebagai berikut:
$$w_y = \frac{\rho(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta})}{(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta})^2} = \frac{\rho(r_i)}{(r_i)^2} \qquad (6)$$
Persamaan (5) dapat ditulis kembali sebagai
$$\hat{\beta}_{M} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} w_y(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta})^2 \qquad (7)$$
Pendugaan koefisien regresi dengan penduga-M dilakukan dengan metode pendugaan OLS dengan pembobot yang dilakukan secara iteratif (Lu 2004). Nilai w_y akan berubah pada tiap iterasinya sehingga diperoleh $\hat{\beta}_M$. Penduga-M hanya resisten terhadap pencilan sisaan. Oleh karena itu, agar resisten terhadap tipe pencilan lain yaitu pengamatan berpengaruh maka bobot pada persamaan (6) akan dikalikan dengan bobot pengamatan berpengaruh w_x (Serneels *et al*. 2005), yaitu :
$$\hat{\beta}_{M} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} w_y\, w_x(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta})^2 \qquad (8)$$
Dengan demikian penduga parameter pada persamaan (8) akan resisten terhadap pencilan sisaan dan pengamatan berpengaruh. Penduga ini selanjutnya disebut Penduga-M kekar.

Regresi-M Kekar Parsial

Jika terdapat masalah multikolinieritas pada kalibrasi ganda, maka model yang sesuai adalah model PLS (Serneels *et al*. 2005). Idenya adalah bahwa cukup meregresikan peubah bebas pada jumlah peubah laten k terbatas. Nilai peubah laten ini ditempatkan bersama dalam matriks skor $\boldsymbol{T}_{n,k}$, yang mempunyai vektor $\mathbf{t}_i$ sebagai kolom, dengan $1 \le i \le n$. Model regresi laten diberikan sebagai berikut:

$$\boldsymbol{y}_i = \mathbf{A}\mathbf{t}_i + \boldsymbol{\varepsilon}_i \quad (9)$$

Karena dimensi **A** rendah, yakni *k*, vektor **A** dapat diduga sama dengan sebelumnya yaitu dengan meregresikan peubah bebas terhadap peubah laten dengan bantuan penduga-M kekar. Perbedaan utamanya adalah bahwa bobot w_y dihitung dari sisaan, yaitu $\boldsymbol{r}_i = \boldsymbol{y}_i - \mathbf{A}\mathbf{t}_i$ dan bobot w_x untuk pengamatan berpengaruh akan dihitung dari skor $\mathbf{t}_i$, sebagai ganti dari peubah bebas asli. Pembobot yang dibutuhkan agar resisten terhadap pencilan sisaan dan pengamatan berpengaruh, adalah :

$w_i = w_y * w_x$ atau $w_i = \min(w_y, w_x)$

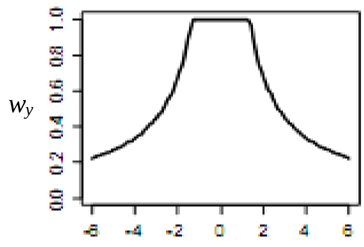
dan menghasilkan penduga yang disebut dengan Penduga-M Kekar Parsial. Selanjutnya adalah untuk memperoleh matriks $\boldsymbol{T}_{n,k}$ yang tidak dapat diamati secara langsung. Matriks $\boldsymbol{T}_{n,k}$ diperoleh dengan menggunakan algoritma SIMPLS. Sewaktu $\widehat{\boldsymbol{A}}$ dihasilkan, maka akhir pendugaan bagi $\boldsymbol{\beta}$ adalah $\widehat{\boldsymbol{\beta}} = \boldsymbol{R}\widehat{\boldsymbol{A}}$.

PLS tampak sebagai kasus khusus jika semua bobot w_i yang diambil sama, sehingga menghasilkan penduga tak kekar. Dengan menganggap bahwa bobot telah ditetapkan, maka tidak akan sulit untuk mendapatkan $\widehat{\boldsymbol{A}}$ yang merupakan penduga PLS yang dihitung dari amatan berbobot $\left(\sqrt{w_i}\boldsymbol{x}_i, \sqrt{w_i}\boldsymbol{y}_i\right)$.

Beberapa fungsi pembobot yang dapat digunakan, antara lain :

1. $w_y = f\left(\frac{r_i}{\hat{\sigma}}, c\right) = f(u, c)$

$$= \begin{cases} 1, & \text{untuk } |u| \leq c \\ \dfrac{c}{|u|}, & \text{untuk } |u| > c \end{cases} \quad (10)$$

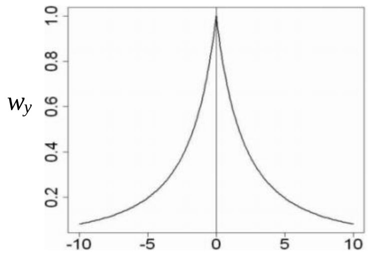


Gambar 1. Fungsi Pembobot Penduga-M Huber

fungsi pembobot *f* disebut fungsi pembobot Huber dan konstanta *c* disebut *tuning constant* dan *u* adalah pengukur simpangan relatif dari Huber.

2. $w_y = f\left(\frac{r_i}{\hat{\sigma}}, c\right) = f(u, c) = \frac{1}{\left(1 + \frac{|u|}{c}\right)^2} \quad (11)$

fungsi pembobot *f* disebut fungsi pembobot Fair, $\hat{\sigma}$ adalah penduga skala sisaan, *c* disebut *tuning constant* dan *u* adalah pengukur simpangan relatif dari Fair.



Gambar 2. Fungsi Pembobot Penduga-M Fair

Cummins dan Andrews (1995) telah melakukan kajian bahwa penduga-M *Fair* akan efektif digunakan ketika *c* = 4, penduga-M Huber akan efektif digunakan ketika *c* = 1.345 (Kuzmic *et al.* 2004) dan sebagai pembanding akan digunakan *c* = 2. Pengukur simpangan relatif *u* pada persamaan (10) dan (11) dihitung dari standar sisaan yaitu sisaan dibagi dengan penduga skala agar prosedur regresi mempunyai skala ragam yang sama (*scale equivariant*). Salah satu penduga skala $\hat{\sigma}$ yang paling kekar dan sederhana adalah MAD (*Median Absoblute Deviation*) yang diperkenalkan oleh Hampel pada tahun 1974. MAD dapat dihitung sebagai berikut :

$$\hat{\sigma} = \underset{i}{\text{median}} \left| r_i - \underset{j}{\text{median}}\, r_j \right| \quad (12)$$

Bobot w_x setelah dilakukan penskalaan setiap vektor skor $\mathbf{t}_i$ dihitung sebagai berikut :

$$w_x = f\left(\frac{\|t_i - \text{med}_{L1}(\text{T})\|}{\text{median}_i \|t_i - \text{med}_{L1}(\text{T})\|}, c\right) \quad (13)$$

dimana $\|\cdot\|$ merupakan norm vektor dan pembobot *f* sama seperti persamaan (10) dan (11). $\text{med}_{L1}(\text{T})$ merupakan median-L_1 yang dihitung dari vektor skor $\{t_1,\ldots,t_n\}$, yaitu penduga kekar dari pusat data awan dari vektor skor berdimensi-*h*. Median-L_1 merupakan median contoh berganda, juga disebut spatial median.

RSIMPLS-M

Kajian mengenai RSIMPLS-M terdiri dari Analisis Komponen Utama Kekar (*Robust Principal Component Analysis*/ROBPCA) dengan menggabungkan konsep *projection pursuit* (PP) dengan penduga ragam kekar, yaitu Determinan Peragam Minimum (*Minimum Covariance Determinant*, MCD) (Ismah 2010).

MCD merupakan penduga yang sangat kekar untuk menduga parameter nilai tengah dan matriks peragam dengan konsep menentukan subhimpunan yang memiliki nilai determinan peragam minimum. Dengan kata lain, MCD bertujuan untuk mendapatkan *h* pengamatan dari *n* pengamatan yang memiliki determinan peragam terkecil. Metode PP bertujuan untuk mendapatkan struktur data peubah ganda dengan memproyeksikan pada subhimpunan berdimensi rendah (Huber 1985). PP tepat digunakan untuk menganalisis data dengan jumlah peubah yang besar. Subhimpunan berdimensi rendah dipilih dengan memaksimumkan indeks proyeksi tertentu.

Langkah-langkah dalam metode RSIMPLS-M adalah sebagai berikut:

1. Pembentukan skor kekar RSIMPLS-M
   Pembentukan skor-skor kekar, $\tilde{\mathbf{t}}_i$ berdimensi *k*, berdasarkan penduga pusat kekar dan ragamnya yang diperoleh menggunakan metode ROBPCA. Vektor bobot RSIMPLS-M, $\mathbf{r}_a$ dan $\mathbf{q}_a$ diperoleh menggunakan metode SIMPLS, tetapi matriks peragam $\boldsymbol{S}$ diganti dengan $\hat{\Sigma}$. Sedangkan vektor loading $\boldsymbol{X}$ didefinisikan $\mathbf{p}_j = \left(\mathbf{r}_j'\hat{\Sigma}_x\mathbf{r}_j\right)^{-1}\hat{\Sigma}_x\mathbf{r}_j$, kemudian $\hat{\Sigma}_{xy}^{(a)}$ diperoleh serupa pada tahap SIMPLS. Dan pada masing-masing tahap skor kekar dihitung :
   $\tilde{\mathbf{t}}_{ia} = \breve{\boldsymbol{x}}_i'\mathbf{r}_a = \left(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_x\right)'\mathbf{r}_a$.
2. Pendugaan model regresi
   Sama seperti pada SIMPLS dimana skor-skor kekar yang diperoleh pada langkah 1 diregresikan dengan peubah tak bebas. Model regresi secara matematis ditulis sebagai berikut:
   $$\boldsymbol{y}_i = \boldsymbol{\alpha}_0 + \mathbf{A}_{q,k}'\tilde{\mathbf{t}}_i + \check{\mathrm{f}}_i$$
   dimana penduga pusat $\boldsymbol{\mu}$ dan peragam $\sum$ dari $(\tilde{\mathbf{t}}, \boldsymbol{y})$ yaitu rataan dan matriks peragam terboboti

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix}\hat{\boldsymbol{\mu}}_{\tilde{\mathbf{t}}}\\ \hat{\boldsymbol{\mu}}_{\boldsymbol{y}}\end{pmatrix} = \frac{\sum_{i=1}^{n} w_i \begin{pmatrix}\tilde{\mathbf{t}}\\ \boldsymbol{y}\end{pmatrix}}{\sum_{i=1}^{n} w_i} \quad \text{dan}$$

$$\hat{\Sigma} = \begin{pmatrix}\hat{\Sigma}_{\tilde{\mathbf{t}}} & \hat{\Sigma}_{\tilde{\mathbf{t}}\boldsymbol{y}}\\ \hat{\Sigma}_{\boldsymbol{y}\tilde{\mathbf{t}}} & \hat{\Sigma}_{\boldsymbol{y}}\end{pmatrix} = \frac{\sum_{i=1}^{n} w_i \begin{pmatrix}\tilde{\mathbf{t}}\\ \boldsymbol{y}\end{pmatrix}\left(\tilde{\mathbf{t}}_i'\ \boldsymbol{y}_i'\right)}{\sum_{i=1}^{n} w_i - 1}$$

dengan w_i = 1 apabila pengamatan ke-*i* tidak diidentifikasi sebagai pencilan dengan metode ROBPCA dalam $(\boldsymbol{x},\boldsymbol{y})$ dan $w_i \approx 0$ untuk lainnya.

Fungsi pembobot pada metode RSIMPLS-M adalah sebagai berikut:

a. $$w_i = \begin{cases} 1 & \text{jika } \mathbf{D}_i \le \sqrt{\chi^2_{k,0.975}} \\ \dfrac{\sqrt{\chi^2_{k,0.975}}}{\mathbf{D}_i} & \text{jika } \mathbf{D}_i > \sqrt{\chi^2_{k,0.975}} \end{cases}$$

dimana $\mathbf{D}_{i(k)} = \sqrt{\left(\tilde{\mathbf{t}}_i^{(\boldsymbol{z})}\right)'(\mathbf{L}^{(\boldsymbol{z})})^{-1}\tilde{\mathbf{t}}_i^{(\boldsymbol{z})}}$

b. $$w_i = \begin{cases} 1 & \text{jika } \mathbf{OD}_i \le \sqrt{\hat{\boldsymbol{\mu}}_{od^2} + \hat{\boldsymbol{\sigma}}_{od^2}\, z_{0.975}} \\ \dfrac{\sqrt{\hat{\boldsymbol{\mu}}_{od^2} + \hat{\boldsymbol{\sigma}}_{od^2}\, z_{0.975}}}{\mathbf{OD}_i} & \text{jika } \mathbf{OD}_i > \sqrt{\hat{\boldsymbol{\mu}}_{od^2} + \hat{\boldsymbol{\sigma}}_{od^2}\, z_{0.975}} \end{cases}$$

c. $$w_i(akhir) = \begin{cases} \mathbf{D}_i & \text{jika } \mathbf{D}_i \le \mathbf{OD}_i \\ \mathbf{OD}_i & \text{jika } \mathbf{D}_i > \mathbf{OD}_i \end{cases} \quad (14)$$

$\mathbf{D}_{i(k)} = \sqrt{\left(\tilde{\mathbf{t}}_i^{(\boldsymbol{z})}\right)'(\mathbf{L}^{(\boldsymbol{z})})^{-1}\tilde{\mathbf{t}}_i^{(\boldsymbol{z})}}$ (jarak kekar)

$\mathbf{OD}_i = \left\|(\boldsymbol{z}_i - \boldsymbol{\mu}_{\boldsymbol{z}}') - \mathbf{P}^{(\boldsymbol{z})}\tilde{\mathbf{t}}_i^{(\boldsymbol{z})}\right\|$ (jarak ortogonal)

dengan vektor ciri $\boldsymbol{Z}$ yaitu $\mathbf{P}_{n,k}^{(\boldsymbol{z})}$ dan akar ciri $\boldsymbol{Z}$ yaitu $\mathbf{diag}(\mathbf{L}_{k,k})$.

$\hat{\boldsymbol{\mu}}_{od^2}$ dan $\hat{\boldsymbol{\sigma}}_{od^2}$ adalah penduga rataan dan ragam dengan MCD (Ismah 2010). **S**etelah $\hat{\boldsymbol{\mu}}$ dan $\hat{\Sigma}$ diperoleh, proses selanjutnya penduga koefisien regresi diperoleh menggunakan metode kuadrat terkecil. Penentuan jumlah komponen *k* menggunakan kriteria RMSECV (*Root Mean Squared Error Cross Validation*). Jumlah komponen ditentukan dari komponen *k* yang memiliki nilai RMSECV minimum.

Pendugaan Parameter Dengan *Jackknife*

Pendekatan *jackknife* diperkenalkan oleh Maurice Henry Quenouille pada tahun 1949 untuk mengoreksi bias suatu penduga. Kemudian pada tahun 1958, John Wilder Tukey mengusulkan penduga ragam. Ide dasar metode *jackknife* dengan menghapus pengamatan ke-*i* untuk *i* = 1,..,*n* dan melakukan pendugaan parameter. Misalkan $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ merupakan contoh acak berukuran *n* pengamatan yang menyebar secara bebas stokastik dan identik dari suatu sebaran peluang F yang tidak diketahui. Dari contoh tersebut dilakukan penarikan contoh kembali (*resample*) sebanyak *n* kali dimana setiap *resample* terdiri dari *n*-1 pengamatan (terhapus 1 pengamatan secara berturut-turut).

Misalkan penaksir $\boldsymbol{\beta}$ yang diperoleh dengan menyisihkan data ke-*i* dan diperoleh penduga $\underline{\mathbf{b}}(i)$, yang disebut statistik *jackknife*, untuk *i* = 1,..,*n* (Dudewicz dan Mishra 1988). Penduga bias *Jackknife* dihitung dengan menggunakan persamaan berikut :

$$\text{Bias}_{jack} = (n-1)\left(\underline{\boldsymbol{b}}(i) - \underline{\boldsymbol{b}}\right) \quad \text{dan}$$

$$\text{Rataan Bias}_{jack} = \frac{(n-1)\left(\underline{\mathbf{b}}(i) - \underline{\mathbf{b}}\right)}{n} \quad (16)$$

$\underline{\mathbf{b}}$ adalah dugaan dari parameter $\boldsymbol{\beta}$ dan $\underline{\mathbf{b}}(i)$ adalah dugaan dari parameter $\boldsymbol{\beta}$ yang dihitung dari penghapusan pengamatan ke-*i* (statistik *jackknife*). Selanjutnya penduga bias *Jackknife* ini digunakan untuk menghasilkan bias terkoreksi penduga J*ackknife* ($\mathbf{b}_j$)yang disebut sebagai penduga *Pseudo*, didefinisikan dengan :

$$\begin{aligned}\underline{\tilde{\mathbf{b}}} &= \underline{\mathbf{b}} - \text{Bias}_{Jack}\\ &= \underline{\mathbf{b}} - (n-1)\left(\underline{\mathbf{b}}(i) - \underline{\mathbf{b}}\right)\\ &= n\underline{\mathbf{b}} - (n-1)\underline{\mathbf{b}}(i) \qquad (17)\end{aligned}$$

sedangkan penduga J*ackknife* ($\underline{\mathbf{b}}_J$) didefinisikan sebagai :

$$\underline{\mathbf{b}}_J = n\underline{\mathbf{b}} - (n-1)\underline{\mathbf{b}}(\cdot)$$

dengan $\underline{\mathbf{b}}(.) = \frac{1}{n}\sum_{i=1}^{n}\underline{\mathbf{b}}(i) \quad (18)$

Penduga ragam *Jackknife* dengan mengapus pengamatan ke-*i* berdasarkan pada nilai *pseudo*, didefinisikan sebagai berikut :

$$\begin{aligned}\underline{\sigma}_{J(i)}^2 &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n-1}\left(\underline{\tilde{\mathbf{b}}}_{(i)} - \frac{1}{n}\sum_{m=1}^{n}\underline{\tilde{\mathbf{b}}}_{(m)}\right)^2\\ &= \frac{n-1}{n}\sum_{i=1}^{n}\left(\underline{\mathbf{b}}_{J(i)} - \underline{\mathbf{b}}(\cdot)\right)^2 \qquad (19)\end{aligned}$$

Selang kepercayaan untuk (1 - α)% bagi koefisien regresi adalah :

$$\left[\underline{\mathbf{b}}_{J(i)} - \mathrm{t}_{\frac{\alpha}{2}(n-1)}\underline{\sigma}_{J(i)}\ ;\ \underline{\mathbf{b}}_{J(i)} + \mathrm{t}_{\frac{\alpha}{2}(n-1)}\underline{\sigma}_{J(i)}\right] \quad (20)$$

dimana $\mathrm{t}_{\frac{\alpha}{2}(n-1)}$ adalah sebaran t dengan derajat bebas *n* – 1.

METODOLOGI

Data

Data yang digunakan dalam penelitian ini yaitu data sekunder yang diperoleh dari Journal Technometrics (Naes 1985), mengenai konsentrasi lemak ikan. Data ini berisi hasil pengukuran konsentrasi lemak 45 contoh ikan dan absorbsinya dari 9 panjang gelombang yang diukur dengan spektrometer NIR. Konsentrasi lemak ikan (%) sebagai peubah respon dan absorban-absorbannya sebagai peubah penjelas. Menurut sumber data, tujuh pengamatan ektrim terletak pada bagian akhir, yakni dari pengamatan ke-39 sampai 45.

Metode Analisis

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

I. Analisis menggunakan seluruh data pengamatan
   1. Algoritma Regresi-M Kekar Parsial (PRM) dengan fungsi pembobot Fair dan Huber
   2. Algoritma *Robust Straightforward Implementation Partial Least Square-M* (RSIMPSL-M).

II. Validasi Model

   Salah satu jenis validasi model yaitu dengan menghitung nilai *Root Mean Squared Error of Prediction* (RMSEP) dengan rumus:

$$\text{RMSEP} = \sqrt{\frac{1}{n_p}\sum_{i=1}^{n_p}(\hat{\boldsymbol{y}}_i - \boldsymbol{y}_i)^2}$$

   $\boldsymbol{y}_i$ menyatakan nilai pengamatan ke-i pada kelompok data validasi, $\hat{\boldsymbol{y}}_i$ menyatakan nilai dugaan pengamatan ke-i dan n_p menyatakan banyak sampel yang digunakan dalam model validasi.

III. Pendugaan selang kepercayaan parameter dengan metode *jackknife.*

1. Algoritma Regresi-M Kekar Parsial (PRM) dengan fungsi pembobot Fair dan Huber adalah sebagai berikut :

a. Tentukan data terpusat peubah $\widetilde{\boldsymbol{X}}_{n,p}$ dan $\widetilde{\boldsymbol{Y}}_{n,q}$
b. Tentukan bobot kekar awal, yaitu $w_i = w_y{*}w_x$ dan $w_i = \min(w_y, w_x)$ dengan menggunakan penduga-M Huber dan penduga-M Fair.
c. Lakukan analisis regresi dengan algortitma SIMPLS pada matriks data $\widetilde{\boldsymbol{X}}_{n,p}$ dan $\widetilde{\boldsymbol{Y}}_{n,q}$ terboboti yang diperoleh dengan mengalikan setiap kolom $\widetilde{\boldsymbol{X}}_{n,p}$ dan $\widetilde{\boldsymbol{Y}}_{n,q}$ dengan $\sqrt{w_i}$.

   c1. Tentukan vektor bobot $\boldsymbol{X}$ ($\mathbf{r}_a$) dan $\boldsymbol{Y}$ ($\mathbf{q}_a$), untuk a = 1.
   c2. Hitung skor SIMPLS ($\mathbf{t}_a$): $\mathbf{t}_1 = \tilde{\boldsymbol{x}}_1'\mathbf{r}_1$, untuk untuk a = 1
   c3. Hitung loading-$\boldsymbol{X}$, $\mathbf{p}_j = \left(\mathbf{r}_j'\boldsymbol{S}_x\mathbf{r}_j\right)^{-1}\boldsymbol{S}_x\mathbf{r}_j$
   c4. Menghitung sebuah basis ortonormal $\{\mathbf{v}_1,\ldots,\mathbf{v}_{a-1}\}$ terhadap loading- $\boldsymbol{X}$ $\{\mathbf{p}_1,\ldots,\mathbf{p}_{a-1}\}$ untuk $2 \le a \le k$
   c5. Hitung matriks peragam silang $\boldsymbol{S}_{XY}^{(1)} = \boldsymbol{S}_{XY}$ dan $\boldsymbol{S}_{XY}^{(a)} = \boldsymbol{S}_{XY}^{(a-1)} - \mathbf{v}_a\left(\mathbf{v}_a'\boldsymbol{S}_{XY}^{(a-1)}\right)$, untuk $2 \le a \le k$
   c6. Hitung vektor bobot $\boldsymbol{X}$ ($\mathbf{r}_a$) dan $\boldsymbol{Y}$ ($\mathbf{q}_a$), untuk $2 \le a \le k$
   c7. Hitung skor SIMPLS ($\mathbf{t}_a$), untuk $2 \le a \le k$
   c8. Ulangi tahap c3.

d. Periksa skor SIMPLS ($\mathbf{t}_a$) dengan membagi ($\mathbf{t}_a$) dengan $\sqrt{w_i}$.
e. Regresikan skor SIMPLS dengan peubah respon (y) : $\boldsymbol{y}_i = \mathbf{A}\mathbf{t}_i + \boldsymbol{\varepsilon}_i$
f. Hitung $\widehat{\boldsymbol{A}}_{k,q} = (\boldsymbol{S}_t)^{-1}\boldsymbol{S}_{ty}$
g. Uji kekonvergenan $\widehat{\boldsymbol{A}}_{k,q}$
h. Jika kekonvergenan tidak tercapai, hitung kembali bobot $w_i = w_y{*}w_t$ dan $w_i = \min(w_y, w_t)$ dengan menggunakan penduga-M Huber dan penduga-M Fair, dimana $\boldsymbol{r}_i = \boldsymbol{y}_i - \mathbf{A}\mathbf{t}_i$
i. Kembali ke langkah 3 sampai $\widehat{\boldsymbol{A}}$ konvergen.
j. Pendugaan akhir dari $\widehat{\boldsymbol{A}}$ secara langsung diperoleh dari langkah RKTP terboboti terakhir.
k. Hitung koefisien regresi SIMPLS terhadap peubah asli.
l. Menghitung PRESS (*Prediction Sum of Squares*).

2. Algoritma RSIMPSL-M

a. Tentukan data terpusat peubah $\widetilde{\boldsymbol{X}}_{n,p}$ dan $\widetilde{\boldsymbol{Y}}_{n,q}$
b. Pembentukan skor kekar

   b1. Tentukan matriks peragam kekar menggunakan metode ROBPCA
   b2. Tentukanvektor bobot kekar $\boldsymbol{X}$ ($\mathbf{r}_a$) dan $\boldsymbol{Y}$ ($\mathbf{q}_a$), untuk a = 1, menggunakan metode SIMPLS, tetapi matriks peragam $\boldsymbol{S}$ diganti dengan $\widehat{\Sigma}$.
   b3. Menghitung skor SIMPLS ($\mathbf{t}_a$): $\mathbf{t}_1 = \tilde{\boldsymbol{x}}_1'\mathbf{r}_1 = \left(\boldsymbol{x}_1 - \hat{\boldsymbol{\mu}}_x\right)'\mathbf{r}_1$ untuk a = 1
   b4. Hitung loading-$\boldsymbol{X}$, $\mathbf{p}_j = \left(\mathbf{r}_j'\widehat{\Sigma}_x\mathbf{r}_j\right)^{-1}\widehat{\Sigma}_x\mathbf{r}_j$
   b5. Menghitung sebuah basis ortonormal $\{\mathbf{v}_1,\ldots,\mathbf{v}_{a-1}\}$ terhadap loading-x $\{\mathbf{p}_1,\ldots,\mathbf{p}_{a-1}\}$ untuk $2 \le a \le k$
   b6. Menghitung matriks peragam silang $\widehat{\Sigma}_{XY}^{(1)} = \widehat{\Sigma}_{XY}$ dan $\widehat{\Sigma}_{XY}^{(a)} = \widehat{\Sigma}_{XY}^{(a-1)} - \mathbf{v}_a\left(\mathbf{v}_a'\widehat{\Sigma}_{XY}^{(a-1)}\right)$, untuk $2 \le a \le k$
   b7. Hitung skor RSIMPLS-M untuk $2 \le a \le k$
   b8. Tentukan vektor bobot kekar $\boldsymbol{X}$ ($\mathbf{r}_a$) dan $\boldsymbol{Y}$ ($\mathbf{q}_a$), untuk $2 \le a \le k$
   b9. Ulangi tahap b3 untuk $2 \le a \le k$

m. Pembentukan regresi kekar : hitung $\widehat{\boldsymbol{A}}_{k,q} = \left(\widehat{\sum}_{\tilde{t}}\right)^{-1}\widehat{\sum}_{\tilde{t}y}$
n. Hitung koefisien regresi RSIMPLS-M terhadap peubah asli.
o. Menghitung RMSECV (*Root Mean Squared Error Cross Validation*).

HASIL DAN PEMBAHASAN

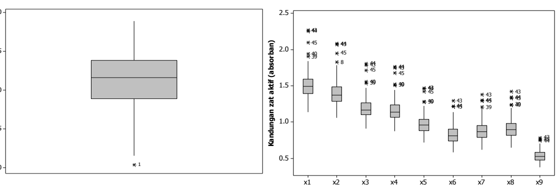
Deskripsi Data

Berdasarkan data yang digunakan dalam penelitian ini, akan dilakukan pengidentifikasian multikolinieritas. Nilai korelasi antar peubah penjelas yang cukup tinggi dengan nilai $p > 0.1$ tertera pada Tabel 1.

Tabel 1 Korelasi (r) antar peubah penjelas data ikan

r	X1	X2	X3	X4	X5	X6	X7	X8
X2	0.91							
X3	0.99	0.91						
X4	0.99	0.91	1.00					
X5	0.99	0.89	0.99	0.99				
X6	0.98	0.89	0.98	0.99	0.99			
X7	0.98	0.89	0.98	0.99	0.99	1.00		
X8	0.99	0.89	0.99	0.99	0.99	1.00	1.00	
X9	0.97	0.87	0.97	0.98	0.98	0.99	0.99	0.99

Adanya data pencilan sering kali memperbesar nilai ragam bagi model, sehingga menyebabkan dugaan bagi selang kepercayaannya makin lebar. Pengidentifikasian pencilan berdasarkan data pengamatan Y menunjukkan bahwa pencilan pada pengamatan ke-1. Sedangkan pengidentifikasian pencilan berdasarkan data pengamatan X menunjukkan terdapat beberapa pengamatan yang merupakan pencilan yaitu pengamatan ke-8, 39, 40, 42, 43, 44, dan 45. Deteksi pengamatan pencilan berdasarkan diagram kotak-garis ditunjukkan pada Gambar 3.



(a) (b)

Gambar 3 Diagram kotak-garis (a) Pengamatan Y dan (b) Pengamatan X

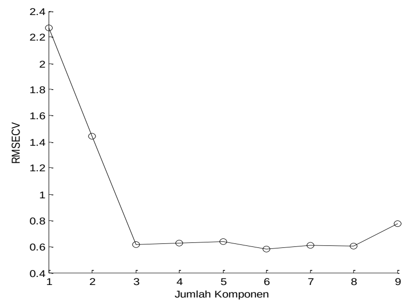
Penentuan Jumlah Peubah Laten

Jumlah peubah laten (komponen) *k* ditentukan berdasarkan analisis menggunakan metode PRM terhadap data tertera pada Tabel 2.

Tabel 2 Hasil Validasi Silang untuk menentukan banyaknya komponen *k*

Banyak Peubah Laten	PRESS	Peluang > PRESS	Pengaruh Model		Peubah Bebas	
			%	Kumulatif	%	Kumulatif
0	1.069	0.0001	-	-	-	-
1	0.807	0.0001	97.4293	97.4293	45.7099	45.7099
2	0.690	0.0470	1.6800	99.1092	20.8394	66.5493
3	0.528	0.1970	0.7861	99.8953	14.9791	81.5284
4	0.463	1.0000	0.0735	99.9688	7.5427	89.0711
5	0.558	0.2400	0.0254	99.9942	1.4707	90.5418
6	1.233	0.3240	0.0055	99.9997	0.2780	90.8198
7	2.185	0.0740	0.0001	99.9999	1.5838	92.4035
8	1.455	0.0280	0.0001	99.9999	0.2844	92.6880
9	1.414	0.0250	0.0001	100.000	0.0665	92.7545

Dari hasil perhitungan PRESS diperoleh model kalibrasi dengan 4 peubah laten (komponen) dengan akar rataan PRESS terkecil, yakni 0.4627. Model terkecil dengan nilai peluang > 0.1, yaitu model dengan 3 peubah laten akan dipilih. Model ini memiliki akar rataan PRESS 0.5281. Tiga komponen tersebut dapat menjelaskan keragaman peubah. Pengaruh model dijelaskan sebesar 99.89% dan untuk peubah bebas dijelaskan sebesar 81.53%. Hal ini menunjukkan kecukupan variasi dari model yang dapat dijelaskan dengan 3 komponen. Sedangkan jumlah peubah laten *k* ditentukan berdasarkan analisis menggunakan RSIMPLS-M terhadap data tampak pada Gambar 4. Dari hasil perhitungan nilai RMSECV diperoleh model kalibrasi dengan 6 komponen dengan nilai RMSECV minimum, yakni 0.5777 dan $R^2 = 0.934$. Selanjutnya, untuk melakukan analisis data akan digunakan jumlah peubah laten (komponen) sebanyak $k = 3$ dan $k = 6$.



Gambar 4 Nilai RMSECV pada beberapa jumlah komponen

Identifikasi Pencilan

Setelah dilakukan pengindetifikasian pencilan pada 45 contoh ikan dengan metode RMKTP dan RSIMPLS-M, dilakukan juga pendugaan model. Hasil identifikasi pencilan pada 45 contoh ikan berdasarkan bobot dan pendugaan model metode prm dan RSIMPLS-M terdapat pada Tabel 3.

Pada Tabel 3 terlihat bahwa jumlah pencilan yang dideteksi oleh RSIMPLS-M lebih kecil daripada jumlah pencilan yang dideteksi oleh PRM secara umum. Hal ini disebabkan karena RSIMPLS-M merupakan metode pendugaan parameter yang didasarkan pada Analisis Komponen Utama Kekar (*Robust Principal*

Component Analysis/ROBPCA). ROBPCA melakukan perhitungan matriks peragam tidak dari semua data, tetapi dari *h* pengamatan dengan nilai keterpencilan terkecil. Sedangkan PRM merupakan metode pendugaan parameter yang menggunakan metode *Iterative Reweighted Partial Least Squares* (IRPLS), Setiap iterasi metode PRM melibatkan keragaman baik pada peubah tak bebas maupun pada peubah bebas dari semua data.

Secara umum, berdasarkan sejajar sumbu vertikal (Y)dan sejajar sumbu horizontal (X) pada seluruh metode PRM untuk k = 3, pengamatan ke-39, 40, 41, 42, 43, 44, dan 45 yang dianggap pengamatan ekstrim menurut Naes (1985), terdeteksi sebagai pencilan. Demikian halnya untuk k = 6, pada seluruh metode PRM, pengamatan ke-39, 40, 41, 42, 43, 44, dan 45, terdeteksi sebagai pencilan. Selanjutnya, berdasarkan sejajar sumbu vertikal (Y) dan sejajar sumbu horizontal (X) pada metode RSIMPLS-M untuk k = 3, pengamatan ke-39, 40, 41, 42, 43, 44, dan 45 yang dianggap pengamatan ekstrim menurut Naes (1985), terdeteksi sebagai pencilan. Sedangkan untuk k = 6, pengamatan ke-39, 40, 41, 42, 43, 44, dan 45 yang dianggap pengamatan ekstrim menurut Naes (1985), hanya pengamatan ke-42 yang tidak terdeteksi sebagai pencilan.

Tabel 3 Perbandingan identifikasi pencilan dan pendugaan model data lemak ikan

Metode	k = 3			k = 6		
	Pengamatan Pencilan			Pengamatan Pencilan		
	Sejajar Sumbu Vertikal (Y)	Sejajar Sumbu Horizontal (X)	R^2 (%)	Sejajar Sumbu Vertikal (Y)	Sejajar Sumbu Horizontal (Y)	R^2 (%)
PRM-Fair (perkalian)	1,5,8,20, 25,32,41, 43,44, 45	2,8,34, 38,39, 40,41, 43,44, 45	83.3	1,2,4, 5,12,15, 20,25,32, 33,34,39, 41,43, 44,45	8,20,39, 41,42,43, 44,45	89.1
PRM-Huber=1.345 (perkalian)	1,4,5,7, 8,12,20, 25,26,32, 33,37,41, 43, 44, 45	1,2,3,8, 27,34, 36,38, 39,40, 41,43, 44,45	75.1	1,2,4,5, 12,15,19, 20,25,29, 33,34,39, 41,43,44, 45	1,8,20, 21,34, 39,40, 41,42, 43,44, 45	88.8
PRM-Huber=2 (perkalian)	1,5,8,12 ,20,25,32, 41,43, 44, 45	8,39,40, 41,43, 44,45	82.7	1,2,12, 15,25,32, 34,41,43, 44	8, 43, 44, 45	90.1
PRM-Fair (Minimum)	1,2,4,5, 7,8,12, 17,20,25, 26,27,32, 35,37,41, 44	1,2,4, 8,34,36, 38,39, 40,41, 43,44, 45	78.5	1,2,4,5,12 ,15,19,25, 32, 32,34,39, 41,43,44	8,20,39, 41,42, 43,44, 45	90.2
PRM-Huber=1.345 (minimum)	1,2,4,5, 7,8,12, 14,17,20, 23,25,26, 27,32,35, 37,41,44, 45	1,2,3,6, 8,22,33, 34,36, 38,39, 40,41,42, 43, 44,45	80.6	1,2,4,5,12 ,15,19,22, 25, 32,33,34, 37,39,41, 43,44	1,8,20, 21,34, 39,40, 41,42, 43,44, 45	90.2
PRM-Huber=2 (minimum)	1,2,4,8, 17,20,25, 26, 27,32, 35,37, 44,45	1,2,8, 36,39, 38,39, 40,41, 43,44, 45	81.4	1,2,12, 15,25,32, 34,41, 43,44	8,42,43, 44,45	90.4
RSIMPLS-M	1,20,39, 41,42, 44, 45	1,8,39, 40, 41, 43,44, 45	86.1	24,40,41, 43, 44,45	1,8,20, 39,40,41,4 3,44,45	90.8

Berdasarkan nilai koefisien determinasi (R^2), model dengan k = 3 memiliki koefisien determinasi yang lebih kecil bila dibandingkan dengan koefisien determinasi untuk k = 6 pada seluruh metode. Akan tetapi, nilai R^2 pada metode RSIMPLS-M lebih baik dibandingkan metode PRM. Selanjutnya untuk menilai baik tidaknya hasil dugaan akan dilakukan validasi model dan pendugaan ragam koefisien regresi dengan menggunakan pendekatan nonparametrik yaitu salah satunya dengan metode *Jackknife*.

Validasi Model

Pada Tabel 3 terlihat bahwa nilai R^2 yang dihasilkan oleh PRM dengan fungsi pembobot Fair (perkalian) dan Fair (minimum) untuk k = 3 dan k = 6 tidak jauh berbeda. Demikian halnya untuk PRM dengan fungsi pembobot Huber (perkalian) dan Huber (minimum) untuk k = 3 dan k = 6, nilai R^2 yang dihasilkan juga tidak jauh berbeda. Oleh karena itu, data dianalisis dengan menggunakan metode yang memiliki nilai R^2 yang terbesar, yakni metode PRM dengan fungsi pembobot Fair (minimum) dan Huber dimana c = 2 (minimum) serta metode RSIMPLS-M.

Dalam validasi, data pengamatan n = 45 dibagi dalam dua kelompok. Kelompok yang pertama sebanyak n_1 = 32 digunakan untuk membentuk model dan kelompok yang kedua sebanyak n_2 = 13 digunakan untuk validasi model. Pemilihan n_2 dilakukan secara acak sebanyak 20 pengambilan tanpa pemulihan. Nilai rata-rata RMSE dari hasil analisis dengan metode PRM dengan fungsi pembobot Fair (minimum) dan Huber dimana c = 2 (minimum) serta metode RSIMPLS-M untuk membentuk model kalibrasi serta rata-rata RMSEP tertera pada Tabel 4.

Tabel 4 Rata-rata RMSE dan RMSEP dari 20 kombinasi pengambilan contoh

Metode	Rata-rata RMSE (n_1 = 32)	Rata-rata RMSEP (n_1 = 13)
	Konsentrasi Lemak	Konsentrasi Lemak
PRM-Fair (minimum)	1.481	2.767
PRM -Huber=2 (minimum)	1.423	2.800
RSIMPLS-M	1.227	2.168

Berdasarkan hasil analisis yang tertera pada Tabel 4, secara umum tampak bahwa untuk konsentrasi lemak ikan (***Y***) pada data dengan metode RSIMPLS-M diperoleh nilai dengan rata-rata RMSEP terkecil, yaitu 2.168. Hal ini berarti metode RSIMPLS-M lebih baik dibandingkan dengan metode PRM.

Pendugaan Selang Kepercayaan Parameter dengan Metode Jackknife

Data dianalisis dengan metode Jackknife menggunakan metode yang memiliki R^2 yang paling besar. Dari Tabel 3 tampak bahwa metode PRM dengan k = 6 dan menggunakan fungsi pembobot Fair (minimum) dan Huber dimana c = 2 (minimum) serta metode RSIMPLS-M dengan k = 6 memiliki R^2 yang paling besar. Nilai rataan bias, simpangan baku dan selang kepercayaan dari metode *Jackknife* untuk setiap metode disajikan pada Tabel 5, Tabel 6, dan Tabel 7.

Pada Tabel 5 dan Tabel 6 terlihat bahwa nilai koefisien **b** antara metode PRM dengan fungsi pembobot Fair (minimum) dan Huber c = 2 (minimum) cenderung mempunyai kesamaan pola, walaupun ada perbedaan dalam besarannya. Berbeda halnya pada Tabel 7, terlihat bahwa nilai **b** yang dihasilkan dengan metode RSIMPLS-M mempunyai pola yang berbeda dan nilai besaran yang dihasilkan lebih besar dibanding metode PRM. Jika ditinjau dari nilai rataan bias, metode PRM dengan fungsi pembobot Fair (minimum) dan Huber c = 2 (minimum) memiliki nilai rataan bias lebih kecil dari metode RSIMPLS-M.

Tabel 5 Bias dan simpangan baku koefisien regresi menggunakan metode PRM dengan fungsi pembobot Fair (minimum)

Koefi-sien	Penduga Titik	Rataan $Bias_{Jack}$	σ_{Jack}	Selang Kepercayaan 95%		Lebar SK
				BB	BA	
$\mathbf{b}_0$	51,43	-12,01	5,83	39,67	63,19	23,52
$\mathbf{b}_1$	30,78	26,36	20,21	-9,95	71,52	81,47
$\mathbf{b}_2$	5,31	-6,80	7,18	-9,16	19,78	28,94
$\mathbf{b}_3$	-121,70	10,06	22,82	-167,68	-75,72	91,96
$\mathbf{b}_4$	-76,61	4,36	11,92	-100,63	-52,59	48,04
$\mathbf{b}_5$	20,46	-2,43	33,33	-46,70	87,63	134,33
$\mathbf{b}_6$	58,36	-29,36	10,42	37,36	79,35	41,99
$\mathbf{b}_7$	86,55	-20,96	11,41	63,57	109,53	45,96
$\mathbf{b}_8$	67,85	-10,67	13,70	40,24	95,47	55,23
$\mathbf{b}_9$	-66,97	36,46	35,57	-138,65	4,71	143,36

BB = Batas bawah; BA = Batas atas ; SK = Selang kepercayaan

Tabel 6 Bias dan simpangan baku koefisien regresi menggunakan metode PRM dengan fungsi pembobot Huber c = 2 (minimum)

Koefi-sien	Penduga Titik	Rataan $Bias_{Jack}$	σ_{Jack}	Selang Kepercayaan 95%		Lebar SK
				BB	BA	
$\mathbf{b}_0$	41,11	-3,05	6,70	27,61	54,60	26,99
$\mathbf{b}_1$	50,70	8,76	20,48	9,43	91,98	82,55
$\mathbf{b}_2$	5,07	-6,39	6,85	-8,74	18,88	27,62
$\mathbf{b}_3$	-119,20	8,84	23,16	-165,87	-72,52	93,35
$\mathbf{b}_4$	-77,06	3,97	12,83	-102,91	-51,21	51,7
$\mathbf{b}_5$	31,49	-12,05	32,96	-34,93	97,91	132,84
$\mathbf{b}_6$	28,52	-3,79	11,88	4,59	52,45	47,86
$\mathbf{b}_7$	71,00	-7,59	13,70	43,39	98,60	55,21
$\mathbf{b}_8$	59,96	-4,22	15,69	28,34	91,57	63,23
$\mathbf{b}_9$	-42,52	16,65	39,12	-121,37	36,34	157,71

BB = Batas bawah; BA = Batas atas ; SK = Selang kepercayaan

Selanjutnya, jika ditinjau dari lebar selang kepercayaan 95% bagi koefisien regresi hasil metode PRM dengan fungsi pembobot Fair (minimum) lebih sempit dibandingkan selang kepercayaan hasil metode PRM menggunakan fungsi pembobot Huber dimana c = 2 (minimum) dan metode RSIMPLS-M. Sedangkan lebar selang kepercayaan 95% bagi koefisien regresi hasil metode PRM dengan fungsi pembobot Huber dimana c = 2 (minimum) lebih sempit dibandingkan selang kepercayaan hasil metode RSIMPLS-M. Ini menunjukkan bahwa hasil yang diperoleh melalui metode PRM baik menggunakan fungsi pembobot Fair ataupun Huber dimana c = 2 (minimum) lebih baik dibanding metode RSIMPLS-M.

Tabel 7 Bias dan simpangan baku koefisien regresi menggunakan metode RSIMPLS-M

Koefi-sien	Penduga Titik	Rataan $Bias_{Jack}$	σ_{Jack}	Selang Kepercayaan 95%		Lebar SK
				BB	BA	
b0	37,37	-2,57	9,24	18,75	55,99	37,24
b1	145,63	-89,11	117,07	-90,31	381,57	471,88
b2	-207,98	207,25	207,45	-626,07	210,11	836,18
b3	-54,52	-42,62	150,02	-356,85	247,84	604,69
b4	116,30	-175,73	41,03	33,61	198,99	165,38
b5	-48,74	50,07	50,02	-149,55	52,07	201,62
b6	199,33	-179,73	211,05	-226,01	624,66	850,67
b7	31,84	10,79	26,41	-21,38	85,05	106,43
b8	-197,48	243,076	208,17	-617,03	222,07	839,1
b9	69,90	-49,39	89,39	-110,26	250,05	360,31

BB = Batas bawah; BA = Batas atas ; SK = Selang kepercayaan

SIMPULAN

Pada data lemak ikan terdapat korelasi antar peubah bebas yang cukup tinggi. Jumlah pencilan pada data lemak ikan yang dideteksi oleh metode PRM menggunakan fungsi pembobot Huber lebih kecil atau sama dengan jumlah pencilan yang dideteksi oleh metode PRM menggunakan Fair. Berdasarkan kriteria R^2 dan RMSEP, hasil yang diberikan oleh metode PRM menggunakan fungsi pembobot Fair tidak berbeda signifikan dengan hasil yang diberikan oleh metode PRM menggunakan Huber.

Jumlah pencilan pada data lemak ikan yang dideteksi oleh RSIMPLS-M lebih kecil dari jumlah pencilan yang dideteksi oleh PRM secara umum. Hal ini disebabkan karena RSIMPLS-M merupakan metode pendugaan parameter yang didasarkan pada Analisis Komponen Utama Kekar (*Robust Principal Component Analysis*/ROBPCA). ROBPCA melakukan perhitungan matriks peragam tidak dari semua data, tetapi dari h pengamatan dengan nilai terkecil. Sedangkan PRM merupakan metode pendugaan parameter yang menggunakan metode *Iterative Reweighted Partial Least Squares* (IRPLS). Setiap iterasi metode PRM melibatkan keragaman baik pada peubah respon maupun pada peubah penjelas dari semua data.

Metode RSIMPLS-M diperoleh nilai R^2 terbesar dan nilai rata-rata RMSEP terkecil. Lebar selang kepercayaan parameter dengan metode PRM lebih sempit dari pada RSIMPLS-M. Sehingga dapat disimpulkan bahwa metode RSIMPLS-M lebih baik dibanding metode PRM untuk prediksi, sedangkan metode PRM lebih baik dibanding metode RSIMPLS-M untuk pendugaan dalam kasus data lemak ikan yang berkorelasi tinggi antar peubah bebas dan adanya pencilan.

DAFTAR PUSTAKA

Cummins DJ, Andrews CW. 1995. Iteratively reweighted partial least squares regression. A performance analysis by Monte Carlo Simultan. *Journal of Chemometrics*. 9:489-507.

Dudewicz EJ, Mishra SN. 1988 *Modern mathematical statistics*, Wiley series in probability and mathematical statistics, John Wiley & Sons. New York, USA.

Huber PJ. 1981. *Robust statistics*. John Wiley & Sons. New York USA.

Geladi P, Kowalski BR. 1986. Partial least squares regression: A tutorial. *Analitycal Chimica Acta*. 185:1-17.

Huber M, Branden KV. 2003. Robust methods for partial least squares regression. *Journal of Chemometrics*. 17:537-549.

Ismah. 2010. Pendekatan regresi kuadrat terkecil parsial kekar dalam pemodelan kalibrasi multirespon. [Tesis]. Bogor: Sekolah Pascasarjana, Institut Pertanian Bogor.

Kuzmic P, Hill C, Janc WJ. 2004. Practical robust fit enzyme inhibition data. *Methods in Enzymology*. 383:366-381.

Liebmann B, Filzmoser P, Varmuza K. 2009. Robust and classical PLS regression. http://www.statistik.tuwien.ac.at.

Martens H, Naes T. 1989. *Multivariate Calibration*. John Willey & Sons. England: Chichester.

Myers RH. 1990. Classical and Modern Regression with Applications, 2nd Edition. PWS-Kent Publishing Company. Boston.

Naes T. 1985. Multivariate calibration when the error covariance matrix is structured. *Technometrics* 27(3): 301-311.

Norliza. 2006. Comparing three method of handling multicollinearity using simulation approach [Tesis]. Universiti Teknologi Malaysia.

Serneels S, Croux C, Filzmoser P, Van Espen PJ. 2005. Partial robust M-regression. *Chemom. Intell. Lab. Syst.*79:55-64.

Sunberg R. 1999. Multivariate calibration-direct and indirect regression methodology. *Board of the Foundation of the Scandinavian Journal of Statistics*. USA Vol 26:161-207.

Turkmen AS. 2008. Robust partial least squares for regression and classification [Disertasi]. Alabama: Auburn University.

Wigena AH, Aunuddin. 1997. Suatu Kajian dan Terapan Metode PLS. Seminar Nasional Statistika IV di ITS Surabaya, 9-10 Desember 1997.