KAJIAN PENANGANAN MULTIKOLENIERITAS DALAM ANALISIS REGRESI MENGGUNAKAN *PARTIAL LEAST SQUARE* REGRESSION

I Gede Nyoman Mindra Jaya

Staf Pengajar Jurusan Statistika Universitas Padjadjaran
Email : jay_komang@yahoo.com

Abstrak

Salah satu tujuan dari analisis regresi adalah meramalkan nilai variabel respon didasarkan pada beberapa variabel independen. Permasalahan yang sering ditemukan dalam analisis regresi adalah adanya korelasi yang tinggi antara variabel independen yang berakibat pada standar error pendugaan dari parameter regresi sangat besar yang berakibat pada model regresi tidak layak digunakan sebagai model peramalan karena modelnya tidak reliable. Beberapa metode diperkenalkan untuk menanggulangi multikolenieritas khususnya yang hampir sempurna, salah satunya adalah *Principal Coponent Regression (*PCR). Namun pendekatan ini memiliki kelemahan dalama proses pereduksian variabel independen tidak mempertimbangkan korelasi antar independen dengan variabel dependen. Kondisi ini menyebabkan pada saat pemodelan regresi, komponen utama pertama yang terbentuk dari proses reduksi tidak dijamin menjadi komponen yang paling mampu menjelaskan keragamana variabel respon, sedangkan komponen utama pertama merupakan komponen yang merangkum paling banyak informasi dari variabel independen. Satu metode baru diperkanalkan sebagai perbaikan dari metode PCR yaitu metode *Partial least square Regression* (PLSR). Metode ini dalam proses reduksi variabel independen telah mengakomodasi korelasi antara variabel respon dengan variabel bebasnya.

Kata Kunci : *Principal Componen Regression* (PCR), *Partial least square Regression* (PLSR), *Singular Value Decomposition* (SVD)

PENDAHULUAN

Salah satu fungsi dari analisis regresi adalah memprediksikan nilai-nilai variabel respon berdasarkan data variabel prediktor. Umumnya semakin banyak variabel yang terlibat dalam model regresi maka semakin akurat dan reliable nilai prediksinya karena tentunya dengan melibatkan banyak vairiabel prediktor, proporsi varians dari variabel respon yang dapat dijelaskan oleh variabel prediktor akan semakin tinggi yang ditunjukkan oleh nilai koefisien determinasi R^2 yang semakin besar. Namun, terdapat satu masalah klasik diantara banyak masalah dalam analisis regresi multiple. Permasalahan tersebut adalah adanya korelasi sempurna atau hampir sempurna antara variabel prediktor. Dalam bahasa regresi persamasalahan ini dikenal dengan nama *multikolenieritas.*

Terdapat beberapa kosekuensi logis dari terjadinya multikolenieritas hampir sempurna dalam analisis regresi diantaranya adalah (1) Mesikpun penaksri OLS masih bisa diperoleh, kesalahan standarnya cenderung semakin besar dengan semakin meningkatknya multikolenieritas, (2) Karena kesalahan standar menjadi besar, maka selang kepercayaan untuk perameter populasi yang relevan cenderung akan menjadi lebih besar, (3) Akibat dari konsekuensi 2 maka peluang untuk menerima hipotesis nol yang seharusnya ditolak (Kesalahan Tipe II) akan semakin besar, (4) Penaksiran koefisien regresi akan sangat sensitive terhadap perubahan data, dan (5) Jika multikolenieritas tinggi, mungkin akan diperoleh R^2 yang tinggi tetapi tidak ada satupun atau sangat sedikit koefisien yang ditaskri penting secara statistik.

Terdapat beberapa langkah praktis yang dapat ditempuh untuk penanggulangan multikolenieritas diantaranya adalah (1) Informasi apriori, (2) Penggabungan data crossection dengan data time series, (3) Mengeluarkan satu variabel, (4) Transformasi variabel dan (5)

Penambahan data baru. Namun metode-metode ini sering sulit dilakukan dalam tataran aplikasi karena keterbatasan informasi dari keterbatasan kemampuan dalam pengumlan data.

Cara lain yang sering ditempuh dalam penanggulangan multikoleniritas adalah dengan teknik statistik. Teknik analisis statistik yang paling umum digunakan adalah *Principal Componen Regression* (PCR). Secara statistik, teknik ini mampu menanggulangi multilkoleniritas dalam variabel prediktor dengan mereduksi variabel independen menjadi beberapa komponen yang saling ortogonal kemudian mentransformasikan kembali kevariabel asal. Namun pendekatan ini dinilai memiliki kekurangan. Teknik PCR mengabaikan korelasi antara variabel respon dengan variabel-variabel prediktor. Hal ini berakibat pada saat pemodelan regresi, komponen utama pertama yang terbentuk dari proses reduksi tidak dijamin menjadi komponen yang paling mampu menjelaskan keragamanan variabel respon, sedangkan komponen utama pertama merupakan komponen yang merangkum paling banyak informasi dari variabel prediktor. Satu metode baru diperkanalkan sebagai perbaikan dari metode PCR yaitu metode *Partial least square* Regression (PLSR). Metode ini dalam proses reduksi variabel independen telah mengakomodasi korelasi antara variabel respon dengan variabel bebasnya.

Regresi Komponen Utama

Telah dijelaskan di atas konsep dari komponen utama sebagai dasar dalam mengatasi masalah multikolenieritas dalam analisis regresi. Selanjutnya bagaimana menggabungkan analisis komponen utama dengan analisis regresi yang disebut dengan regresi komponen utama.
Model Regresi Komponen Utama

$$\mathbf{y=Z\alpha+\varepsilon} \qquad \mathbf{(1)}$$

dimana,

$$\mathbf{Z=XT,\quad \alpha=T`\beta,\ T`X'XT=Z`Z=\Lambda} \qquad \mathbf{(2)}$$

dan **T** matriks dari vektor eigen untuk setiap nilai eigen yang bersesuaian. Sedangkan $\mathbf{\Lambda}$ adalah matriks diagonal nilai eigen.
Parameter regresi duduga dengan menggunakan meode *ordinary least square* sebagai berikut .

$$\hat{\boldsymbol{\alpha}} = \mathbf{(Z`Z)^{-1}Z`y} \qquad \mathbf{(3)}$$

dengan

$$V(\hat{\boldsymbol{\alpha}}) = \hat{\sigma}^2\mathbf{(Z`Z)^{-1}}\,\mathbf{.} \qquad \mathbf{(4)}$$

Nilai $\hat{\sigma}^2$ diperoleh dari Means Square Error model regresi pada persamaan (). Selanjutnya untuk untuk transformasi kemodel awal dengan *standardized coefficent* dilakukan sebagai berikut :

$$\hat{\boldsymbol{\beta}} = \mathbf{T}\hat{\boldsymbol{\alpha}}$$

$$\mathbf{V}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\sigma}}^2\mathbf{T\Lambda^{-1}T`} \qquad \mathbf{(5)}$$

dan untuk transformasi ke model *unstandardized coefficient* ($\hat{\boldsymbol{\beta}}^*$) dilakukan sebagai berikut :

$$\hat{\boldsymbol{\beta}}^* = \mathbf{S}\hat{\boldsymbol{\beta}}, \qquad \hat{\beta}_0^* = \bar{y} - \sum_{i=1}^{p}\hat{\beta}_i\bar{x}_i\,,$$

$$\mathbf{V}(\hat{\boldsymbol{\beta}}^*) = (\hat{\boldsymbol{\sigma}}^2\mathbf{T\Lambda^{-1}T`})\mathbf{S}^2 \qquad \mathbf{(6)}$$

$$\mathbf{S} = D\left(\frac{s_y}{s_{xi}}\right)$$

Di bawah ini dijelaskan beberapa tahapan dalam analisis regresi omponen utama.
Langkah 1 : Lakukan perhitungan Jordan Dekomposisi dari matriks korealsi dan dapatkan eigen vector ($Ti=\{t_{1i}, t_{2i},\ldots, t_{pi}\}$) dari setiap nilai eigen yang bersesuaian (λi).
Langkah 2 : Membangun komponen utama dari variabel yang sudah disetandarkan melalui kombinasi linier dengan vektor eigen sebagai koefisiennya.

$$\mathbf{KU=ZT} \qquad \mathbf{(7)}$$

Dengan **KU** adalah komponen utama, **Z** merupakan variabel yang distandarkan dan **T** adalah eigen vektor dari matriks korelasi antara variabel prediktor.

Langkah 3 : Lakukan pemodelan regresi komponen utama dengan banyak komponen utama pertama yang dipilih berdasarkan metode *scree plot* atau berdasarkan nilai proporsi varians paling tidak lebih besar dari 75%.

Langkah 4 : Transformasi model regresi komponen utama ke model asal.

***Partial least square Regression* (PLSR)**

Salah satu kelemahan dari analisis komponen utama dalam kaitan penanggulangan multikolenieritas, adalah reduksi variabel dalam analisis komponen utama melalui SCL hanya menangkap karakteristik dari variabel X atau prediktor tanpa memperhatikan bagaimana relasinya dengan variabel y sebagai respon. Melalui *Partial least square*, memungkinkan kita melakukan reduksi pada variabel prediktor dengan mempertimbangkan relasinya dengan variabel responden.

Misalkan **X** adalah matriks prediktor dengan dimensi n x p, dan **Y** adalah matriks dependen dengan dimensi n x q. Metode *partial least square* bekerja secara iteratif mengekstrak faktor dari X dan Y sehingga kovarians antara faktor yang diekstrasi maksimum. *Partial least square* dapat bekerja untuk respon merupakan variabel multivariate. Dalam penelitian ini kita spesifikan hanya untuk respon univariate.

Metode PLS mencoba untuk menemukan kombinais linier dari X dan Y sehingga $\mathbf{X=TP^{`}+E}$ dan $\mathbf{Y=UQ+F}$ dimana

$\mathbf{T}_{nxr}$ =**X**-Skor	$\mathbf{U}_{nxr}$=**Y**-Skor
$\mathbf{P}_{pxr}$ =**X**-Loading	$\mathbf{Q}_{1xr}$= **Y**-Loading
$\mathbf{E}_{nxp}$=**X**-Residual	$\mathbf{F}_{nx1}$=**Y**-Residual

Proses dekomposisi dilakukan untuk memaksimumkan kovarians antara T dan U. Ada beberapa algoritma yang dikembangkan untuk PLS seperti NIPLS, SIMPLS. Untuk NIPLS terkenal algorita Wold dan Matens. Namun kedua metode ini dilakukan secara iteratif untuk melakukan ekstrasi skor X dan skor Y. Skor X dan Y diekstrasi secara berurutan dan banyaknya faktor yang diekstrak (r) tergantung pada rank dari X dan Y. Dalam penelitian ini, Y adalah vektor, dan semua kemungkan faktor X akan diekstrasi.

Singular Value Dekomposisi

Setiap ekstrasi x-skor adalah kombinasi linier dari X. Sebagai contoh, ekstrasi pertama dari x-skor yaitu t memiliki bentuk t=Xw, dimana w adalah eigen vektor yang bersesuaian dengan nilai eigen dari matriks **X`YY`X**. Hal yang sama juga berlaku untuk y-skor yaitu u memiliki bentuk u=Yc dengan c adalah vektor eigen yang bersesuaian dengan nilai eigen dari matriks **Y`XX`Y**. Perlu diingat bahwa **X`Y** menyatakan kovarians dari **X** dan **Y**

Setelah faktor pertama dapat diekstraksi menyatakan nilai X dan Y sebagai :

$$\mathbf{X_1=X}\text{-}\boldsymbol{tt}\mathbf{`X} \text{ dan } \mathbf{Y_1=Y}\text{-}\boldsymbol{tt}\mathbf{`Y} \qquad \mathbf{(8)}$$

Proses di atas diulangi untuk mendapatkan PLS faktor kedua dan seterusnya sehingga matriks X menjadi matriks Nol.

Dugaan parameter regresi (β) dapat ditulis :

$$\mathbf{b_A} = \mathbf{W_A}(\mathbf{P_A}\mathbf{`}\mathbf{W_A})^{-1}\mathbf{q_A} \qquad \mathbf{(9)}$$

Dengan W_A adalah matriks pembobot yang merupakan hasil perkalian dari $E_{a-1}F_{a-1}$. Indeks A menunjukkan banyak faktor laten yang dibentuk yang artinya sama dengan bayak iterasi.

Metode Bootstrap

Pendugan parmeter regresi dengan menggunakan PLS tidak memperhatikan sebaran data sehingga untuk pengujian hipotesis signifikansi paramter menggunakan pendekatan Bootstrap. Pada metode bootstrap dibentuk B buah sampel Bootstrap, masing-masing merupakan sampel acak berukuran n yang diambil dengan pengembalian dari populasi n pengamatan. Pengamatan ke-i (i=1,2,...,n) dari sampel awal mungkin muncul beberapa kali pada sampel Bootstrap ke-r (r=1,2,..,B). Sedangkan pengamatan lain mungkin tidak muncul sama sekali.

Bila **b**(r) adalah dugaan parameter β yang diperoleh dari sampel boostrap ke-r (r=1,2,...,B) maka dugaan simpangan baku dari parameter β adalah :

$$s\hat{e}_B = \left\{ \sum_{r=1}^{B} [b(r) - b(.)]^2 / (B-1) \right\}^{1/2}$$
$$b(.) = \sum_{r=1}^{B} b(r) / B \tag{10}$$

Kebaikan Model

Kebaikan model regresi dilihat dari beberapa ukuran statistik yaitu Kuadrat Tengan Galat (KTG) dan koefisien determinasi (R^2)

$$KTG = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 / (n-p) \tag{11}$$

$$R^2 = 1 - \frac{(n-p)KTG}{(n-1)S^2 Total} \tag{12}$$

dengan

$$S^2 Total = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 / (n-1)$$

Selanjutnya untuk melihat apakah PLSR lebih baik dibandingkan dengan PCR dilakukan dari perbandingkan Bias pendugaan parameter. Yaitu perbedaan absolute antara dugaan parameter PLSR dan PCR dengan OLS.

Contoh Aplikasi

Untuk lebih memperjelas pembahasan yang telah diuraikan pada bagian sebelumnya, maka pada bagian ini akan dijelaskan mengenai penerapan dari metode *Partial least square* Regression PLSR dan perbandingannya dengan metode pendugaan OLS dan PCR dalam penanggulangan multikoleniritas. Adapun data yang digunakan merupakan data kualitas batubara di daerah Girimulya, Kalimantan Selatan yang diteliti oleh sebuah lembaga yaitu Pusat Penelitian Teknologi Mineral (PPTM). Endapan batubara yang ditemukan di lokasi tersebut mempunyai 9 lapisan dan data yang diambil sebagai sampel adalah data untuk lapisan batubara yang pertama sebanyak 30 buah. Kualitas batubara yang diukur adalah nilai kalori kotor (CGV=*Gross Calorivie Value*) dalam setiap kkal/gr sebagai variabel Y, berdasarkan kadar air (IM=*Inherent Moisture*) sebagai variabel X_1, kadar abu (*Ash*) sebagai variabel X_2, kadar zat terbang (VM=*Volatille Matter*) sebagai variabel X_3, kadar karbon tertambang (FC=*Fixed Carbon*) sebagai variabel X_4 dan kadar total sulpur (TS=*Total Sulfur*) sebagai variabel X_5. Data yang digunakan dapat dilihat pada table 1 berikut ini:

Tabel 1.
Data Kualitas Batubara Untuk Lapisan 1 (Seam W0) Di Daerah Girimulya

No	IM (X_1)	ASH (X_2)	VM (X_3)	FC (X_4)	TS (X_5)	CGV (Y)
1	24.7	4.2	36.9	34.2	0.11	4698
2	21	15.7	32.9	30.4	0.3	4177
3	23.8	4.7	38.5	33	0.15	4781
4	18.7	4.5	37.8	39	0.17	5023
5	16.7	3.6	40	39.3	0.16	5339
6	18	8.5	38.8	34.7	0.65	4753
7	19.2	3.9	42	34.9	0.6	5189
8	18.7	4.2	41	36.1	0.75	5159
9	18.2	3.5	40.3	38	0.1	5128
10	17.4	3.7	40.4	38.5	0.1	5058

No	IM (X_1)	ASH (X_2)	VM (X_3)	FC (X_4)	TS (X_5)	CGV (Y)
11	18.3	3.3	40	38.4	0.09	5073
12	18.2	3.5	40.3	38	0.1	5128
13	17.6	3.2	40.6	38.6	0.11	5137
14	17.8	3.1	40.9	32.2	0.09	5149
15	16.7	3.9	41.8	37.6	0.1	5289
16	24.2	4.3	36.7	34.8	0.09	4610
17	24	4.1	37.9	34	0.09	4664
18	24.7	2.9	36.6	35.8	0.09	4639
19	24.4	2.9	36.7	36	0.09	4667
20	24.7	3.1	36.9	35.3	0.09	4670
21	25	4.6	35.7	34.7	0.1	4615
22	25.9	4.9	36.8	32.4	0.22	4519
23	24.4	5.2	36.7	33.7	0.16	4571
24	24.4	3.2	36.7	35.7	0.1	4659
25	24.3	4.1	38.7	32.9	0.12	4847
26	22.9	7.8	36.1	33.2	0.12	4533
27	25.7	4	36.9	33.4	0.1	4658
28	20.8	7.6	37.3	34.3	0.26	4599
29	23.3	5.4	38.1	33.2	0.14	4584
30	23.4	3.4	39.4	33.8	0.1	4746

Regresi Kuadrat Terkecil Biasa

Tahap awal dalam analisis data di atas adalah melakukan perhitungan model regresi linier multipel dengan menggunakan pendugaan Ordinary Least Square. Hasilnya adalah sebagai berikut :

$$Y=4384.35808-47.81260X_1-45.04494X_2+39.62497X_3+4.35036X_4+ 51.88481X_5$$

Untuk mendeteksi apakah terjadi pelanggaran multikolenieritas dilihat dari nilai Variance Inflation Fator (VIF). Nilai VIF yang lebih besar dari 5 mengindikasikan terjadinya permasalahan multikoleniritas.

Tabel 2.
Hasil Pengujian Koefisien Regresi dan nilai VIF.

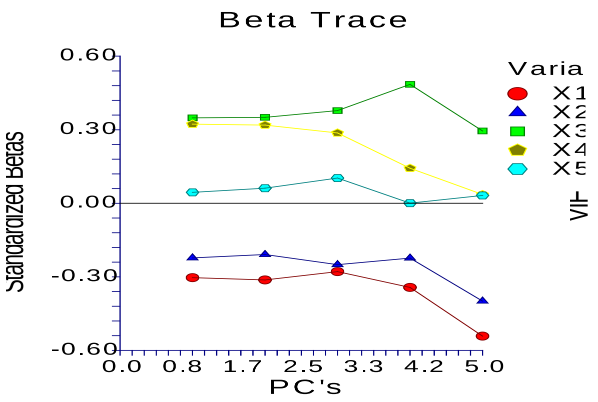
Variabel	Nilai Taksiran Parameter	Standard Error	t-hitung	P-value	VIF
Intersep	4384.358	1589.430	2.758	0.011	-
X1	-47.813	15.044	-3.178	0.004	12.475
X2	-45.045	17.133	-2.629	0.015	9.887
X3	39.625	22.245	1.781	0.088	11.759
X4	4.350	13.169	0.330	0.744	5.028
X5	51.885	101.848	0.509	0.615	1.673
KTA = 5411.777					
R^2 = 0.9441					

Tabel diatas pada kolom VIF ditunjukkan bahwa hanya untuk variabel X_5 nilai VIF nya lebih kecil dari 5 sedangkan untuk yang lainnya lebih besar dari 5. Ini mengindikasikan adanya permasalahan multikolenieritas. Hal ini berakibat pada hanya sedikti variabel independen yang signifikan walau koefisien determinasinya sangat besar. Efek yang lain dari adanya multikolenieritas adalah ketidakstabilan pendugaan parameter beta.

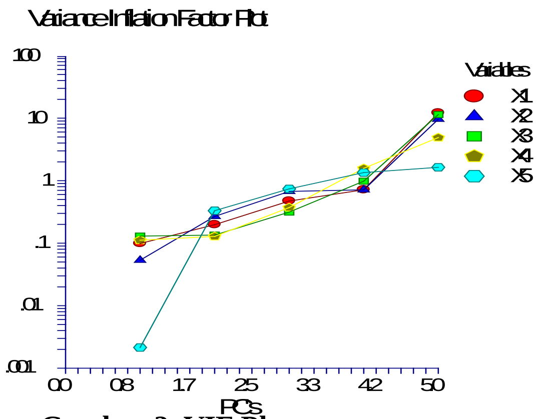
Regresi Komponen Utama

Seperti yang telah dijelaskan metode Regresi Komponen Utama (PCR) merupakan sebuah solusi dari adanya permasalahan multikolenieritas. Di bawah ini disajikan hasil analisis dengan metode PCR.

Banyaknya komponen utama yang dipilih adalah sebanyak 2. Ini didasarkan pada Plot Beta Trace dan Plot VIF seperti yang disajikan di bawah ini:



Gambar 1. Beta Trace



Gambar 2. VIF Plot

Dari kedua gambar di atas, disimpulkan bahwa banyaknya komponen utama yang optimal adalah sebanyak 2 komponen. Dari dua komponen utama ini sudah mampu menjelaskan kergamana data sebanyak 79,4%. Secara legkap dapat dilihat dari table di abwah ini :

Tabel 3.
Nilai Eigen

No.	Eigenvalue	Incremental Percent	Cumulative Percent
1	2.4902	49.803	49.803
2	1.4805	29.61	79.413
3	0.6481	12.962	92.375
4	0.353	7.059	99.435
5	0.0283	0.565	100

Tabel 4
Pendugaan Paramter dengan PCR

Independent Variable	Regression Coefficient	Standard Error	t Value	VIF
Intercept	2341.549			
X1	-27.5912	11.427	-2.414	0.2008
X2	-23.6046	17.021	-1.387	0.2723
X3	47.07757	14.369	3.276	0.1369
X4	38.76649	12.691	3.055	0.1303
X5	100.1429	273.539	0.366	0.3367

KTA= 7758.081
R^2 = 0.9198

Terlihat dari tabel 4 di atas secara umum nilai standar error mengecil dibandingkan dengan pendugaan OLS, namun khusus untuk X5 nilai standard errornya tinggi karena rasio keragaman X_5 sangat kecil dibandingkan dengan keragaman Y. Nilai VIF yang sangat kecil menandakan sudah tidak adanya pelanggaran asumsi non multikolenieritas dalam regresi.

Partial least square Regression

Seperti yang telah dijelaskan di atas, metode *Partial least square Regression* adalah satu metode yang diusulkan sebagai metode penanganan masalah multikolenieritas dalam analisis regresi. Hasil pengolahan diperoleh rangkuman sebagai berikut :

Tabel 5
Pendugaan Paramter dengan PLSR

Independent Variable	Regression Coefficient	Standard Error	T Value
Intercept	2576.53		
X1	-30.2624	4.637	-6.5261
X2	-27.3095	10.538	-2.5916
X3	61.8735	5.818	10.6352
X4	18.2083	9.319	1.954
X5	82.9863	117.526	0.7061
KTG = 5859.713			
R^2 = 0.94			

Dari model di atas diperlihatkan bahwa standard error pendugaan parameter yang diperoleh dari prosedur Bootstrap secara umum lebih kecil dibandingkan dengan standard error pendugaan parameter dengan OLS ataupun PCR. Jika diambil nilai t table pada alpha 0.05 dan derajat bebas 24 yaitu t-tabel = 2.06. Ini artinya variabel X1, X2 dan X3 memiliki hubungan signifikan dengan Y, dan untuk X4 memiliki hubungan signifikan pada nilai alpha lebih kecil dari 0.1.

Perbandingan Model

Di bawah ini rangkuman perbandingan metode pendugaan dengan OLS, PCR, dan PLSR.

Tabel 6
Pendugaan Parameter

Beta	Koefisien Regresi			Standard Error			T		
	OLS	PCR	PLS	OLS	PCR	PLS	OLS	PCR	PLS
b0	4384.36	2341.55	2576.53						
b1	-47.81	-27.59	-30.26	15.04	11.43	4.64	-3.18	-2.41	-6.53
b2	-45.04	-23.60	-27.31	17.13	17.02	10.54	-2.63	-1.39	-2.59
b3	39.63	47.08	61.87	22.25	14.37	5.82	1.78	3.28	10.63
b4	4.35	38.77	18.21	13.17	12.69	9.32	0.33	3.05	1.95
b5	51.88	100.14	82.99	101.85	273.54	117.53	0.51	0.37	0.71

Tabel 7
Bias Pendugaan Terhadap OLS

Beta	Bias	
	PCR	PLSR
b0	2042.81	1807.83
b1	20.22	17.55
b2	21.44	17.74
b3	7.45	22.25
b4	34.42	13.86

b5	48.26	31.10

Telah diketahui bahwa walaupun terjadi pelangaran asumsi non-multikolenieritas hampir sempurna, pendugaan parameter regresi tetap tak berbias namun memiliki standar error yang tinggi. Hasil pendugana *partial least square* regression (PLSR) memiliki bias yang relative lebih kecil dibandingkan dengan PCR terhadap pendugaan OLS. PLSR juga memiki koefisein determinasi R^2 yang lebih baik dibandingkan dengan PCR dan nilainya sama dengan nilai koefisien determinasi model dengan pendugaan OLS. Ini menujukkan PLSR untuk kasus dalam penelitian ini lebih baik dibandingkan PCR.

KESIMPULAN

Hasil pengolahan data menunjukkan bahwa untuk data kualitas batubara untuk lapisan 1 (Seam W0) di daerah Girimulya, metode PLSR dalam penanganan multikolenieritas memberikan hasil pendugaan yang lebih baik dibandingkan dengan regresi komponen utama baik dari nilai standar error pendugaan, koefisien determinasi sebagai ukuran kebaikan model dan juga dari bias pendugaan parameter terhadap OLS.

DAFTAR PUSTAKA

Indrawati Kumala, Pendugaan Model Regresi dengan Metode Kuadrat Terkecil Partial, Thesis (1997), Institut Pertanian Bogor.

Norliza Adnan, Maizah Hura Ahmad, Robiah Adnan. A Comparative Study On Some Methods For Handling Multicollinearity Problems, Journal MATEMATIKA, Volume 22 (2006), Number 2, pp. 109–119

Saikat Maitra and Jun Yan, Principle Component Analysis and *Partial least squares*: Two Dimension Reduction Techniques for Regression, Casualty Actuarial Society, 2008 Discussion Paper Program