S - 8

Regresi Komponen Utama, Regresi Ridge, dan Regresi Akar Laten dalam Mengatasi Masalah Multikolinieritas

Dian Agustina

Jurusan Matematika, FMIPA, Universitas Bengkulu
dianagustina117@yahoo.com

***Abstrak*—** Uji mengenai multikolinieritas merupakan salah satu tahapan yang harus dilewati pada analisis regresi. Variabel-variabel yang mengalami multikolinieritas menyebabkan regresi tidak efisien atau penyimpangannya besar. Penelitian ini betujuan untuk melihat bagaimana metode regresi komponen utama, regresi ridge, dan regresi akar laten mengatasi masalah multikolinieritas. Data yang digunakan adalah data Hald, dengan banyaknya trikalsium alumina, banyaknya trikalsium silika, banyaknya tetrakalsium alumino besi(2), dan banyaknya dikalsium silika sebagai variabel-variabel bebas, serta evolusi panas dalam kalori per gram semen sebagai variabel tak bebasnya. Dari ketiga metode regresi yang digunakan pada data tersebut, metode regresi ridge dan regresi akar laten mempunyai MSE yang lebih kecil daripada yang dihasilkan oleh metode regresi komponen utama.

Kata kunci: Regresi Komponen Utama, Regresi Ridge, Regresi Akar Laten.

I. PENDAHULUAN

Analisis regresi merupakan salah satu teknik statistika yang digunakan untuk menyelidiki hubungan atau pengaruh antara suatu variabel dengan variabel lainnya. Variabel-variabel regresi yang berhubungan secara linier disebut regresi linier. Regresi linier antara satu variabel respon dengan satu variabel bebas disebut regresi linier sederhana, sedangkan regresi linier antara satu variabel respon dengan dua atau lebih variabel bebas disebut regresi linier berganda [1].

Pada regresi linier berganda yang mempunyai banyak variabel bebas, sering muncul masalah karena terjadinya hubungan linier antara dua atau lebih variabel bebas. Variabel bebas yang saling berkorelasi disebut kolinieritas ganda (multikolinieritas). Salah satu asumsi yang harus dipenuhi dalam model regresi linier berganda adalah tidak adanya multikolinieritas di antara variabel bebas yang masuk ke dalam model. Multikolinieritas merupakan suatu kondisi dimana terjadi korelasi yang tinggi antar variabel-variabel bebas yang mengakibatkan determinan dari matriks $X'X$ akan mendekati 0 sehingga menyebabkan matriks tersebut hampir *singular* yang mengakibatkan nilai dari penduga parameternya tidak dapat ditemukan [2].

Untuk mengetahui adanya multikolinieritas dapat dilakukan dengan menghitung koefisien korelasi sederhana antara sesama variabel bebas, jika terdapat korelasi sederhana yang hampir mendekati ± 1 maka ini menunjukkan adanya masalah multikolinieritas [3]. Selain itu, untuk mengukur adanya multikolinieritas yaitu dengan melihat nilai $VIF > 10$. VIF (*Variance Inflation Factor*) merupakan suatu faktor yang mengukur seberapa besar kenaikan ragam dari koefisien penduga regresi dibandingkan terhadap variabel bebas yang orthogonal jika dihubungkan secara linier. Gejala multikolinieritas menimbulkan masalah dalam model regresi. Korelasi antar variabel bebas yang sangat tinggi menghasilkan penduga model regresi yang berbias, tidak stabil, dan mungkin jauh dari nilai prediksinya [4].

Beberapa metode yang dapat digunakan untuk mengatasi masalah multikolinieritas ini diantaranya adalah Regresi Komponen Utama, Regresi Ridge, dan juga Regresi Akar Laten. Dalam metode Regresi Komponen Utama, hasil transformasi persamaan regresi komponen utama dari komponen utama $(W) \rightarrow Z$ memberikan persamaan regresi komponen utama yang tidak lagi mengandung multikolinieritas. Metode Regresi Ridge akan menghasilkan penduga yang bias karena adanya penambahan nilai θ^* pada proses perhitungannya tetapi mempunyai varian yang minimum yang ditunjukan dari nilai Kuadrat Tengah Galat (*Mean Square Error*) yang minimum. Sementara, metode regresi akar laten merupakan perluasan dari metode regresi komponen utama, yaitu dengan menggabungkan matriks data yang berasal dari variabel bebas dengan variabel respon yang telah dibakukan [5].

Berdasarkan uraian di atas, penulis tertarik untuk melihat masalah multikolinieritas dengan ketiga metode tersebut di atas yang dapat mengatasinya.

II. METODE PENELITIAN

A. Regresi Linier

Bentuk umum dari persamaan regresi linier diberikan oleh

$$Y_i = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \varepsilon_i \ , \qquad\qquad i = 1, 2, \ldots, n \tag{1}$$

dimana

$n > k + 1,$
$Y_i =$ respon ke $- i,$
$\beta_0 =$ intersep,
$\beta_i =$ koefisien regresi variabel bebas ke $- i,$'
$X_i =$ variabel bebas ke $- i,$ dan
$\varepsilon_i =$ galat dari pengamatan ke $- i.$

Dengan notasi matrik, (1) dapat dituliskan sebagai

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ 1 & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_3 \end{bmatrix} \tag{2}$$
$$\mathbf{Y}_{(n\times 1)} = \mathbf{X}_{(n\times(k+1))}\boldsymbol{\beta}_{((k+1)\times 1)} + \boldsymbol{\varepsilon}_{(n\times 1)}$$

Regresi linier berganda mempunyai asumsi-asumsi yang harus dipenuhi, salah satunya ialah multikolinieritas.

B. Multikolinieritas

Istilah multikolinearitas pertama kali dikemukakan oleh Ragnar Frisch yang berarti adanya hubungan linear yang “sempurna” atau pasti diantara beberapa atau semua variable bebas dari model regresi berganda. Multikolinearitas berkenaan dengan terdapatnya lebih dari satu hubungan linear pasti. Multikolinieritas menyebabkan regresi tidak efisien atau penyimpangannya besar [6].

Salah satu cara untuk mendeteksi multikolinearitas dengan menggunkan *Variance Inflation Factor* (VIF), yang diberikan oleh

$$VIF_j = \frac{1}{1 - R_j^2} \tag{3}$$

VIF merupakan unsur-unsur diagonal utama matriks korelasi $\mathbf{C} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}$ dengan R_j^2 merupakan koefisien determinasi yang didapat dari variabel bebas X_j diregresikan terhadap p variabel bebas lain. Jika X_j tidak berkorelasi dengan variabel bebas lain, maka R_j^2 akan bernilai kecil dan VIF_j mendekati 1. Sebaliknya jika X_j mempunyai korelasi dengan variabel bebas lain, maka R_j^2 akan mendekati 1 dan VIF_j menjadi besar. Jika nilai VIF_j lebih dari 10,maka ini menunjukkan data mengalami masalah multikolinearitas [7].

C. Regresi Komponen Utama

Regresi komponen utama didasarkan pada analisis komponen utama. Melalui analisis komponen utama akan dihasilkan variabel-variabel baru (komponen utama) yang merupakan kombinasi linier dari variabel-variabel bebas asal, dan antar variabel bebas baru ini bersifat saling bebas. Selanjutnya, variabel bebas baru ini diregresikan dengan variabel respon asal [5].

Pada awalnya, berdasarkan matrik **X** pada (2) yang telah melalui prosedur *centering* dan *rescaling*, diperoleh matrik X yang terpusat dan terskalakan, yang kemudian disebut matrik **Z**. Matrik korelasinya $\mathbf{Z}'\mathbf{Z}$ dan akar ciri (disebut juga akar laten atau nilai eigen) matrik korelasi ini adalah *r* solusi $\lambda_1, \lambda_2, \ldots, \lambda_r$ dari persamaan determinan $|\mathbf{Z}'\mathbf{Z} - \lambda\mathbf{I}| = 0$. Jumlah akar ciri matrik korelasi ini sama dengan trace matrik $\mathbf{Z}'\mathbf{Z}$.

Untuk setiap akar ciri λ_j terdapat vektor ciri γ_j yang memenuhi sistem persamaan homogen $\left(\mathbf{Z}'\mathbf{Z} - \lambda_j\mathbf{I}\right)\gamma_j = \mathbf{0}$. Vektor ciri solusinya $\gamma_j = \left(\gamma_{1j}, \gamma_{2j}, \ldots, \gamma_{rj}\right)'$, yang dipilih dari sekian banyak solusi sebanding yang ada untuk setiap *j*, merupakan solusi yang ternormalkan sedemikian rupa sehingga $\gamma_j{}'\gamma_j = 1$. Jika semua λ_j berbeda, maka setiap pasang vektor ciri akan ortogonal sesamanya. Vektor γ_j digunakan untuk mengucapkan kembali Z ke dalam suku-suku komponen utama W dalam bentuk

$$W_j = \gamma_{1j}z_1 + \gamma_{2j}z_2 + \ldots + \gamma_{rj}z_r \tag{4}$$

sehingga jumlah kuadrat setiap peubah W_j, yang unsur-unsurnya W_{ji} untuk $i = 1, 2, \ldots, n$, adalah λ_j. Dengan kata lain W_j mengambil sejumlah λ_j dari keragaman totalnya. Yaitu $\sum_{j=1}^{r} \lambda_j = r$ sehingga jumlah kuadrat totalnya $\sum_{j=1}^{r}\sum_{i=1}^{n} W_{ji}^2 = r$ seperti semula.

Jadi, prosedur ini menciptakan variabel-variabel baru W_j dari variabel-variabel Z_j, melalui suatu transformasi linier pada (4), sedemikian sehingga vektor-vektor *W* itu ortogonal sesamanya. Variabel W_j padanan nilai λ_j yang terbesar disebut komponen utama pertama. Komponen ini menjelaskan bagian terbesar dari keragaman yang dikandung oleh gugusan data yang telah dibakukan. Komponen-komponen W_j yang lain menjelaskan proporsi keragaman yang semakin lama semakin kecil sampai semua keragaman datanya terjelaskan, jadi $\sum_{j=1}^{p} \lambda_j = r$. Biasanya kita tidak menggunakan semua W_j, melainkan mengikuti suatu aturan tertentu. Sebagian peneliti hanya mengambil akar ciri yang lebih besar dari 1. Referensi [8] menyarankan “komponen-komponen dapat dihitung sampai sejumlah tertentu proporsi keragaman data yang cukup besar (mungkin 75 persen atau lebih) telah terjelaskan”.

Selanjutnya prosedur kuadrat terkecil digunakan untuk memperoleh persamaan peramalan bagi Y sebagai fungsi dari variabel-variabel W_j yang terpilih itu. Urutan masuknya W_j tidak ada pengaruhnya dalam hal ini, sebab semuanya ortogonal satu sama lain. Bila persamaan regresi dalam W_j telah diperoleh, persamaan ini dapat dikembalikan menjadi fungsi variabel semula Z_i bila dikehendaki, atau ditafsirkan berdasarkan variabel-variabel W_j tadi.

D. Regresi Ridgen

Misalkan **F** adalah padanan matrik **X** yang terpusat dan terskalakan. Jika model asal diberikan oleh

$$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_r Z_r + \varepsilon, \tag{5}$$

variabel bebas terpusat dan terskalakan yang baru adalah $f_{ju} = \frac{Z_{ju} - \bar{Z}_j}{S_{jj}^{1/2}}$, dimana $\bar{Z}_j$ adalah rata-rata Z_{ju}, $u = 1, 2, \ldots, n$, dan $S_{jj} = \sum_u \left(Z_{ju} - \bar{Z}_j\right)^2$. Jadi

$$\mathbf{F} = \begin{bmatrix} f_{11} & f_{21} & \cdots & f_{r1} \\ \vdots & \vdots & & \vdots \\ f_{1u} & f_{2u} & \cdots & f_{ru} \\ \vdots & \vdots & & \vdots \\ f_{1n} & f_{2n} & \cdots & f_{rn} \end{bmatrix} \tag{6}$$

Dan $\mathbf{F}'\mathbf{F}$ adalah matrik korelasi dari $\mathbf{Z}$ yang berukuran $r \times r$, model mempunyai ($r+1$) parameter, termasuk β_0.

Dalam regresi ridge, biasanya **Y** hanya dipusatkan, tetapi jika diinginkan dapat pula diskalakan. Jika **Y** tidak diskalakan, model (5) dapat dituliskan sebagai

$$Y_u - \bar{Y} = \beta_{1F} f_{1u} + \beta_{2F} f_{2u} + \cdots + \beta_{rF} f_{ru} + \varepsilon_u \tag{7}$$

dengan vektor estimasi kuadrat terkecil ditulis sebagai

$$b_F(\theta) = \left(\mathbf{F}'\mathbf{F} + \theta \mathbf{I}_r\right)^{-1} \mathbf{F}'\mathbf{Y} \tag{8}$$

dimana θ bilangan positif yang berada dalam rentang (0,1). Untuk mendapatkan kembali estimasi parameter dalam (5) digunakan

$$b_j(\theta) = \frac{b_{jF}(\theta)}{S_{jj}^{1/2}}, \quad j = 1, 2, \cdots, r, \tag{9}$$

dan

$$b_0(\theta) = \bar{Y} - \sum_{j=1}^{r} b_j(\theta) \bar{Z}_j , \tag{10}$$

sehingga terdapat vektor $\mathbf{b}(\theta) = \left\{b_0(\theta), b_1(\theta), \cdots, b_r(\theta)\right\}'$ yang berukuran $(r+1) \times 1$. Jika $\theta = 0$ maka $b_j(0),\ j = 0, 1, 2, \ldots, r$, merupakan estimasi kuadrat terkecil dari parameter-parameter dalam (5).

Plot $b_{jF}(\theta)$ atau plot $b_j(\theta)$ terhadap θ untuk $j = 1, 2, \ldots, r$ disebut plot *ridge trace*. Plot ini digambarkan dalam unit-unit korelasi, dimaksudkan untuk memungkinkan perbandingan langsung diantara efek-efek relatif dari berbagai koefisien, dan untuk membuang efek-efek dari penskalaan Z yang mempengaruhi ukuran estimasi koefisien-koefisien ini. θ yang meningkat mendekati 1 menyebabkan estimasi menjadi kecil, dan akan cenderung menuju nol bila θ menuju tak hingga. Kemudian dipilih sebuah nilai θ, sebut saja θ^*. Saat θ telah dipilih (sama dengan θ^*), $b_j\left(\theta^*\right)$ digunakan dalam persamaan peramalan. Persamaan yang dihasilkan bukan dari estimasi kuadrat terkecil dan berbias, tetapi lebih stabil dan memberikan keseluruhan kuadrat tengah galat yang lebih kecil, pengurangan keragaman galat yang dicapai metode ini memiliki nilai lebih ketimbang bias yang ada [9].

Salah satu metode pemilihan θ^* diajukan oleh [10], yaitu dengan menggunakan

$$\theta^* = \frac{rs^2}{\left\{\mathbf{b}_F(0)\right\}' \left\{\mathbf{b}_F(0)\right\}} \tag{11}$$

dimana

r = banyaknya parameter dalam model tanpa β_0

s^2 = kuadrat tengah residual dalam tabel analisis varian yang diperoleh berdasarkan (4) atau berdasarkan (6)(hasilnya sama)

$$\{\mathbf{b}_F(0)\}' = \{b_{1F}(0), b_{2F}(0), \ldots, b_{rF}(0)\}$$
$$= \left\{\sqrt{S_{11}}b_1(0), \sqrt{S_{22}}b_2(0), \ldots, \sqrt{S_{rr}}b_r(0)\right\}$$

E. Regresi Akar Laten

Pada referensi [11] dilakukan perluasan regresi komponen utama untuk pemeriksaan persamaan peramalan alternatif dan untuk pembuangan peubah peramal. Matrik data yang berasal dari variabel bebas yang telah dibakukan digandengkan dengan variabel respon yang telah dibakukan, dengan menempatkan variabel respon yang telah dibakukan ini di depan, untuk memperoleh

$$\mathbf{Z}^* = (\mathbf{y}, \mathbf{Z}) \tag{12}$$

dimana

$\mathbf{Z}$ = matrik $\mathbf{X}$ yang telah dipusatkan dan diskalakan (dibakukan),

$$\mathbf{y} = \frac{(\mathbf{Y} - \mathbf{1}\bar{Y})}{S_{YY}^{1/2}},$$

$$S_{YY} = \sum (Y_i - \bar{Y})^2.$$

Dengan demikian, $\mathbf{Z}^{*\prime}\mathbf{Z}^*$ merupakan matrik korelasi gandengannya (*augmented correlation matrix*).

Seperti halnya dalam metode komponen utama, akar laten dan vektor latennya kemudian dihitung, namun di sini unsur pertamanya, yaitu koefisien-Y, setiap vektor laten digunakan sebagai tolak ukur keteramalan responnya oleh vektor laten tersebut. Semakin besar unsur pertama vektor laten tersebut, semakin bermanfaatlah vektor itu di dalam meramalkan responnya dan begitu pula sebaliknya. Adanya akar laten yang kecil menandakan kemungkinan adanya ketergantungan atau ketidakbebasan linier di antara variabel-variabel bebas. Semakin kecil akar laten, semakin kuat ketidakbebasan linier tersebut. Persamaan kuadrat terkecil untuk suatu model tertentu adalah kombinasi linier terbaik (dalam pengertian kuadrat terkecil) dari semua vektor laten tersebut. Dengan membuang vektor laten yang akar laten dan unsur pertama vektor laten padanannya kecil, maka akan diperoleh persamaan kuadrat terkecil yang telah termodifikasi. Prosedur termodifikasi ini menghasilkan penduga yang berbias. Setelah persamaan kuadrat terkecil termodifikasinya diperoleh, prosedur eliminasi langkah mundur dapat digunakan untuk mengeluarkan variabel bebas dari persamaan itu.

F. Data dan teknik analisis

Data yang digunakan adalah data yang mengandung multikolinieritas, dalam hal ini digunakan data Hald. Data Hald pada awalnya merupakan data dari *Jurnal Industrial and Engineering Chemistry* tahun 1932 dengan judul *Effect of composition of Portland heat evolved during hardening*. Dengan 13 sampel, variabel bebasnya adalah banyaknya trikalsium alumina, banyaknya trikalsium silika, banyaknya tetrakalsium alumino besi(2), dan banyaknya dikalsium silika, serta evolusi panas dalam kalori per gram semen sebagai variabel tak bebasnya [9].

Prosedur penelitian yang dilakukan adalah sebagai berikut:
1. Mengambil data contoh dari referensi.
2. Menguji multikolinieritas.
3. Menerapkan masing-masing metode regresi pada data contoh dengan menggunakan bantuan software NCSS untuk regresi komponen utama dan regresi ridge.
4. Merekapitulasi hasil estimasi.
5. Membuat kesimpulan berdasarkan hasil estimasi yang diperoleh.

III. HASIL DAN PEMBAHASAN

Dengan mengasumsikan bahwa data yang digunakan telah memenuhi asumsi regresi selain asumsi nonmultikolonieritas, berikut adalah hasil uji asumsi nonmultikolieritas.

TABEL 1. NILAI VIF MASING-MASING VARIABEL BEBAS

Variabel bebas	VIF
X_1	38,4962
X_2	254,4232
X_3	282,5129
X_4	46,8684

Tabel 1 menunjukkan nilai VIF > 10 untuk setiap variabel bebas. Artinya pada data ini terdapat adanya multikolinieritas. Adanya multikolinieritas pada variabel-variabel bebas mengakibatkan model regresi yang diperoleh jauh dari akurat. Apabila tetap dilakukan analisis regresi linier akan diperoleh estimasi seperti pada Tabel 2, dengan model $\hat{Y} = 62,405 + 1,551X_1 + 0,510X_2 - 0,144X_3 + 0,102X_4$ dan R^2 = 0,982. Dapat dilihat bahwa R^2 besar namun dari uji-t ternyata variabel-variabel bebas tidak signifikan.

TABEL 2. HASIL ANALISIS REGRESI LINIER

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	62,405	70,071		,891	,399
	x1	1,551	,745	,607	2,083	,071
	x2	,510	,724	,528	,705	,501
	x3	-,144	,709	-,160	-,203	,844
	x4	,102	,755	,043	,135	,896

Sehingga diperlukan cara yang dapat mengatasi masalah multikolinieritas. Seperti yang telah dituliskan di atas, yang akan digunakan untuk mengatasi masalah multikolinieritas ini adalah Regresi Komponen Utama, Regresi Ridge, dan Regresi Akar Laten.

Dari analisis komponen utama sebagai langkah awal dari regresi komponen utama, Tabel 3 memberikan dua variabel baru, W_1 dan W_2, dengan total keragaman yang dapat dijelaskan oleh keduanya adalah sebesar 95,29%.

TABEL 3. TOTAL KERAGAMAN

Komponen	Eigenvalue	Individual %	Cumulative %
1	2,235704	55,89	55,89
2	1,576066	39,40	95,29
3	0,186606	4,67	99,96
4	0,001624	0,04	100,00

Sehingga dengan menggunakan kedua komponen utama ini diperoleh model regresi komponen utama $\hat{Y} = 95,423 - 14,777W_1 - 1,967W_2$. Model regresi ini mempunyai R^2 = 0,982. Apabila model ini dikembalikan ke variabel asal maka diperoleh model $\hat{Y}$=88,956+0,7889X_1+0,361X_2-0,327X_3-0,596X_4 dengan R^2 = 0,9650 dan MSE = 11,87298.

Sementara itu dengan menggunakan $\theta^* = 0,0131$ dari (11), diperoleh Tabel 4 yang memberikan hasil sebagai berikut.

TABEL 4. KOEFISIEN REGRESI RIDGE

Independent Variable	Regression Coefficient	Standard Error	Standardized Regression Coefficient	VIF
Intercept	83,41824			
X_1	1,299391	0,2385869	0,5081	2,8320
X_2	0,2998926	0,1044032	0,3102	3,7948
X_3	-0,3486462	0,09799983	-0,3879	3,8687
X_4	-0,1420839	0,2155091	-0,0605	2,7396

Semua variabel bebas mempunyai nilai VIF < 10, artinya sudah tidak terdapat masalah multikolinieritas lagi. Berdasarkan tabel 4, dapat dibuat model regresi ridgenya, yaitu adalah $\hat{Y} = 83,418 + 1,299X_1 + 0,299X_2 - 0,348X_3 - 0,142X_4$. Model ini mempunyai R^2 = 0,9754, dan MSE = 8,346168.

Sedangkan metode regresi akar laten dimulai dengan langkah pertama yaitu membuat matrik korelasi gandengan,

$$\begin{array}{c} \\ Y \\ Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{array} \begin{array}{c} \begin{array}{ccccc} Y & Z_1 & Z_2 & Z_3 & Z_4 \end{array} \\ \begin{bmatrix} 1,0000 & & & & \\ 0,7307 & 1,0000 & & & \\ 0,8163 & 0,2286 & 1,0000 & & \\ -0,5347 & -0,8241 & -0,1392 & 1,0000 & \\ -0,8213 & -0,2454 & -0,9730 & 0,0295 & 1,0000 \end{bmatrix} \end{array}.$$

Selanjutnya menghitung akar laten λ_j dari matrik gandengan berikut vektor laten padanannya γ_j, sebagai berikut.

Akar Laten λ_j	Y γ_{0j}	Z_1 γ_{1j}	Z_2 γ_{2j}	Z_3 γ_{3j}	Z_4 γ_{4j}
$\lambda_4 = 3,2116$	$0,5534$	$0,4012$	$0,4682$	$-0,3189$	$-0,4603$
$\lambda_3 = 1,5761$	$-0,0034$	$0,5125$	$-0,4096$	$-0,6080$	$0,4471$
$\lambda_2 = 0,1990$	$0,2112$	$0,5809$	$-0,3899$	$0,6747$	$-0,1039$
$\lambda_1 = 0,0117$	$-0,8047$	$0,4129$	$0,1884$	$-0,0531$	$-0,3791$
$\lambda_0 = 0,0016$	$0,0408$	$-0,2617$	$-0,6523$	$-0,2657$	$-0,6586$

Terlihat bahwa $\lambda_0 = 0,0016 < 0,05$ dan $\gamma_{00} = 0,0408 < 0,10$, artinya bahwa vektor baris ini hendaknya dibuang saja. Akar laten berikutnya $\lambda_1 = 0,0117 < 0,05$ menandakan adanya masalah kesingularan, namun $\gamma_{01} = -0,8047 > 0,10$ menandakan keteramalan tinggi sehingga vektor baris ini dipertahankan. Begitupun dengan λ_j lainnya dan vektor laten padanannya γ_j, dipertahankan.

Langkah keempat, berdasarkan langkah sebelumnya telah diketahui vektor mana yang dipertahankan. Vektor koefisien kuadrat terkecil termodifikasi diberikan oleh

$$\mathbf{b}^* = \begin{bmatrix} b_1^* \\ b_2^* \\ \vdots \\ b_r^* \end{bmatrix} = c\sum_j{}^* \gamma_{0j}\lambda_j^{-1} \begin{bmatrix} \lambda_{1j} \\ \lambda_{2j} \\ \vdots \\ \lambda_{rj} \end{bmatrix}$$

dimana c adalah konstanta, $c=-\left\{\sum_j{}^{*}\gamma_{0j}\lambda_j^{-1}\right\}^{-1}\left\{\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2\right\}^{1/2}$. Σ^* menyatakan penjumlahan untuk vektor yang dipertahankan. Sehingga didapat $c=-0,9356$, $b_1^*=25,9489$, $b_2^*=12,4420$, $b_3^*=-24,2319$, dan $b_4^*=-4,0399$. Jadi modelnya $Z_y=95,4231+25,9485Z_1+12,4420Z_2-24,2319Z_3-4,0399Z_4$, yang bila dikembalikan ke variabel asal model akan mempunyai R^2 = 0,982 dengan MSE = 3,7506 dan berbentuk $\hat{Y}=89,4870+1,2734X_1+0,2308X_2-0,4179X_3-0,1821X_4$. Apabila menerapkan langkah mundur pada tahap berikutnya akan menyebabkan keluarnya variabel X_4 memberikan persamaan regresi kuadrat terkecil termodifikasi sebagai $\hat{Y}=95,947+1,435X_1+0,1993X_2-0,4396X_3$.

IV. SIMPULAN DAN SARAN

Metode regresi komponen utama, regresi ridge, dan regresi akar laten yang diterapkan pada data Hald untuk mengatasi masalah multikolinieritas memberikan hasil yang tidak terlalu jauh berbeda. Regresi kompone utama memberikan model $\hat{Y}=88,956+0,7889X_1+0,361X_2-0,327X_3-0,596X_4$ dengan R^2 = 0,9650 dan MSE = 11,87298. Metode regresi ridge menghasilkan model $\hat{Y}=83,418+1,299X_1+0,299X_2-0,348X_3-0,142X_4$ dengan R^2 = 0,9754, dan MSE = 8,3468. Sedangkan model yang diperoleh dari metode regresi akar laten mempunyai R^2 = 0,982 dengan MSE = 3,7506 dan dapat dituliskan sebagai $\hat{Y}=83,418+1,299X_1+0,299X_2-0,348X_3-0,142X_4$. Tampak bahwa MSE yang timbul dari metode regresi akar laten lebih kecil dibandingkan dengan yang ditimbulkan oleh dua metode lainnya.

DAFTAR PUSTAKA

[1] Kutner, M. H., C. J, Nachtsheim., J. Neter dan W. Li, Applied Linear Statistical Model, Fifth Edition, McGraw-Hill, New York, 2004.

[2] Hoerl, A. E dan R. W. Kennard, 1970, “Ridge Regression: Biased Estimationfor Non Orthogonal Problem”, Technometrics, vol. 12, No. 1, Hal: 55-67,1970.

[3] Sembiring, R. K, Analisis Regresi, ITB, Bandung, 2003.

[4] Bilfarsah, A, “Efektifitas Metode Aditif Spline Kuadrat Terkecil Parsial dalam Pendugaan Model Regresi”, Makara, Sains, Vol. 9, No.1, Hal: 28-33, 2005.

[5] Draper, N. R dan H. Smith, Analisis Regresi Terapan, Edisi Kedua, Diterjemahkan oleh Bambang Sumantri, Jurusan Statistika FMIPA IPB, Bogor, 1992.

[6] Gujarati, D. N, Basic Econometrics, Fourth Edition, McGraw-Hill Higher Education, New York, 2003.

[7] Montgomery, D. C., Peck, E. A., *Introduction to Linear Regression Analysis*, John Wiley & Sons, New York, 1991.

[8] Morrison, D. F, Multivariate Statistical Methods, Second Edition, McGraw-Hill, New York, 1976.

[9] Draper, N. R dan H. Smith, Applied Regression Analysis, Third Edition, John Wiley and Son, Inc., New York, 1998.

[10] Hoerl, A.E., Kennard, R. W., dan Baldwin, K. F., “Ridge Regression: some simulations”, Communication in Statistics – Theory and Methods 4, Hal:105-123, 1975.

[11] Webster, J. T., Gunst, R. F., dan Mason, R. L., “Latent root regression analysis”, Technometrics 16, Hal: 513-522, 1974.