Metode Penanganan *Multikolinieritas* pada RLB:
Perbandingan *Partial Least Square dengan Ridge Regression*

Yulia Atma Putri, Margaretha Ari Anggorowati
Sekolah Tinggi Ilmu Statistik
yulia.atma@bps.go.id, m.ari@stis.ac.id

ABSTRACT

Multicolinierity between variable predictor in multiple regression is assuming violation for ordinary least square estimator (OLS). Ridge Regression (RR) and Partial Least Square Regression (PLSR were used to handle the multicolinierity problem. RR modify OLS by adding subjective bias consatant, while PLSR, generalize and combine Principal Component Analisis and multiple regression. The efficiency of these two methods will be compared based on the value of RMSE. This study simulated generating data in different level of multicolinearity, the number of variabel, and number of observation were controlled. This study results that, overall, both method equally efficient.

Keywords: RLB, OLS, *Multikolinieritas*, RR, PLSR.

ABSTRACT

Multikolinieritas antar variabel prediktor merupakan pelanggaran asumsi pada Regresi Linier Berganda (RLB) ketika estimasi dilakukan dengan menggunakan estimator *Ordinary Least Square* (OLS). *Ridge Regression* (RR) dan *Partial Least Square Regression* (PLSR) adalah metode yang umum digunakan untuk menangani masalah tersebut. RR memodifikasi metode OLS dengan menambahkan suatu konstanta bias yang bersifat subjektif, sedangkan PLSR mengeneralisasi dan mengkombinasikan metode Analisis Komponen Utama dengan metode RLB. Efisiensi kedua metode akan dibandingkan berdasarkan nilai RMSE. Data yang akan digunakan adalah data *generate* berdasarkan tingkat *multikolinieritas*, jumlah variabel, dan jumlah observasi. Perbandingan memberikan hasil bahwa secara keseluruhan kedua metode memiliki tingkat efisiensi yang sama.

Keywords: RLB, OLS, *Multikolinieritas*, RR, PLSR.

1. PENDAHULUAN

Menurut Netter (1998), analisis regresi adalah alat analisis statistik yang memanfaatkan hubungan antara dua atau lebih variabel kuantitatif sehingga suatu variabel dapat diprediksi berdasarkan variabel lainnya. Pada analisis Regresi Linier Berganda (RLB), satu variabel respon dijelaskan oleh beberapa variabel prediktor dengan menggunakan model berbentuk linier. Sebagai alat analisis statistik, regresi banyak digunakan untuk, mendeskripsikan hubungan antar variabel, mengontrol hasil observasi sesuai dengan nilai estimasinya, dan memprediksi suatu variabel berdasarkan variabel lain.

Berdasarkan teori Gauss Markov, *Ordinary Least Square* (OLS) merupakan estimator yang baik untuk mengestimasi parameter regresi. Hal ini dikarenakan estimator tersebut memiliki sifat *Best Linier Unbias Estimator* (BLUE) yaitu estimator yang tidak bias dan memiliki *variance* minimum sehingga presisi estimator ini lebih baik dari estimator lainnya.

Bowerman and O'Connell (1990), menyatakan bahwa m*ultikolinieritas* adalah masalah pada analisis regresi yang terjadi ketika variabel prediktor saling berhubungan dan saling mempengaruhi satu sama lain. Permasalahan ini sering terjadi terutama pada data yang berkaitan dengan variabel bisnis, ekonomi dan sosial.

Beberapa masalah yang muncul karena adanya *multikolinieritas* yaitu:

1. *Multikolinieritas* sempurna menyebabkan koefisien regresi tidak unik sedangkan *multikolinieritas* yang mendekati sempurna menyebabkan metode OLS tidak minimum *variance* walaupun tetap *unbias*, hal ini akan menyebabkan selang kepercayaan pendugaan parameter melebar.
2. Koefisien regresi tidak lagi signifikan jika diuji secara individu meskipun terdapat hubungan antara variabel respon dengan seluruh variabel prediktor. Dengan kata lain, terjadi kontradiksi antara hasil pengujian hipotesis parameter regresi secara individu menggunakan uji t dengan hasil pengujian secara serentak menggunakan uji F.
3. Interpretasi koefisien regresi sebagai perubahan variabel respon ketika salah satu variabel prediktor berubah satu satuan dan variabel prediktor yang lain dianggap konstan, tidak lagi dapat diterapkan.

Menurut B.M Golam Kibria (2003), ada beberapa metode yang dapat digunakan untuk mengatasi masalah *multikolinieritas. Ridge regression* (RR) adalah metode yang paling populer dan banyak digunakan dilapangan. RR menangani masalah *multikolinieritas* dengan cara memodifikasi metode OLS, yakni menambahkan persamaan estimator OLS dengan suatu konstanta bias sehingga dihasilkan koefisien yang stabil dengan *variance* yang minimum.

Paul H. Garthwate (1994) memperkenalkan *Partial Least Square Regression* (PLSR) sebagai metode yang juga dapat mengatasi masalah *multikolinieritas*. Dari penelitian yang dilakukannya disimpulkan bahwa PLSR memiliki akurasi yang lebih baik dari metode lain. Selain itu, metode ini juga dapat digunakan pada kondisi data dengan jumlah observari terbatas. PLSR merupakan kombinasi metode Analisis Komponen Utama (AKU) dengan metode RLB.

Aplikasi statistik seperti Minitab dan SPSS yang dapat digunakan untuk mengaplikasikan metode RR maupun PLSR merupakan aplikasi yang berbayar. R-*software* sudah menyediakan *packages* yang dapat digunakan untuk menerapkan kedua metode tersebut dan *software* ini dapat diperoleh secara bebas. Namun, R masih berbasis *Command Line Interface* (CLI). Penggguna harus mengetahui *code* untuk menerapkan kedua metode tersebut. Selain itu, pada *packages* PLSR masih terdapat keterbatasan, yakni pengguna harus menentukan sendiri jumlah komponen yang akan digunakan.

Penelitian ini bertujuan membandingkan efisiensi metode PLSR dengan metode RR dalam mengatasi multikolinieritas berdasarkan nilai RMSE yang dihasilkan. Untuk membantu proses simulasi akan dibuat suatu aplikasi dengan menggunakan *software* R.

2. METODOLOGI

2.1. TINJAUAN REFERENSI

Regresi Linier Berganda

Pada RLB, model linier digunakan untuk menjelaskan hubungan lebih dari satu variabel prediktor dengan satu variabel respon. Untuk data dengan (p-1) variabel prediktor, model regresi linier yang terbentuk adalah:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{ip-1} + \varepsilon_i \quad \text{.................................... (1)}$$

Y_i = nilai variabel respon pada observasi ke − i

$\beta_i, \ldots\ldots.., \beta_{p-1}$ = parameter regresi

$X_{i1}, \ldots, X_{ip-1}$ = nilai variabel prediktor ke k pada observasi ke − i

ε_i = random error ke − i, dimana $\varepsilon_i \sim \text{independent } N(0, \sigma^2)$

I = 1,2,3, … … . . , n

Model RLB dengan menggunakan pendekatan matriks menjadi:

$$Y = X\beta + \varepsilon \quad \text{... (2)}$$

Y = vektor dari variabel respon

X = matriks dari variabel prediktor, kolom pertama adalah vektor 1

X = matriks dari variabel prediktor, kolom pertama adalah vektor 1

β = vektor dari parameter regresi

ε = vektor random error

Ordinary Least Square

Estimasi koefisien regresi dari estimator OLS dengan pendekatan matriks adalah:

$$b = \left(X'X\right)^{-1}(X'Y) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

dimana $b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ \vdots \\ b_n \end{bmatrix}$ = vektor dari estimasi koefisien regresi

Penerapan metode OLS dalam mengestimasi parameter regresi harus memenuhi beberapa asumsi yaitu kenormalan *error*, konstan *variance* dari *error (homokedastisitas)*, tidak ada *multikolinieritas* antar variabel prediktor, tidak terjadi *otokorelasi* pada *error*, dan *linearitas* fungsi regresi. Jika terjadi pelanggaran pada asumsi tersebut, misalnya terjadi *multikolinieritas*, *heterokedastisitas*, atau terdapat *otokorelasi*, salah satu konsekuensinya akan menyebabkan metode OLS menghasilkan estimasi yang tetap *unbias*, tetapi tidak lagi memiliki *variance* yang minimum. Konsekuensi lain juga akan muncul jika terjadi pelanggaran pada asumsi lainnya.

Multikolinieritas

Multikolinieritas oleh Ragnar Frisch diartikan sebagai adanya hubungan linier yang pasti diantara beberapa atau semua variabel prediktor dari model RLB. Salah satu cara formal yang dapat digunakan untuk mendeteksi adanya *multikolinieritas* adalah dengan menghitung nilai *Variance Inflation Factor (VIF).* Nilai VIF menunjukan seberapa besar *variance* dari koefisien regresi meningkat akibat adanya *multikolinieritas*. Nilai maksimum VIFk yang lebih dari 10 mengindikasikan adanya *multikolinieritas* yang serius yang akan mempengaruhi estimasi *least square* sehingga diperlukan penanganan terhadap *multikolinieritas*.

Interpretasi lain dari besarnya nilai VIF yaitu:

1. VIF = 1 : Tidak terdapat *multikolinieritas*
2. 1<VIF<5 : *Multikolinieritas* sedang
3. 5<VIF<10 : *Multikolinieritas* tinggi
4. VIF > 10 : *Multikolinieritas* sangat tinggi

Besarnya nilai VIF dapat dihitung menggunakan persamaan berikut ini,

$$VIF_k = \frac{1}{1-R_k{}^2} \quad (4)$$

$R_k{}^2$ = koefisien determinasi ketika variabel prediktor ke-k diregresikan dengan p-2 variabel prediktor lainnya.

Ridge Regression

RR memodifikasi metode OLS dengan menambahkan matriks (X'X) dengan suatu konstanta bias sehingga dihasilkan koefisien regresi *ridge* yang bias dengan *variance* yang minimum. Meskipun koesifien regresi *ridge* bias, tetapi koefisien ini lebih stabil dan memiliki presisi lebih tinggi daripada koefisien regresi *unbias* pada metode *least square.* Estimator *ridge* adalah sebagai berikut

$$b^R = (X'X+cI)^{-1}X'y \quad (5)$$

b^R = vektor koefisien regresi ridge

c = konstanta ridge

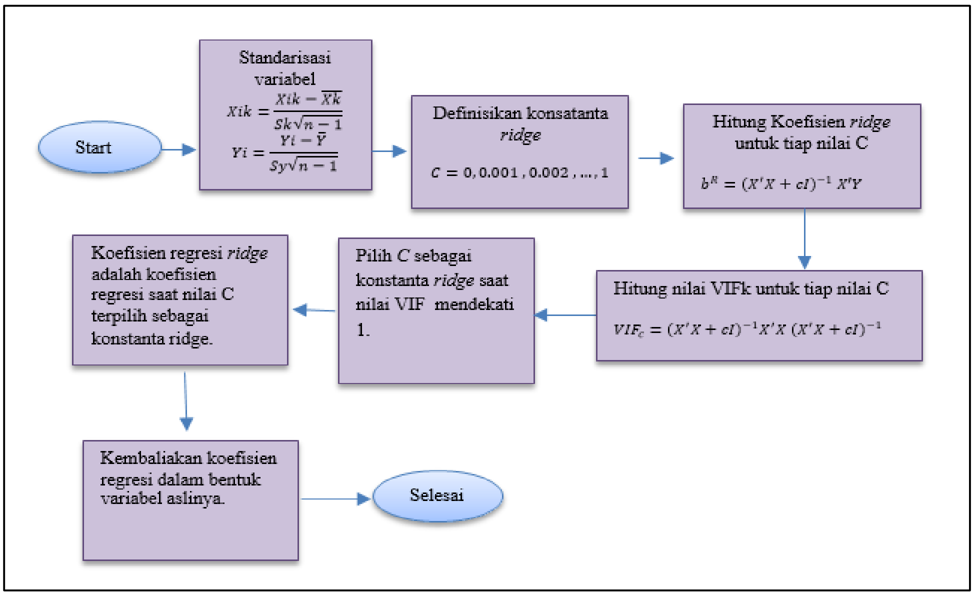
I = matrik identitas (p-1) x (p-1)

Metode yang digunakan untuk menentukan konstanta *ridge* menurut John Netter dalam bukunya *Applied Linear Regression Models* didasarkan pada *ridge trace* dan VIF. *Ridge trace* adalah plot secara simultan (p-1) koefisien regresi ridge (b^R) untuk setiap nilai konstanta (c) yang berbeda dengan nilai c yang terletak antara nol dan satu. VIF adalah nilai *variance inflation factor* untuk setiap nilai c. Adapun rumus VIF yang digunakan pada metode ini adalah:

$$VIF_c = (X'X+cI)^{-1}X'X\,(X'X+cI)^{-1} \quad (6)$$

Nilai c yang terpilih sebagai konstanta *ridge* adalah nilai c terkecil ketika b^R pertama kali dianggap menjadi stabil pada *ridge trace* dan nilai VIF pada c yang bersesuaian mendekati satu. Kemudian, koefisien regresi *ridge* adalah koefisien regresi pada saat c terpilih sebagai konstanta *ridge*. Namun, dengan cara ini koefisien regresi *ridge* yang terpilih masih bersifat subjektif. Kemungkinan akan terdapat perbedaan penentuan nilai c untuk setiap peneliti yang berbeda, sehingga akan berdampak pula pada perbedaan nilai koefisien regresi yang dihasilkan.

Adapun *pseudocode* dari metode RR digambarkan sebagai berikut:



Gambar 1. *Pseudocode* metode RR

Partial Least Square

PLSR adalah teknik yang mengeneralisasi dan mengkombinasikan AKU dengan RLB dengan tujuan untuk memprediksi atau menganalisis variabel respon berdasarkan sekumpulan variabel prediktor (Herve Abdi, 2007). Pada metode analisis komponen utama, dibentuk beberapa komponen yang merupakan kombinasi linier dari variabel prediktor kemudian dipilih beberapa komponen awal yang mampu secara maksimal menjelaskan variabilitas dari variabel prediktor. Kelemahan dari metode ini adalah komponen yang terpilih merupakan komponen yang menjelaskan variabel prediktor saja tetapi tidak ada jaminan komponen tersebut juga relevan dengan variabel respon. PLS membentuk komponen dari variabel prediktor (X) yang juga relevan dengan variabel respon (Y). Komponen yang terbentuk tersebut menampilkan secara simultan dekomposisi dari variabel X dan Y, dan semaksimal mungkin mampu menjelaskan *covariance* dari variabel X dan Y. Tahap ini disebut tahap generalisasi komponen utama. Kemudian, RLB diterapkan untuk membentuk model regresi yang dapat digunakan untuk memprediksi Y.

Secara garis besar, PLSR mencari sekumpulan w (bobot variabel X) dan c (bobot variabel Y) untuk membuat kombinasi linier dari X dan Y sehingga menghasilkan co*variance* X Y yang maksimal. Secara spesifik, tujuannya adalah untuk memperoleh komponen pertama $t = Xw$ dan $u = Yc$ dimana $w'w = I$, $c'c = I$ dan $t'u$ menjadi maksimal. Kemudian komponen tersebut dikurangi dari variabel X dan Y.

Algoritma yang digunakan untuk menghasilkan koefisien regresi PLSR adalah sebagai berikut:

1. Data variabel X dan Y distandarisasi.
2. Herve Abdi (2007) memberikan beberapa langkah algoritma PLSR, yakni sbb:
   ($\propto$ berarti normalisasi)
   Langkah 1:w $\propto$ X'u (estimasi weight X)
   Langkah 2: t $\propto$ Xw (estimasi score X / komponen)
   Langkah 3: c $\propto$ Y't (estimasi loading Y)
   Langkah 4: u = Yc (estimasi score Y)
3. Hitung b = t' u untuk memprediksi Y dari t , dan juga hitung nilai p = X't sebagai vector loading dari X.
4. Selanjutnya kurangkan efek dari komponen t pada X dan Y
   X = X-tp'
   Y = Y-tbc'
5. Nilai dari vector t,u,w,c,dan p disimpan pada matrix T,U,W,C, dan P dan nilai b disimpan sebagai elemen diagonal matriks B.
6. Ulangi langkah pada point 2 – 5 hingga matrix X dan Y mendekati 0.
7. Untuk menentukan banyak komponen yang cukup untuk menjelaskan variasi dari X dan Y, digunakan indikator $\frac{p'p}{P'P}$. Jika $\frac{p'p}{P'P} \geq 80\%$, banyaknya komponen
   sudah cukup untuk menjelaskan variasi dari variabel X dan Y.

Abdi, Valentin, dan Edelmen (1999) menerangkan bahwa algoritma iteratif diatas sama dengan metode mencari *eigenvectors.* PLSR terkait erat dengan *eigen* dan *singular value decompositions,* sehingga setelah didekomposisi didapatkan bahwa *t* adalah *eigenvector* pertama dari XX'YY'.

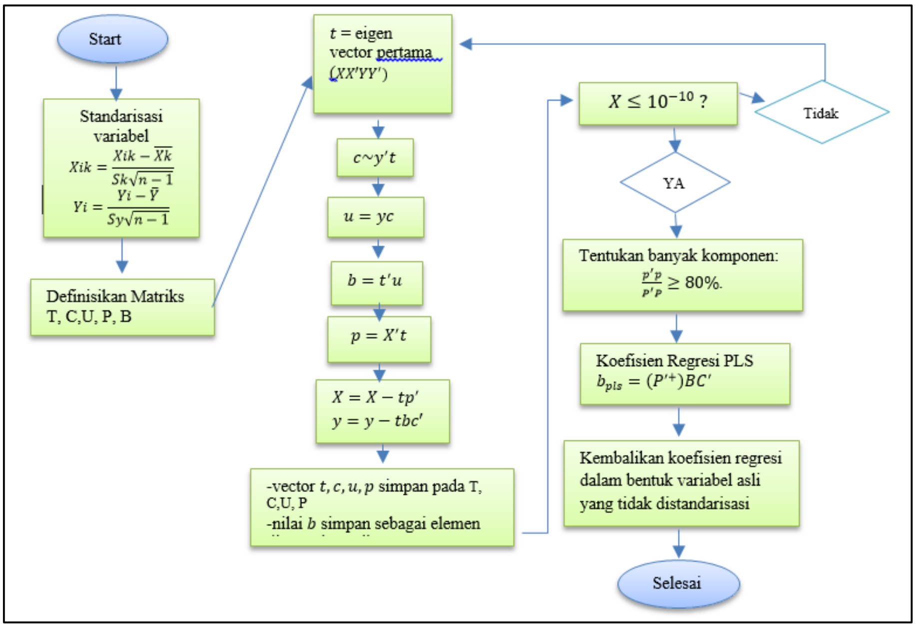
Tahap selanjutnya untuk mendapatkan koefisien regresi PLS adalah menghitung koefisien regresi PLS sesuai banyak komponen yang digunakan.

$$b_{pls} = (P'^{+})BC' \quad \text{...} (7)$$

(P'^{+})= Moore-Penrose pseudo-invers of P'

dan model regresi yang dihasilkan adalah $\hat{y}_{pls} = X\, b_{pls}$ .... (8)

Berdasarkan penjelasan diatas, adapun langkah-langkah algoritma PLSR untuk mendapatkan koefisien regresi digambarkan oleh pseudocode berikut ini:



Gambar 2. *Pseudocode* metode PLSR

Ukuran Perbandingan Metode

Root Mean square error (RMSE) merupakan akar dari rata-rata jumlah kuadrat penyimpangan antara nilai observasi dengan nilai estimasinya. Nilai ini menggambarkan seberapa dekat nilai observasi dengan nilai estimasinya. Adapun rumusnya adalah sebagai berikut:

$$RMSE = \sqrt{\frac{\sum(Y_i - \hat{Y}_i)^2}{n}} \qquad (9)$$

Jika nilai RMSE yang dihasilkan semakin kecil, maka nilai prediksi yang dihasilkan metode tersebut semakin mirip dengan nilai observasinya.

Bahasa Pemrograman R

R adalah *software* untuk komputasi statistik dan grafis yang dapat digunakan secara luas untuk pengembangan aplikasi statistik dan analisis data. Bahasa pemrograman yang digunakan dalam *software* ini adalah bahasa S dengan struktur penulisan *functional programming*. Aplikasi R bersifat *open source* yang berbasis *command line interface* dengan *code* yang dapat diperoleh secara umum.

Packages *'tcltk'* adalah *packages* yang menyediakan fungsi-fungsi untuk membuat *user interface*. *Packages* ini terdiri dari *tcl* dan *tk*. *Tcl* digunakan untuk membuat *command* dan *tk* digunakan untuk membuat *widgets*. Salah satu contoh implementasi *'tcltk'* pada R-*software* adalah *Rcommander*.

2.2 METODE ANALISIS

Sumber Data

Data yang digunakan dalam penelitian ini adalah data hasil *generate software* R yang diatur skenarionya berdasarkan tingkat *multikolinieritas* (*multikolinieritas* sedang, tinggi, sangat tinggi, dan *multikolinieritas* mendekati sempurna), jumlah variabel (2, 3, 5), dan jumlah observasi (10, 50, 100, 500). Terdapat empat skenarion data. Skenario I: dengan tingkat *multikolinieritas* sedang (0,15), skenario II: dengan tingkat *multikolinieritas* tinggi (0,5), skenario III: dengan tingkat *multikolinieritas* sangat tinggi (0,8), dan skenario IV: tingkat *multikolinieritas* mendekati sempurna (0,95). Kemudian dari tiap skenario tersebut di-*generate* data dengan jumlah variabel 2, 3, 5 dan jumlah observasi 10, 50, 100, 500.

Metode Analisis

Metode analisis yang digunakan dalam penelitian ini adalah membandingkan tingkat efisiensi metode PLSR dan RR dalam mengatasi *multikolinieritas* melalui nilai RMSE yang dihasilkan masing-masing metode dari simulasi tiap skenario data.

3. HASIL DAN PEMBAHASAN

Membandingkan efisiensi metode PLSR dengan metode RR dilihat berdasarkan nilai RMSE. Metode yang lebih efisien adalah metode yang memiliki nilai RMSE lebih kecil dari metode lain. Perbandingan tingkat efisiensi metode penanganan *multikolinieritas* ini dilakukan dengan simulasi data yang terdiri dari empat skenario berbeda yaitu skenario tingkat *multikolinieritas* sedang, tinggi, sangat tinggi, dan *multikolinieritas* mendekati sempurna. Untuk tiap skenarionya, data yang di-*generate* adalah data dengan kombinasi jumlah variabel 2, 3, 5 dan jumlah observasi 10, 50, 100, 500. Rincian hasil simulasi dari tiap skenario terdapat pada penjelasan berikut ini.

Skenario I : Tingkat *Multikolinieritas* Sedang

Pada skenario dengan tingkat *multikolinieritas* sedang, nilai RMSE yang dihasilkan dari masing-masing metode berdasarkan jumlah variabel dan jumlah observasi adalah sebagai berikut:

Tabel 1. Nilai RMSE kedua metode pada tingkat multikolinieritas sedang

p	Multiko sedang							
	PLSR				RR			
	n=10	n=50	n=100	n=500	n=10	n=50	n=100	n=500
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2	0.209	0.068	0.045	0.024	0.209	0.068	0.045	0.024
3	0.156	0.054	0.039	0.019	0.156	0.054	0.039	0.019
5	0.039	0.031	0.024	0.012	0.042	0.031	0.024	0.012

Tabel 2. Kesimpulan metode yang lebih efisien berdasarkan nilai RMSE pada tingkat multikolinieritas sedang

	n=10	n=50	n=100	n=500
(1)	(2)	(3)	(4)	(5)
p=2	PLSR=RR	PLSR=RR	PLSR=RR	PLSR=RR
p=3	PLSR=RR	PLSR=RR	PLSR=RR	PLSR=RR
p=5	PLSR	PLSR=RR	PLSR=RR	PLSR=RR

Berdasarkan nilai RMSE pada Tabel 1 dan kesimpulannya pada Tabel 2, untuk semua kondisi dengan tingkat multikolinieritas sedang, baik yang berdasarkan jumlah variabel maupun berdasarkan jumlah observasi, metode PLSR memiliki tingkat efisiensi yang sama kuat dengan metode RR. PLSR terlihat lebih efisien ketika data yang disimulasikan terdiri dari 5 variabel prediktor dan 10 observasi.

Jika efisiensi metode RR dan PLSR dibandingkan dengan metode OLS sebagai estimator pembentuk koefisien regresi saat terjadi *multikolinieritas* sedang, nilai RMSE metode OLS pada Tabel 3 menunjukan nilai yang lebih besar daripada metode RR dan PLSR pada Tabel 1, hal ini membuktikan untuk data dengan tingkat *multikoinieritas* sedang, masalah *multikolinieritas* tetap harus ditangani dengan menggunakan metode PLSR atau RR.

Tabel 3. Nilai RMSE metode OLS pada tingkat *multikolinieritas* sedang

RMSE				
	n=10	n=50	n=100	n=500
p=2	0.742	1.115	0.973	1.014
p=3	0.808	0.913	1.104	1.006
p=5	0.389	0.835	0.973	0.935

Skenario II : Tingkat *Multikolinieritas* Tinggi

Pada skenario dengan tingkat *multikolinieritas* tinggi, nilai RMSE yang dihasilkan dari masing-masing metode berdasarkan jumlah variabel dan jumlah observasi adalah sebagai berikut:

Tabel 4. Nilai RMSE kedua metode pada tingkat multikolinieritas tinggi

	Multiko tinggi							
	PLSR				RR			
	n=10	n=50	n=100	n=500	n=10	n=50	n=100	n=500
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2	0.169	0.061	0.039	0.020	0.167	0.062	0.038	0.020
3	0.148	0.033	0.024	0.013	0.138	0.034	0.025	0.013
5	0.062	0.022	0.012	0.006	0.058	0.022	0.012	0.007

Tabel 5. Kesimpulan metode yang lebih efisien berdasarkan nilai RMSE pada tingkat multikolinieritas tinggi

	n=10	n=50	n=100	n=500
(1)	(2)	(3)	(4)	(5)
p=2	RR	PLSR	RR	PLSR=RR
p=3	RR	PLSR	PLSR	PLSR=RR
p=5	RR	PLSR=RR	PLSR=RR	PLSR

Bersadarkan nilai RMSE pada Tabel 4 dan kesimpulannya pada Tabel 5, untuk kondisi 2 variabel prediktor, saat data yang disimulasikan sebanyak 10 observasi dan 100 observasi, nilai RMSE yang dihasilkan metode PLSR lebih besar daripada metode RR, ini menunjukan bahwa metode RR lebih efisien. Ketika data yang disimulasikan sebanyak 50 observasi, PLSR memiliki nilai RMSE yang lebih kecil sehingga lebih efisien. Kemudian ketika data yang digunakan sebanyak 500 observasi, kedua metode menunjukan tingkat efisiensi yang sama dalam menangani *multikoliniritas*. Selanjutnya, simulasi data dengan kondisi 3 variabel prediktor menunjukan metode RR lebih efisien untuk data sebanyak 10 observasi, untuk data dengan 50 dan 100 observasi efisiensi yang lebih baik diberikan oleh metode PLSR, dan untuk data dengan 500 observasi kedua metode memberikan RMSE yang sama besar sehingga sama efisiennya. Berlanjut untuk simulasi data dengan 5 variabel prediktor, RR lebih efisien untuk jumlah observasi 10, untuk 50 dan 100 observasi kedua metode sama efisiennya, dan untuk observasi 500 PLSR lebih efisien.

Dengan demikian, dapat disimpulkan untuk tingkat *multikolinieritas* tinggi, dengan berbagai kondisi data yang disimulasikan memberikan kesimpulan bahwa sebagian besar kondisi menunjukan metode PLSR lebih efisien. Terdapat pula sebagian kondisi lain yang menunjukan metode RR lebih efisien. Namun, perbedaan efisiensi berdasarkan nilai RMSE yang terdapat pada Tabel 4 tersebut tidak jauh berbeda.

Jika dibandingkan efisiensi metode RR dan PLSR dengan metode OLS sebagai estimator pembentuk koefisien regresi ketika terdapat *multikolineritas* tinggi, nilai RMSE metode OLS pada Tabel 6 menunjukan nilai yang lebih besar daripada metode RR dan PLSR pada Tabel 4, hal ini membuktikan untuk data dengan tingkat *multikoinieritas* tinggi, masalah *multikolinieritas* tetap harus ditangani dengan menggunakan metode PLSR atau RR.

Tabel 6. Nilai RMSE metode OLS pada tingkat multikolinieritas tinggi

	RMSE			
	n=10	n=50	n=100	n=500
p=2	0.932	0.919	0.953	1.001
p=3	1.037	0.829	0.959	0.979
p=5	0.894	1.028	0.836	0.951

Skenario III : Tingkat *Multikolinieritas* Sangat Tinggi

Pada skenario dengan tingkat *multikolinieritas* sangat tinggi, nilai RMSE yang dihasilkan dari masing-masing metode berdasarkan jumlah variabel dan jumlah observasi adalah sebagai berikut:

Tabel 7. Nilai RMSE kedua metode pada tingkat multikolinieritas sangat tinggi

	Mutiko = 0,8							
	PLSR				RR			
	n=10	n=50	n=100	n=500	n=10	n=50	n=100	n=500
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2	0.082	0.056	0.032	0.017	0.079	0.057	0.032	0.017
3	0.104	0.027	0.020	0.010	0.091	0.027	0.020	0.010
5	0.027	0.016	0.009	0.004	0.026	0.016	0.009	0.005

Tabel 8. Kesimpulan metode yang lebih efisien berdasarkan nilai RMSE pada tingkat multikolinieritas sangat tinggi

	n=10	n=50	n=100	n=500
(1)	(2)	(3)	(4)	(5)
p=2	RR	PLSR	PLSR=RR	PLSR=RR
p=3	RR	PLSR=RR	PLSR=RR	PLSR=RR
p=5	RR	PLSR=RR	PLSR=RR	PLSR

Bersadarkan nilai RMSE pada Tabel 7 dan kesimpulannya pada Tabel 8, untuk kondisi 2 variabel prediktor, metode RR lebih efisien untuk data dengan 10 observasi, PLSR lebih efisien saat jumlah data sebanyak 50 observasi, dan efisiensi kedua metode sama untuk data dengan observasi 100 dan 500. Ketika kondisi 3 variabel prediktor, metode RR lebih efisien untuk data dengan 10 observasi, untuk jumlah observasi lainnya kedua metode memiliki tingkat efisiensi yang sama. Begitu pula untuk data dengan variabel prediktor sebanyak 5, RR efisien untuk observasi sebanyak 10, tingkat efisiensi kedua metode sama saat observasi sebanyak 50 dan 100, dan PLSR lebih unggul untuk observasi 500.

Dari rincian tabel diatas dapat disimpulkan bahwa untuk tingkat *multikolinieiritas* sangat tinggi, dengan berbagai kondisi data yang disimulasikan beberapa kondisi menunjukan metode PLSR lebih efisien. Sebagian kondisi lainnya menunjukan metode RR yang lebih efisien. Namun, sebagian besar kondisi menunjukan kedua metode memiliki tingkat efisiensi yang sama kuat. Meskipun demikian, perbedaan efisiensi berdasarkan nilai RMSE yang terdapat pada tabel diatas tidak jauh berbeda.

Jika dibandingkan efisiensi metode RR dan PLSR dengan metode OLS sebagai estimator pembentuk koefisien regresi ketika *multikolinieritas* sangat tinggi, maka berdasarkan nilai RMSE yang ditunjukan oleh Tabel 9, metode OLS menghasilkan nilai RMSE yang lebih besar, hal ini membuktikan buntuk data dengan tingkat *multikoinieritas* sangat tinggi, masalah *multikolinieritas* tetap harus ditangani dengan menggunakan metode PLSR atau RR.

Tabel 9. Nilai RMSE metode OLS pada tingkat *multikolinieritas* sangat tinggi

RMSE				
	n=10	n=50	n=100	n=500
p=2	0,539	1,154	0,984	1,010
p=3	0,578	0,904	0,994	1,028
p=5	0,467	0,999	0,944	0,967

Skenario IV: *Multikolinieritas* Mendekati Sempurna

Pada skenario dengan tingkat *multikolinieritas* yang mendekati sempurna, nilai RMSE yang dihasilkan dari masing-masing metode berdasarkan jumlah variabel dan jumlah observasi adalah sebagai berikut:

Tabel 10. Nilai RMSE kedua metode pada tingkat multikolinieritas mendekati sempurna

p	Mutiko = 0,95							
	PLSR				RR			
	n=10	n=50	n=100	n=500	n=10	n=50	n=100	n=500
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2	0.134	0.052	0.030	0.015	0.134	0.052	0.030	0.015
3	0.090	0.024	0.016	0.009	0.090	0.024	0.016	0.009
5	0.029	0.014	0.008	0.004	0.029	0.014	0.008	0.004

Tabel 11. Kesimpulan metode yang lebih efisien berdasarkan nilai RMSE pada tingkat multikolinieritas mendekati sempurna

	n=10	n=50	n=100	n=500
(1)	(2)	(3)	(4)	(5)
p=2	PLSR=RR	PLSR=RR	PLSR=RR	PLSR=RR
p=3	PLSR=RR	PLSR=RR	PLSR=RR	PLSR=RR
p=5	PLSR=RR	PLSR=RR	PLSR=RR	PLSR=RR

Berdasarkan nilai RMSE pada Tabel 10 dan kesimpulan pada Tabel 11, metode PLSR memiliki nilai RMSE yang sama besar dengan metode RR untuk semua kondisi data yang disimulasikan. Dengan demikian, ketika terjadi multikolinieritas yang mendekati sempura, baik metode PLSR maupun metode RR sama efisiennya.

Jika dibandingkan efisiensi metode RR dan PLSR dengan metode OLS sebagai estimator pembentuk koefisien regresi ketika *multikolinieritas* mendekati sempurna, nilai RMSE metode OLS lebih besar daripada metode RR dan PLSR seperti yang ditunjukan oleh Tabel 12, hal ini membuktikan untuk data dengan tingkat *multikoinieritas* mendekati sempurna, masalah *multikolinieritas* tetap harus ditangani dengan menggunakan metode PLSR atau RR.

Tabel 12. Nilai RMSE metode OLS pada tingkat multikolinieritas sangat tinggi

RMSE				
	n=10	n=50	n=100	n=500
p=2	0,978	1,105	0,921	0,941
p=3	0,838	0,902	0,847	1,024
p=5	0,639	1,003	0,943	0,977

4. KESIMPULAN DAN SARAN

Berdasarkan simulasi yang telah dilakukan, untuk tingkat *multikolinieritas* sedang, tinggi, dan sangat tinggi, beberapa kondisi menunjukan salah satu metode lebih efisien dan sebagian besar kondisi lainnya menunjukan tingkat efisien yang sama untuk kedua metode. Sedangkan tingkat *multikolnieritas* mendekati sempurna, hasil simulasi menunjukan kedua metode sama efisien untuk semua kondisi data. Meskipun untuk beberapa kondisi salah satu metode lebih efisien, tetapi selisih perbedaan nilai RMSE dari kedua metode tidak berbeda jauh, sehingga dapat dikatakan kedua metode memiliki tingkat efisiensi yang sama. Jika kompleksitas dari metode penanganan multikolinieritas tidak diperhatikan dalam perhitungan koefisien regresi, metode PLSR lebih baik digunakan karena merupakan metode yang tidak bias.

Untuk penelitian lebih lanjut, metode penanganan *multikolinieritas* yang lain dapat dikaji tingkat efisiensinya dengan metode PLSR dan RR untuk menemukan metode lain yang mungkin lebih efisien dalam mengatasi masalah *multikolinieritas* pada RLB. Selain itu, proses simulasi data dapat dikembangkan untuk mencari *cutting point* atau nilai batas sehingga suatu metode dikatakan lebih efisien dari metode lainnya.

5. DAFTAR PUSTAKA

Abdi, H. 2007. Partial Least Square Regression (PLS-Regression). *Ensyclopedia of Measurement and Statistics*, 1-13. http://plstools.googlecode.com/svn-history/r13/trunk/Documentation/Abdi-PLSR2007-pretty.pdf (Diakses 21 Maret, 2015).

Adnan, N., Ahmad, M. H., dan Adnan, R. 2006. A Comparative Study On Some Methods For Handling Multicolinierity Problems. *Journal of Matematics*, Vol. 22 No. 2, 109-119. http://www.matematika.utm.my/index.php/matematika/article/viewFile/179/174. (Diakses 19 Maret, 2015).

Garthwaite, P. H. 1994. An Intepretation of Partial Least Squares. *Journal of the American Statistical Association*, Vol. 89 No. 425, 122-127. http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1994.10476452 . (Diakses, 20 Maret 2015).

Kibria, B. M. 2003. Performance of Some New Ridge Regression Estimators. *Communications in Statistics-Simulation and Computation*, Vol. 32 No. 2, 419-435. http://www.tandfonline.com/doi/abs/10.1081/SAC-120017499. (Diakses, 24 Maret 2015).

Netter, J., Wasserman, W., dan H.Kutner, M. 1989. *Applied Linier Regression Model (2th ed).* Homewood: IRWIN Book Team.

Gujarati, D. N. 2010. *Dasar-Dasar Ekonometrika* . Jakarta: Salemba Empat.

Soemartini. 2008. *Penyelesaian Multikolinieritas Melalui Metode Ridge Regression [Skripsi]*. Jatinangor: Universitas Padjadjaran.