

# Metode Regresi Ridge dengan Iterasi HKB dalam Mengatasi Multikolinearitas

Nurhasanah\*

## Abstrak

Regresi berganda dengan peubah bebas saling berkorelasi (multikolinearitas) merupakan masalah yang sering terjadi dalam melakukan analisis data. Salah satu akibat yang terjadi adalah ragam penduga parameter regresi yang besar sehingga dugaan koefisien regresi cenderung menjadi tidak signifikan. Salah satu pendekatan yang sering digunakan adalah regresi komponen utama. Penelitian ini mengkaji metode regresi *ridge* dengan iterasi HKB dibandingkan dengan regresi komponen utama dengan menggunakan data simulasi. Hasil penelitian menunjukkan bahwa metode regresi *ridge* dengan iterasi HKB memberikan hasil yang lebih baik, sehingga metode regresi *ridge* dengan iterasi HKB dapat dijadikan alternatif dalam penggunaan untuk mengatasi multikolinearitas.

**Kata kunci:** Korelasi, multikolinearitas, regresi komponen utama, regresi ridge, iterasi HKB.

## Abstract

Multiple regressions with independent variables have correlation (multicollinearity) is a cases that often done in data analysis. The variance of hinger parameter regression is one of the result, thus the test of hypothesis is dispose in significant. One approach frequently used is the principal component regression. This research consider the ridge regression method with HKB (Hoerl, Kennard, and Balwin) iteration compared with the principal komponen regression using simulated data. The results showed that the ridge regression method with HKB iteration give better results, so that the ridge regression method with HKB iteration can be an alternative in use to solve multicollinearity.

**Keywords:** *Correlation, multicollinearity, principal component regression, ridge regression, HKB iteration.*

## 1. Pendahuluan

Analisis regresi adalah aplikasi statistika yang digunakan untuk membantu mengungkapkkan hubungan atau pengaruh satu atau lebih peubah penjelas terhadap peubah respons. Dalam analisis regresi linear terdapat beberapa asumsi untuk menghasilkan penduga yang tak bias dan terbaik. Asumsi-asumsi tersebut di antaranya adalah galat menyebar normal dengan rata-rata nol dan ragamnya  $\sigma_\epsilon^2$ , serta tidak terdapat multikolinearitas antar peubah bebas. Masalah multikolinearitas penting diperhatikan dan perlu diatasi karena akan berdampak terhadap pendugaan dan pengujian koefisien regresi.

---

\*Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Syiah Kuala, Nangroe Aceh Darussalam

Masalah multikolinieritas muncul akibat terdapat fungsi linear hampir konstan atau sering disebut dengan tingginya korelasi linear di antara dua atau lebih peubah bebas. Multikolinieritas di antara beberapa peubah bebas berarti adanya kebergantungan linear antara dua atau lebih peubah bebas tersebut. Apabila terdapat Multikolinieritas di antara peubah bebas maka pendugaan parameter model regresi dengan menggunakan metode kuadrat terkecil akan menghasilkan penduga yang tak bias tetapi mungkin penduga tersebut mempunyai ragam yang besar [9]. Ragam yang besar ini akan mengakibatkan pengujian hipotesis mengenai signifikansi koefisien regresi cenderung menerima  $H_0$ , yang berarti koefisien regresi tersebut tidak berbeda nyata dengan nol. Ada beberapa metode untuk memeriksa adanya masalah multikolinieritas di antaranya menggunakan matriks korelasi, *variance inflation factors* (VIF) dan analisis sistem eigen pada matriks  $\mathbf{X}^t\mathbf{X}$ .

Penyelesaian masalah multikolinieritas ini telah dikembangkan berbagai pendekatan. Suatu pendekatan yang mungkin dilakukan adalah dengan menggunakan regresi himpunan terbaik (best subset regression), regresi stepwise (stepwise regression), lihat [2], namun jika seluruh peubah bebas berkorelasi tinggi, hal ini sulit dilakukan dan tidak akan memperoleh solusi yang baik. Pendekatan kedua dalam mengatasi multikolinieritas disebut penggunaan penduga regresi berbias. Beberapa metode yang berkenaan dengan pendekatan regresi berbias ini diantaranya adalah regresi gulud (*ridge regression*), regresi akar laten (latent root regression), dan regresi komponen utama (*principal component regression*).

Pada penelitian ini akan dikaji penggunaan regresi ridge dalam mengatasi multikolinieritas. Regresi ridge merupakan modifikasi dari metode kuadrat terkecil dengan cara menambah tetapan bias  $q$  yang kecil kepada nilai diagonal matriks  $\mathbf{X}^t\mathbf{X}$ . Besarnya tetapan bias  $q$  mencerminkan besarnya bias dalam koefisien penduga ridge,  $q$  yang bernilai nol merupakan implementasi dari metode kuadrat terkecil. Beberapa peneliti lain telah mencoba untuk membandingkan satu metode dengan yang lain, di antaranya [1] yang menyatakan bahwa *regresi ridge* dengan pemilihan nilai  $k$  menggunakan rumusan HKB (Hoerl, Kennard dan Baldwin) yang ditemukan tahun 1975, lihat [8], tanpa iterasi lebih unggul daripada regresi komponen utama untuk masalah multikolinieritas dengan derajat rendah. Kemudian [5] menyimpulkan bahwa metode kuadrat kecil parsial lebih unggul daripada metode regresi komponen utama dan regresi ridge trace. Kubokawa dan Srivastava (lihat [4]) memperkenalkan regresi ridge dengan penduga Bayes yang merupakan metode yang lebih unggul daripada regresi ridge yang dalam pemilihan nilai  $k$ -nya melalui *ridge trace*. Selanjutnya [4] melakukan pengkajian metode regresi *ridge* dengan menggunakan penduga Bayes sebagai metode alternatif lain untuk menanggulangi multikolinieritas, yang kemudian dibandingkan dengan metode regresi *ridge* dengan pemilihan nilai  $k$  baik melalui *ridge trace* maupun dengan iterasi HKB dimana dari hasil penelitian tersebut disimpulkan regresi ridge dengan iterasi HKB lebih unggul dari regresi ridge dengan cara lain. Pada penelitian ini dilakukan pengkajian regresi ridge dengan iterasi HKB dan regresi komponen utama dengan menggunakan data simulasi.

## 2. Metodologi Penelitian

Data yang digunakan dalam penelitian ini adalah data simulasi. Simulasi yang dilakukan terdiri dari dua cara pembangkitan data. Data simulasi I dibangkitkan dari rancangan matriks korelasi. Pada simulasi I ini ingin dilihat multikolinieritas berdasarkan nilai korelasi sederhananya, dimana korelasi tinggi menunjukkan adanya multikolinieritas. Data simulasi I dirancang dengan tiga tingkat korelasi yaitu korelasi rendah (0,3), sedang (0,6) dan tinggi (0,9). Dalam data kondisi korelasi rendah dan sedang tidak dapat dijadikan dasar adanya multikolinieritas tetapi belum tentu tidak ada multikolinieritas. Berdasarkan kondisi tersebut, simulasi II dirancang dengan cara membuat satu peubah bebas merupakan fungsi dari peubah

bebas lainnya. Simulasi II dikondisikan pada multikolinearitas rendah, sedang dan tinggi yang dibangun dengan cara menentukan nilai  $R^2$  (0,3; 0,6; 0,9) dari regresi fungsi peubah bebas.

Data simulasi dibangkitkan dengan menggunakan bantuan program Matlab. Banyaknya data yang diteliti dengan ukuran sampel  $n = 10, 20, 30, 50, 100, 1.000$  dan  $1.500$ , terdiri dari tiga peubah bebas ( $X_1, X_2$  dan  $X_3$ ) menyebar normal ganda dan sebuah peubah tak bebas ( $Y$ ) serta galat ( $\varepsilon$ ) menyebar Normal (0,1).

## 2.1 Prosedur Simulasi

Simulasi data yang dilakukan adalah untuk memperoleh data dengan masalah multikolinearitas dalam berbagai kondisi yaitu tingkatan korelasi rendah, sedang dan tinggi, dimana ketiga tingkatan tersebut ditandai dengan nilai korelasi 0,3, 0,6 dan 0,9 serta multikolinearitas rendah, sedang dan tinggi yang ditandai dengan nilai  $R^2$  berada sekitar nilai 0,3, 0,6 dan 0,9. Pembangkitan data untuk setiap ukuran sampel dengan program Matlab menggunakan perintah *mvnrand*.

### Simulasi I

1. Membangkitkan peubah bebas :  $X_1, X_2, X_3 \sim \text{MVN}(\mu, \Sigma)$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} \text{ dan } \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}$$

Pembangkitan data untuk kondisi korelasi rendah, sedang dan tinggi dilakukan dengan menetapkan nilai  $\mu$  dimana dalam penelitian ini ditentukan sebarang yaitu  $\mu = (2,34 ; 3,45 ; 4,56)$  dan  $\sigma_{11} = \sigma_{22} = \sigma_{33} = 1$  serta untuk  $\sigma_{ij} = r_{ij}$ ,  $i, j = 1, 2, 3$  dan  $i \neq j$  adalah nilai korelasi 0,3, 0,6, dan 0,9.

2. Membangkitkan peubah tak bebas ( $Y$ ) :  $\hat{Y} = X\beta + \varepsilon$ , dimana  $\varepsilon \sim N(0,1)$  dengan langkah pengerjaan menetapkan nilai  $\beta$  yaitu (2,89 ; 1,54 ; 1,76 ; 0,32).

### Simulasi II

1. Membangkitkan  $X_1, X_2 \sim N(2,5)$ .
2. Menentukan  $X_3$  yang merupakan fungsi dari  $X_1$  dan  $X_2$  yaitu:  
 $X_3 = 2X_1 + 3X_2 + \varepsilon_1$ , dimana  $\varepsilon_1 \sim N(0, \sigma^2)$ . Pembangkitan data untuk kondisi multikolinearitas rendah, sedang, dan tinggi ditandai dengan mengubah nilai  $\sigma^2$ , yaitu  $\sigma^2 = 120$ ,  $\sigma^2 = 35$ , dan  $\sigma^2 = 5$ .

Tabel. 1 Nilai  $R^2$  berdasarkan  $\sigma^2$  ditetapkan

Ukuran contoh (n)	$\sigma^2$		
	120	35	5
	$R^2$ rendah	$R^2$ sedang	$R^2$ tinggi
10	0,31399	0,63641	0,91178
20	0,35037	0,61094	0,90977
30	0,32916	0,60023	0,90893
50	0,31408	0,59476	0,90919
100	0,30373	0,59183	0,90951
1000	0,29459	0,58833	0,90903

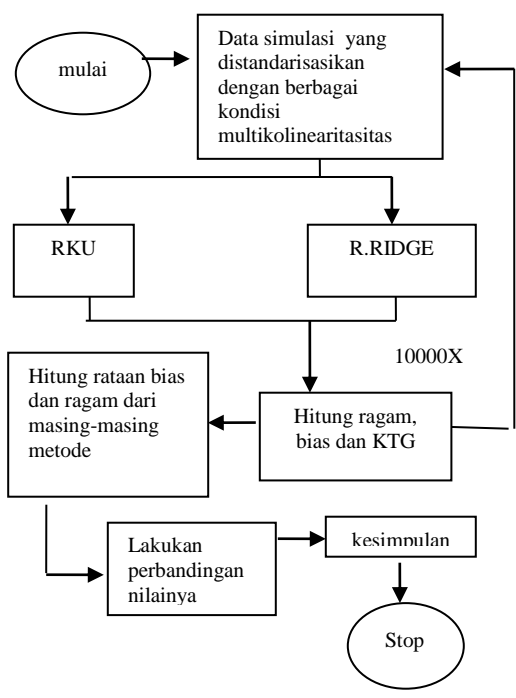
---

1500	0,29471	0,58862	0,90915
------	---------	---------	---------

---

Membangkitkan peubah tak bebas ( $Y$ ) :  $Y = X\beta + \varepsilon_2$ , dimana  $\varepsilon_2 \sim N(0,1)$  dengan langkah pengerjaan menetapkan nilai  $\beta$  yaitu (2,85 ; 1,35 ; 1,23 ; 1,12).

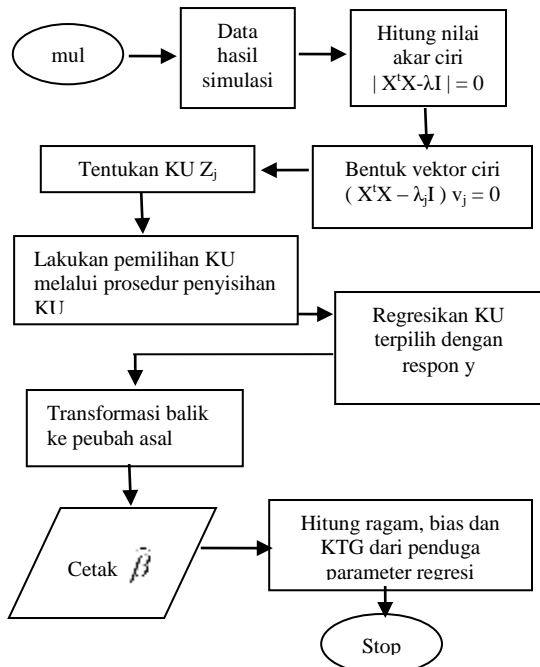
**Diagram Alur Penelitian**



Gambar 1. Diagram Alur Penelitian.

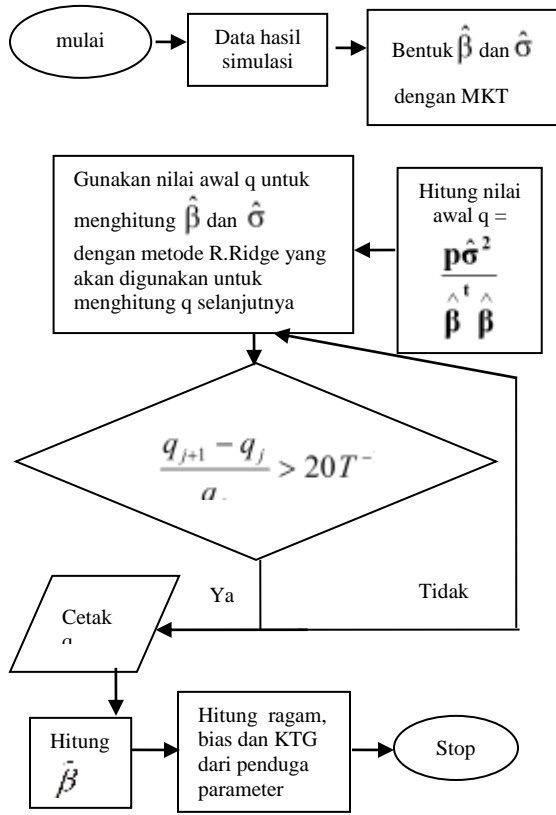
**2.2 Algoritma Metode yang Diteliti  
Metode Regresi Komponen Utama (RKU)**

Nurhasanah



Gambar 2. Algoritma Metode Regresi Komponen Utama (RKU).

Metode Regresi Ridge Iterasi HKB



Gambar 3. Algoritma Metode Regresi Ridge Iterasi HKB (R.Ridge).

### 3. Hasil dan Pembahasan

Dalam mengatasi multikolinearitas banyak metode yang dapat digunakan untuk mengatasinya. Penelitian ini menggunakan beberapa metode dimana untuk melihat kebaikan metode-metode tersebut digunakan nilai bias, ragam dan kuadrat tengah galat (KTG/MSE) penduga parameter sebagai kriterianya. Kriteria nilai bias dan ragam penduga parameter untuk melihat metode terbaik adalah nilai yang minimum, tetapi kadang-kadang kedua kriteria ini terjadi baku timbang (*trade off*) dimana bias yang kecil menghasilkan ragam yang besar. Kedua kriteria tersebut kemudian digabungkan menjadi nilai kuadrat tengah galat yang merupakan kriteria relevan dalam menentukan suatu metode pendugaan adalah mempunyai bias, ragam dan MSE penduga parameter yang minimum.

Simulasi yang dilakukan adalah simulasi pembangkitan data multikolinearitas dengan menggunakan distribusi normal ganda untuk beberapa ukuran sampel yaitu 10, 20, 30, 50, 100, 1.000 dan 1.500, serta simulasi dilakukan sebanyak 10.000 kali. Data yang dibangkitkan dengan kondisi multikolinearitas ini digunakan pada metode regresi komponen utama (RKU) dengan cara tanpa menghilangkan komponen utama ( $u=0$ ), menghilangkan satu komponen utama ( $u=1$ ) dan dengan menghilangkan dua komponen utama ( $u=2$ ) dan regresi *ridge* dengan iterasi HKB. Penentuan metode terbaik dalam mengatasi multikolinearitas ini digunakan nilai rata-rata dari ragam, bias dan MSE dugaan parameter karena proses yang dilakukan berulang-ulang. Simulasi yang dilakukan adalah data tingkat multikolinearitas rendah, sedang dan tinggi.

Hasil penelitian untuk data simulasi tidak ditampilkan dalam pembahasan ini, namun dari ke empat metode yang dilakukan dapat dikatakan untuk simulasi RKU0 tidak termasuk dalam metode untuk mengatasi multikolinearitas karena RKU0 sama dengan analisis regresi dengan MKT. RKU1 dan RKU2 adalah bagian dari RKU. Secara umum untuk simulasi pada kondisi multikolinearitas tinggi, jika ditinjau dari nilai rata-rata ragam, bias dan MSE dugaan parameter minimum dipengaruhi oleh ukuran contoh. Pada ukuran sampel besar dan kecil RKU mempunyai rata-rata ragam dugaan parameter minimum diikuti oleh regresi *ridge*. Untuk bias dan MSE dugaan parameter, jika ukuran sampel kecil terjadi ketidak-konsistenan sehingga sulit menentukan urutan metode berdasarkan nilai rata-rata minimum, namun tidak untuk ukuran sampel besar. Ukuran sampel besar menunjukkan nilai rata-rata bias dan MSE dugaan parameter paling minimum pada metode regresi *ridge* diikuti oleh RKU.

### 4. Kesimpulan

Pada dasarnya banyak metode yang dapat digunakan untuk mengatasi multikolinearitas, di antaranya yang digunakan dalam penelitian ini yaitu regresi komponen utama dan regresi *ridge* dengan iterasi HKB. Dari kedua metode yang digunakan tersebut, regresi *ridge* dengan iterasi HKB memberikan hasil yang lebih baik. Hasil kajian menunjukkan metode regresi *ridge* dengan iterasi HKB dapat dijadikan alternatif dalam penggunaan untuk mengatasi multikolinearitas.

### Daftar Pustaka

- [1] Anwar C., 1986. Studi Simulasi Masalah Multikolinearitas pada Regresi Linear Ganda dengan Beberapa Metode Pendugaan dan Penerapan. *Tesis*. IPB Bogor.
- [2] Draper N. dan Smith., 1992. *Analisis Regresi Terapan*, Terjemahan Bambang Sumantri, Edisi Ke-2. Gramedia, Jakarta.
- [3] Gujarati D., 1995. *Ekonometrika Dasar*, Terjemahan Sumarno Zain. Erlangga, Jakarta.

*Nurhasanah*

- [4] Gusriani N., 2004. Regresi ridge dengan penduga Bayes untuk Mengatasi Multikolinearitas. *Tesis*. IPB Bogor.
- [5] Herwindiati D.E., 1997. Pengkajian Regresi Komponen Utama, Regresi ridge dan Regresi Kuadrat Terkecil Parsial untuk Mengatasi Multikolinearitas. *Tesis*. IPB Bogor.
- [6] Jolliffe I.T., 1986. *Principal Component Anaalysis*. Spinger-Verlag, New York.
- [7] Naes T., 1985. Multivariate Calibration When the Error Covariance Matrix is Structured. *Technometrics*. V-27, no.3:301-311.
- [8] Ryan T.P., 1996. *Modern Regression Method*. John Wiley & Sons, New York.
- [9] Walpole R.E. dan Myers, R.H. 1995. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan*, Terjemahan Bambang Sumantri, Edisi ke-4. Penerbit ITB. Bandung