ISBN:1858-1382

J M S K
Jurnal Matematika, Statistika, & Komputasi
http://journal.unhas.ac.id/index.php/jmsk
Vol. 10, No. 1, 6-13, Juli 2013

Perturbasi Nilai Eigen dalam Mengatasi Multikolinearitas

Andi Yuni Deviyanti[1], Andi Kresna Jaya[1], Anisa[1]

Abstrak

Multikolinieritas adalah salah satu pelanggaran asumsi dalam regresi linier berganda. Multikolinieritas terjadi ketika variabel bebas memiliki korelasi dengan variabel bebas yang lain. Apabila dalam suatu data terdapat multikolinieritas maka koefisien regresi yang dihasilkan tidak dapat memberikan hasil analisis yang mewakili sifat atau pengaruh variabel bebas terhadap variabel terikat, selain itu variansi yang dihasilkan juga besar. Jika terjadi multikolinieritas maka rasio antara nilai eigen terbesar dan nilai eigen terkecil akan melebihi 100 bahkan bisa lebih dari 1000. Untuk itu, dalam penelitian ini ditawarkan suatu solusi untuk mengatasi multikolinearitas yaitu dengan menggunakan perturbasi pada nilai eigen. Perturbasi (τ) diberikan dengan menambahkan nilai τ, yang nilainya berkisar antara 0 dan 1, pada diagonal matriks nilai eigen.

Kata Kunci:Multikolinieritas, nilai eigen, regresi ridge, perturbasi, variansi, bias.

1. Pendahuluan

Analisis regresi linier berganda digunakan untuk mengetahui hubungan antara variabel terikat Y dengan beberapa variabel bebas X_i, $i =$ (1, 2, …,k). Dari model akan diukur hubungan tersebut menggunakan metode estimasi.

Salah satu metode estimasi yang dapat digunakan dalam mengestimasi parameter adalah metode kuadrat terkecil biasa atau *Ordinary Least Square* (*OLS*). Parameter yang akan diestimasi adalah $\beta_1, \beta_2, \dots, \beta_k$. Salah satu asumsi yang harus dipenuhi dalam regresi linier berganda di atas adalah tidak ada hubungan linier antara kolom-kolom atau variabel bebas X, di mana $X = (X_1, X_2, \dots, X_k)$. Namun dalam kenyataannya, hal ini tidak selamanya dipenuhi.

Jika dalam suatu model terjadi multikolinieritas, maka akan mengakibatkan penggunaan metode kuadrat terkecil dalam mengestimasi parameter akan terganggu, dan galat yang dihasilkan akan menjadi besar. *Montgomery* dan *Peck* (1992) dalam [6] menyatakan bahwa adanya multikolinieritas pada variabel bebas X juga menyebabkan rasio antara nilai eigen terbesar dan nilai eigen terkecil lebih dari 100 atau melebihi 1000.

2. Multikolinearitas dan Cara Mengatasinya Melalui Perturbasi Nilai Eigen

Multikolinieritas artinya adanya hubungan linier antara variabel-variabel bebas. Multikolinieritas dapat dideteksi dengan beberapa cara, antara lain menghitung koefisien korelasi sederhana antara sesama variabel bebas, menghitung nilai *Variation Inflation Factor* (*VIF*), dan sistem nilai eigen dari $X^*{}'X^*$. Akar-akar karakteristik atau nilai eigen dari $X^*{}'X^*$adalah $\lambda_1, \lambda_2, \dots, \lambda_k$ yang dapat digunakan untuk mengukur adanya multikolinieritas. Jika ada satu atau

[1]*Jurusan Matematika FMIPA Universitas Hasanuddin, email: andikresna@yahoo.com, nkalondeng@gmail.com*

lebih hubungan linier di dalam data, maka terdapat nilai eigen yang lebih kecil dibandingkan nilai eigen yang lain Multikolinieritas dapat diukur dalam bentuk rasio dari nilai terbesar dan terkecil dari nilai eigen, yaitu $\varphi = \frac{\lambda_{maks}}{\lambda_{min}}$ yang disebut nilai kondisi dari matriks korelasi. Montgomery dan Peck (1992) dalam [6] memberikan kategori multikolinieritas berdasarkan harga $\varphi = \frac{\lambda_{maks}}{\lambda_{min}}$, jika:

- $\varphi < 100$ maka disebut multikolinieritas rendah atau tidak ada multikolinieritas
- $100 \leq \varphi < 1000$ maka disebut multikolinieritas agak kuat
- $\varphi \geq 1000$ maka disebut multikolinieritas sangat kuat

Apabila dalam suatu data terdapat multikolinieritas, maka solusi dari metode kuadrat terkecil $\hat{\beta} = (X'X)^{-1}X'Y$ tidak akan menghasilkan hasil estimasi yang tepat karena invers dari $X'X$ tidak ada. Untuk itu, dari metode kuadrat tekecil tersebut ditambahkan suatu nilai pada diagonal $X'X$ agar determinannya tidak nol dan menghasilkan invers. Metode ini dinamakan metode regresi Ridge, pertama kali diperkenalkan oleh Hoerl dan Kennard pada tahun 1970 ([3] dan [4]).

Berdasarkan persamaan linier

$$Y = X\beta + \varepsilon \qquad (1)$$

maka

$$\varepsilon = Y - X\beta$$

Dengan menggunakan metode pengali lagrange yang meminimumkan fungsi $\varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta)$, dengan fungsi kendala $\beta'\beta = l^2$. Sehingga didapatkan fungsi baru

$$G = (Y - X\beta)'(Y - X\beta) + \tau(\beta'\beta - l^2)$$

yang memenuhi syarat

$$\left.\frac{\partial G}{\partial \beta}\right|_{\hat{\beta}} = 0$$

Sehingga

$$G = (Y - X\beta)'(Y - X\beta) + \tau(\beta'\beta - l^2)$$
$$= Y'Y - 2\beta'X'Y + \beta'X'X\beta + \tau(\beta'\beta - l^2)$$

$$\left.\frac{\partial G}{\partial \beta}\right|_{\hat{\beta}} = 0$$

$$-2X'Y + 2X'X\hat{\beta} + 2\tau\hat{\beta} = 0$$
$$\hat{\beta} = (X'X + \tau I)^{-1}X'Y \qquad (2)$$

Persamaan (2) merupakan estimator dari regresi ridge.

Jika $X'X$ adalah matriks yang dapat didiagonalisasi, maka terdapat P yangmerupakan matriks vektor eigen sehingga $X'X = PDP'$.

Selanjutnya, $X'X + \tau I$ dari dapat diubah menjadi $PDP' + \tau I$

Maka,

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\boldsymbol{p}} &= (PDP' + \tau I)^{-1}X'\boldsymbol{Y} \\ &= P(D + \tau I)^{-1}P'X'\boldsymbol{Y} \qquad (3)\end{aligned}$$

Karena P adalah matriks vektor eigen, maka $\boldsymbol{p_i} = [p_{1i} \quad p_{2i} \quad \cdots \quad p_{ni}]'$, dimana $\boldsymbol{p_i}$ adalah vektor eigen yang bersesuaian dengan λ_i. Maka persamaan (3) menjadi

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{p}} = P(D + \tau I)^{-1}P'X'\boldsymbol{Y}$$

$$= \sum(\lambda_i + \tau)^{-1}\boldsymbol{p_i}\boldsymbol{p_i'}X'Y$$

Misalkan $\boldsymbol{p_i'}X'\boldsymbol{Y} = c_i$, maka

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{p}} = \sum(\lambda_i + \tau)^{-1}c_i\boldsymbol{p_i} \qquad (4)$$

Persamaan (4) merupakan estimator perturbasi nilai eigen yang diperoleh dengan menambahkan gangguan pada nilai eigen.

3. Sifat Estimator Perturbasi Nilai Eigen

Sifat dari perturbasi nilai eigen yaitu

1. Bias

   Estimator perturbasi nilai eigen merupakan estimator bias karena estimasinya ditambahkan dengan gangguan sebesar τI pada matriks diagonal nilai eigen. Sehingga bias yang dihasilkan oleh estimator perturbasi nilai eigen adalah

   $$\begin{aligned}E\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{p}}\right) &= E\{P(D + \tau I)^{-1}P'X'\boldsymbol{Y}\} \\ &= \boldsymbol{\beta} - \tau P(D + \tau I)^{-1}P'\boldsymbol{\beta}\end{aligned}$$

   Jadi, bias pada estimator perturbasi nilai eigen sebesar

   $$\boldsymbol{\beta} - \tau P(D + \tau I)^{-1}P'\boldsymbol{\beta}$$

2. Variansi

Variansi dari penaksir perturbasi nilai eigen lebih kecil dibandingkan dengan variansi penaksir OLS.

$$Var(\widehat{\boldsymbol{\beta}}_p) = E\{(\widehat{\boldsymbol{\beta}} - E(\widehat{\boldsymbol{\beta}}))(\widehat{\boldsymbol{\beta}} - E(\widehat{\boldsymbol{\beta}}))'\}$$
$$= \sigma^2 PD((D+\tau I)^2)^{-1}P'$$

Jadi, entri diagonal dari variansi penaksir dari perturbasi nilai eigen $\sigma^2 PD((D+\tau I)^2)^{-1}P'$ lebih kecil dibandingkan $\sigma^2\left(X'X\right)^{-1}$ atau bisa ditulis dengan $\sigma^2 PD^{-1}P$.

$$\sigma^2 PD((D+\tau I)^2)^{-1}P' < \sigma^2 PD^{-1}P'.$$

4. Penerapan Metode Perturbasi Nilai Eigen

Dalam menerapkan konsep perturbasi nilai eigen, maka akan dijelaskan dengan sebuah data sekunder dari Badan Pusat Statistik (BPS) MakassartentangPertumbuhan Ekonomi di kota Makassar. Data ini diamati selama 10 tahun dari tahun 2000–2009 dengan empat variabel bebas X, yaitu X_1 = besarnya Pendapatan Regional dari Segi Lapangan Usaha Listrik, Gas, dan Air (persen), X_2 = Besarnya Pendapatan Regional dari Segi Lapangan Usaha Bangunan (persen), X_3= Besarnya Pendapatan Regional dari Segi Lapangan Usaha Angkutan dan Komunikasi (persen), X_4= Besarnya Pendapatan Regional dari Segi Lapangan Usaha Keuangan, Persewaan dan Jasa Perusahaan (persen), dan Y = Pertumbuhan Ekonomi Menurut Harga konstan (persen).

Tabel 1. Data Pertumbuhan Ekonomi di Kota Makassar.

Y	X_1	X_2	X_3	X_4
15.37	3.04	7.52	13.38	5.02
7.3	2.35	7.38	14.95	7.1
7.14	2.27	7.51	14.35	7.08
8.6	2.1	7.64	15.29	7.97
10.17	1.82	7.68	15.91	10.18
7.16	2.06	8.02	13.38	10.15
8.09	1.95	7.8	15.92	10.19
8.11	1.93	7.85	16.2	10.47
10.52	1.99	8.34	16.14	10.55
9.2	1.99	8.6	16.17	10.79

Dalam metode perturbasi nilai eigen, dilakukan proses standarisasi terlebih dahulu. Nilai eigen yang akan diganggu adalah nilai eigen dari matriks $X^{*\prime}X^{*}$ yang merupakan hasil transformasi variabel bebas dari proses standarisasi. Transformasi untuk variabel yang distandarisasi adalah

$$X_{ki}^{*} = \frac{1}{\sqrt{n-1}}\left(\frac{X_{ki}-\bar{X}_k}{s_k}\right) \tag{5}$$

$$Y_i^* = \frac{1}{\sqrt{n-1}}\left(\frac{Y_i-\bar{Y}}{S_Y}\right) \tag{6}$$

dimana

$$S_k = \sqrt{\frac{\sum_{i=1}^{n}(X_{ki}-\bar{X}_k)^2}{n-1}} \tag{7}$$

$$S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{n-1}} \tag{8}$$

Dari transformasi diatas bentuk persamaan regresi

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

menjadi

$$Y_i^* = \beta_1^* X_{1i}^* + \beta_2^* X_{2i}^* + \cdots + \beta_1^* X_{ki}^* + \varepsilon_i \tag{9}$$

Dari hasil proses standarisasi, maka dihasilkan variabel-variabel X yang telah distandarisasi. Bentuk $X^{*\prime}X^*$ yang telah distandarisasi sama dengan bentuk matriks korelasinya.

$$X^{*\prime}X^* = R_{xx} = \begin{bmatrix} 1 & r_{12} & r_{13} & \dots & r_{ik} \\ r_{21} & 1 & r_{23} & \dots & r_{ik} \\ r_{31} & r_{32} & 1 & \dots & r_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & r_{k3} & \dots & 1 \end{bmatrix}$$

$$X^{*\prime}X^* = \begin{bmatrix} 1 & -0.4630 & -0.7133 & -0.9010 \\ -0.4630 & 1 & 0.4375 & 0.7316 \\ -0.7133 & 0.4375 & 1 & 0.6704 \\ -0.9010 & 0.7316 & 0.6704 & 1 \end{bmatrix}$$

Dari matriks $X^{*\prime}X^*$ maka dapat diperoleh matriks nilai eigen dan matriks vektor eigen sebagai berikut

$$P = \begin{bmatrix} -0.5911 & 0.5051 & 0.3458 & -0.5254 \\ 0.3072 & 0.2576 & 0.8072 & 0.4333 \\ -0.0598 & 0.7434 & -0.4686 & 0.4735 \\ -0.7434 & -0.3549 & 0.0963 & 0.5587 \end{bmatrix}$$

$$D = \begin{bmatrix} 0.0353 & 0 & 0 & 0 \\ 0 & 0.3469 & 0 & 0 \\ 0 & 0 & 0.6350 & 0 \\ 0 & 0 & 0 & 2.9828 \end{bmatrix}$$

Dari matriks nilai eigen tersebut, maka akan ditambahkan dengan suatu gangguan sebesar τI, dimana I adalah matriks identitas.

Pemilihan Nilai Perturbasi$\boldsymbol{\tau}$

Dalam estimasi melalui perturbasi nilai eigen, bias yang diinginkan tidak terlalu besar dan dapat memperkecil variansi dari penaksir. Jika dalam suatu data terdapat multikolinieritas, maka terdapat suatu nilai eigen yang kecil daripada nilai eigen yang lainnya dan dapat menyebabkan rasio antara nilai eigen terbesar dan nilai eigen terkecil lebih dari 100 atau bahkan lebih dari 1000. Maka nilai eigennya ditambahkan dengan suatu gangguan yang relatif kecil antara $0 < \tau < 1$.

Berdasarkan penjelasan sebelumnya, sifat dari penaksir perturbasi nilai eigen memiliki variansi yang minimum, dan nilai dari *Variance Inflation Factor* (VIF) dapat dituliskan dalam bentuk matriks diagonal dari

$$PD((D+\tau I)^2)^{-1}P' = \sum_{i=1}^{n} \frac{\lambda_i}{(\lambda_i+\tau)^2}\boldsymbol{p_i}\boldsymbol{p_i'}$$

Dari nilai VIF diatas, maka diharapkan nilai τ yang diperoleh adalah nilai yang relatif kecil dan dapat memperkecil VIF, selain itu nilai τ yang diharapkan dapat memperkecil rasio antara nilai eigen terbesar dan nilai eigen terkecil.

Dalam tulisan ini digunakan metode penentuan nilai τ menurut Farebrother (1975) dalam [6], yaitu dengan

$$\tau = \frac{\alpha^2}{\widehat{\boldsymbol{\beta}}'\,\widehat{\boldsymbol{\beta}}}$$

dimana:

$\alpha^2 = (\boldsymbol{Y} - X\boldsymbol{\beta})'(\boldsymbol{Y} - X\boldsymbol{\beta})$
$\widehat{\boldsymbol{\beta}} =$ estimator OLS yang distandarisasi

Dari rumus di atas didapatkan

$$\tau = \frac{\alpha^2}{\widehat{\boldsymbol{\beta}}'\,\widehat{\boldsymbol{\beta}}} = 0.0832$$

Nilai τ yang didapatkan menurut metode Farebrother yaitu 0.0832.

Setelah nilai τ didapatkan, maka akan dilihat nilai VIF dan rasio nilai eigen untuk beberapa nilai τ, dan akan dilihat bagaimana nilai VIF dan rasio nilai eigen saat $\tau = 0.0832$

Nilai VIF untuk $\tau = 0.0832$memberikan nilai VIF kurang dari 10 yaitu $VIF(\hat{\beta}_1) = 1.5909$, $VIF(\hat{\beta}_2) = 1.2232$, $VIF(\hat{\beta}_3) = 1.3867$, $VIF(\hat{\beta}_4) = 1.7353$, dan rasio nilai eigen saat $\tau = 0.0832$ yaitu 25.8734

Nilai $\widehat{\boldsymbol{\beta}}_{\boldsymbol{p}}$ berdasarkan estimator perturbasi nilai eigen untuk $\tau = 0.0832$, yaitu

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{p}} = P(D+\tau I)^{-1}P'X'\boldsymbol{Y}$$

$$= \begin{bmatrix} 0.9197 \\ 0.2275 \\ 0.2836 \\ 0.0549 \end{bmatrix}$$

Jadi, persamaan regresi yang diperoleh melalui perturbasi nilai eigen pada saat $\tau = 0.0832$ adalah

$$\hat{Y}^* = 0.9197X_1^* + 02275X_2^* + 0.2836X_3^* + 0.0549X_4^*$$

5. Kesimpulan

Berdasarkan pembahasan mengenai perturbasi nilai eigen dalam mengatasi multikolinieritas, maka diperoleh beberapa simpulan berikut:

1. Estimator perturbasi nilai eigen didapatkan dengan menambahkan gangguan sebesar τ pada matriks diagonal nilai eigen, sehingga estimatornya adalah

   $$\widehat{\boldsymbol{\beta}}_p = P(D+\tau I)^{-1}P'X'Y$$

   atau dapat ditulis juga dalam bentuk

   $$\widehat{\boldsymbol{\beta}}_p = \sum_{i=1}^{n} (\lambda_i + \tau)^{-1}\, c_i \boldsymbol{p_i}$$

   dimana $c_i = \boldsymbol{p}_{\boldsymbol{i}}'X'Y$.

2. Dari data dapat dilihat entri diagonal dari variansi perturbasi nilai eigen lebih kecil daripada variansi OLS.

   $$\sigma^2 PD((D+\tau I)^2)^{-1}P' < \sigma^2 PD^{-1}P'$$

   $$Var(\widehat{\boldsymbol{\beta}}_p) = \begin{bmatrix} \mathbf{0.1588} & 0.0058 & 0.0513 & 0.0716 \\ 0.0058 & \mathbf{0.1220} & -0.0087 & -0.0572 \\ 0.0513 & -0.0087 & \mathbf{0.1383} & -0.0354 \\ 0.0716 & -0.0572 & -0.354 & \mathbf{0.1733} \end{bmatrix}$$

   Sedangkan variansi dari OLS,

   $$Var(\widehat{\boldsymbol{\beta}}) = \begin{bmatrix} \mathbf{29.6359} & -16.4939 & 0.3332 & 6.7590 \\ -16.4939 & \mathbf{10.6910} & 0.9422 & -3.7972 \\ 0.3332 & 0.9422 & \mathbf{0.4412} & -0.0021 \\ 6.7590 & -3.7972 & -0.0021 & \mathbf{1.7040} \end{bmatrix}$$

   Dapat dilihat bahwa variansi perturbasi nilai eigen lebih kecil dibandingkan variansi dari OLS.

Nilai VIF dari perturbasi nilai eigen juga lebih kecil dari OLS dan kurang dari 10, yaitu $VIF(\hat{\beta}_1) = 1.5909$, $VIF(\hat{\beta}_2) = 1.2232$, $VIF(\hat{\beta}_3) = 1.3867$, $VIF(\hat{\beta}_4) = 1.7353$.Sehingga dapat dikatakan metode perturbasi nilai eigen dapat mengatasi mutikolinieritas

Daftar Pustaka

[1] Anton H., 1987. *Aljabar Linier Elementer.* Erlangga, Indonesia.

[2] Harville D.A., 2008. *Matrix Algebra from a Statistician's Perspective.* Springer Science+Business Media LLC, New York.

[3] Hoerl A.E. dan Kennard R.W., 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, Vol. 12, No. 1. pp. 55-67.

[4] Hoerl A.E. dan Kennard R.W., 1970. Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics,* Vol. 12, No. 1. pp. 69-82.

[5] Lancaster P. dan Miron T., 1985. *The Theory of Matrices Second Edition with Application.* Academic Press Inc., Florida.

[6] Nduka E.C. dan Ijomah M.A., 2012. The Effects Of Perturbing Eigenvalues In The Presence Of Multicollinierity. *Electron. J. App. Stat. Anal*., Vol. 5. pp. 304–311.

[7] Pradipta N., 2009. *Metode Regresi Ridge untuk Mengatasi Model Regresi Linier Berganda yang Mengandung Multikolinieritas*. Universitas Sumatera Utara, Medan.

[8] Putri A.P., 2011. *Penggunaan Metode Ridge Trace dan Variance Inflation Factors* (*VIF*) *pada Regresi Ridge*. Universitas Negeri Yogyakarta, Yogyakarta.

[9] Rawling J.O., 1998. *Applied Regression Analysis: A Research Tool., 2nd Ed*. Springer-Verlag Inc., New York.

[10] Singh R. *et al.*, 2012. Solution of Multicollinearity by Ridge Regression. *International Journal Of Research In Computer Application & Management*, Vol. 2, Issue 28.