Analisis Regresi *Ridge Robust* (RR) untuk Mengatasi Masalah Multikolinearitas dan Pencilan pada Data Proksimat di Muara Niru, Jelawatan, dan Enim

[1]Asti Rahmatika, [2]Suliadi, [3]Teti Sofia Yanti
[1,2,3] Program Studi Statistika, Fakultas MIPA, Universitas Islam Bandung, Jl. Tamansari No. 1 Bandung 40116
e-mail:[1]asti.rahmatika@yahoo.com, [2] suliadi@gmail.com, [2] tetisofiayanti@gmail.com

Abstrak. Metode Kuadrat Terkecil (MKT) merupakan salah satu metode penaksiran parameter regresi. Metode MKT mudah terpengaruh terhadap kehadiran pencilan dan terjadinya multikolinearitas. Metode *ridge* digunakan untuk mengatasi masalah multikolinearitas. Sedangkan metode *robust* digunakan untuk mengatasi masalah pencilan. Pada skripsi ini dilakukan penggabungan antara metode regresi *ridge* dan *robust* agar dapat menangani masalah multikolinearitas dan pencilan. Dalam skripsi ini akan diterapkan metode regresi *ridge robust* pada data proksimat. Dari hasil analisis yang telah dilakukan, dapat diambil kesimpulan bahwa metode regresi *ridge robust* LAV memiliki hasil yang lebih baik dibandingkan dengan regresi *ridge robust* LMS dan LTS dalam menangani masalah multikolinearitas.

Kata Kunci : Regresi Metode Kuadrat Terkecil (MKT), Pencilan, Multikolinearitas, Regresi *Ridge*, Regresi *Robust*, Regresi *Ridge Robust*.

1. Pendahuluan

Dalam suatu penelitian, analisis regresi dapat digunakan untuk membantu melihat pengaruh antara satu atau lebih variabel bebas terhadap variabel tak bebas. Jika dalam analisis hanya melibatkan sebuah variabel bebas, maka analisis yang digunakan adalah analisis regresi linear sederhana. Tetapi pada kenyataannya dalam kehidupan sehari-hari permasalahan yang bisa diatasi adalah dengan regresi linear berganda.

Salah satu cara untuk mendapatkan koefisien regresi linear berganda adalah menggunakan Metode Kuadrat Terkecil (MKT). Penaksiran parameter model regresi dengan menggunakan MKT akan menghasilkan penaksir yang tak bias tetapi penaksir tersebut mempunyai varians yang besar jika ada kekolinearan antar variabel bebas (Walpole dan Myers, 1995). Dikarenakan variabel bebasnya lebih dari dua variabel maka dimungkinkan akan terjadi masalah multikolinearitas. Akibat adanya pengaruh yang ditimbulkan oleh multikolinearitas tersebut diperlukan solusi untuk mengatasinya. Salah satu teknik analisis yang digunakan untuk mengatasi masalah tersebut adalah dengan menggunakan regresi *ridge*. Penaksir regresi *ridge* memiliki MSE (*Mean Square Error*) lebih kecil dari penaksir MKT (Hoerl dan Kennard, 1970).

Salah satu asumsi MKT yang harus terpenuhi lainnya adalah asumsi normalitas. Salah satu penyebab ketidaknormalan data adalah adanya pencilan. Pengaruh pencilan dalam analisis data dapat dibedakan berdasarkan asal pencilan tersebut, yaitu yang berasal dari variabel tak bebas (y-pencilan; titik *influence*) dapat dideteksi dengan melihat nilai TRES atau berasal dari variabel bebasnya (x-pencilan) dengan melihat *leverage value*. Secara umum pencilan tidak selalu merupakan pengamatan berpengaruh ataupun sebaliknya. Pendeteksian pengamatan berpengaruh ditentukan oleh ukuran nilai *DFFITS*, *DFBETAS*, *Cook's Distance* dan *Covratio*. Untuk mengatasi masalah pengamatan berpengaruh, salah satu metode yang dapat digunakan yaitu metode regresi *robust*. Regresi *robust* diperkenalkan oleh Andrews (1972) sebagai model regresi yang digunakan apabila distribusi dari galat tidak normal. Dalam regresi *robust* terdapat beberapa metode penaksiran parameter seperti penaksir *Least Absolute Value* (LAV), penaksir *Least Median Square* (LMS), dan penaksir *Least Trimmed Square* (LTS) (Chen, 2002).

Apabila dalam model regresi linear berganda terdapat multikolinearitas antar vaiabel bebas, dan pencilan pada variabel bebas dan tak bebas maka metode yang digunakan untuk mengatasi hal tersebut adalah metode regresi *ridge robust*. Dikarenakan metode regresi *ridge* dan *robust* tidak dapat menangani masalah pencilan dan multikolinearits secara bersamaan, akan lebih baik jika menggabungkan kedua metode tersebut (Myers, 1990).

Dalam penelitian ini akan diterapkan metode regresi *ridge robust* yang digunakan untuk mengestimasi mengenai data pertambangan yaitu data proksimat. Dimana variabel yang akan digunakan adalah variabel *gross calorfic value* sebagai variabel tak bebas, *total moisture, moisture in air dried sample, ash, volatile matter, fixed carbon,* HGI (*Hardgrove grindability index*) dan TSG (*True Specific Gravit*iy) sebagai variabel bebas.

2. Tinjauan Pustaka

2.1. Metode Kuadrat Terkecil (MKT)

Analisis regresi adalah analisis statistika yang bertujuan untuk memodelkan hubungan antara variabel bebas dengan variabel tak bebas. Apabila kita dihadapkan pada suatu masalah penaksiran atau peramalan nilai suatu variabel, katakanlah Y, berdasarkan variabel lain, X. Secara umum, variabel tak bebas dapat dihubungkan oleh k buah variabel bebas, X_1, X_2, …, X_k, maka model yang digunakan adalah:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

Model diatas yang disebut sebagai model regresi linear berganda karena melibatkan lebih dari satu variabel bebas. Nilai parameter dari persamaan diatas dapat ditentukan menggunakan MKT sehingga diperoleh,

$$\hat{\beta} = (X^t X)^{-1}(X^t Y)$$

MKT merupakan metode penaksiran parameter yang meminimalkan jumlah kuadrat sisaan (galat). Metode ini merupakan kelas penaksir yang memiliki sifat BLUE. Menurut teorema Gauss-Markov, setiap penaksir MKT yang asumsinya terpenuhi akan bersifat BLUE (*Best Liniear Unbiased Estimator*).

2.2. Multikolinearitas

Istilah multikolinearitasitas pertama kali ditemukan oleh Ragnar Frisch pada tahun 1934 yang berarti adanya hubungan linear diantara beberapa atau semua variabel bebas dalam model regresi. Masalah multikoliniearitas hanya akan muncul pada model regresi linear berganda. Suatu model yang bebas multikolinearitas adalah model yang memiliki nilai Faktor *Variance Inflation Factors* (VIF) > 10 mengindikasikan terdapatnya multikolinearitas (Myers, 1990).

2.3. Pemeriksaan Data Berpengaruh

Istilah pencilan (*outliers*) merujuk pada suatu pengamatan yang dalam beberapa hal tidak konsisten dengan observasi lainnya yang ada dalam suatu data. Sedangkan istilah pencilan dalam galat merujuk pada titik data yang galat pengamatannya lebih besar daripada apa yang diharapkan dari keragaman acak itu sendiri. Kemudian, istilah data yang berpotensi sebagai data berpengaruh digunakan pada suatu pengamatan yang merupakan data pencilan dalam satu atau lebih variabel bebas. Pendeteksian pencilan dapat dilakukan dengan melihat *leverage value* dan nilai TRES.

Metode yang digunakan dalam mengidentifikasi pencilan terhadap variabel X adalah nilai pengaruh (*leverage* value). Nilai pengaruh (h_{ii}) dari pengamatan (X_i, Y) menunjukkan besarnya peranan$Y$ terhadap $\hat{Y}$ dan didefinisikan sebagai,

$$h_{ii} = x_i^t (X^t X)^{-1} x_i$$

Dengan i = 1,2, . . . , n, $\boldsymbol{x_i^T} = [X_{i1} \;\; X_{i2} \; \dots \; X_{ik}]$ adalah vektor baris yang berisi nilai – nilai dari k variabel bebas pada pengamatan ke-i. Nilai h_{ii} berada diantara 0 dan 1, yaitu $0 \le h_{ii} \le 1$. Jika h_{ii} lebih besar dari $\frac{2p}{n}$, dengan $p = k + 1$ maka pengamatan ke-i dikatakan pencilan terhadap X.

Menurut Draper dan Smith (1998) metode yang digunakan dalam mengidentifikasi pencilan terhadap variabel Y adalah *Studentized Deleted Residual* (TRES) yang didefinisikan sebagai:

$$TRES_i = \varepsilon_i \left[\frac{n-k}{JKS(1-h_{ii})-\varepsilon_i^2}\right]^{\frac{1}{2}}; \qquad i = 1, 2, \dots, n$$

Dimana, $\varepsilon_i = Y_i - \hat{Y}_i$, n = banyaknya pengamatan, k = banyaknya variabel bebas, JKS = Jumlah Kuadrat Sisa.

Hipotesis untuk menguji adanya pencilan:

H_0 : Pengamatan ke – i bukan pencilan

H_1 : Pengamatan ke – i merupakan pencilan

TRES adalah statistik uji untuk mengetahui pencilan terhadap Y. Kriteria uji yang melandasi keputusan adalah tolak H_0 jika nilai $|TRES_i| \le t_{(\frac{\alpha}{2}, n-k-1)}$, dan terima H_0 jika nilai $|TRES_i| > t_{(\frac{\alpha}{2}, n-k-1)}$. Dimana $t_{\frac{\alpha}{2}}$ adalah distribusi t-student.

Secara umum pencilan tidak selalu merupakan pengamatan berpengaruh ataupun sebaliknya. Pendeteksian pengamatan berpengaruh ditentukan oleh ukuran nilai *DFFITS*, *DFBETAS*, *Cook's Distance* dan *Covratio*. DFFITS digunakan untuk mengetahui pengaruh suatu pengamatan ke-i terhadap model regresi yang ditinjau dari nilai taksirannya. Besarnya nilai DFFITS adalah:

$$\text{DFFITS}_i = (RStudent)_i \sqrt{\frac{h_{ii}}{1-h_{ii}}}$$

Dalam rumus diatas R-Student merupakan ukuran pencilan (dalam variabel y atau variabel tak bebas) dan h_{ii} yang merupakan indikator pencilan dalam variabel X atau variabel bebas. Suatu pengamatan ke-i dikatakan berpengaruh apabila pengamatan tersebut memiliki nilai $|\text{DFFITS}_i| > 2/\sqrt{n}$ (Hajarisman, 2010).

DFBETAS digunakan untuk menyatakan pengaruh suatu pengamatan ke-i terhadap koefisien ke-k. Besarnya nilai DFBETAS adalah:

$$\text{DFBETAS}_{k,i} = \frac{b_k - b_{k,-i}}{s_i \sqrt{c_{kk}}}$$

di mana c_{kk} adalah unsur diagonal ke-k matrik $(X^t X)^{-1}$ Karena $b_{k,-i}$ adalah koefisien regresi variabel bebas ke-k yang diperoleh tanpa mengikutsertakan pengamatan ke-i, maka $\text{DFBETAS}_{k,i}$ dapat diartikan sebagai besarnya perubahan yang terjadi terhadap koefisien regresi b_k jika pengamatan ke-i tidak diikutsertakan dalam pendugaan model regresi. Suatu pengamatan ke-i dikatakan berpengaruh terhadap koefisien ke-k apabila pengamatan tersebut memiliki nilai $|\text{DFBETAS}_{k,i}| > 2\sqrt{p/n}$ (Hajarisman, 2010).

2.4. Penaksir Regresi Ridge

Salah satu masalah utama dalam metode penaksir regresi adalah multikolinearitas. Terdapat beberapa teknik atau metode untuk mengatasi masalah multikolinearitas. Model regresi *ridge* telah dianjurkan dalam literatur sebagai alternatif penaksir MKT untuk masalah multikolinearitasitas (Hoerl & Kennard, 1970). Metode regresi *ridge* dikembangkan oleh Hoerl dan Kennard dengan cara menambahkan konstanta yang bernilai positif λ terhadap elemen diagonal X^tX.

Penaksir regresi *ridge* bagi $\hat{\beta}$ untuk MKT adalah:

$$\hat{\beta}_R = (X^{*t}X^* + \lambda_{LS}I)^{-1}X^{*t}Y^*$$

Dimana *I* adalah matriks identitas berukuran (k x k) dan λ adalah sebuah bilangan yang positif atau $\lambda \geq 0$, umumnya λ terletak antara interval $0 < \lambda < 1$. Salah satu penaksir λ diusulkan oleh Hoerl *et. al.* (1975) seperti berikut ini,

$$\lambda_{LS} = \frac{ps_{LS}^2}{\hat{\beta}_{LS}^T\hat{\beta}_{LS}}$$

Dimana $$s_{LS}^2 = \frac{(Y - X\hat{\beta}_{LS})^T(Y - X\hat{\beta}_{LS})}{n-K}$$

Transformasi persamaan Regresi *Ridge Robust* LAV, LMS, dan LTS ke dalam bentuk awal dapat menggunakan rumus,

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}_1 + \hat{\beta}_2\bar{X}_2 + \cdots + \hat{\beta}_k\bar{X}_k$$

$$\hat{\beta}_0 = \bar{Y} - \sum_{k=1}^{K}\beta_k\bar{X}_k$$

$$\hat{\beta}_k = \left(\frac{s_y}{s_{x_k}}\right)\beta_k^* \quad ; k = 1,2,\ldots.K$$

Setelah nilai $\hat{\beta}$ didapatkan, maka model regresi berganda yang siap digunakan untuk penaksir (Neter hal. 414).

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1X_1 + \hat{\beta}_2X_2 + \cdots + \hat{\beta}_kX_k$$

2.5. Penaksir Regresi *Robust*

Regresi *robust* diperkenalkan oleh Andrews (1978) sebagai model regresi yang digunakan apabila distribusi dari galat tidak normal atau adanya beberapa pencilan yang berpengaruh pada model. Dalam regresi *robust* terdapat beberapa metode penaksiran parameter seperti penaksir *Least Absolute Value* (LAV), penaksir *Least Median Square* (LMS), dan penaksir *Least Trimmed Square* (LTS) (Chen, 2002).

Least Absolute Value dikenal dengan berbagai nama, yaitu *Minimum Absolute Deviation regression*, regresi *Least Absolute Deviation* (LAD), dan regresi *Minimum Sum of Absolute Errors*. Dielman (1984) menyatakan bahwa penaksir LAV untuk mendapatkan penaksir β adalah meminimalkan jumlah nilai mutlak dari galat (ε_i) yaitu:

$$\hat{\beta} = \min \sum_{i=1}^{n}|\varepsilon_i|$$

$$= \min \sum_{i=1}^{n}\left|y_i - x_k^t\beta\right|$$

dengan k = 1, 2, ..., K dan k adalah banyak variabel bebas. Jika $k \geq 2$ maka untuk mendapatkan β adalah dengan menggunakan metode regresi LAV berganda. LAV kuat untuk sebuah pencilan dalam *y*. Tetapi, LAV tidak dapat melindungi terhadap pencilan *x* (*leverage*).

Metode Least Median Square (LMS) merupakan salah satu jenis regresi robust dengan high breakdown point. Menurut Venables dan Ripley (1999), Algoritma ini meminimumkan median kuadrat galat dari i pengamatan untuk mendapatkan koefisien regresi β , yaitu:

$$\hat{\beta} = \min\ median\left(\varepsilon_i^2\right) = \min\ median\left(y_i - \hat{y}_i\right)^2 \quad , i = 1,2,3,\dots,n$$

Metode *Least Trimmed Square* (LTS) merupakan salah satu metode penaksiran parameter model regresi yang *Robust* terhadap kehadiran pencilan. LTS digunakan untuk mendapatkan parameter dengan meminimalisasi jumlah kuadrat galatnya dari h pengamatan. Penaksir LTS adalah sebagai berikut:

$$\hat{\beta} = \min\left(\sum_{i=1}^{h} \varepsilon_i^2\right)$$

$$= \min\left(\sum_{i=1}^{h} (y_i - \hat{y}_i)^2\right) \quad , \frac{(3n+p+1)}{4} \le h \le n$$

ε_i^2 = kuadrat galat (sisaan kuadrat) yang terurut dari terkecil hingga terbesar.

$$\varepsilon_1^2 < \varepsilon_2^2 < \varepsilon_3^2 < \cdots < \varepsilon_i^2$$

2.6. Penaksir Regresi *Ridge Robust*

Dalam hal ini regresi *ridge* merupakan metode alternatif dalam menangani masalah multikolieritas, tetapi jika terdapat pencilan dan pengamatan yang berpengaruh besar, maka regresi *ridge* yang biasa tidak dapat digunakan. Dikarenakan metode regresi *ridge* dan *robust* tidak dapat menangani masalah pencilan dan multikolinearits secara bersamaan. Holland (1973) memberikan rumus untuk dari metode regresi *ridge* ketika beban yang terkait dengan masing-masing pengamatan, dan mengusulkan kombinasi regresi *ridge* dengan metode regresi yang robust.

Pfaffenberger dan Dielman (1984) dan Lawrence dan Arthur (1990) menyarankan regresi ridge robust dengan cara menggabungkan sifat-sifat Least Absolute Value (LAV) dan penaksir regresi ridge itu disebut sebagai LAV. Penaksir regresi RLAV bagi β adalah:

$$\hat{\beta}_{RLAV} = (X^tX + \lambda_{LAV}{}^{*}I)^{-1}X^tY$$

dengan mengganti λ_{LS} dengan $\lambda_{LAV}*$ dan $s_{LS}^2 = \frac{\left(Y - X\hat{\beta}_{LS}\right)^T\left(Y - X\hat{\beta}_{LS}\right)}{n-K}$ dengan $s_{LAV}^2 = \frac{\left(Y - X\hat{\beta}_{LAV}\right)^T\left(Y - X\hat{\beta}_{LAV}\right)}{n-K}$.

Diusulkan LMS *robust* didasarkan pada konsep statistik *robust*, dimana LMS meminimalkan galat sebagai pengganti kuadrat terkecil biasa. Penaksir regresi LMS bagi β adalah:

$$\hat{\beta}R_{LMS} = (X^tX + \lambda_{LMS}{}^{*}I)^{-1}X^tY$$

dengan mengganti λ_{LS} dengan $\lambda_{LMS}*$ dan $s_{LS}^2 = \frac{\left(Y - X\hat{\beta}_{LS}\right)^T\left(Y - X\hat{\beta}_{LS}\right)}{n-K}$ dengan $s_{LMS}^2 = \frac{\left(Y - X\hat{\beta}_{LMS}\right)^T\left(Y - X\hat{\beta}_{LMS}\right)}{n-K}$.

Peter Rousseeuw memperkenalkan penaksir regresi robust *Least Trimmed Square* (LTS) adalah metode *high breakdown point* diperkenalkan oleh Rousseeuw (1984). Breakdown point adalah ukuran dari proporsi pencemaran prosedur tersebut bahwa dapat menahan dan masih mempertahankan kekokohannya. Penaksir regresi LTS bagi β adalah:

$$\hat{\beta}_{LTS} = (X^tX + \lambda_{LTS}{}^{*}I)^{-1}X^tY$$

dengan mengganti λ_{LS} dengan $\lambda_{LTS}*$ dan $s_{LS}^2 = \frac{\left(Y - X\hat{\beta}_{LS}\right)^T\left(Y - X\hat{\beta}_{LS}\right)}{n-K}$ dengan $s_{LTS}^2 = \frac{\left(Y - X\hat{\beta}_{LTS}\right)^T\left(Y - X\hat{\beta}_{LTS}\right)}{n-K}$.

2.7. Contoh Aplikasi

Analisis yang akan digunakan menggunakan regresi ridge robust akan diaplikasikan pada persentase hasil analisis pada data proksimat di daerah Muara Niru, Muara Jelawatan, dan Muara Enim. Bahan yang digunakan menggunakan data sekunder yang diperoleh di PUSLITBANG Tek-MIRA.

Dalam menganalisis data pertama-tama melakukan asumsi normalitas, multikolinearitas dan pemeriksaan data berpengaruh, transformasi data. Berikutnya menkasir regresi *ridge robust* LAV, LMS dan LTS. Terakhir Transformasi persamaan Regresi *Ridge Robust* LAV, LMS, dan LTS ke dalam bentuk awal.

3. Hasil dan Pembahasan

3.1. Pemodelan Regresi MKT

Hasil penaksiran parameter berdasarkan metode kuadrat terkecil untuk regresi linear berganda didapatkan sebagai berikut,

$$\hat{Y} = 5919.047 - 4.279\,X_1 - 50.716\,X_2 - 67.849\,X_3 + 14.214\,X_4 + 1.527\,X_5 + 2.941\,X_6 - 97.756\,X_7$$

Setelah didapatkan model diatas dapat diuji asumsi – asumsi normalitas, multikolinearitas dan pengamatan yang berpengaruh.

Asumsi normalitas dimaksudkan untuk mengetahui apakah galat berdistribusi normal atau tidak. Pada peneltian ini pengujian normalitas akan menggunakan uji Kolmogorov-Smirnov.

Hipotesis:

H_0: Galat berdistribusi normal

H_1: Galat tidak berdistribusi normal

Dari hasil analisis diperoleh nilai *p-value* sebesar 0.001. Dengan menggunakan α = 0.05, maka diputuskan untuk menolak H_0. Karena nilai *p-value* < α, 0.001 < 0.05. Sehingga dapat disimpulkan bahwa galat tidak berdistribusi normal.

3.1.1. Multiolinearitas

Untuk mengetahui apakah terdapat masalah multikolinearitas dalam data proksimat, maka dapat menggunakan metode *variance inflation factors* (VIF). Dari hasil yang diperoleh bahwa variabel X_2, X_3, X_4, dan X_5 memiliki nilai VIF lebih dari 10 dengan nilai 14.16378, 74.37769, 16.09991, dan 19.20357, maka dapat disimpulkan bahwa diantara variabel bebas terdapat masalah multikolinearitas, sehingga dapat disimpulkan bahwa dalam data proksimat terdapat kasus multikolinearitas dalam variabel bebasnya.

3.1.2. Pemeriksaan Pengamatan Berpengaruh

Berdasarkan statistik uji untuk mengetahui pencilan terhadap X yaitu dapat menggunakan h_{ii} apabila nilai $h_{ii} > \frac{2p}{n}$. Hasil yang diperoleh bahwa pengamatan ke 9, 19, 20, 82, 89, 99, 104, 108, 111 dan 124 mempunyai nilai h_{ii} berturut-turut 0.4606, 0.2501, 0.1905, 0.2046, 0.1929, 0.8771, 0.1338, 0.4507, 0.28 dan 0.1322 karena nilai $h_{ii} > \frac{2p}{n} > \frac{2x8}{124} = 0.129$, maka pengamatan tersebut merupakan pencilan.

Untuk mengetahui pencilan terhadap Y dapat digunakan metode TRES. Dengan menarik statistik uji tolak H_0 apabila nilai |TRES| > $t_{\frac{\alpha}{2}}$. Dari hasil perhitungan diperoleh bahwa pengamatan ke 2, 18, 19, dan 86 mempunyai nilai TRES 8.3622, 7.8722, 2.0300, dan 7.8722 karena nilai |TRES| > $t_{\frac{\alpha}{2}}$= 1,6602 maka H_0 ditolak yang berarti bahwa pengamatan tersebut merupakan pencilan.

Nilai leverage terbesar diberikan oleh pengamatan ke-99 (h_{ii}= 0.8771), dimana nilai tersebut jauh lebih besar daripada $h_{ii} > \frac{2p}{n} > \frac{2x8}{124} = 0.129$. Hal yang perlu diperhatikan juga adalah pengamatan yang ke-9 yang memberikan nilai leverage yang cukup besar, yaitu h_{ii} = 0.4606. Hal ini juga dapat dilihat pada nilai DFFITS dari kedua data tersebut, yang masing-masing memberikan nilai (DFFITS)9 = -0.2312 dan (DFFITS)99 = 3-.2369. Hal ini mengindikasikan bahwa data ke-99 merupakan data yang berpengaruh pada nilai dugaan y karena pengamatan tersebut mempunyai nilai DFFITS yang lebih besar dari $\frac{2}{\sqrt{n}} = \frac{2}{\sqrt{124}} = 0.1796$.

Pengamatan ke-2, 18, 19 dan 99 mempunyai nilai leverage yang lebih kecil dari $2\sqrt{p/n} = 2\sqrt{7/124} = 0.5080$, dimana pengamatan ke-99 mempunyai nilai leverage yang lebih besar dibandingkan pengamatan ke-2, ke-18, dan ke-19 oleh karena itu pengaruhnya terhadap model relatif besar.

Dilihat dari hasil diatas terlihatada maslah terkait dengan data yaitu adanya data berpengaruh dan adanya multikolinearitas diantara variabel bebas, maka diputuskan untuk menggunakan regresi *ridge robust* untuk mengatasi masalah tersebut.

3.2. Regresi *Ridge Robust*

Dalam proses pengestimasian model regresi *ridge* pemilihan tetapan bias λ merupakan hal yang paling penting dalam penelitian ini. Setelah diperoleh nilai λ maka dicari penaksir *ridge robust* untuk LAV, LMS dan LTS.

Tabel 3.1 Nilai λ dan Persamaan Regresi Ridge Robust

Metode	λ	Persamaan
***Ridge Robust* LAV**	0. 004358	$y^* = -0.04734\ x_1^* - 0.32198\ x_2^* - 0.96623\ x_3^* + 0.09157\ x_4^* + 0.01223\ x_5^* + 0.06420\ x_6^* - 0.01408\ x_7^*$
***Ridge Robust* LMS**	0. 003527	$y^* = -0.04733\ x_1^* - 0.32220\ x_2^* - 0.96676\ x_3^* + 0.09135\ x_4^* + 0.01199\ x_5^* + 0.06420\ x_6^* - 0.01404\ x_7^*$
***Ridge Robust* LTS**	0. 005862	$y^* = -0.04735\ x_1^* - 0.32159\ x_2^* - 0.96528\ x_3^* + 0.09197\ x_4^* + 0.01268\ x_5^* + 0.06420\ x_6^* - 0.01415\ x_7^*$

Setelah diperoleh persaman diatas dilakukan pengembalian model persamaan regresi *ridge* ke dalam bentuk awal.

Tabel 3.2 Hasil Estimasi Koefisien Regresi, λ dan MSE

Metode	λ	Persamaan	MSE
LAV	0.00436	$\hat{Y} = 5902.92136 - 4.28395\ X_1 - 50.53638\ X_2 - 67.65517\ X_3 + 14.39636\ X_4 + 1.70809\ X_5 + 2.9412\ X_6 - 99.08612\ X_7$	27773.2
LMS	0.00353	$\hat{Y} = 5902.92136 - 4.28305\ X_1 - 50.57059\ X_2 - 67.69207\ X_3 + 14.36181\ X_4 + 1.67358\ X_5 + 2.94124\ X_6 - 98.83283\ X_7$	32684.362
LTS	0.0059	$\hat{Y} = 5897.39279 - 4.28549\ X_1 - 50.47469\ X_2 - 67.58858\ X_3 + 14.45874\ X_4 + 1.77033\ X_5 + 2.94114\ X_6 - 99.54345\ X_7$	29897

Dengan membandingkan nilai MSE yang paling kecil dari ketiga metode tersebut dapat disimpulkan bahwa pada data proksimat yang mengandung pencilan dan multikolinearitas pada variabel *Gross Calorfic Value* terhadap variabel *total moisture*, *moisture in air dried sample*, *ash*, *volatile matter*, *fixed carbon*, HGI dan TSG bahwa metode regresi *ridge robust* LAV memiliki hasil yang lebih baik dibandingkan dengan regresi *ridge robust* LMS dan regresi *ridge robust* LTS.

4. Kesimpulan

Berdasarkan hasil kajian pada bagian sebelumnya menunjukkan bahwa Adanya masalah multikolinearitas yang dapat dilihat dari besarnya nilai VIF yaitu nilainya lebih dari 10 pada variabel *moisture in air dried sample, ash, volatile matter,* dan *fixed carbon*.

Pada pengujian DFFITS pengamatan ke 2, 5, 9, 10, 11, 12, 18, 19, 20, 28, 34, 60, 64, 67, 70, 82, 86, 99, 102, 108 dan 124 merupakan pengamatan yang berpengaruh pada nilai dugaan y. Pada pengujian DFBETA dimana pengamatan ke 2, 18, 19 dan 99 merupakan pengamatan berpengaruh. Dapat disimpulkan bahwa pengamtan ke 2, 5, 9, 10, 11, 12, 18, 19, 20, 28, 34, 60, 64, 67, 70, 82, 86, 89, 99, 102, 104, 108, 111 dan 124 merupakan pengamatan berpengaruh.

Dalam mengatasi masalah multikolinearitas dengan menggunakan regersi *ridge robust* dapat menggunakan tiga buah model yaitu *ridge robust* LAV, LMS, dan LTS. Dari hasil analisis yang telah dilakukan, dapat diambil kesimpulan bahwa metode regresi *ridge robust* LAV memiliki hasil yang lebih baik dibandingkan dengan regresi *ridge robust* LMS dan regresi *ridge robust* LTS karena memiliki nilai *Mean Square Error* terkecil.

DAFTAR PUSTAKA

Andrews, D. F., (1972). *Robust Estimates of Location: Survey and Advances*, Princeton University Press.

Chen, C. (2002). *Robust regression and outlier detection with the ROBUSTREG procedure*. Proceedings of the Twenty-seventh Annual SAS Users Group International Conference. Cary, NC: SAS Institute Inc.

Hoerl, A. E. and R. W. Kennard., (1970). *Ridge Regression: Applications to Nonorthogonal Problems. Technometrics*.

Holland, P. W., (1973). *Weighted ridge regression: Combining ridge and robust regression methods*. NBER Working Paper Series.

Lawrence, K. D. & Arthur, J. L. (1990). *Robust Regression:Analysis and Application*. New York: Marcel Dekker.

Myers, R. (1990). *Classical and Modern Regression with Applications*. Boston, MA: Duxburry.

Hajarisman, N. (2010). Analisis Regresi Lanjut. Program Studi Statistika Universitas Islam Bandung.

Rousseeuw, P.J. (1984). *Least median of squares regression, Journal of the American Statistical Association*.

Walpole, R.E & Raymond, H. (1995). Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan. Terjemahan: Sembiring, R.K. Edisi 4. Jakarta: Gramedia Pustaka.