Nama: Harmi Sugiarti & Andi Megawarni

Judul: **Penggunaan Metode Regresi Robust untuk Mencari Selang Kepercayaan Koefisien Garis Regresi jika Ragam Galat Tidak Homogen**

Tanggal: September, 2003

Abstract

The assumption of homogeneous error variance underlying the OLS method is very important to get the best linear unbiased estimation of the regression coefficients. This paper aims to compare the width of confidence interval which is resulted by the OLS, the WLS, and the robust regression methods when the error term is not homogen. Comparing the three methods indicate that the width of confidence interval by the robust regression method is narrower than the OLS method but the WLS method is the narrowest.

PENDAHULUAN

Asumsi ragam galat homogen ($\sigma^2(\varepsilon_i) = \sigma^2$) diperlukan oleh metode kuadrat terkecil *(ordinary least square, OLS)* untuk mendapatkan penduga parameter yang bersifat tak bias linear terbaik *(best linear unbiased estimator, BLUE)* dari model regresi $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \varepsilon_i, \quad i = 1, 2, \ldots, n$. Y_i adalah nilai peubah respons pada pengamatan ke-i, X_i adalah nilai peubah bebas pada pengamatan ke-i dan $\beta_0, \beta_1, \ldots, \beta_p$ adalah koefisien regresi yang tidak diketahui nilainya dan akan dicari nilai taksirannya.

Penanganan kasus ketidakhomogenan ragam galat kadangkala diikuti munculnya penyimpangan asumsi lainnya, diantaranya adalah munculnya pengamatan pencilan *(outlier)* dalam data. Adanya pengamatan pencilan dalam data dapat mengakibatkan penduga koefisien garis regresi yang diperoleh tidak tepat. Namun demikian tindakan membuang (menolak) begitu saja suatu pengamatan pencilan bukanlah tindakan yang bijaksana, karena adakalanya pengamatan pencilan memberikan informasi yang cukup berarti.

Dalam hal asumsi ragam galat homogen tidak dipenuhi, salah satu metode alternatif yang dapat dicoba adalah metode kuadrat terkecil tertimbang *(weighted least square, WLS)*. Metode ini menghasilkan lebar selang kepercayaan untuk koefisien garis regresi lebih sempit dibanding metode OLS (Sugiarti, 2001).

Adanya faktor kesulitan penentuan pembobot $w_i = \frac{1}{\sigma^2}$ karena besarnya σ^2 tidak diketahui, penggunaan metode WLS masih belum dapat mereduksi munculnya pengamatan pencilan. Untuk mengatasi kelemahan-kelemahan dari metode yang ada, baik metode OLS maupun metode WLS, perlu dicoba metode lain yang bersifat tidak sensitif terhadap pelanggaran asumsi-asumsi, yaitu metode regresi robust *(robust regression)*. Tulisan ini bertujuan untuk membandingkan lebar selang kepercayaan koefisien garis regresi yang diperoleh metode regresi robust dengan lebar selang kepercayaan koefisien garis regresi yang diperoleh metode OLS dan metode WLS.

SELANG KEPERCAYAAN

Selang kepercayaan adalah suatu kisaran nilai yang dianggap mengandung nilai parameter populasi yang sebenarnya. Besaran B dan A dikatakan menentukan selang kepercayaan (1-a) 100% bagi suatu parameter apabila memenuhi kriteria berikut:

$P[B \leq$ nilai parameter yang sebenarnya $\leq A] \geq (1-\alpha)$ dan

a. nilai-nilai B dan A dapat dihitung dari sampel yang telah diambil dari populasi.

Selang kepercayaan yang cukup baik adalah selang kepercayaan yang mempunyai lebar selang yang sempit dan persentase selang yang memuat parameter cukup besar (Koopmans, 1987).

METODE REGRESI ROBUST

Prosedur regresi robust dirancang untuk mengurangi pengaruh dari pengamatan-pengamatan yang mempunyai pengaruh tinggi jika metode kuadrat terkecil digunakan. Oleh karena itu, prosedur regresi robust cenderung untuk mengabaikan sisaan-sisaan yang berhubungan dengan pencilan-pencilan yang besar. Disamping tidak sensitif jika terdapat kasus pencilan, prosedur regresi robust mempunyai tingkat efisiensi yang sama 90%-95% dibanding kuadrat terkecil jika dibawah distribusi normal (Montgomery & Peck, 1992).

Menurut Staudte & Sheather (1990), jika hubungan linear antara satu peubah respon dengan peubah-peubah bebasnya dimodelkan sebagai: $Y_i = X_i^T \beta + \varepsilon_i$, dengan X_i^T menyatakan baris ke-i dari matriks rancangan X, β menyatakan parameter model dan ε_i menyatakan suku galat; maka penaksir kemungkinan maksimum atau penaksir-M *(M-estimator)* $\hat{\beta}_p$ untuk model dengan p parameter diperoleh dengan cara meminimumkan $\sum_i \rho(x_i, e_i) = \sum_i \rho\left(x_i, y_i - x_i^T\hat{\beta}_p\right)$ atau mencari penyelesaian dari $\sum_i x_i \Psi\left(x_i, \frac{y_i - x_i^T\hat{\beta}_n}{\sigma}\right) = \sum_i \left(y_i - x_i^T\hat{\beta}_n\right) x_i\, w_i = 0$, dengan w_i adalah fungsi pembobot yang bernilai antara 0 dan 1 dengan $w_i = \frac{\Psi\left(x_i, \frac{e_i}{\sigma}\right)}{\frac{e_i}{\sigma}}$. Secara umum pembobot w_i dirumuskan sebagai $w_i = \frac{\sigma v(x_i)}{e_i} \Psi_c\left(\frac{e_i}{\sigma v(x_i)}\right)$, dengan Ψ_c adalah *influence function* dan $v(x_i)$ adalah suatu fungsi yang tidak diketahui dan tergantung pada x melalui leveragenya. Dengan menentukan nilai $v(x_i) = \frac{(1-h_{ii})}{\sqrt{h_{ii}}}$ dan $\hat{\sigma} = s_{-i}$ serta memilih fungsi Huber Ψ_c yang berbentuk:

$$\Psi_c\left(\frac{e_i}{\sigma v(x_i)}\right) = \begin{cases} \varsigma \text{ jika } \frac{e_i}{\sigma v(x_i)} > c \\ \frac{e_i}{\sigma v(x_i)}, \text{ jika } \left|\frac{e_i}{\sigma v(x_i)}\right| \leq c \\ -\varsigma \text{ jika } \frac{e_i}{\sigma v(x_i)} < -c \end{cases},$$

maka nilai pembobot w_i menjadi bergantung pada kombinasi besarnya leverage dan studentized residual melalui DFFITS. Sebagai salah satu ukuran yang dapat digunakan untuk mendeteksi adanya pengamatan yang berpengaruh, DFFITS dirumuskan sebagai $(\text{DFFITS})_i = \frac{\hat{y}_i - \hat{y}_{-i}}{s_{-i}\sqrt{h_{ii}}}$.

Secara singkat nilai pembobot w_i dapat dinyatakan dalam bentuk: $w_i = w\left(x_i, \frac{e_i}{\sigma}\right) = \min\left(\frac{2\sqrt{p/n}}{|DFFITS_i|}, 1\right)$.

Jadi persamaan $\sum_i \left(y_i - x_i^T\hat{\beta}_\eta\right) w_i x_i = 0$ dapat ditulis dalam bentuk matriks $X^T W X \beta = X^T W Y$ yang kita kenal sebagai persamaan normal kuadrat terkecil tertimbang dengan W adalah matriks diagonal yang berisi pembobot. Solusi persamaan normal tersebut akan memberikan taksiran untuk β yaitu: $\hat{\beta} = \left(X^T W X\right)^{-1}\left(X^T W Y\right)$ dan penaksir-M untuk β diperoleh dengan cara melakukan iterasi sampai diperoleh suatu hasil yang konvergen, cara ini biasa dikenal sebagai metode kuadrat terkecil tertimbang secara iteratif *(iteratively reweighted least square).*

Berdasarkan pembobot w_i dan penaksir-M parameter $\hat{\beta}$, matriks kovarians untuk parameter $\hat{\beta}$ yakni Σ_n dapat didekati dengan persamaan berikut: $\Sigma_n = \frac{1}{n-p}\left(X^T D_1 X\right)^{-1}\left(X^T D_2 X\right)\left(X^T D_1 X\right)^{-1}$, dengan D_1 adalah matriks diagonal dengan elemen-elemen diagonalnya $\Psi'_c\left(\frac{e_i}{\sigma v(x_i)}\right)$ dan D_2 adalah matriks diagonal dengan elemen diagonalnya $w_i^2 e_i^2$. Selang kepercayaan (1-α)100% untuk parameter β diperoleh sebagai: $\hat{\beta} \pm z_{(\alpha/2)}\, s\left(\hat{\beta}\right)$ (Staudte,1990).

METODOLOGI

Data yang dipergunakan dalam tulisan ini adalah data simulasi, yaitu data yang dibangkitkan dengan bantuan paket program MINITAB versi 11.12 dan data eksperimen, yaitu berupa data rata-rata panjang daun (cm) tanaman temulawak *(curcuma xanthorrhiza roxb.)* pada umur 17 minggu yang diberi pupuk kandang pada berbagai taraf yaitu: tanpa pupuk, 0,5 kg/lubang, 1 kg/lubang dan ditanam pada variasi jarak tanam 60 x 40 cm dan 60 x 60 cm (Priono, 1988).

Ada beberapa langkah yang harus ditempuh untuk melakukan simulasi, yaitu pertama adalah mencari galat (ε) yang memenuhi kriteria nilai rataan nol dan ragam tidak homogen. Galat tersebut dapat diperoleh dengan cara membangkitkan data. Pembangkitan galat dikelompokkan ke dalam tiga kelompok dengan setiap kelompok mempunyai rataan nol dan ragam berbeda-beda.

Langkah kedua adalah menentukan nilai-nilai peubah bebas X_1 dan X_2 , karena X_1 dan X_2 adalah konstanta yang diketahui. Nilai-nilai dari koefisien regresi yaitu β diasumsikan dengan nilai tertentu. Dari nilai-nilai yang telah diketahui, dapat dicari nilai peubah respons sebagai $Y = X\beta + \varepsilon$.

Langkah ketiga adalah menguji apakah dari pasangan data X_1 , X_2 dan Y diperoleh ragam galat yang tidak homogen. Apabila hasil uji menyatakan ragam galat tidak homogen, maka pasangan data X_1 , X_2 dan Y dengan rataan galat nol dan ragam galat tidak homogen telah diperoleh.

Langkah keempat, setelah pasangan data X_1 , X_2 dan Y yang mempunyai rataan galat nol dan ragam galat tidak homogen telah diperoleh, selanjutnya dilakukan pendugaan koefisien regresi dengan metode regresi robust dan mencari selang kepercayaan bagi β.

Langkah kelima, membandingkan antara lebar selang kepercayaan yang diperoleh metode regresi robust dengan lebar selang kepercayaan yang diperoleh metode OLS dan metode WLS .

Langkah keenam adalah mengulang langkah keempat dan kelima untuk data eksperimen.

HASIL DATA SIMULASI

Dengan mengasumsikan $\beta_0 = 0$, $\beta_1 = 1$ dan $\beta_2 = 1$, maka pasangan data X_1, X_2 dan Y dapat dilihat pada Lampiran 1. Metode OLS, metode WLS, dan metode regresi robust memberikan penduga bagi koefisien regresi dan simpangan baku penduga koefisien regresinya seperti pada Tabel 1.

Tabel 1. Penduga Koefisien Regresi dan Simpangan Baku untuk Data Simulasi

OLS		WLS		Robust		
$\hat{\beta}$	$s(\hat{\beta})$	$\hat{\beta}$	$s(\hat{\beta})$	$\hat{\beta}$	$s(\hat{\beta})$	
$\hat{\beta}_0$	2.39600	1.55700	-0.93610	0.76870	2.28129	1.72501
$\hat{\beta}_1$	0.96714	0.02383	1.01526	0.01140	0.96812	0.02226
$\hat{\beta}_2$	0.98137	0.02049	1.00305	0.01005	0.98205	0.01736

Jika dibandingkan nilai-nilai penduga bagi koefisien regresi yang diperoleh metode regresi robust dengan metode OLS dan metode WLS, tampak nilai-nilai penduga bagi koefisien regresi yang diperoleh dengan metode regresi robust terletak diantara nilai-nilai penduga bagi koefisien regresi yang diperoleh dengan metode OLS dan metode WLS.

Demikian juga dengan simpangan baku penduga koefisien regresinya, pada Tabel 1 ditunjukkan bahwa nilai-nilai simpangan baku penduga koefisien regresi yang diperoleh dengan metode regresi robust terletak diantara nilai-nilai simpangan baku penduga bagi koefisien regresi yang diperoleh dengan metode OLS dan metode WLS.

Lebar selang kepercayaan 95% dan 99% untuk koefisien regresi yang diperoleh dengan metode OLS, metode WLS dan metode regresi robust disajikan dalam Tabel 2. Pada tabel tersebut tampak bahwa selang untuk koefisien regresi $\hat{\beta}_1$ dan $\hat{\beta}_2$ yang dihasilkan oleh metode regresi robust lebih sempit dibanding metode OLS sedangkan metode WLS menghasilkan selang yang paling sempit diantara ketiga metode. Sedangkan selang kepercayaan 95% untuk koefisien garis regresi lebih sempit dibanding selang kepercayaan 99% untuk koefisien garis regresi yang dihasilkan metode OLS, metode WLS dan metode regresi robust.

Tabel 2. Lebar Selang Kepercayaan untuk Koefisien Regresi Data Simulasi

Selang Kepercayaan 95%			Selang Kepercayaan 99%			
OLS	WLS	Robust	OLS	WLS	Robust	
	6.3899	3.1548	7.0794	8.6289	4.2601	9.56000

$\hat{\beta}_0$						
$\hat{\beta}_1$	0.0978	0.0468	0.0914	0.1321	0.1321	0.1234
$\hat{\beta}_2$	0.0841	0.0413	0.0713	0.1136	0.0557	0.0962

HASIL DATA EKSPERIMEN

Dengan menggunakan data eksperimen pada Lampiran 2 tentang rata-rata panjang daun tanaman temulawak berumur 17 minggu (Y) yang diberi pupuk kandang pada tiga taraf (X_1) dan ditanam pada dua variasi jarak tanam (X_2), metode OLS, metode WLS, dan metode regresi robust memberikan penduga bagi koefisien regresi dan simpangan baku penduga bagi koefisien regresi seperti pada Tabel 3. Nilai-nilai simpangan baku penduga koefisien regresi yang diperoleh dengan metode regresi robust terletak diantara nilai-nilai simpangan baku penduga bagi koefisien regresi yang diperoleh dengan metode OLS dan metode WLS.

Tabel 3. Penduga Koefisien Regresi dan Simpangan Baku untuk Data Eksperimen

dengan Konsentrasi Pupuk sebagai Pembobot

OLS		WLS		Robust		
$\hat{\beta}$	$s(\hat{\beta})$	$\hat{\beta}$	$s(\hat{\beta})$	$\hat{\beta}$	$s(\hat{\beta})$	
$\hat{\beta}_0$	39.0400	27.2500	54.0710	4.2910	61.5178	12.8986
$\hat{\beta}_1$	31.5900	12.7300	34.3810	3.2870	19.3309	6.7357
$\hat{\beta}_2$	0.2693	0.5197	-0.0406	0.0822	-0.1394	0.2455

Lebar selang kepercayaan 95% dan 99% untuk koefisien regresi yang diperoleh dengan metode OLS, metode WLS dan metode regresi robust disajikan dalam Tabel 4. Pada tabel tersebut tampak bahwa selang kepercayaan untuk koefisien regresi $\hat{\beta}_1$ dan $\hat{\beta}_2$ yang dihasilkan oleh metode regresi robust lebih sempit dibanding metode OLS sedangkan metode WLS menghasilkan selang yang paling sempit diantara ketiga metode. Selang kepercayaan 95% untuk koefisien garis regresi lebih sempit dibanding selang kepercayaan 99% untuk koefisien garis regresi yang dihasilkan metode OLS, metode WLS dan metode regresi robust.

Tabel 4. Lebar Selang Kepercayaan untuk Koefisien Regresi Data Eksperimen

dengan Konsentrasi Pupuk sebagai Pembobot

Selang Kepercayaan 95%			Selang Kepercayaan 99%			
OLS	WLS	Robust	OLS	WLS	Robust	
$\hat{\beta}_0$	116.1395	18.2882	54.9738	160.6115	25.2912	76.0244
$\hat{\beta}_1$	54.2553	14.0092	28.7076	75.0306	19.3736	39.7002

$\hat{\beta}_2$	2.2150	0.3502	1.0463	3.0631	0.4845	1.4470

Jika digunakan jarak tanam sebagai pembobot pada metode WLS, maka nilai-nilai penduga bagi koefisien regresi dan simpangan baku penduga koefisien regresi yang diperoleh dengan metode OLS, metode WLS, dan metode regresi robust disajikan dalam Tabel 5 berikut ini.

Tabel 5. Penduga Koefisien Regresi dan Simpangan Baku untuk Data

Eksperimen dengan Jarak Tanam sebagai Pembobot

OLS		WLS		Robust		
$\hat{\beta}$	$s(\hat{\beta})$	$\hat{\beta}$	$s(\hat{\beta})$	$\hat{\beta}$	$s(\hat{\beta})$	
$\hat{\beta}_0$	39.0400	27.2500	44.8700	20.9600	61.5178	12.8986
$\hat{\beta}_1$	31.5900	12.7300	19.9330	2.5750	19.3309	6.7357
$\hat{\beta}_2$	0.2693	0.5197	0.2693	0.5198	-0.1394	0.2455

Nilai-nilai simpangan baku penduga koefisien regresi yang diperoleh dengan metode regresi robust terletak diantara nilai-nilai simpangan baku penduga bagi koefisien regresi yang diperoleh dengan metode OLS dan metode WLS. Khusus untuk $\hat{\beta}_0$, metode regresi robust menghasilkan simpangan baku paling kecil diantara ketiga metode.

Lebar selang kepercayaan 95% dan 99% untuk koefisien regresi yang diperoleh dengan metode OLS, metode WLS dan metode regresi robust disajikan dalam Tabel 6. Pada tabel tersebut tampak bahwa selang untuk koefisien regresi yang dihasilkan oleh metode regresi robust lebih sempit dibanding metode OLS sedangkan metode WLS menghasilkan selang yang paling sempit diantara ketiga metode kecuali pada selang kepercayaan untuk $\hat{\beta}_2$, dengan metode regresi robust menghasilkan selang paling sempit diantara ketiga metode tersebut.

Tabel 6. Lebar Selang Kepercayaan untuk Koefisien Regresi Data Eksperimen

dengan Jarak Tanam sebagai Pembobot

Selang Kepercayaan 95%			Selang Kepercayaan 99%			
OLS	WLS	Robust	OLS	WLS	Robust	
$\hat{\beta}_0$	116.1395	89.3315	54.9738	160.6115	123.5382	76.0244
$\hat{\beta}_1$	54.2553	10.9747	28.7076	75.0306	15.1771	39.7002
$\hat{\beta}_2$	2.2150	2.2154	1.0463	3.0631	3.0637	1.4470

Selang kepercayaan 95% untuk koefisien garis regresi lebih sempit dibanding selang kepercayaan 99% untuk koefisien garis regresi yang dihasilkan metode OLS, metode WLS dan metode regresi robust.

KESIMPULAN

Lebar selang yang dihasilkan oleh metode regresi robust secara keseluruhan lebih sempit dibanding selang yang dihasilkan metode OLS tetapi sedikit lebih lebar dibanding selang yang dihasilkan metode WLS.

DAFTAR PUSTAKA

Koopmans, L.H. 1987. *Introduction to Contemporary Statistical Methods*. 2nd ed. Boston: PWS.

Montgomery,D.C.&Peck,E.A. 1992. *Introduction to Linear Regression Analysis.* 2nd ed. Homewood, Illinois:Irwin

Priono, M. 1988. *Pengaruh Pemberian Pupuk Kandang & Jarak Tanam Terhadap Pertumbuhan Tanaman Temulawak (Curcuma Xanthorrhiza Roxb.)*. Skripsi yang tidak dipublikasikan. Universitas Jenderal Soedirman, Purwokerto.

Staudte, R.G. & Sheather, S.J. 1990. *Robust Estimation and Testing*. New York: Wiley.

Sugiarti, H. & Megawarni, A. 2001. *Selang Kepercayaan untuk Koefisien Garis Regresi Jika Ragam Galat Tidak Homogen dengan Metode OLS dan WLS*. Jakarta: Lembaga Penelitian Universitas Terbuka.

Lampiran 1. Data Simulasi

Obs.	Y	X_1	X_2	ε
1.	100.972	48	53	-0.02759

2.	121.659	95	26	0.65870
3.	41.731	42	1	-1.26902
4.	109.832	62	46	1.83219
5.	55.494	34	22	-0.50565
6.	26.970	17	11	-1.03007
7.	111.646	23	90	-1.35382
8.	112.610	62	50	0.60968
9.	107.549	83	24	0.54867
10.	170.493	79	91	0.49285
11.	103.917	42	62	-0.08276
12.	32.976	18	13	1.97609
13.	121.193	89	36	-3.80707
14.	53.497	8	40	5.49658
15.	88.483	76	14	-1.51694
16.	34.910	2	36	-3.09022
17.	55.939	37	23	-4.06147
18.	95.829	5	89	1.82904
19.	50.646	21	28	1.64600
20.	138.479	48	89	1.47855
21.	170.862	83	88	-0.13793
22.	54.293	40	11	3.29348
23.	77.655	60	24	-6.34511

Lampiran 2. Data Eksperimen

Obs.	Y	X_1	X_2
1.	52.90	0.0	40
2.	55.07	0.0	40
3.	55.73	0.0	40
4.	46.97	0.0	60
5.	48.00	0.0	60
6.	53.50	0.0	60
7.	69.00	0.5	40
8.	70.60	0.5	40
9.	64.40	0.5	40
10.	69.27	0.5	60
11.	71.03	0.5	60
12.	71.20	0.5	60
13.	72.13	1.0	40
14.	72.73	1.0	40

24.	67.161	30	28	9.16097
25.	134.472	37	100	-2.52823
26.	134.850	66	74	-5.15037
27.	144.231	69	82	-6.76911
28.	81.048	67	11	3.04840
29.	83.743	55	26	2.74334
30.	100.464	37	61	2.46426

15.	77.90	1.0	40
16.	46.73	1.0	60
17.	73.93	1.0	60
18.	158.30	1.0	60

Kembali