PENERAPAN METODE *WEIGTHED LEAST SQUARE* UNTUK MENGATASI HETEROSKEDASTISITAS PADA ANALISIS REGRESI LINEAR

Nurul Hanifah, Nar Herrhyanto, Fitriani Agustina
Departemen Pendidikan Matematika FPMIPA UPI
Correspondent author: hanifazz@yahoo.com

ABSTRAK Analisis regresi merupakan analisis statistik yang mempalajari bagaimana memodelkan regresi linear. Jika model regresi linear memenuhi uji asumsi klasik dengan metode OLS maka mempunyai sifat BLUE (*Beast Linear Unbiased Estimator)*. Uji heteroskedastisitas,yaitu varian error pada setiap nilai variabel bebas bernilai tidak konstan. Akibat dari heteroskedastisitas yaitu nilai parameter yang diperoleh tetap tidak bias tetapi varian penaksir yang diperoleh menjadi tidak efisien, artinya uji hipotesis yang dilakukan tidak akan memberikan hasil yang baik (tidak valid) atau prediksi koefisien-koefisien populasinya akan keliru. Oleh karena itu untuk mengetahui apakah terdapat heteroskedastisitas dilakukan uji White. Karena terdapat heteroskedastisitas pada skripsi ini, maka harus dilakukan transformasi dengan metode kuadrat terkecil tertimbang (*Weighted Least Square*).
Kata Kunci: Uji Asumsi Klasik, *Weighted least Square*, Uji White.

ABSTRACT Regression analysis is a statistical analysis that learn how to model linear regression. If a linear regression model meets the Classic Assumption Test by OLS method, it has the nature of BLUE (Best Linear Unbiased Estimator). Error variance at each independent variable value is not constant. It means that heteroskedasticity test is unfulfilled and the classical assumption is not met.The result of heteroskedastisitas is that the parameter value remains biased but variance estimator becomes inefficient. It means thata hypothesis test wouldn't give good results (not valid) or predictions coefficients of the population would be mislead. Therefore, to know whether there are heteroskedasticity, White test is conducted. Because heteroskedasticity exists in this thesis, transformation with weighted least squares method (Weighted Least Square) must be carried out.
Keyword: Classic Assumption Test, Weighted least Square, White Test.

1. Pendahuluan

Gujarati (Sarwono, 2013:1)[1] mendefinisikan analisis regresi sebagai kajian terhadap hubungan satu variabel yang dinamakan variabel yang diterangkan (variabel tidak bebas) dengan satu atau lebih variabel yang menerangkan (variabel bebas). Untuk mendapatkan model regresi linier dapat diperoleh dengan melakukan penaksiran terhadap parameter-parameternya menggunakan metode kuadrat terkecil (OLS).

Menurut Gauss (dalam Nachrowi dan Hardius, 2002;19)[2] mengemukakan jika model regresi linear memenuhi asumsi-asumsi berikut maka taksiran yang diperoleh dengan metode OLS mempunyai sifat BLUE (*Best Linear Unbiased Estimator*):

1. E($\varepsilon_i|X_i$) = 0
2. Var ($\varepsilon_i|X_i$) = σ^2 (*homoscedasticity*)
3. Cov ($\varepsilon_i|\varepsilon_j$) = 0 ; i≠j (*nonautocorrelation*)
4. Antar X saling *independent* (*collinnearity*)
5. $\varepsilon \sim N(0, \sigma^2)$

Jika terdapat pelanggaran asumsi di atas dapat dilakukan dengan uji asumsi klasik, yaitu: uji normalitas, uji linearitas, uji heteroskedastisitas, uji autokrelasi dan uji multikolinearitas (Setyadharma, 2010: 1)[3]. Salah satu asumsi klasik yang harus dipenuhi dalam estimasi OLS, yaitu varian error ε_i pada setiap nilai – nilai variabel bebas adalah sama (homoskedastisitas) Var ($\varepsilon_i|X_i$) = σ^2. Pelanggaran terhadap asumsi homoskedastisitas disebut heteroskedastisitas, maka nilai parameter yang diperoleh tetap tidak bias karena sebagai penaksir tidak bias tidak memerlukan asumsi bahwa varian error harus konstan, tetapi varians penaksir yang diperoleh akan menjadi tidak efisien mengakibatkan uji hipotesis yang dilakukan tidak akan memberikan hasil yang baik (tidak valid). Sehingga jika varian penaksir model tidak memenuhi asumsi homoskedastisitas, maka prediksi mengenai koefisien-koefisien populasinya akan keliru.

Untuk mengetahui adanya heteroskedastisitas, dapat diketahui dengan mendeteksi persamaan regresi tersebut dengan menggunakan uji *White*. Apabila terdapat Heteroskedastisitas dapat dilakukan dengan Metode Kuadrat Terkecil Tertimbang (*Weighted Least Square).* Hal ini dikarenakan WLS memiliki kemampuan untuk menetralisasi akibat dari pelanggaran asumsi heteroskedatisitas dan dapat menghilangkan sifat ketidakbiasan dan konsistensi dari model taksiran OLS. Pada penelitian ini penulis ingin mengetahui bagaiamana langkah-langkah penerapan metode *Weighted Least Square* untuk mengatasi heteroskedastiistas pada analisis regresi linear serta pada kasus anak terlantar. Pada penelitian ini digunakan data hubungan anak yang bekerja dan tidak/belum sekolah dengan anak terlantar di kabupaten pulau Jawa tahun 2011.

2. Landasan Teori

2.1 Metode Kuadrat Terkecil (OLS)

Model regresi linear berganda dapat ditulis sebagai berikut:

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

Atau persamaan regresi dalam bentuk matrik sebagai berikut:

$$Y = X \quad + \varepsilon \text{ dengan } \hat{Y} = X$$

Untuk memperoleh taksiran parameter pada regresi dapat dilakukan dengan metode kuadrat terkecil. Metode kuadrat terkecil(OLS) bertujuan meminimumkan jumlah kuadrat error ($\sum_{i=1}^{n} \varepsilon_i^2$) sehingga nilai regresinya akan mendekati nilai yang sebenarnya. Semakin kecil nilai *error*, hal ini menunjukkan semakin tingginya ketepatan model yang dihasilkan untuk menjelaskan nilai vaiabel tak bebas. Dengan OLS penaksir koefisien regresi Sehingga diperoleh

$$b_0 = \bar{Y}_i - b_1\bar{X}_{1i} - b_2\bar{X}_{2i}$$

$$b_1 = \frac{(\sum x_{1i}y_i)(\sum x_{2i}{}^2) - (\sum x_{2i}y_i)(\sum x_{1i}x_{2i})}{(\sum x_{1i}{}^2)(\sum x_{2i}{}^2) - (\sum x_{2i}x_{1i})}$$

$$b_2 = \frac{(\sum x_{2i}y_i)(\sum x_{1i}{}^2) - (\sum x_{1i}y_i)(\sum x_{1i}x_{2i})}{(\sum x_{1i}{}^2)(\sum x_{2i}{}^2) - (\sum x_{2i}x_{1i})}$$

Dalam bentuk matrik $b = X'Y(X'X)^{-1}$ dan var(b)=σ^2 $(X'X)^{-1}$

2.2 Heteroskedastisitas

Gujarati dan Zain (1999: 177)[4] menjelaskan bahwa salah satu asumsi klasik yang penting adalah bahwa varians setiap unsur , terantung pada nilai yang dipilih dari variabel yang menjelaskan adalah suatu angka konstan yang sama dengan σ^2. Sehingga uji heteroskedastisitas dilakukan untuk melihat apakah terdapat ketidaksamaan varian dari satu pengamatan terhadap pengamatan lain. Jika terdapat keberagaman varian pada model maka terdapat heteroskedastisitas yang mengakibatkan penaksir parameter pada model regresi tidak bersifat BLUE. Adanya heteroskedastisitas akan mengakibatkan sebagai berikut:

a. Penaksir OLS yang diperoleh tetap memenuhi tidak bias
b. Namun varian yang diperoleh menjadi tidak efisien, artinya cenderung membesar sehingga tidak lagi merupakan varian terkecil. Kecenderungan semakin membesar varian akan menyebabkan standar error juga membesar mengakibatkan nilai uji t atau uji F terlalu besar maka kesimpulan dari model regresi yang dibuat dapat menyesatkan.

Untuk mengetahui apakah terdapat heteroskedastisitas dapat dilakukan dengan Uji statistik yang dapat digunakan adalah uji Glejser, uji Park atau uji White.

2.3 Uji White

Untuk mengetahui adanya heteroskedastisitas pada model dapat dilakukan dengan uji White. Langkah-langkah uji White adalah sebagai berikut:

a. Estimasi persamaan dibawah dengan OLS dan ε_i

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

b. Selanjutnya regresikan model dibawah

$$\varepsilon_i^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{1i}^2 + \alpha_4 X_{2i}^2 + \alpha_5 X_{1i} X_{2i} + \mu_i$$

c. Perumusan hipotesis
   H_0 : tidak terdapat heteroskedastisitas pada model
   H_1 : terdapat heteroskedastisitas pada model

d. Statistik Uji
   Hitung nR^2 dengan n jumlah observasi dan R^2 koefisien determinasi dari model

e. Kriteria Pengujian
   Dengan mengambil taraf nyata α, maka H_0 ditolak, jika $nR^2 > \chi^2_{\alpha;k}$ dengan $\chi^2_{\alpha;k}$ diperoleh dari Tabel Distribusi Chi-Kuadrat dengan peluang = α dan dk = k

f. kesimpulan
   Penafsiran dari H_0 diterima atau ditolak

2.4 Metode Weighted Least Square

Menurut Montgomery *et al*[5] mengatakan untuk mengatasi model regresi dengan varian error tidak konstan dapat dilakukan dengan Metode Kuadrat Terkecil Tertimbang (*Weighted Least Square Method).* WLS memiliki kemampuan untuk menetralisasi akibat dari pelanggaran asumsi heteroskedatisitas dan dapat menghilangkan sifat ketidakbiasan dan konsistensi dari model taksiran OLS. Pada metode ini digunakan *"weight"* atau pembobot yang proporsional terhadap *inverse*(kebalikan) dari varians variabel respon sehingga diperoleh error baru yang memiliki sifat seperti pada regresi dengan OLS.

Ketika model regresi $Y = X\beta + \varepsilon$ dengan $\text{var}(\varepsilon_i) = W\sigma^2$ dengan W matriks diagonal dengan elemen diagonal bernilai tidak sama, maka Y pengamatan tidak berkorelasi tetapi memiliki varians tidak sama dengan elemen non-diagonalnya bernilai 0. $W\sigma^2$ matriks varcovar dengan W matriks diagonal (n x n) definit positif dan nonsingular. Sehingga W dapat difaktorisasi sebagai berikut:

$$W = C \Lambda C'$$

Dengan C_{nxn} kolom-kolomnya eigen vektor dari W dan Λ_{nxn} matriks diagonal dengan unsur diagonal eigen value dari W.

$$\Lambda = \begin{bmatrix} \lambda_i & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

Didefinisikan T = $C\,\Lambda^{1/2}$ dengan $\Lambda^{1/2}$ merupakan matriks diagonal yang entri pada diagonal ke-i adalah $\sqrt{\lambda_i}$ sehingga W= TT' . Misalkan $P' = C\,\Lambda^{-1/2}$ dimana $\Lambda^{\frac{1}{2}}$ matriks diagonal yang entri pada diagonal ke-i nya adalah $\sqrt{\lambda_i}$

$PP' = C\Lambda^{\frac{1}{2}}\left(C\Lambda^{\frac{1}{2}}\right)'$

$= C\Lambda\, C'$

$= W$

Selanjutnya ubah vektor error ε menjadi

$$f = P^{-1}\varepsilon$$

Bertujuan agar

$$E(f) = E(P^{-1}\varepsilon) = P^{-1}E(\varepsilon) = 0$$

Dan

$$\begin{aligned} var(f) &= E(ff') \\ &= E[(P^{-1}\varepsilon)(P^{-1}\varepsilon)'] \\ &= E[(P^{-1}\varepsilon)\varepsilon'(P')^{-1}] \\ &= P^{-1}E(\varepsilon\varepsilon')P' \\ &= I\sigma^2 \end{aligned}$$

Oleh karena itu vektor Y dikalikan dengan P^{-1} menjadi:

$$P^{-1}Y = P^{-1}X\beta + P^{-1}\varepsilon$$

Atau

$$Z = V\beta + f \qquad \ldots(2.2)$$

Sehingga terdapat notasi baru:

- variabel respon Z = $P^{-1}Y$,
- variabel prediktor **V**$_0$ dan **V**$_1$, yang terhimpun di dalam matrik $V = P^{-1}X$,
- residual $f = P^{-1}\varepsilon$

Persamaan (2.2) memenuhi asumsi OLS, sehingga persamaan menjadi:

$$Z = V$$

$$V'Z = V'V$$

$$(P^{-1}X)'(P^{-1}Y) = (P^{-1}X)'(P^{-1}X)b$$

$$X'(P^{-1})'\,P^{-1}Y = X'\,(P^{-1})'\,P^{-1}X\,b \qquad \ldots(2.3)$$

Diketahui bahwa $P \quad = P \quad ' = W$, sehingga

$$(PP')^{-1} = W^{-1}$$
$$(P^{-1})'P^{-1} = W^{-1}$$

Maka persamaan (2.3) menjadi

$$X'W^{-1}Y = X'W^{-1}X$$

$$(X'W^{-1}X)^{-1}X'W^{-1}Y = I \qquad \ldots(2.4)$$

Persamaan (2.4) menghasilkan nilai penduga b metode kuadrat terkecil. Sehingga jika $\varepsilon \sim N(0, W\sigma^2)$, maka distribusi error menjadi $f \sim N(0, I\sigma^2)$

Misalkan $V = W^{-1}$ dimana W adalah matrik diagonal, maka V matriks diagonal dengan elemen diagonalnya $w_1, w_2, \ldots, w_n$ disebut pembobot. Maka estimasi *Weighted Least Square* adalah:

$$(X'V \quad)^{-1}X'V \quad = I$$

Dan varians b sebagai berikut:

$$v \quad (b) = E(b'b) = E\{(b - \beta)(b - \beta)'\}$$

Diketahui bahwa

$b = (X'V \quad)^{-1}X'V \quad = (X'V \quad)^{-1}X'V(X \quad + \varepsilon)$

$= (X'V \quad)^{-1}(X'V \quad)\beta + (X'V \quad)^{-1}X'V$

$= I\beta + (X'V \quad)^{-1}X'V$

Sehingga

Var (b) $= E\{[(X'V \quad)^{-1}X'V \quad][(X'V \quad)^{-1}X'V \quad]'\}$

$= E\{[(X'V \quad)^{-1}X'V \quad][((X'V \quad)')^{-1}X \quad '\varepsilon']\}$

$= E\{[(X'V \quad)^{-1}X'V \quad]\;[(X'V'X)^{-1}X \quad '\varepsilon']\}$

$= E\{[(X'V\quad)^{-1}(X'V\quad)(\varepsilon\varepsilon')\ (X'V'X)^{-1}V'\}$
$= E(\varepsilon\varepsilon')\ (X'V\quad)^{-1}(X'V\quad)\ (X'V'X)^{-1}V'\}$
$= \sigma^2 W\ \ I\ (X'X)^{-1}\ (V')^{-1}V'$
$= \sigma^2 V^{-1}\ \ I\ (X'X)^{-1}\ I$

Var (b) $= \sigma^2 (X'V\quad)^{-1}$

3. Pembahasan

3.1 Uji White dengan Eviews

Tabel 1 hasil estimasi OLS dengan eviews

Dependent Variable: ANAK_TERLANTAR
Method: Least Squares
Date: 06/16/15 Time: 21:17
Sample: 1 119
Included observations: 119

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	24531.80	2558.938	9.586713	0.0000
BEKERJA	1.953404	1.360411	1.432704	0.1546
TIDAK_SEKOLAH	3.680349	0.080128	45.93096	0.0000

Dari hasil tabel 1 diketahui bahwa persamaan regresi sebagai berikut:

$$Y = 24531.80 + 1.953404\ X_1 + 3.680349\ X_2$$

Dengan:
Y = anak terlantar
X_1= anak yang bekerja
X_2= anak tidak sekolah

Pada model diatas dilakukan uji white dengan eviews, berikut hasil output

Tabel 2. Hasil Uji White

Heteroskedasticity Test: White			
F-statistic	16.56111	Prob. F(2,116)	0.0000
Obs*R-squared	26.43163	Prob. Chi-Square(2)	0.0000

pada Tabel 2, nilai nR^2= 26,43163 dan prob. Chi square 0,000.. Karena nR^2= 26,43163 > $\chi^2_{2(0,0\ )}$ = 5,99148 dan prob. Chi square 0,000 ≤ 0,05, maka dapat disimpulkan terdapat heteroskedastisitas pada model regresi anak terlantar. Karena terdapat heteroskedastisitas pada model, maka transformasi data tersebut dengan metode *Weighted Least Square*

3.2 Metode Wighted Least Square dengan Eviews

Pemilihan pembobot bergantung pada ε^2 berkorelasi terhadap X. Pada data kali ini pembobot untuk metode *Weighted Least Square* yaitu variabel tidak sekolah.Berikut model regresi setelah dilakukan pembobotan

Tabel 3. Hasil pembobotan

Included observations: 11[illegible]
Weighting series: (TIDAK_SEKOLAH)^-0.5
Weight type: Inverse standard deviation (EViews default scaling)
White heteroskedasticity consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7880.635	1910.239	4.125472	0.0001
BEKERJA	3.095680	1.596720	1.938771	0.0550
TIDAK_SEKOLAH	4.376621	0.130004	33.66523	0.0000

Weighted Statistics

R-squared	0.932302	Mean dependent var	71332.43
Adjusted R-squared	0.931136	S.D. dependent var	23662.32
S.E. of regression	15867.70	Akaike info criterion	22.20685
Sum squared resid	2.92E+10	Schwarz criterion	22.27691
Log likelihood	-1318.307	Hannan-Quinn criter.	22.23530
F-statistic	798.7465	Durbin-Watson stat	1.404715
Prob(F-statistic)	0.000000	Weighted mean dep.	35060.78

Unweighted Statistics

R-squared	0.914042	Mean dependent var	110816.5
Adjusted R-squared	0.912560	S.D. dependent var	82256.14
S.E. of regression	24323.29	Sum squared resid	6.86E+10
Durbin-Watson stat	1.513115		

Pada Tabel 3 WLS atau *Weighted Statistics* nilai *R-squared* yaitu 0,932302 dan *standard error* yaitu 15867,7 sedangkan sebelum dilakukan pembobotan nilai *R squared* dan *standard error* yaitu 0,914042 dan 24323,29 . maka dilakukan pengecekan ulang pada regresi yang telah diboboti apakah masih terdapat heteroskedastisitas dengan uji White.

Tabel 4. Output Uji White Pada Data yang Telah Diboboti

Heteroskedasticity Test: White			
F-statistic	2.687269	Prob. F(3,115)	0.0498
Obs*R-squared	7.795718	Prob. Chi-Square(3)	0.0504
Scaled explained SS	7.491722	Prob. Chi-Square(3)	0.0578

Pada Tabel 4 hasil uji white pada data yang telah diboboti menggunakan eviews tidak terdapat heteroskedastisitas karena nilai nR^2= 7,795718 < $\chi^2_{3(0,0\ )}$= 7,81472 dan prob. Chi square 0,0504 $\geq$ 0,05. Dapat disimpulkan regresi yang telah diboboti tidak terdapat heteroskedastisitas pada model regresi anak terlantar. Sehingga model regresi regresi yang telah dilakukan pembobotan dapat dipakai sebagai model regresi anak terlantar.Dengan nilai *R-squared* yaitu 0,932302 atau 9,32% aritnya variabel anak terlantar dapat diterangkan oleh variabel anak yang bekerja dan anak tidak sekolah. Sedangkan sisanya diterangkan variabel lain. *standard error* yaitu 15867,7.

Persamaan regresi yang telah ditransormasi sebagai berikut:

$$Y = 7880.635 + 3.095680\, X_1 + 4.376621\, X_2$$

Dapat disimpulkan bahwa:

- Konstanta sebesar 7880,635 menyatakan bahwa jika tidak ada anak yang bekerja dan tidak ada anak yang tidak sekolah, maka anak telantar berjumlah 7880,635 jiwa.
- Koefisien regresi X_1 sebesar 3,095680 menyatakan bahwa setiap penambahan (karena tanda +) satu jiwa anaka yang bekerja akan meningkatkan jumlah anak terlantar sebesar 7883,73068 jiwa.
- Koefisien regresi X_2 sebesar 4,376621 menyatakan bahwa setiap penambahan satu jiwa anak tidak sekolah akan meningkatkan jumalah anak terlantar sebesar 7885,01162 jiwa.

4. Kesimpulan dan Saran

Metode ini digunakan *"weight"* atau pembobot yang proporsional terhadap *inverse*(kebalikan) dari varians variabel respon. Pembobot pada WLS dinotasikan $P=C\Lambda^{\frac{1}{2}}$ dimana C matriks nxn dengan kolom-kolomnya vektor ciri (eigen vector) dan Λ matriks diagonal nxn dengan unsur diagonalnya akar ciri (eigen value) dan $\Lambda^{\frac{1}{2}}$ matriks diagonal yang entri pada diagonal ke-i nya adalah $\sqrt{\lambda_i}$, sehingga W=

PP' . Pada kasus anak terlantar, pembobot yaitu variabel tidak sekolah. Sehingga model regresi yang telah diboboti menjadi Y = 24531.80 + 1.953404 X_1 +3.680349 X_2.

Pada penelitian dalam studi kasus selanjutnya diharapkan dapat mengetahui pembobot yang terbaik untuk mengatasi heteroskeadstisitas pada data.

DAFTAR PUSTAKA

[1] Sarwono, J. (2013). *12 Jurus Ampuh SPSS untuk Riset Skripsi: Kupas Tuntas Prosedur-Prosedur Regresi dan 'Decision Trees' dalam IBM SPSS*. Jakarta: PT Elex Media Komputindo.

[2] Nachrowi D. N dan Usman, H. (2002). *Penggunaan teknik ekonometri*. Jakarta:
Raja Grafindo Persada.

[3] Setyadharma, A. (2010). *Uji Asumsi Klasik Dengan SPSS 16.0*. Semarang : FE UNNES.

[4] Gujarati, D dan Zain, S. (1999). *Ekonomi Dasar*. Jakarta: Erlangga.

[5] Montgomery, D.C., Peck, E.A. dan Vining, G.G. (2012). *Introduction to Linear Regression Analysis (fourth Edition).* New York: Wiley