

HETEROSKEDASTISITAS

Uji Heteroskedastisitas

Pengertian Heteroskedastisitas

- Heteroskedastisitas muncul apabila kesalahan (e) atau residual dari model yang diamati tidak memiliki varians yang konstan dari satu observasi ke observasi lainnya

- rumus regresi diperoleh dengan asumsi bahwa variabel pengganggu (*error*) atau e , diasumsikan memiliki variabel yang konstan (rentangan e kurang lebih sama). Apabila terjadi varian e tidak konstan, maka kondisi tersebut dikatakan tidak *homoskedastik* atau mengalami *heteroskedastisitas*

Konsekuensi Heteroskedastisitas

1. Standard error menjadi bias
2. Nilai b bukan nilai yang terbaik.
3. Munculnya masalah heteroskedastisitas yang mengakibatkan nilai se menjadi bias, akan berdampak pada nilai t dan nilai F yang menjadi tidak valid.

Pendeteksian Heteroskedastisitas

- Untuk mendeteksi ada tidaknya heteroskedastisitas, dapat dilakukan dengan berbagai cara seperti uji grafik, uji Arch, uji Park, Uji Glejser, uji *Spearman's Rank Correlation*, dan uji White menggunakan Lagrange Multiplier
- Membandingkan sebaran data pada scatter plot

menggunakan uji Arch

- $e^2 = a + b\hat{Y}^2 + u$.
- Cari R^2
- Kalikan R^2 dengan n (sampel)/ $R^2 \times n$.

- Jika R^2 x n lebih besar dari *chi-square* (χ^2) tabel, maka *standar error* mengalami heteroskedastisitas.
- Sebaliknya, jika R^2 x n lebih kecil dari *chi-square* (χ^2) tabel, maka *standar error* telah bebas dari masalah heteroskedastisitas, atau telah homoskedastis.

Pendeteksian

- Sifat Dasar Masalah
- Metode Grafik
- Pengujian Park

Memformalkan metode grafik dengan σ^2 merupakan fungsi X : $\ln \sigma^2 = \beta \ln X + u$

- Pengujian Glejser

Meregresi e_i terhadap X yang diperkirakan mempunyai hubungan dengan σ^2

- Pengujian rank korelasi dari Spearman

- Masalah heteroskedastisitas lebih sering muncul dalam data *cross section* dari pada data *time series*
- Karena dalam data *cross section* menunjukkan obyek yang berbeda dan waktu yang berbeda pula. Antara obyek satu dengan yang lainnya tidak ada saling keterkaitan, begitu pula dalam hal waktu.

- Sedangkan data *time series*, antara observasi satu dengan yang lainnya saling mempunyai kaitan. Ada trend yang cenderung sama. Sehingga *variance* residualnya juga cenderung sama.

Asumsi Homoskedastisitas

Memiliki varians yang sama, dengan lambang:

$$E(ui^2) = \sigma^2$$

Contoh Analisis untuk Heteroskedastisitas

- Data harga rumah dari 88 sampel rumah di London
- Price : harga rumah dalam Poundsterling
- Rooms : jumlah kamar setiap rumah
- Sqfeet : luas rumah dalam square feet
- Ingin diketahui apakah jumlah kamar dan luas rumah memberikan pengaruh nyata dalam menentukan harga rumah
- Indikasi:
 - semakin banyak jumlah kamar semakin beragam harga
 - Semakin luas rumah semakin beragam harga

Penduga Model Regresi

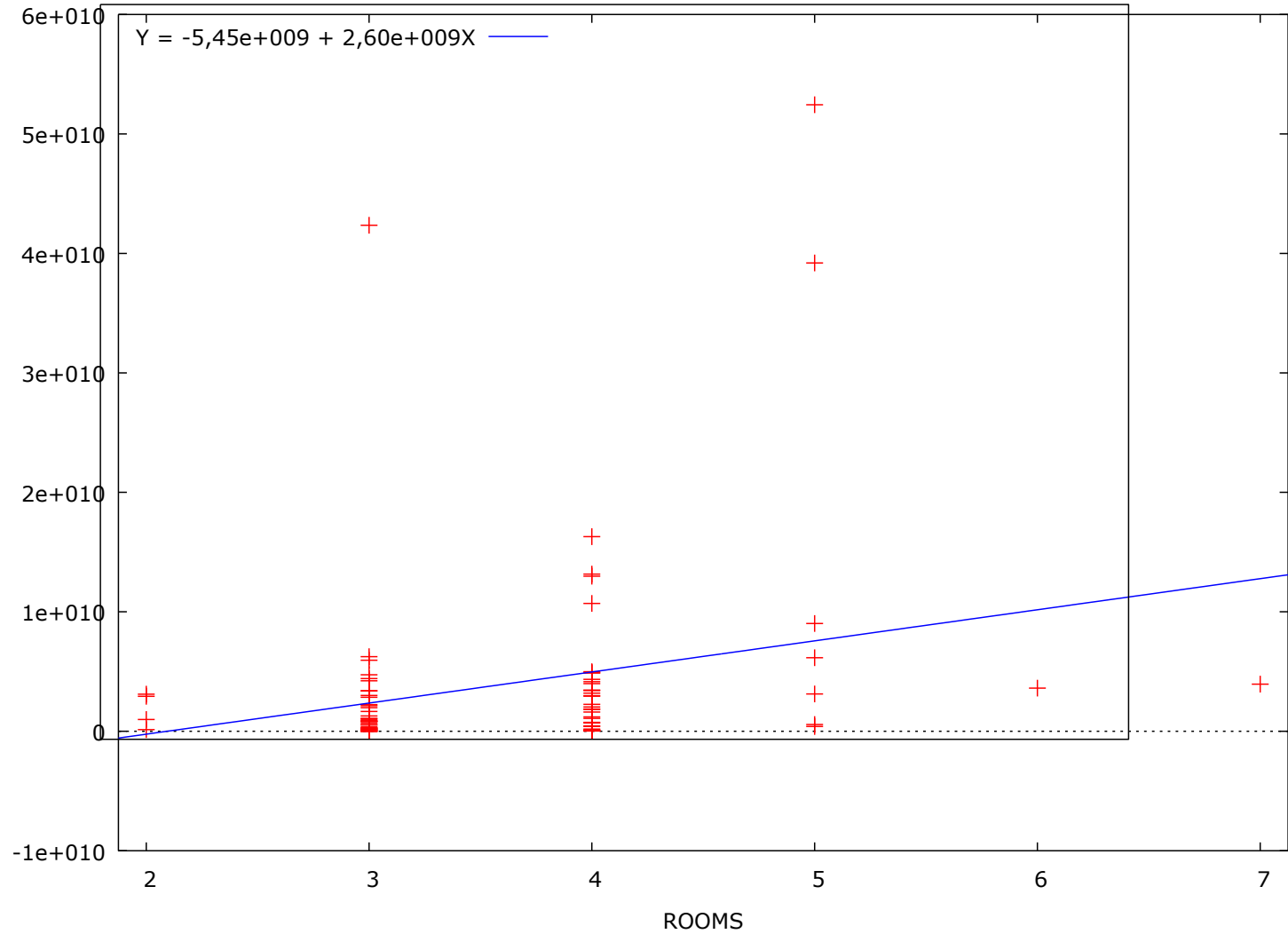
Model 1: OLS, using observations 1-88
Dependent variable: PRICE

	coefficient	std. error	t-ratio	p-value
-----	-----	-----	-----	-----
const	-19315,0	31046,6	-0,6221	0,5355
ROOMS	15198,2	9483,52	1,603	0,1127
SQFEET	128,436	13,8245	9,291	1,39e-014 ***
Mean dependent var	293546,0	S.D. dependent var	102713,4	
Sum squared resid	3,38e+11	S.E. of regression	63044,84	
R-squared	0,631918	Adjusted R-squared	0,623258	
F(2, 85)	72,96353	P-value(F)	3,57e-19	
Log-likelihood	-1095,881	Akaike criterion	2197,763	
Schwarz criterion	2205,195	Hannan-Quinn	2200,757	

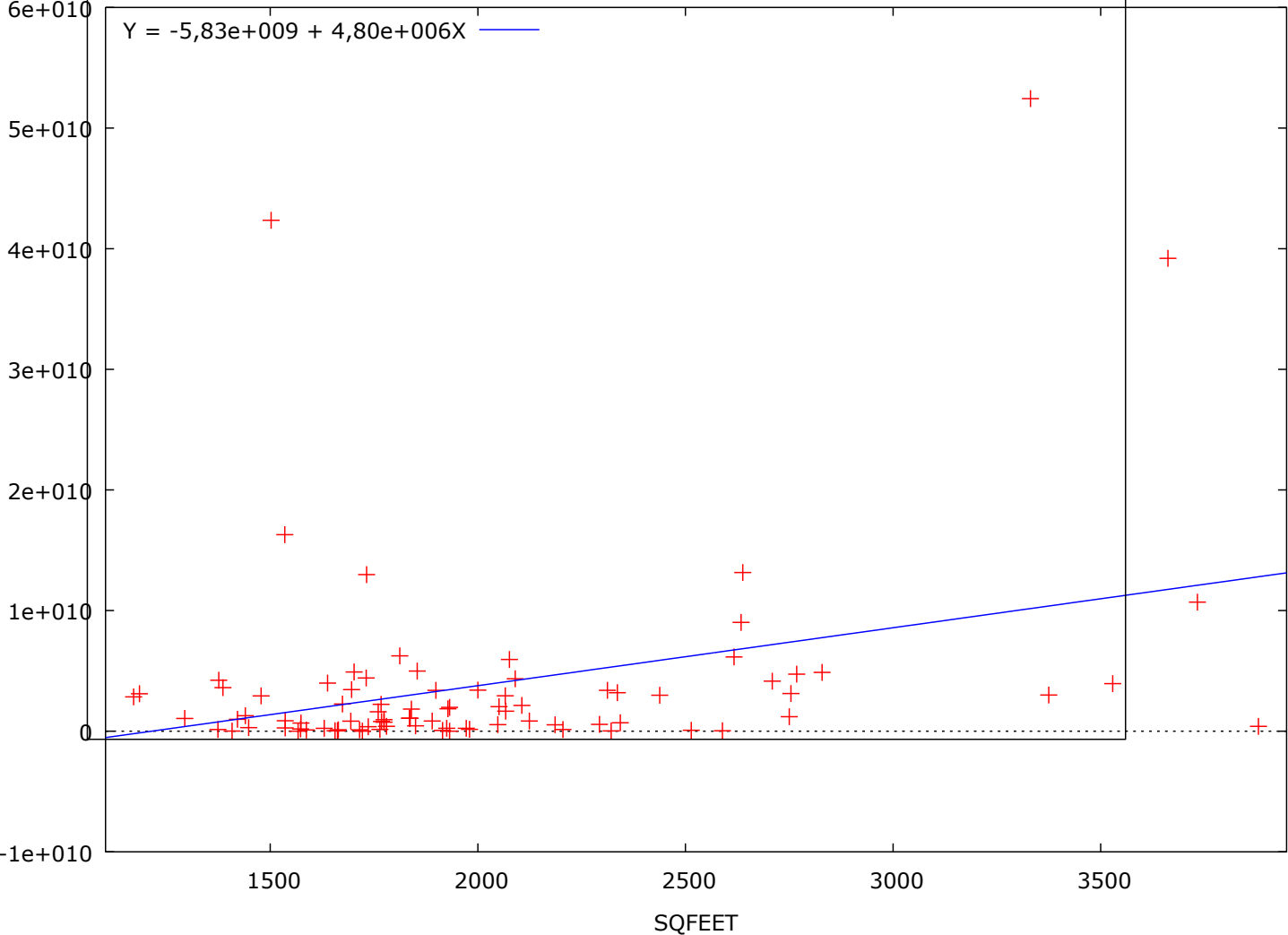
Jika indikasi heteroskedastisitas benar maka standar deviasi koefisien parameter tidak mengukur standar deviasi yang sebenarnya

Metode Grafis untuk mendeteksi Heteroskedastisitas

usq1 versus ROOMS (with least squares fit)



usq1 versus SQFEET (with least squares fit)



Terdapat Heteroskedastisitas

- Dari plot residual per peubah bebas diamati
- Berdasarkan jumlah kamar
 - Keragaman terbesar ada pada jumlah kamar yang medium
 - Ragam kecil pada jumlah kamar sedikit
 - Ragam kembali kecil pada jumlah kamar banyak
- Berdasarkan luas rumah
 - Ragam meningkat seiring luas rumah
- vt

Breusch-Pagan Test

- Langkah 1. Menduga model harga sebagai fungsi dari jumlah kamar dan luas rumah (model sebelumnya)
- Langkah 2. Menduga residual dari model tersebut dan menduga model:

$$\hat{u}_i^2 = a_1 + a_2 \mathit{Rooms}_i + a_3 \mathit{Sqfeet}_i + v_i$$

Model 2: OLS, using observations 1-88

Dependent variable: u_2

	coefficient	std. error	t-ratio	p-value	
const	-8.21788e+09	3.90705e+09	-2.103	0.0384	**
bdrms	1.18840e+09	1.19345e+09	0.9958	0.3222	
sqrft	3.88172e+06	1.73974e+06	2.231	0.0283	**
Mean dependent var	3.84e+09	S.D. dependent var	8.36e+09		
Sum squared resid	5.35e+21	S.E. of regression	7.93e+09		
R-squared	0.120185	Adjusted R-squared	0.099484		
F(2, 85)	5.805633	P-value(F)	0.004331		
Log-likelihood	-2129.248	Akaike criterion	4264.496		
Schwarz criterion	4271.928	Hannan-Quinn	4267.490		

$$LM = nR^2 = 88 \times 0.120185 = 10.58 \sim \chi_2^2 \quad \text{p-value: } 0.005^{**}$$

Goldfeld-Quant Test

- Untuk sorting dan pendugaan parameter dan JK setiap sub sampel
- Sorting dilakukan berdasarkan peubah jumlah kamar
- $1/6$ dari 88 pengamatan yang berada ditengah dihapuskan

White Test

- Menduga residual berdasarkan model awal: harga fungsi dari jumlah kamar dan luas rumah
- Menggunakan penduga residual untuk menduga model berikut:

$$\hat{u}_i^2 = a_1 + a_2 Rooms_i + a_3 Sqfeet_i + a_4 Rooms_i^2 + a_5 Sqfeet_i^2 + a_6 Rooms_i Sqfeet_i + v_i$$

- ▶ Hasil uji secara spesifik dapat menentukan variabel mana yang paling mempengaruhi ragam atau varian

White's test for heteroskedasticity
 OLS, using observations 1-88
 Dependent variable: uhat^2

	coefficient	std. error	t-ratio	p-value	
const	1,08097e+010	1,31453e+010	0,8223	0,4133	
ROOMS	6,99749e+09	5,66659e+09	1,235	0,2204	
SQFEET	-2,34047e+07	1,00764e+07	-2,323	0,0227	**
sq_ROOMS	-1,27783e+09	8,38898e+08	-1,523	0,1316	
X2_X3	1,97915e+06	1,81940e+06	1,088	0,2799	
sq_SQFEET	4020,88	2198,69	1,829	0,0711	*

Warning: data matrix close to singularity!

Unadjusted R-squared = 0,195741

Test statistic: $TR^2 = 17,225190$,
 with p-value = $P(\text{Chi-square}(5) > 17,225190) = 0,004092$

Sqfeet: luas rumah lebih mempengaruhi ragam/varian

Weighted Least Square

- Jika diasumsikan bahwa ragam berhubungan dengan luas rumah

$$\text{var}(u_i) = \sigma^2 Sqfeet_i$$

$$\frac{Price_i}{\sqrt{Sqfeet_i}} = \beta_1 \frac{1}{\sqrt{Sqfeet_i}} + \beta_2 \frac{Rooms_i}{\sqrt{Sqfeet_i}} + \beta_3 \frac{Sqfeet_i}{\sqrt{Sqfeet_i}} + v_i \quad v_i = \frac{u_i}{\sqrt{Sqfeet_i}}$$

$$\text{var}(v_i) = \frac{\text{var}(u_i)}{Sqfeet_i} = \sigma^2 \frac{Sqfeet_i}{Sqfeet_i} = \sigma^2$$

Dilakukan WLS
dengan:

$$weight = \frac{1}{\sqrt{Sqfeet_i}}$$

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	0	#N/A	#N/A	#N/A
1/sqfeet ^{0.5}	8008.411898	30728.95299	0.260615	0.79502
bdrms.sqfeet ^{0.5}	11578.29955	8961.647359	1.291983	0.199865
5				
<u>sqft.sqfeet^{0.5}</u>	<u>121.2817059</u>	<u>14.44144647</u>	<u>8.398169</u>	<u>8.92E-13</u>

Model 1: OLS, using observations 1-88

Dependent variable: PRICE

	coefficient	std. error	t-ratio	p-value
-----	-----	-----	-----	-----
const	-19315,0	31046,6	-0,6221	0,5355
ROOMS	15198,2	9483,52	1,603	0,1127
SQFEET	128,436	13,8245	9,291	1,39e-014 ***

White, Heteroskedasticity-corrected Regression

- Standar deviasi setiap parameter dihitung berdasarkan penduga varian/ragam yang berbeda untuk variabel eksogen yang berbeda
- Nilai t dan F mencerminkan sifat data yang sebenarnya
 - Tidak underestimated atau overestimated

Ragam
terkoreksi/robust:

$$\text{var}(\hat{\beta}_2) = \frac{\sum x_i^2 \hat{e}_i^2}{(\sum x_i^2)^2}$$

Model 3: Heteroskedasticity-corrected, using observations 1-88

Dependent variable: PRICE

	coefficient	std. error	t-ratio	p-value	
const	59618,6	31711,7	1,880	0,0635	*
ROOMS	2707,17	8977,31	0,3016	0,7637	
SQFEET	108,210	15,4204	7,017	5,13e-010	***

Statistics based on the weighted data:

Sum squared resid	279,6099	S.E. of regression	1,813706
R-squared	0,418814	Adjusted R-squared	0,405139
F(2, 85)	30,62634	P-value(F)	9,62e-11
Log-likelihood	-175,7332	Akaike criterion	357,4663
Schwarz criterion	364,8983	Hannan-Quinn	360,4605

Statistics based on the original data:

Mean dependent var	293546,0	S.D. dependent var	102713,4
Sum squared resid	3,74e+11	S.E. of regression	66352,61

Model 1: OLS, using observations 1-88

Dependent variable: PRICE

	coefficient	std. error	t-ratio	p-value	
const	-19315,0	31046,6	-0,6221	0,5355	
ROOMS	15198,2	9483,52	1,603	0,1127	
SQFEET	128,436	13,8245	9,291	1,39e-014	***

Cara lain mengatasi heteroskedastisitas

- Melakukan transformasi dengan membagi model regresi dengan salah satu variabel bebas
- Melakukan transformasi log