

Asumsi-asumsi OLS

Asumsi-asumsi yg mendasari berlakunya metode estimasi OLS atau disebut juga asumsi-asumsi model regresi linier klasik atau asumsi Gaussian, atau Asumsi Standar; ada 10 items:

Asumsi 1. Model Regresi Linier

Model persamaan regresi adalah linier dalam parameter

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Asumsi 2: Nilai X bersifat tetap dalam repeated sampling.

Nilai-nilai yang *dipilih* untuk variabel X bersifat tetap dalam repeated sampling. Atau X diasumsikan bersifat *non-stochastic*.

X : bersifat fixed
 Y : bersifat random

Repeated sampling in X

Obs	X	Y	obs	X	Y	obs	X	Y	obs	X	Y
1	20	35.92	11	30	43.15	21	40	49.09	31	50	59.05
2	20	35.08	12	30	43.39	22	40	50.47	32	50	57.64
3	20	35.48	13	30	40.70	23	40	49.57	33	50	57.15
4	20	35.47	14	30	43.82	24	40	50.01	34	50	57.69
5	20	36.29	15	30	42.97	25	40	50.24	35	50	57.53
6	20	34.82	16	30	43.53	26	40	51.00	36	50	58.31
7	20	36.61	17	30	43.14	27	40	49.50	37	50	58.05
8	20	36.81	18	30	41.92	28	40	50.35	38	50	56.20
9	20	35.53	19	30	43.32	29	40	50.29	39	50	56.51
10	20	34.36	20	30	42.85	30	40	48.99	40	50	57.88

Mean	35.637	42.879	49.951	57.601
Stdev	0.782	0.918	0.645	0.834

Asumsi 3: Nilai rerata disturbance u_i adalah nol.

Untuk nilai X tertentu, rerata atau expected value dari u_i adalah nol. Secara teknis disimbolkan : $E(u_i|X_i) = 0$

The regression equation is
 $Y = 21.0 + 0.730 X$

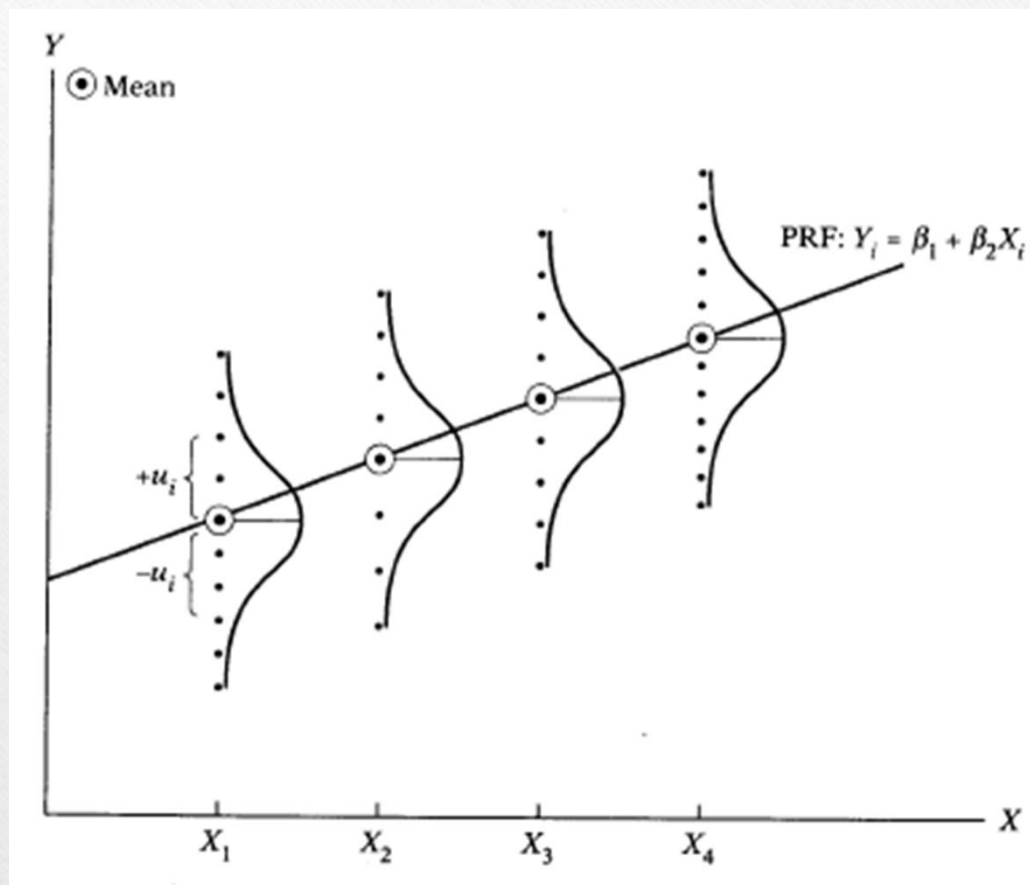
Predictor	Coef	SE Coef	T	P
Constant	20.9796	0.4110	51.04	0.000
X	0.72964	0.01119	65.22	0.000

S = 0.7910 R-Sq = 99.1% R-Sq(adj) = 99.1%

Rerata Error: **0.0002**

Obs	Y	Y-cap	Error	Obs	Y	Y-cap	Error
1	35.92	35.57	0.347	21	49.09	50.17	-1.076
2	35.08	35.57	-0.493	22	50.47	50.17	0.304
3	35.48	35.57	-0.093	23	49.57	50.17	-0.597
4	35.47	35.57	-0.103	24	50.01	50.17	-0.159
5	36.29	35.57	0.717	25	50.24	50.17	0.079
6	34.82	35.57	-0.753	26	51.00	50.17	0.834
7	36.61	35.57	1.037	27	49.50	50.17	-0.661
8	36.81	35.57	1.237	28	50.35	50.17	0.188
9	35.53	35.57	-0.043	29	50.29	50.17	0.129
10	34.36	35.57	-1.213	30	48.99	50.17	-1.176
11	43.15	42.87	0.285	31	59.05	57.46	1.584
12	43.39	42.87	0.524	32	57.64	57.46	0.180
13	40.70	42.87	-2.165	33	57.15	57.46	-0.312
14	43.82	42.87	0.954	34	57.69	57.46	0.227
15	42.97	42.87	0.098	35	57.53	57.46	0.068
16	43.53	42.87	0.665	36	58.31	57.46	0.845
17	43.14	42.87	0.270	37	58.05	57.46	0.593
18	41.92	42.87	-0.951	38	56.20	57.46	-1.260
19	43.32	42.87	0.454	39	56.51	57.46	-0.955
20	42.85	42.87	-0.016	40	57.88	57.46	0.416

- Visualisasi dari asumsi $E(u_i|X_i)=0$, atau disebut juga *conditional distribution of u_i*



X[1]	Error	X[2]	Error	X[3]	Error	X[4]	Error
20	0.34	30	0.29	40	-1.08	50	1.58
20	-0.49	30	0.52	40	0.30	50	0.18
20	-0.09	30	-2.16	40	-0.60	50	-0.31
20	-0.10	30	0.95	40	-0.16	50	0.23
20	0.72	30	0.10	40	0.08	50	0.07
20	-0.76	30	0.66	40	0.83	50	0.85
20	1.03	30	0.27	40	-0.66	50	0.59
20	1.23	30	-0.95	40	0.19	50	-1.26
20	-0.05	30	0.45	40	0.13	50	-0.95
20	-1.22	30	-0.02	40	-1.18	50	0.42
Mean	0.063		0.012		-0.214		0.138
Standard Deviation	0.782		0.918		0.646		0.833

Asumsi 4: *Homoscedasticity* atau *equal variance of u_i*

Untuk setiap nilai X_i tertentu (*given*), varians dari u_i adalah σ^2 . Secara matematis dapat disimbolkan:

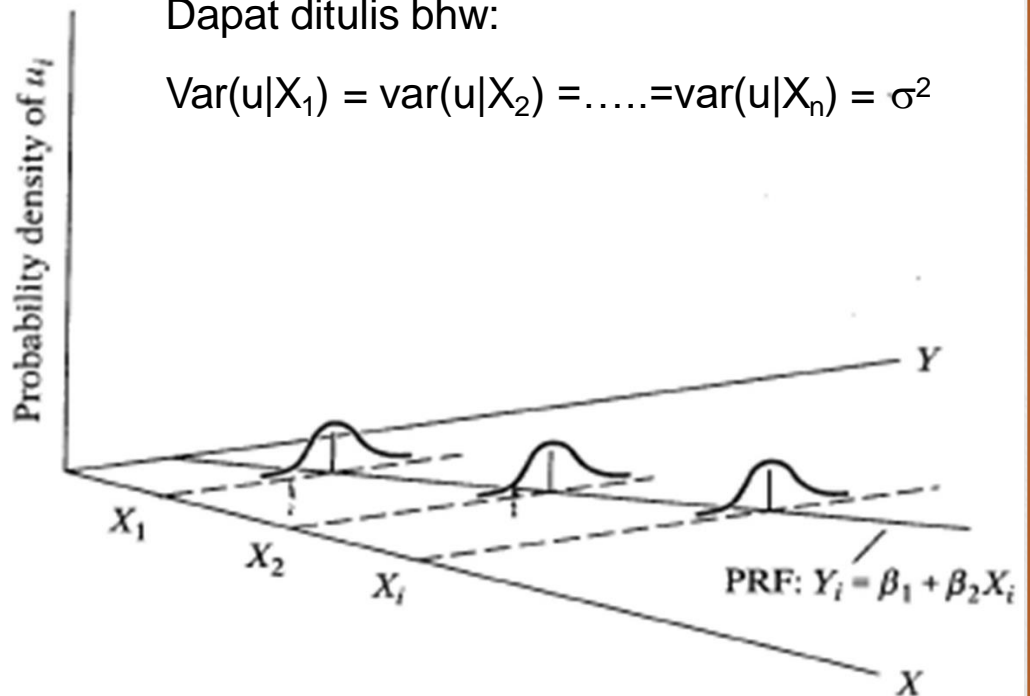
$$\begin{aligned} \text{Var}(u_i|X_i) &= E[(u_i|X_i) - E(u_i|X_i)]^2 \\ &= E(u_i^2|X_i) \\ &= \sigma^2 \end{aligned}$$

X[1]	Error	X[2]	Error	X[3]	Error	X[4]	Error
20	0.34	30	0.29	40	-1.08	50	1.58
20	-0.49	30	0.52	40	0.30	50	0.18
20	-0.09	30	-2.16	40	-0.60	50	-0.31
20	-0.10	30	0.95	40	-0.16	50	0.23
20	0.72	30	0.10	40	0.08	50	0.07
20	-0.76	30	0.66	40	0.83	50	0.85
20	1.03	30	0.27	40	-0.66	50	0.59
20	1.23	30	-0.95	40	0.19	50	-1.26
20	-0.05	30	0.45	40	0.13	50	-0.95
20	-1.22	30	-0.02	40	-1.18	50	0.42
Mean	0.063		0.012		-0.214		0.138
Var $u X$	0.611		0.843		0.417		0.693

$f(u)$

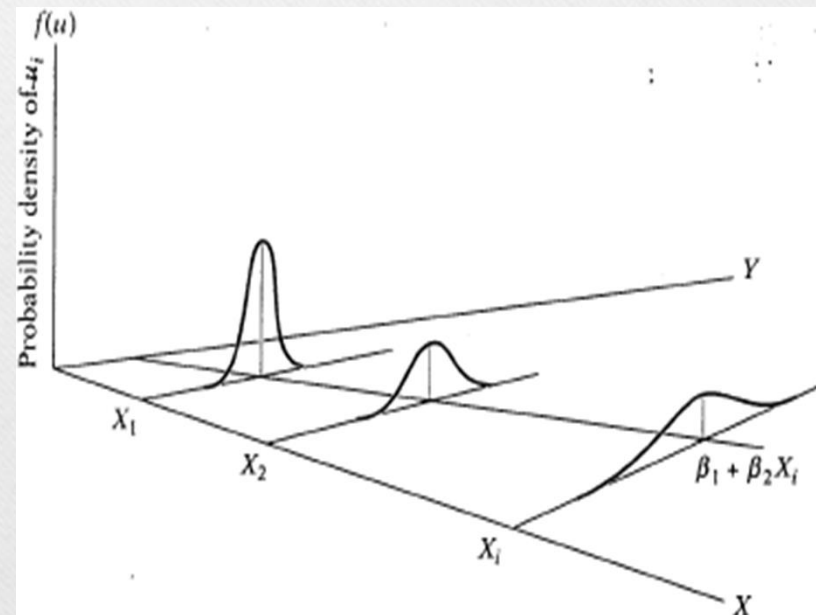
Dapat ditulis bhw:

$$\text{Var}(u|X_1) = \text{var}(u|X_2) = \dots = \text{var}(u|X_n) = \sigma^2$$



- Lawan dari *homoscedasticity* adalah *heteroscedasticity*, yaitu kondisi dimana di antara varians -varians *disturbance* antar-kelompok nilai X, ada yang tidak sama atau ada yang berbeda.
 - kondisi minimum
Minimal ada sepasang $\text{var}(u|X_i) \neq \text{var}(u|X_j)$ untuk $i \neq j$.
 - Kondisi paling ekstrim adalah: $\text{var}(u|X_1) \neq \text{var}(u|X_2) \neq \dots \neq \text{var}(u|X_n)$

$$\text{var}(u | X_1) < \text{var}(u | X_2) < \dots < \text{var}(u | X_n)$$



Data dengan gejala heteroskedastik

Obs	X	Y	obs	X	Y	obs	X	Y	obs	X	Y
1	20	35.92	11	30	43.15	21	40	52.89	31	50	63.28
2	20	35.08	12	30	43.39	22	40	42.70	32	50	58.41
3	20	35.48	13	30	40.70	23	40	45.43	33	50	54.59
4	20	35.47	14	30	43.82	24	40	51.66	34	50	59.45
5	20	36.29	15	30	42.97	25	40	52.03	35	50	62.26
6	20	34.82	16	30	43.53	26	40	56.51	36	50	56.56
7	20	36.61	17	30	43.14	27	40	46.90	37	50	57.34
8	20	36.81	18	30	41.92	28	40	54.44	38	50	57.54
9	20	35.53	19	30	43.32	29	40	51.87	39	50	59.53
10	20	34.36	20	30	42.85	30	40	49.50	40	50	53.75
						40		50.39			58.27
Mean	35.63			42.88						18.142	
Var	0.611			0.843						9.149	

Heteroskedastik
Dalam Nilai Y

The regression equation is
 $Y = 20.4 + 0.754 X$

Predictor	Coef	SE Coef	T	P
Constant	20.399	1.358	15.02	0.000
X	0.7541	0.03697	20.40	0.000

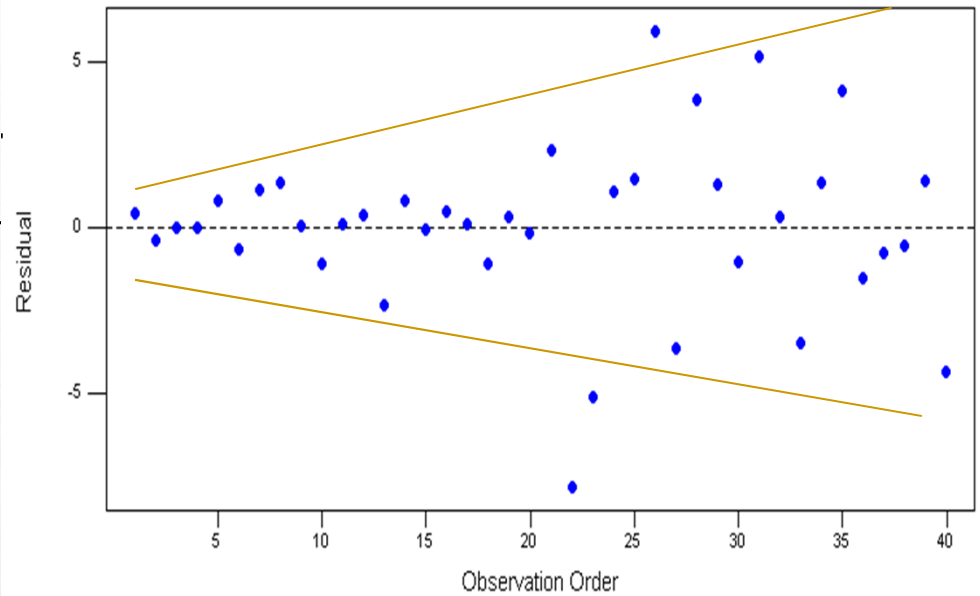
S = 2.614 R-Sq = 91.6% R-Sq(adj) = 91.4%

Y Actual dan
Y-Fitted

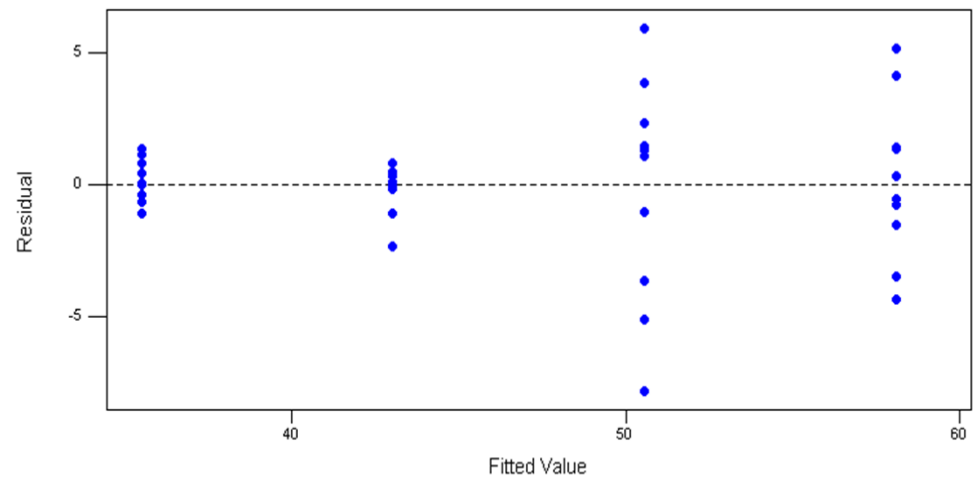
Y	Y-cap	Y	Y-cap	Y	Y-cap	Y	Y-cap
35.92	35.48	43.15	43.02	52.89	50.57	63.28	58.11
35.08	35.48	43.39	43.02	42.70	50.57	58.41	58.11
35.48	35.48	40.70	43.02	45.43	50.57	54.59	58.11
35.47	35.48	43.82	43.02	51.66	50.57	59.45	58.11
36.29	35.48	42.97	43.02	52.03	50.57	62.26	58.11
34.82	35.48	43.53	43.02	56.51	50.57	56.56	58.11
36.61	35.48	43.14	43.02	46.90	50.57	57.34	58.11
36.81	35.48	41.92	43.02	54.44	50.57	57.54	58.11
35.53	35.48	43.32	43.02	51.87	50.57	59.53	58.11
34.36	35.48	42.85	43.02	49.50	50.57	53.75	58.11

Distribusi Error yang heteroskedastik

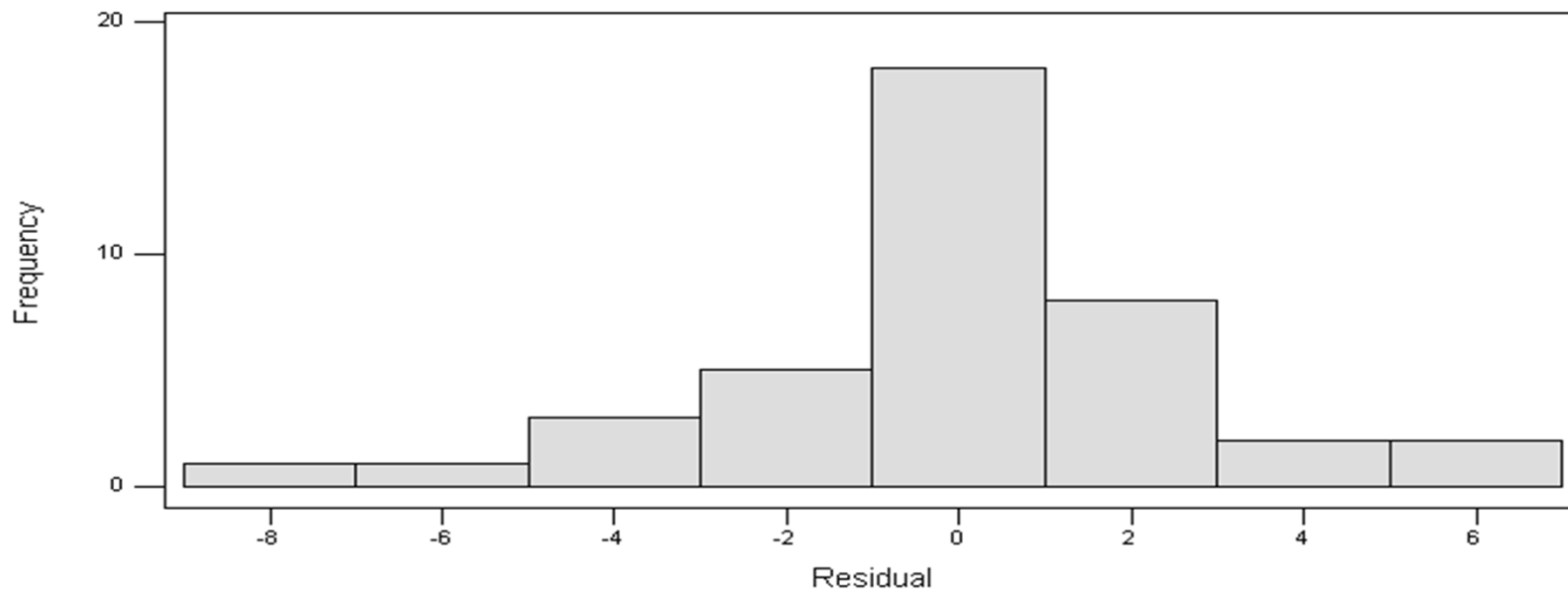
X[1]	Error X[2]	Error X[3]	Error X[4]	Error			
20	0.43	30	0.13	40	2.32	50	5.18
20	-0.40	30	0.37	40	-7.87	50	0.30
20	0.00	30	-2.32	40	-5.14	50	-3.52
20	-0.01	30	0.80	40	1.09	50	1.34
20	0.81	30	-0.06	40	1.47	50	4.16
20	-0.67	30	0.51	40	5.94	50	-1.55
20	1.12	30	0.11	40	-3.66	50	-0.77
20	1.32	30	-1.11	40	3.87	50	-0.57
20	0.04	30	0.30	40	1.31	50	1.42
20	-1.13	30	-0.17	40	-1.07	50	-4.36
Mean	0.153		-0.143		-0.173		0.163
Var	0.611		0.843		18.142		9.149



Residuals Versus the Fitted Values
(response is Y2)



Histogram of the Residuals
(response is Y2)



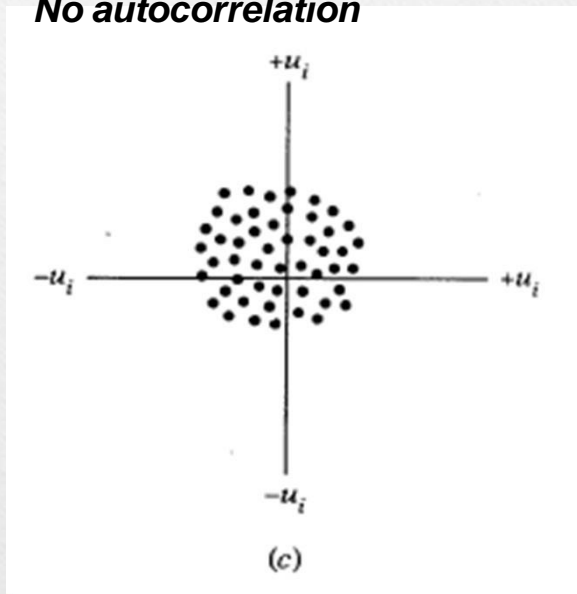
Distribusi Error atau Residual harus mengikuti distribusi Peluang Normal

Asumsi 5: *No autocorrelation between the disturbances.*

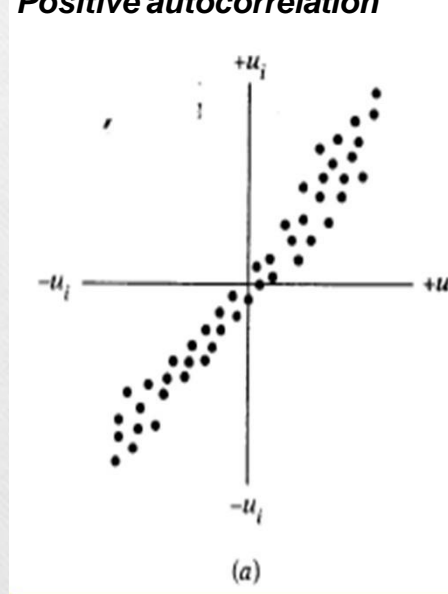
Untuk dua nilai X tertentu, X_i dan X_j ($i \neq j$), korelasi antara dua u_i dan u_j ($i \neq j$) adalah nol. Secara matematis:

$$\begin{aligned} \text{Cov}(u_i, u_j | X_i, X_j) &= E\{ (u_i - E(u_i) | X_i) \} \{ (u_j - E(u_j) | X_j) \} \\ &= E\{ (u_i | X_i)(u_j | X_j) \} \\ &= 0 \end{aligned}$$

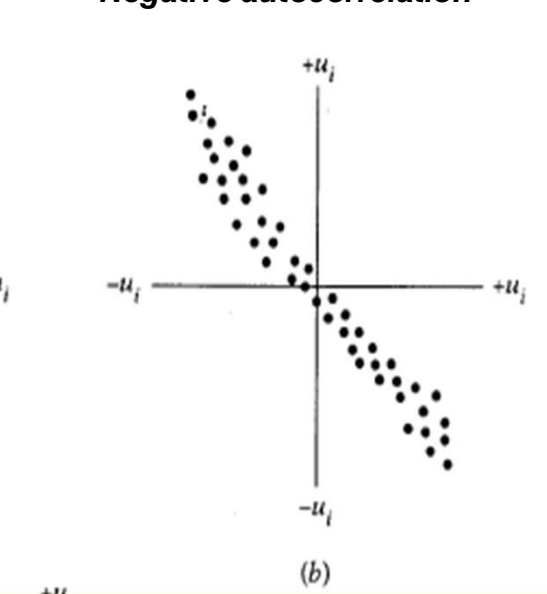
No autocorrelation



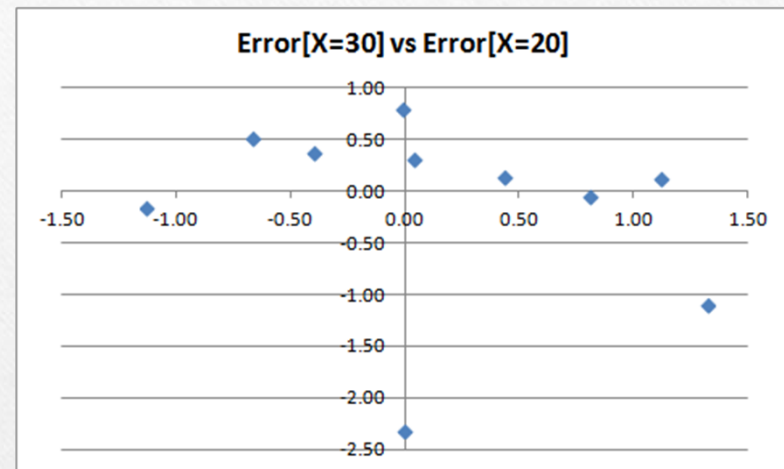
Positive autocorrelation



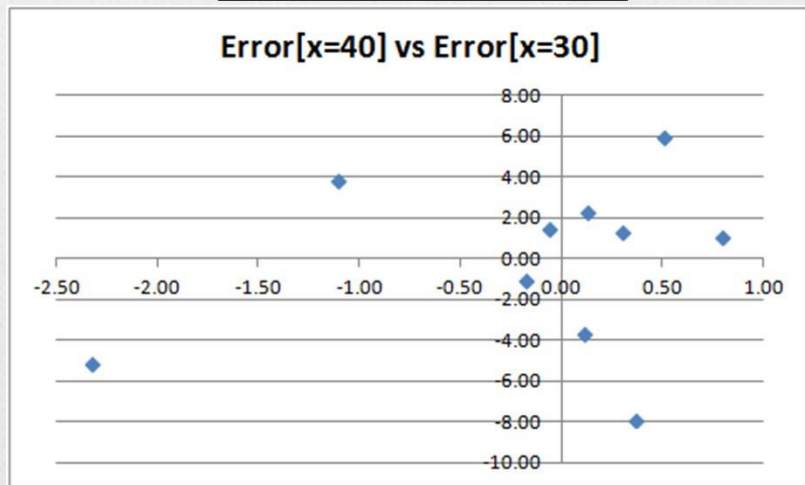
Negative autocorrelation



X[1]	Error X[2]	Error X[3]	Error X[4]	Error			
20	0.43	30	0.13	40	2.32	50	5.18
20	-0.40	30	0.37	40	-7.87	50	0.30
20	0.00	30	-2.32	40	-5.14	50	-3.52
20	-0.01	30	0.80	40	1.09	50	1.34
20	0.81	30	-0.06	40	1.47	50	4.16
20	-0.67	30	0.51	40	5.94	50	-1.55
20	1.12	30	0.11	40	-3.66	50	-0.77
20	1.32	30	-1.11	40	3.87	50	-0.57
20	0.04	30	0.30	40	1.31	50	1.42
20	-1.13	30	-0.17	40	-1.07	50	-4.36
Mean	0.153		-0.143		-0.173		0.163
Var	0.611		0.843		18.142		9.149



Tidak ada korelasi antar Error



Plot disamping menunjukkan tidak ada korelasi antara Residual[i] dan X[j], untuk $i \neq j$

Asumsi 6: Zero covariance between u_i and X_i ,

$$\text{Cov}(u_i, X_i) = E\{ (u_i - E(u_i))(X_i - E(X_i)) \}$$

$$= E\{ (u_i(X_i - E(X_i))) \}$$

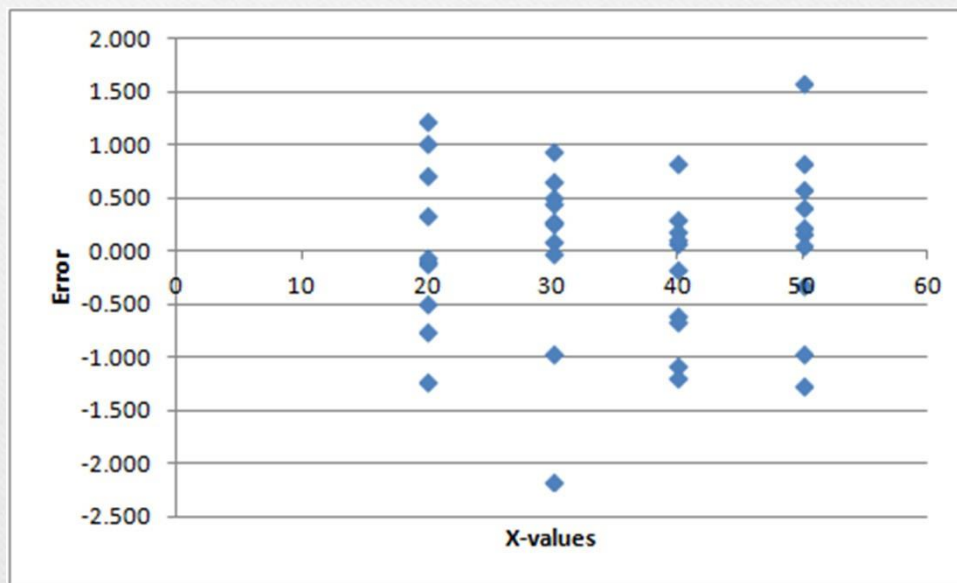
$$= E(u_i X_i) - E(X_i) E(u_i)$$

$$= E(u_i X_i)$$

$$= 0$$

Plot antara :

- Sumbu-vertikal: error atau residual model
- Sumbu-horisontal: Nilai-nilai X



Tampak tidak ada koralasi antara Nilai X dan Error/Residual . Jadi tidak melanggar asumsi

Asumsi 7: banyak pengamatan n harus lebih besar daripada banyaknya parameter yang diestimasi. Alternatif: banyak observasi harus jauh lebih besar daripada banyaknya variable penjelas.

Asumsi 8: Variabilitas X . Nilai-nilai dalam variable X tidak boleh sama. Alternatif: $\text{varians}(X)$ harus benar-benar bilangan positif.

Asumsi 9: Model regresi dispesifikasikan dengan benar. Alternatif: tidak ada bias spesifikasi dalam model yang digunakan dalam analisis empirik.

Asumsi 10: Tidak ada Multikolinieritas Sempurna. Tidak ada hubungan linier sempurna diantara variable penjelas.

Asumsi: Non Perfect Multicollinearity

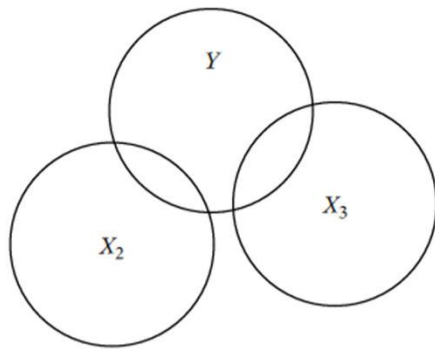
- Asumsi 10 dari model regresi linier klasik (CLRM) adalah tidak ada multikolinieritas di antara regressors yang dimasukkan ke dalam model regresi.
- Istilah multikolinieritas berasal dari Ragnar Frisch. Awalnya berarti adanya hubungan linear antara beberapa atau semua variabel penjelas dari model regresi yang "sempurna", atau tepat, .
- Untuk k-variabel regresi yang melibatkan variable penjelas X_1, X_2, \dots, X_k (di mana $X_1 = 1$ untuk semua pengamatan untuk memungkinkan suku intercept), hubungan tepat dikatakan linear:

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k = 0$$

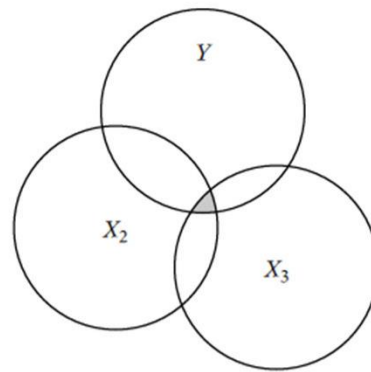
- Kasus di mana variabel X saling berhubungan tetapi tidak begitu sempurna:

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k + v_i = 0$$

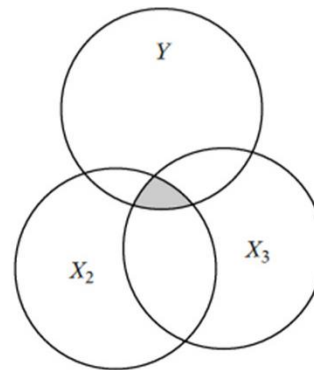
■ The Ballentine view of multicollinearity.



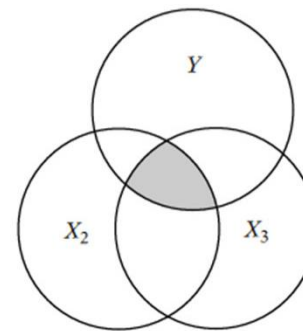
(a) No collinearity



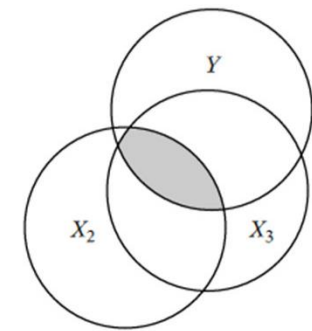
(b) Low collinearity



(c) Moderate collinearity



(d) High collinearity



(e) Very high collinearity

Why multicollinearity?

Jika multikolinieritas bersifat sempurna :

- koefisien regresi dari variable X tidak bisa ditentukan
- Standar error koefisien dugaan menjadi tidak terhingga,

Jika multikolinieritas kurang sempurna:

- Koefisien regresi walaupun bisa ditentukan, tetapi memiliki standar error (relatif terhadap koefisien itu sendiri) berarti koefisien-koefisien tidak dapat diestimasi dengan presisi dan akurasi yang tinggi.

Sumber Multikolinieritas

1. Metode pengumpulan data yang diterapkan. Terbatasnya kisaran nilai pada variable X ketika mengambil sampel dari populasi.
2. Kendala-kendala pada model atau populasi yang diambil sampel nya. Regresi konsumsi listrik oleh pendapatan (X_2) dan luas rumah (X_3). Secara fisik keluarga dengan pendapatan semakin besar, memiliki rumah semakin besar.
3. Spesifikasi model. Menambahkan suku polynomial (X^2 , X^3) ketika kisaran nilai X sangat pendek.
4. Overdetermined Model. Terlalu banyak variable penjelas.

Konsekuensi Praktis dari Multicollinierity

Dalam kasus multikolinieritas sedang atau tinggi, kemungkinan akan menghadapi konsekuensi berikut:

1. Walaupun BLUE, penduga OLS memiliki variance dan covariance yang terlalu besar, menyulitkan memperoleh estimasi yang akurat
2. Oleh karena konsekuensi 1, selang keyakinan cenderung lebih lebar, lebih sering mengarahkan untuk menerima “Hipotesis null”
3. Oleh karena konsekuensi 1, rasio-t dari satu atau lebih variable secara statistic tidak signifikan.
4. Walaupun rasio-t dari satu atau lebih variable tidak signifikan, namun R^2 sangat tinggi.
5. Penduga OLS dan standar error menjadi sensitive terhadap perubahan kecil dalam data.

Bagaimana mendeteksi

1. High R^2 but few significant t ratios.
2. High pair-wise correlations among regressors.
3. Examination of partial correlations.
4. Auxiliary regressions.

Ilustrasi Multikolinieritas

Y	X1	X2	X3	Koefisien korelasi antar-variabel
---	----	----	----	-----------------------------------

167.349	50	110	25.1672
158.921	50	110	23.5555
158.945	50	110	23.8647
168.751	50	110	24.7598
164.548	50	110	25.0701
198.235	60	100	30.5601
198.677	60	100	30.5268
191.024	60	100	28.8592
197.178	60	100	30.5772
190.825	60	100	28.7815
245.532	80	90	40.0588
245.885	80	90	40.6819
253.976	80	90	42.0141
248.709	80	90	40.3368
249.185	80	90	38.4482
300.875	100	70	47.8078
302.493	100	70	50.8256
308.509	100	70	51.2236
305.338	100	70	50.2455
304.256	100	70	50.6173

penjelas

	X1	X2	X3
X1	1		
X2	-0.99027	1	
X3	0.994819	-0.98364	1

The regression equation is

$$Y = 90.1 + 1.21 X1 - 0.417 X2 + 2.45 X3$$

Predictor	Coef	SE Coef	T	P	Paramater
Constant	90.09	46.72	1.93	0.072	50.00
X1	1.2112	0.4302	2.82	0.012	1.90
X2	-0.4171	0.3152	-1.32	0.204	-0.15
X3	2.4488	0.6441	3.80	0.002	1.50

S = 2.885 R-Sq = 99.8% R-Sq(adj) = 99.7%

Dugaan
berbias

Ilustrasi Output Regresi dengan Regressor Lain

Regression Analysis: Y versus X1, X2

The regression equation is
 $Y = 68.8 + 2.57 X1 - 0.290 X2$

Predictor	Coef	SE Coef	T	P
Constant	68.82	62.09	1.11	0.283
X1	2.5651	0.3231	7.94	0.000
X2	-0.2901	0.4195	-0.69	0.499

S = 3.86149 R-Sq = 99.6% R-Sq(adj) = 99.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	57265	28632	1920.20	0.000
Residual Error	17	253	15		
Total	19	57518			

Source	DF	Seq SS
X1	1	57257
X2	1	7

Regression Analysis: Y versus X1, X3

The regression equation is
 $Y = 28.3 + 1.58 X1 + 2.36 X3$

Predictor	Coef	SE Coef	T	P
Constant	28.346	2.659	10.66	0.000
X1	1.5758	0.3377	4.67	0.000
X3	2.3583	0.6545	3.60	0.002

S = 2.94822 R-Sq = 99.7% R-Sq(adj) = 99.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	57370	28685	3300.18	0.000
Residual Error	17	148	9		
Total	19	57518			

Source	DF	Seq SS
X1	1	57257
X3	1	113

Unusual Observations

Obs	X1	Y	Fit	SE Fit	Residual	St Resid
16	100	300.875	298.672	2.005	2.203	1.02 X

X denotes an observation whose X value gives it large leverage.

Variance Inflation Factor

- A general rule is if the correlation between two independent variables is between -0.70 and 0.70 there likely is not a problem using both of the independent variables.
- A more precise test is to use the **variance inflation factor (VIF)**.
- A $VIF > 10$ is unsatisfactory. Remove that independent variable from the analysis.
- The value of VIF is found as follows:

$$VIF = \frac{1}{1 - R_j^2}$$

The term R_j^2 refers to the coefficient of determination, where the selected independent variable is used as a dependent variable and the remaining independent variables are used as independent variables.

Variance Inflation Factor: Regresi X1 oleh X2 dan X3

Y	X1	X2	X3
167.349	50	110	25.1672
158.921	50	110	23.5555
158.945	50	110	23.8647
168.751	50	110	24.7598
164.548	50	110	25.0701
198.235	60	100	30.5601
198.677	60	100	30.5268
191.024	60	100	28.8592
197.178	60	100	30.5772
190.825	60	100	28.7815
245.532	80	90	40.0588
245.885	80	90	40.6819
253.976	80	90	42.0141
248.709	80	90	40.3368
249.185	80	90	38.4482
300.875	100	70	47.8078
302.493	100	70	50.8256
308.509	100	70	51.2236
305.338	100	70	50.2455
304.256	100	70	50.6173

	Coefficients	Standard Error	t Stat	P-value
Intercept	90.094	46.725	1.928	0.072
X1	1.211	0.430	2.815	0.012
X2	-0.417	0.315	-1.324	0.204
X3	2.449	0.644	3.802	0.002

Regression Statistics	
Multiple R	0.998842
R Square	0.997685
Adjusted R Square	0.99725
Standard Error	2.885111
Observations	20

$$VIF = \frac{1}{1 - 0.998} = 431.88 > 10$$

VIF terlalu besar, korelasi antar variable bebas sangat besar, mengganggu validitas dan akurasi dugaan. Perlu mengeluarkan satu atau lebih variable bebas

Bagaimana mengatasi ?

- A priori information.
- Combining cross-sectional and time series data.
- Dropping a variable(s) and specification bias (*But in dropping a variable from the model specification bias or specification error*)
- Transformation of variables.
- Reducing colinearity in polynomial regressions.

Asumsi: Homoskedastisitas (Homoscedasticity atau non Heteroskedastisitas (Heteroscedasticity)

- Homoskedastisitas?
- Asumsi homoskedastisitas, homo=sama, scedasticity =variance, atau “kesamaan variance”:

$$E(u_i^2) = \sigma^2 \quad i = 1, 2, \dots, n$$

- Variance Y_i tidak sama, jadi disebut heteroskedastisitas:

$$E(u_i^2) = \sigma_i^2$$

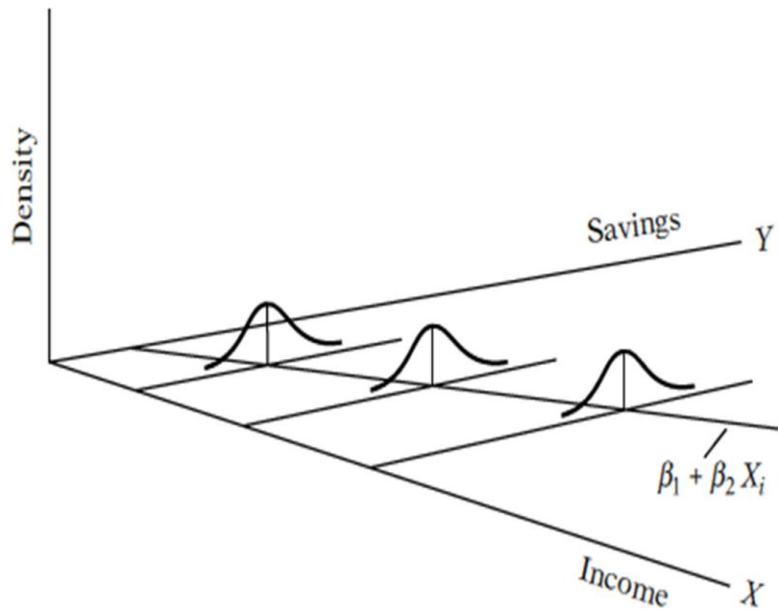


FIGURE 11.1 Homoscedastic disturbances.

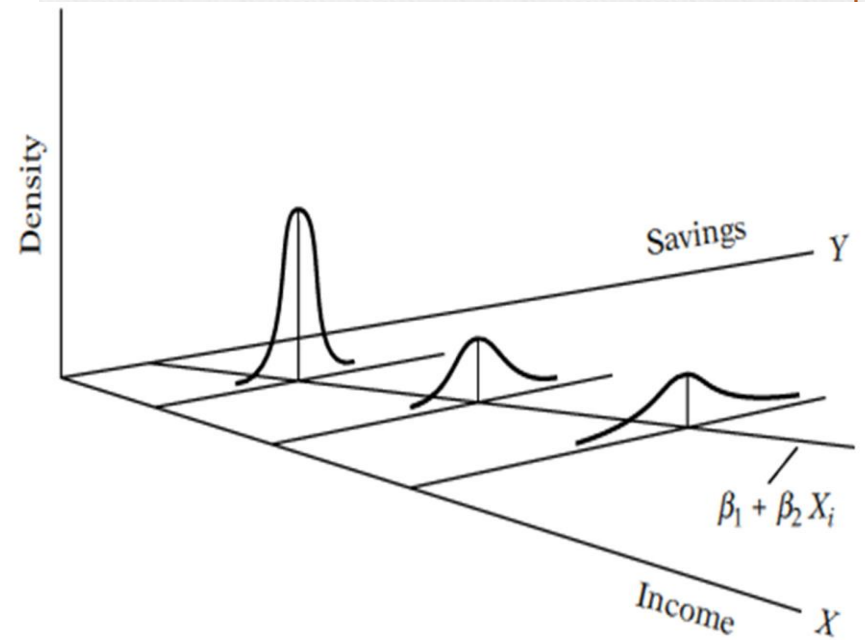


FIGURE 11.2 Heteroscedastic disturbances.

- Heteroscedasticity

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

- Homoscedasticity

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

Penyebab?

- Mengikuti Error-Learning model, seperti orang belajar, kesalahan perilaku mereka menjadi lebih kecil dari waktu ke waktu

- Dengan tumbuhnya pendapatan, orang memiliki lebih banyak pendapatan dan karenanya cakupan pilihan belanja semakin banyak.
- Sejalan dengan perbaikan pengumpulan data, σ_i^2 semakin menurun.
- Heteroskedastisitas juga bisa timbul sebagai akibat dari kehadiran outlier.
- Sumber lain dari heteroskedastisitas muncul dari melanggar Asumsi 9 dari CLRM, yaitu, bahwa model regresi ditentukan dengan benar.
- Sumber lain dari heteroskedastisitas adalah skewness dalam distribusi satu atau lebih regressors dimasukkan dalam model

Konsekuensi?

- Singkatnya, jika kita bertahan dalam menggunakan prosedur pengujian yang biasa meskipun hadir gejala heteroskedastisitas, kesimpulan apa pun yang kita buat mungkin sangat menyesatkan.

Cara Mendeksi?

- Graphical Method

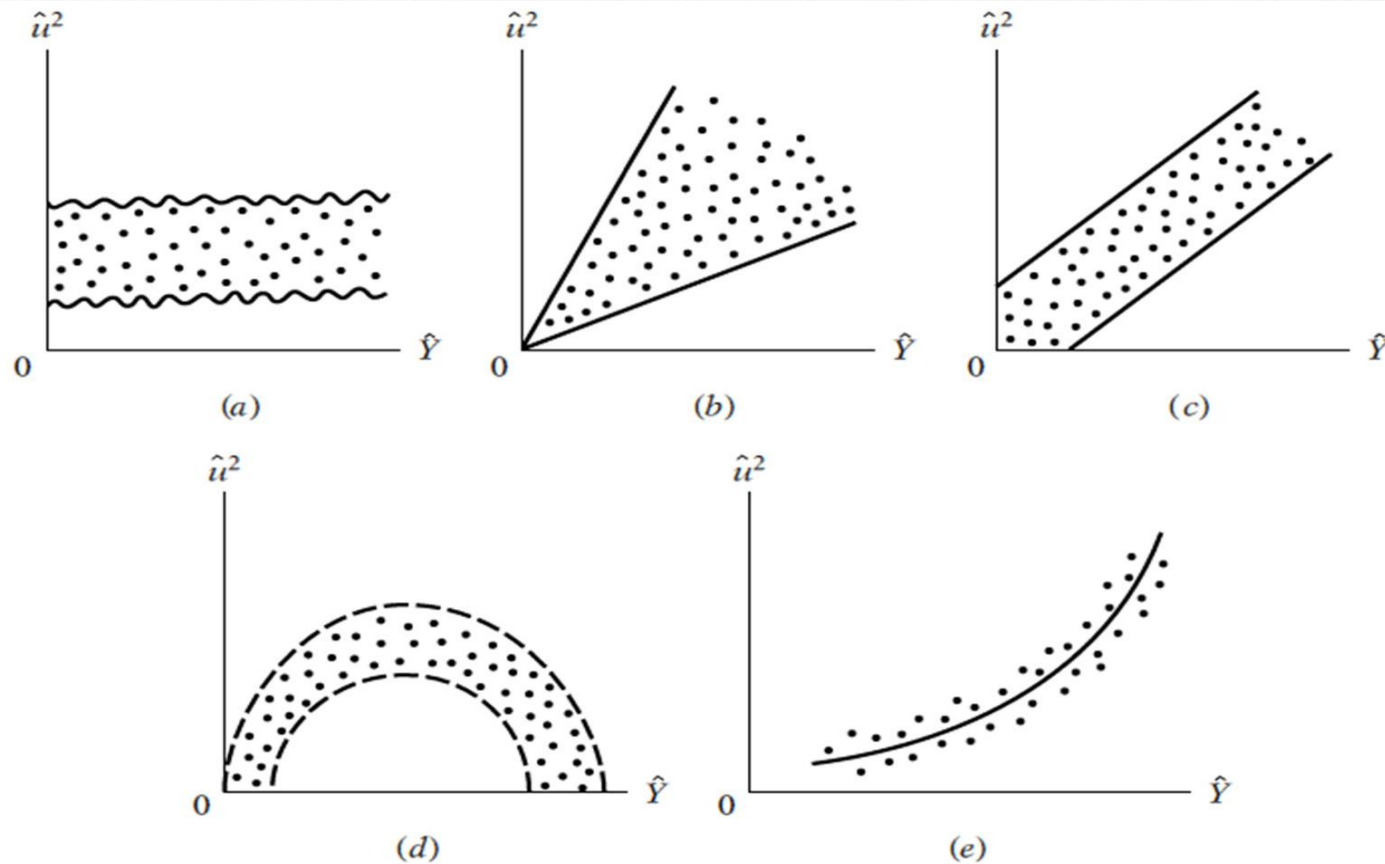


FIGURE 11.8 Hypothetical patterns of estimated squared residuals.

No Serial Autocorrelation
(Autokorelasi serial): Konsekuensi,
Cara Mendeteksi dan Mengatasinya

Pengertian Autokorelasi

- *autocorrelated error* menggambarkan korelasi antara error e_i dengan e_j untuk $i \neq j$.
- Di dalam literatur ekonometrika persoalan ini sering disebut dengan otokorelasi yang merupakan persoalan yang umum ditemukan dalam data *time series*.

Penyebab adanya Autokorelasi

1. Adanya *shocks* yang seringkali pengaruhnya tetap muncul dalam suatu periode waktu yang cukup lama.
2. *inertia* atau *psychological conditioning*.
3. Salah spesifikasi (Adanya variabel yang penting tidak masuk dalam model dan bentuk fungsi tidak tepat)

Penyebab adanya Autokorelasi

4. Lag

Dalam model *autoregressive* terdapat variable bebas yang nilainya merupakan lag dari variable terikat.

5. Manipulasi data

Misalkan seseorang dapat memperoleh data kuartalan dari data bulanan dengan merata-ratakan data secara 3 bulanan. Sedang data untuk kuartal ke dua diperoleh dengan merata-ratakan data secara 3 bulanan selanjutnya. Jika kita melakukan ini, maka kita akan mendapatkan *smoothness*/kehalusan dalam data yang tidak ada sebelumnya. Selanjutnya ini akan mempengaruhi *error term*

6. Fenomena CobWeb

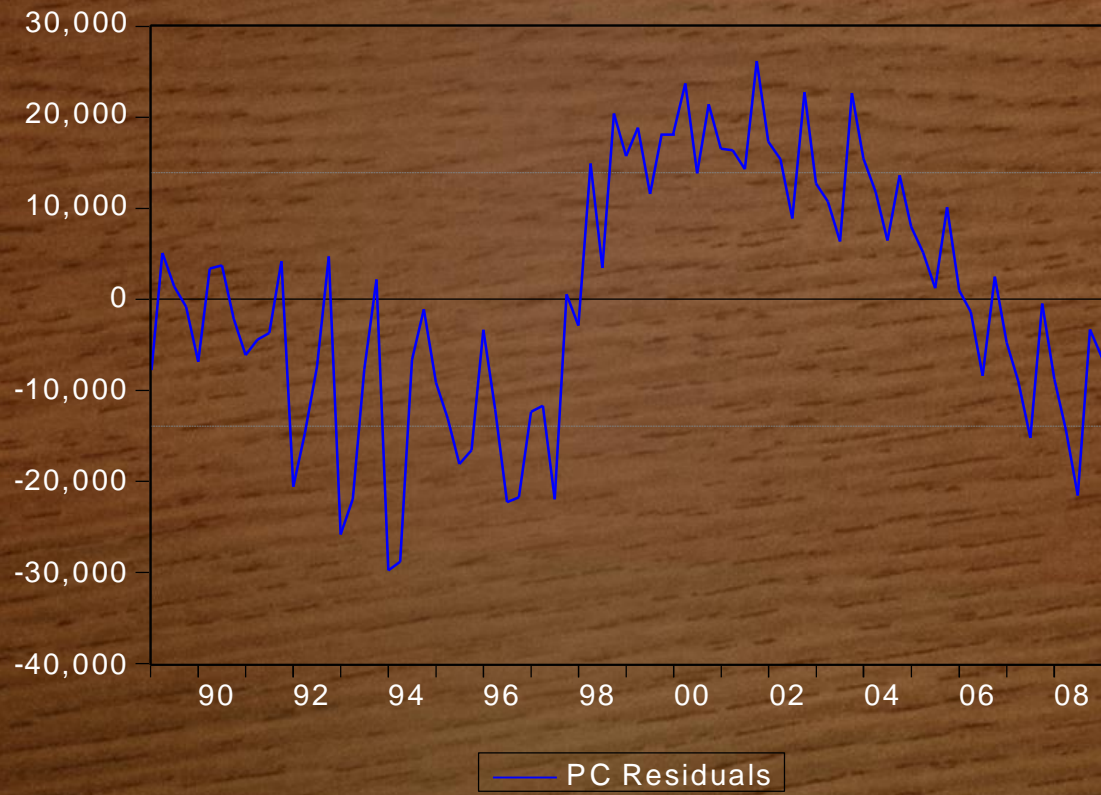
Jika pada akhir t , harga pertanian saat t lebih kecil dibanding $t-1$, maka supply pertanian saat $t+1$ lebih kecil dibanding saat t . Sehingga, *error* pada saat t (u_t), tidak akan *random*, karena jika petani memproduksi hasil pertanian berlebih (*overproduce*) pada saat t , maka mereka akan mengurangi produksi saat $t+1$, sehingga membentuk pola *Cobweb*.

Konsekuensi Adanya Autokorelasi

1. Estimasi OLS tetap linear dan tidak bias namun tidak lagi efisien/BBLUE (variannya tidak minimum).
2. Interval keyakinan akan semakin lebar, menyebabkan kita menerima hipotesa H_0 (koefisien tidak signifikan).
3. R^2 juga akan *over estimate*.
4. t-stat dan F-ratio akan tidak valid; yang jika digunakan akan menyebabkan kesimpulan yang salah.

Cara Mendeteksi Keberadaan Autokorelasi : Cara Visual

- Cara Grafik :Residualnya berpola



Cara Mendeteksi Keberadaan Autokorelasi: DW test

Durbin Watson Test

H0: NoAutocorrelation

H1: Autocorrelation

Terdapat Autokorelasi Positif	Inconclusive Interval	Ho diterima Tidak ada Autokorelasi	Inconclusive Interval	Terdapat Autokorelasi Negatif
	(dl)	(du) 2	(4-du)	(4-dl)

Jika nilai Durbin-Watson Statistik terletak diantara nilai (du) dan (4-du) maka tidak terdapat autokorelasi di dalam model. Nilai dl dan du didapat dari tabel statistik Durbin-Watson dengan *significance level* tertentu (biasanya 5%) dan *degree of freedom* (didasarkan dari besarnya sampel dikurang dengan jumlah parameter yang diestimasi di dalam model).

Simple Rule of Thumb:

Jika nilai Durbin Watson statistik ada disekitar angka dua maka tidak terdapat autokorelasi di dalam model

Cara Mendeteksi Keberadaan Autokorelasi : BG-test

■ The Breusch-Godfrey (BG) Test

Digunakan Jika model regresi yang kita gunakan terdapat unsur lag variabel dependen pada variabel independennya, maka nilai Durbin Watson statistik ini tidak valid lagi. Jika ini terjadi gunakan tes statistik lain untuk mencek autokorelasi (misal dengan BG-test).

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	46.76927	Prob. F(2,78)	0.0000
Obs*R-squared	44.71392	Prob. Chi-Square(2)	0.0000

Test Equation:

Dependent Variable: RESID

Method: Least Squares

Date: 03/09/10 Time: 14:07

Sample: 1989Q1 2009Q2

Included observations: 82

Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1761.852	4606.631	0.382460	0.7032
PDB	-0.005328	0.012476	-0.427027	0.6705
RESID(-1)	0.552783	0.110516	5.001831	0.0000
RESID(-2)	0.239940	0.110732	2.166864	0.0333

R-squared	0.545292	Mean dependent var	-3.87E-11
Adjusted R-squared	0.527803	S.D. dependent var	13867.13
S.E. of regression	9529.023	Akaike info criterion	21.20962
Sum squared resid	7.08E+09	Schwarz criterion	21.32702
Log likelihood	-865.5945	Hannan-Quinn criter.	21.25676
F-statistic	31.17951	Durbin-Watson stat	2.100832
Prob(F-statistic)	0.000000		

**TERIMA
KASIH**