

Bahan Praktikum Ekonometrika I

Pengujian Asumsi Klasik: Normalitas dan Multikolinieritas

IMAM AWALUDDIN

Review Asumsi-asumsi yang mendasari Metode OLS

Ada 10 asumsi yang mendasari **model Gaussian, model standar**, atau **CLRM** (Classical Linear Regression Model).

1. Regresi linier pada parameter
2. Nilai peubah bebas (eksogen) dianggap non stokastik (fixed)
3. Galat mempunyai nilai harapan nol $\rightarrow E(u) = 0 \rightarrow$
4. Homoskedastisitas: ragam yang sama pada galat
5. Galat tidak saling berkorelasi (tidak ada otokorelasi)
6. Peubah bebas (eksogen) dan galat saling bebas

Review Asumsi-asumsi yang mendasari Metode OLS

7. Jumlah pengamatan harus lebih besar daripada jumlah parameter yang akan diduga
8. Nilai peubah bebas harus bervariasi
9. Model regresi harus dispesifikasikan dengan tepat: *no specification bias*
10. Tidak ada multikolinieritas sempurna

Catatan: tambahan asumsi no. 10 adalah perluasan asumsi model regresi majemuk

⇒ (Perfect Multicollinearity)

⇒ (exact multicollinearity)

↓
for the estimation

Pengujian Asumsi-asumsi di dalam Regresi Linier

- Galat menyebar normal
- Multicollinearity
- Heteroscedasticity
- Autocorrelation
- Misspecification:
 - Peubah bebas yang kurang tepat
 - Measurement errors
 - Bentuk fungsional yang salah

Data yang digunakan untuk Praktikum

Data Buku Gujarati & Porter 5th Edition:

1. Table 10_2
2. Table 10_5
3. Table 10_7

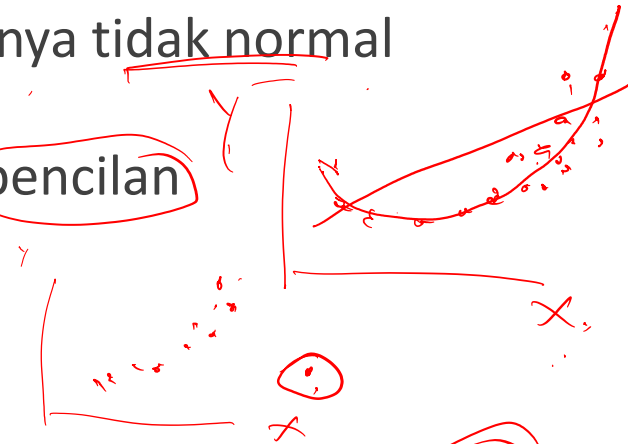
Data $\rightarrow$ Cy. ① sampel besar ($n \geq 30$).
normal. ② ada penculan (outlier).

Statistika $\rightarrow$ Parametrik
Nonparametrik

Asumsi kenormalan Residual

► Pelanggaran, dengan kemungkinan penyebab:

1. Sebaran peubah eksogen atau endogennya tidak normal
2. Pelanggaran asumsi linieritas variabel
3. Sebaran galat menjulur karena adanya penculan
4. Ukuran sampel yang terlalu kecil



► Efek pelanggaran:

- Penculan berpengaruh besar terhadap penduga parameter (bias)
- Hasil pengujian tidak sah
- Selang kepercayaan terlalu lebar atau terlalu sempit

Petai Kaca $\rightarrow$ 1000 KK $\rightarrow$ 3% $\rightarrow$ 50 $\rightarrow$ 50

Uji Normalitas Residual

- Bertujuan untuk menguji apakah dalam model regresi , variabel pengganggu atau residual memiliki distribusi normal.
- Model regresi yang baik harus memenuhi asumsi bahwa residual data terdistribusi normal.

Asumsi Kenormalan

▶ Bagaimana mendeteksinya?

- ▶ Normal probability plot ✓
- ▶ Histogram dari sisaan ✓
- ▶ Chi square goodness test of fit
- ▶ Anderson Darling normality test
- ▶ Jarque-Berra normality test ✓



Deteksi Normalitas Residual

Analisis Grafik

Salah satu cara termudah untuk melihat normalitas dari residual data adalah dengan melihat grafik histogram yang membandingkan antara data observasi dengan data distribusi.

- Jika data menyebar disekitar garis diagonal dan mengikuti arah garis diagonal atau grafik histogramnya menunjukkan pola distribusi normal, maka model regresi memenuhi asumsi normalitas.
- Jika data menyebar jauh dari diagonal dan/atau tidak mengikuti arah garis diagonal atau grafik histogram tidak menunjukkan pola distribusi normal, maka model regresi tidak memenuhi asumsi normalitas.

Uji Normalitas Residual

Uji {
- normalitas
- residual
- heteroskedastisitas
- autokorelasi

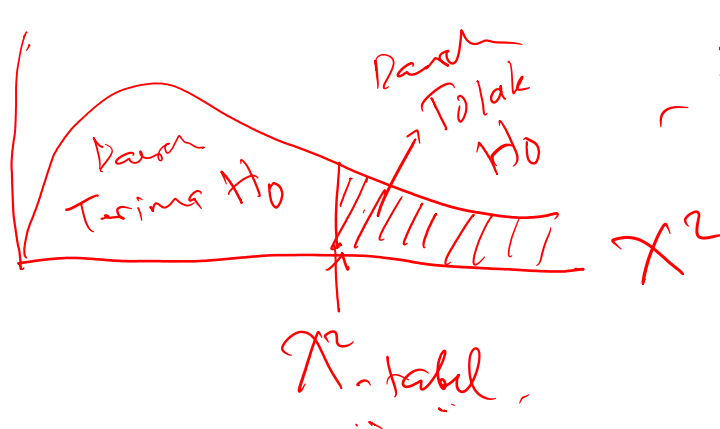
Uji secara formal

→ Dengan menggunakan uji Jarque-Berra

χ^2

H_0 : Residu tersebar secara normal

H_a : residu tidak tersebar secara normal



Kolmogorov-Smirnov

$$JB = n \left[\frac{s^2}{6} + \frac{(k-3)^2}{24} \right]$$

S = skewness.
 K = kurtosis.

Contoh Table 10_7 Buku Gujarati

Masukkan data dalam Table 10_7

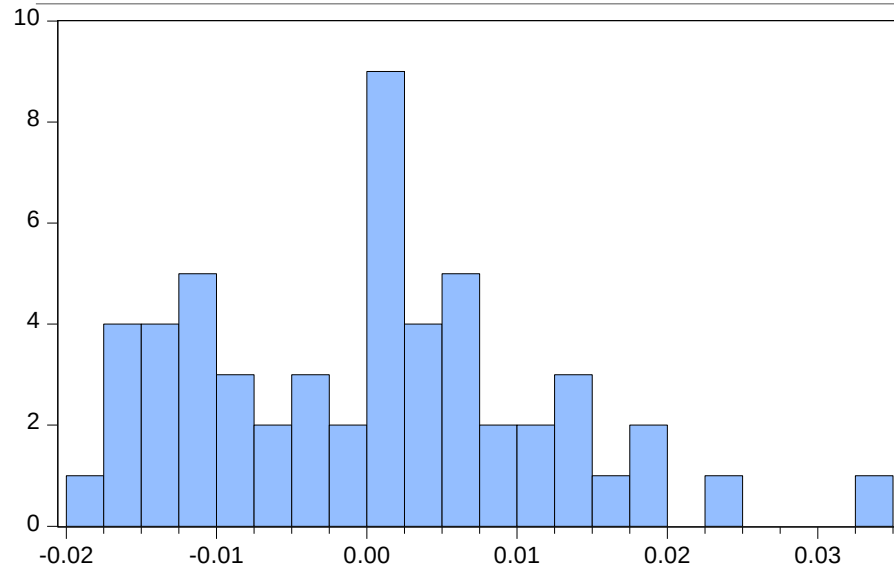
Estimasi dengan persamaan: $\log(\text{cons}) = c + \log(yd) + \log(w)$ I

Catatan log di sini adalah logaritma natural

Simpan persamaan estimasinya dengan memberi nama missal eq01.

Pada persamaan eq01, klik view, pilih residual diagnostic, pilih Histogram – Normality test.

Hasil Uji Normalitas Residual dengan JB



Series: Residuals
Sample 1947 2000
Observations 54

Mean	1.19e-15
Median	0.000337
Maximum	0.032578
Minimum	-0.018441
Std. Dev.	0.011591
Skewness	0.437812
Kurtosis	2.810938

Jarque-Bera	1.805543
Probability	0.405444

Dengan menggunakan uji
Jarque-Berra

H_0 : Residu tersebar
secara normal

H_a : residu tidak
tersebar secara
normal

Asumsi Kenormalan

► Jika dilanggar, bagaimana memperbaikinya?

- Transformasi non linier pada penyebab 1 atau 2
- Pada penyebab 3, pencilan harus dievaluasi penyebabnya
 - Murni kesalahan: pencilan dapat dibuang *utk cross-section*
 - Apa adanya: pencilan memberikan informasi tambahan pada hasil analisis
- Perbesar ukuran sampel untuk penyebab 4

→ *obat generik*

► Transformasi: sesuaikan dengan permasalahan teori ekonomi yang ingin dianalisis

- Ukuran sampel yang diperbesar dapat memperbaikinya

Handwritten notes:

Regresi
Debet (R)

2000	900
2010	1500
19	1050
20	200
21	250

(Note: The last two rows are circled in red in the original image.)

MULTIKOLINIERITAS

- **Asumsi Regresi Klasik:** Tidak ada hubungan exact collinearity antara variabel Independen (variabel-variabel X);
→ Misalnya: antara X_2 dan X_3 tidak ada hubungan **exact linear** ($X_2 = 4X_3$) ✓
- Multikolinieras: hubungan linier antar variabel independen.
- Ada hubungan linier yang sempurna (*perfect multicollinearity*) dan hubungan linier yang tak sempurna (*imperfect multicollinearity*)

Contoh Multikolinieritas

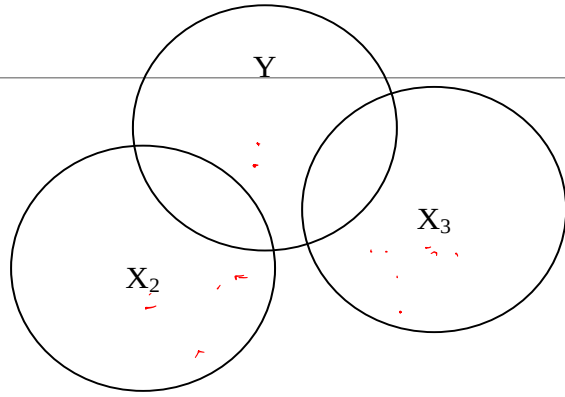
Obs.

	X_2	X_3	X^*_3
1	10	50	52
2	15	75	75
3	18	90	97
4	24	120	129
5	30	150	152

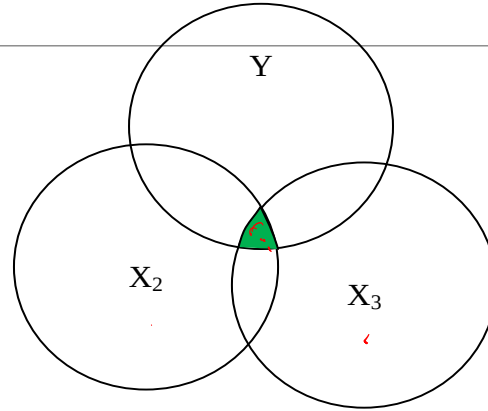
→ $X_{3i} = 5X_{2i}$ → hubungan **exact linear** (multikolinearitas sempurna).

→ antara X^*_{3i} tidak lagi memiliki hubungan **exact linear** dengan X_{2i} namun tetap memiliki korelasi yang kuat sebesar 0,9959.

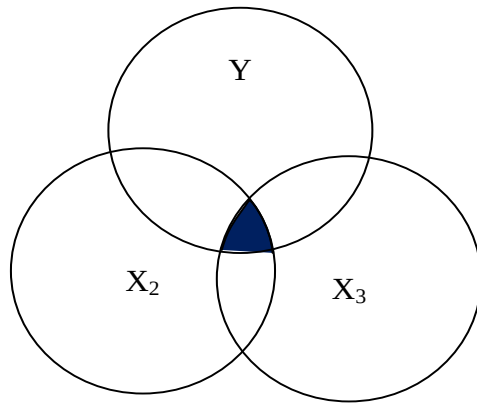
Gambaran Multikolinieritas



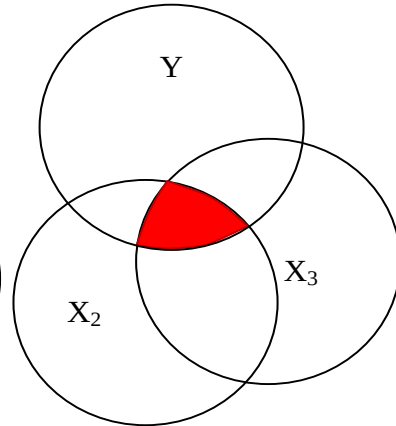
(a) Tidak ada kolinieritas



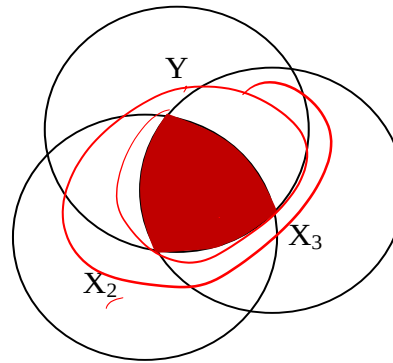
(b) Kolinieritas rendah



(c) Kolinieritas sedang



(d) Kolinieritas tinggi



(e) Kolinieritas sangat tinggi

Multikolinieritas Sempurna

- ▶ Terdapat hubungan linier di antara peubah bebas
- ▶ Multikolinieritas sempurna:
 - ▶ Satu peubah eksogen adalah fungsi linier dari peubah eksogen yang lain

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$X_3 = \delta_1 + \delta_2 X_2$$

Multikolinieritas Sempurna

- Efek dari multikolinieritas:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u \quad X_3 = \delta_1 + \delta_2 X_2$$

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 (\delta_1 + \delta_2 X_2) + u$$

$$Y = (\beta_1 + \beta_3 \delta_1) + (\beta_2 + \beta_3 \delta_2) X_2 + u$$

$$Y = v_1 + v_2 X_2 + u$$

- Sampel dipakai untuk menduga koefisien v_1 dan v_2

Multikolinieritas Sempurna

- ▶ Untuk memperoleh penduga β_1 dan β_2 : solusi dari persamaan berikut:

$$\hat{v}_1 = \hat{\beta}_1 + \hat{\beta}_3 \delta_1$$

$$\hat{v}_2 = \hat{\beta}_2 + \hat{\beta}_3 \delta_2$$

- ▶ 2 persamaan untuk 3 peubah
- ▶ Tidak ada solusi unik bagi penduga parameter populasi
- ▶ Efek dari struktur matriks akibat satu kolom yang merupakan fungsi linier dari kolom yang lain:

$X'X$

Matriks singular

Multikolinieritas Sempurna

Karena determinan matriks singular = 0

Singular matrix,

$$|\mathbf{X}'\mathbf{X}| = 0$$

- ▶ Tidak dapat diperoleh inverse dari $\mathbf{X}'\mathbf{X}$ pada:

$$\hat{\beta}_2 = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

Karena: $(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{|\mathbf{X}'\mathbf{X}|} \text{adj}(\mathbf{X}'\mathbf{X})$

*1/0 → B2
definisi*

- ▶ Koefisien regresi menjadi 'indeterminate'

Estimasi pada Multikolinearitas Sempurna

- Tidak dapat diperoleh solusi yang unik (tunggal) bagi koefisien regresi individual 
 - Varians dan standard error dari koefisien regresi secara individual bernilai tidak terhingga
- 

Multikolinieritas tak sempurna

- ▶ Terjadi jika terdapat hubungan linier yang tidak sempurna antar peubah eksogen

$$X_3 = X_2 + v$$

- ▶ Dengan v sebagai galat acak yang tidak sama dengan nol
- ▶ Kasus ini sering terjadi pada kasus terapan
- ▶ Bagaimana mengidentifikasi seberapa serius derajat multikolinieritas yang terjadi.

Estimasi pada Multikolinearitas Tidak Sempurna

- Masih dapat diperoleh solusi yang unik (tunggal) bagi koefisien regresi individual
- Varians dan standard error dari koefisien regresi secara individual bernilai besar.

Efek dari Multikolinieritas tak sempurna

- ▶ Penduga OLS tetap dapat diduga
- ▶ Penduga OLS tetap bersifat BLUE
- ▶ Penduga OLS tetap efisien (ragam dari penduga paling kecil dari semua penduga yang mungkin)
 - ▶ Akan tetapi pada nilai yang cukup besar
 - ▶ Relatif lebih besar jika tidak ada multikolinieritas

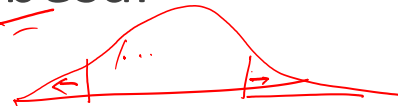
varians ↑

$$\text{var}(\hat{\beta}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

$$|\mathbf{X}'\mathbf{X}| \rightarrow 0 \quad (\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{|\mathbf{X}'\mathbf{X}|} \text{adj}(\mathbf{X}'\mathbf{X}) \rightarrow \infty$$

Efek dari Multikolinieritas tak sempurna

- ▶ *varians* dan *(covarians)* dari penduga OLS relatif besar
- ▶ Selang kepercayaan menjadi lebih besar
 - ▶ Lebih banyak menerima hipotesis nol (koefisien tidak nyata)
- ▶ Statistik uji t dari satu atau beberapa koefisien menjadi tidak nyata
 - ▶ Walaupun R^2 secara keseluruhan besar
- ▶ Tanda bagi penduga koefisien berkebalikan dengan teori a priori



Konsekuensi Teoritis dari Multikolinearitas

- Dalam kasus multikolinieritas tak sempurna (*near multicollinearity*), estimator OLS tidak bias.
- Multokolinearitas tidak merusak sifat varians minimum, namun tidak berarti bahwa varians dari estimator OLS akan kecil.
- Multikolinearitas pada dasarnya merupakan fenomena (regresi) sebuah sampel. Meskipun variabel-variabel X tidak berhubungan secara linear pada populasi, bisa jadi mereka berhubungan pada suatu sampel tertentu.

Konsekuensi Praktis Multicollinearity

Konsekuensi adanya multicollinearity yang tinggi:

- Meskipun masih BLUE, namun estimator OLS mempunyai varians dan covarians yg besar → sulit utk menentukan estimasi yg tepat.
- Konfiden interval lebih melebar. The confidence intervals tend to be much wider, leading to the acceptance of the “zero null hypothesis” (i.e., the true population coefficient is zero) more readily.
- t-ratio cenderung secara statistik tidak signifikan ($H_0: \beta_k = 0$ diterima). Namun demikian R^2 dapat sangat tinggi.
- Estimator OLS dan standard errors-nya sensitif thd perubahan kecil dlm data.

VIF (Variance Inflation Factor)

$$\text{cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\sigma^2}{(1-r_{23}^2)\sqrt{\sum x_{2i}^2 \sum x_{3i}^2}}$$

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$X_2 = \alpha_1 + \alpha_2 X_3 + v$$

- Struktur Korelasi dinamakan dengan Variance Inflation Factor (VIF)

$$VIF = \frac{1}{(1-r_{23}^2)}$$

koef. Determinasi regresi
d.t. auxiliary auxiliary

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{(\sum x_{2i}^2)} VIF \quad \text{VIF} = \frac{\text{var}(\hat{\beta}_3)}{(\sum x_{3i}^2)} VIF$$

- Semakin besar multikolinieritas maka semakin besar VIF
- Semakin besar VIF semakin besar ragam penduga OLS

VIF = 5
 $r_{23}^2 = 0.8$
 $r_{23}^2 = 0.9 \rightarrow VIF = 10$
 r_{23}

VIF (Variance Inflation Factor)

Kecepatan meningkatnya varians dan kovarians dpt dilihat dari VIF

$$VIF = \frac{1}{(1 - r_{23}^2)}$$

Auxiliary.

Kebalikan dari VIF adalah TOL (toleransi)

$$TOL_j = \frac{1}{VIF_j} (1 - R_j^2)$$

VIF lebih dari 2 Variabel Bebas

- ▶ Untuk regresi lebih dari 2 peubah definisi dari VIF:

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

R_j^2 : Koefisien determinasi dari *auxiliary regression*

- ▶ *Auxiliary regression*: regresi dengan X_j sebagai peubah endogen, dan X selainnya sebagai peubah eksogen

Nilai VIF berdasarkan Koefisien Determinasi dari Auxiliary Regression

R_j^2	VIF
0	1
0.5	2
0.8	5
0.9	10
0.95	20
0.975	40
0.99	100
0.995	200
0.999	1000

- ▶ VIF yang naik seiring dengan kenaikan koefisien determinasi
- ▶ VIF yang lebih dari 10: bukti cukup untuk multikolinieritas

Pendeteksian Multikolinieritas

- Dari koefisien korelasi sederhana *→ antar var. bebas*
 - Efektif untuk regresi dengan 2 peubah eksogen (*bebas*), *→ $r > 0,8$*
 - Dari VIF, multikolinieritas serius jika $r \geq 0.9$
- Dari koefisien determinasi *auxiliary regression*
 - Efektif untuk regresi dengan 3 peubah eksogen atau lebih
 - Peubah eksogen pada *auxiliary regression* : peubah yang mempunyai masalah multikolinieritas
 - Hasil dari *auxiliary regression*:
 - Standar error yang kecil
 - Statistik uji t yang nyata bagi masing-masing koefisien

Contoh:

- ▶ Model regresi dengan 2 peubah eksogen,
- ▶ Dua peubah eksogen tsb mempunyai korelasi tinggi:
 - ▶ Dari matrix korelasi berikut:

	X2	X3	Y
X2	1		
X3	0.999995	1	
Y	0.857369	0.857438	1

- ▶ Kedua X berkorelasi positif dengan Y
- ▶ Antar X berkorelasi positif

Contoh Tabel 10.5 Buku Basic Econometrics Gujarati & Porter

Dependent Variable: Y

Method: Least Squares

Date: 05/11/12 Time: 08:04

Sample: 1 10

Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	24.77473	6.752500	3.668972	0.0080
X2	0.941537	0.822898	1.144172	0.2902
X3	-0.042435	0.080664	-0.526062	0.6151
R-squared	0.963504	Mean dependent var	111.0000	
Adjusted R-squared	0.953077	S.D. dependent var	31.42893	
S.E. of regression	6.808041	Akaike info criterion	6.917411	
Sum squared resid	324.4459	Schwarz criterion	7.008186	
Log likelihood	-31.58705	F-statistic	92.40196	
Durbin-Watson stat	2.890614	Prob(F-statistic)	0.000009	

Dependent Variable: X3

Method: Least Squares

Date: 05/11/12 Time: 08:06

Sample: 1 10

Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.545455	29.47581	0.255988	0.8044
X2	10.19091	0.164262	62.04047	0.0000
R-squared	0.997926	Mean dependent var		1740.000
Adjusted R-squared	0.997667	S.D. dependent var		617.7312
S.E. of regression	29.83972	Akaike info criterion		9.806415
Sum squared resid	7123.273	Schwarz criterion		9.866932
Log likelihood	-47.03207	F-statistic		3849.020
Durbin-Watson stat	2.077534	Prob(F-statistic)		0.000000

Dependent Variable: Y

Method: Least Squares

Date: 05/11/12 Time: 08:07

Sample: 1 10

Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	24.45455	6.413817	3.812791	0.0051
X2	0.509091	0.035743	14.24317	0.0000
R-squared	0.962062	Mean dependent var		111.0000
Adjusted R-squared	0.957319	S.D. dependent var		31.42893
S.E. of regression	6.493003	Akaike info criterion		6.756184
Sum squared resid	337.2727	Schwarz criterion		6.816701
Log likelihood	-31.78092	F-statistic		202.8679
Durbin-Watson stat	2.680127	Prob(F-statistic)		0.000001

Dependent Variable: Y
Method: Least Squares
Date: 05/11/12 Time: 08:08
Sample: 1 10
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	24.41104	6.874097	3.551164	0.0075
X3	0.049764	0.003744	13.29166	0.0000
R-squared	0.956679	Mean dependent var		111.0000
Adjusted R-squared	0.951264	S.D. dependent var		31.42893
S.E. of regression	6.938330	Akaike info criterion		6.888856
Sum squared resid	385.1233	Schwarz criterion		6.949373
Log likelihood	-32.44428	F-statistic		176.6681
Durbin-Watson stat	2.417419	Prob(F-statistic)		0.000001

Deteksi Multikolinieritas

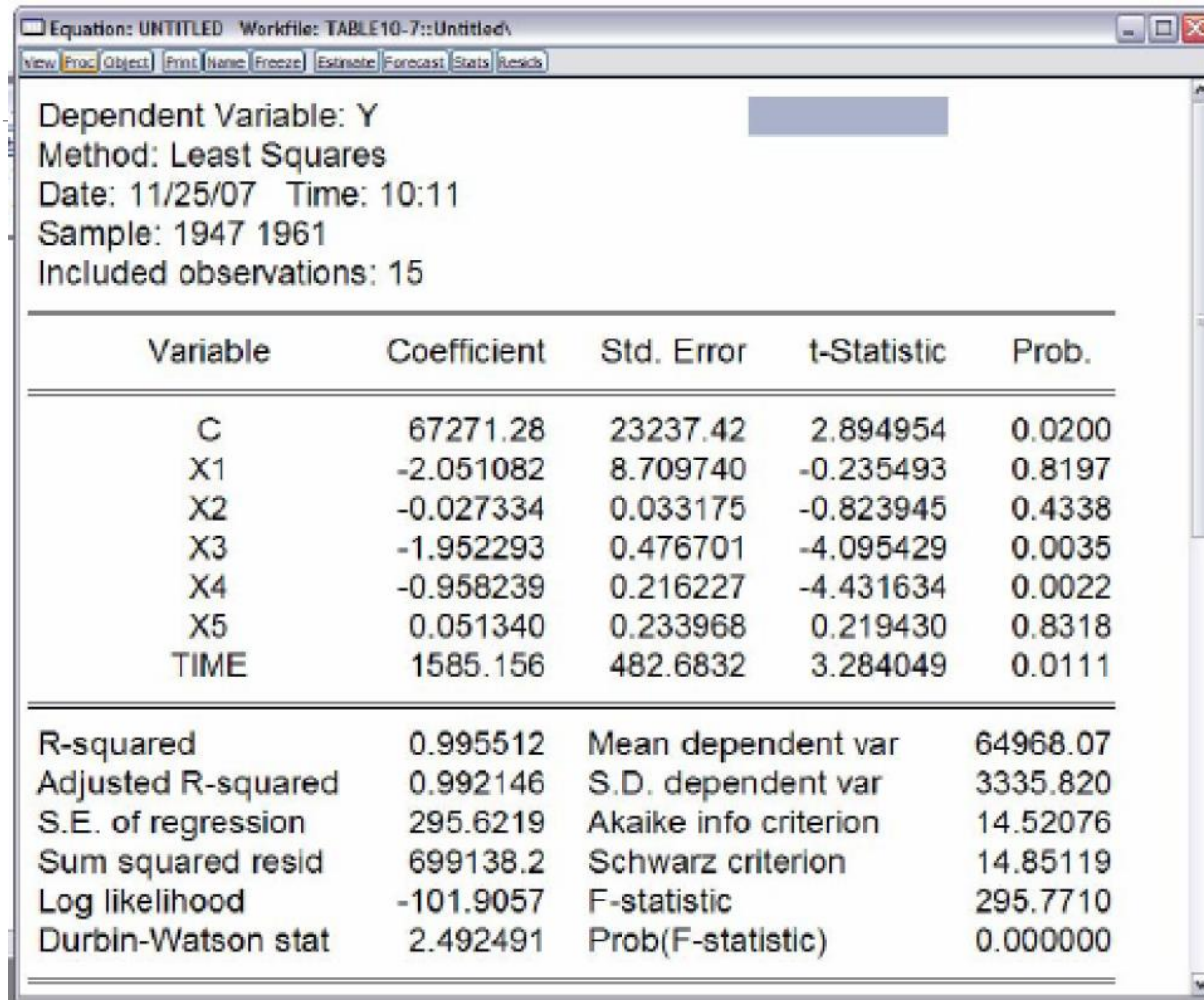
- Nilai R^2 tinggi namun hanya sedikit variabel independen yang signifikan
- Korelasi parsial antar variabel independen $\rightarrow R > 0,5$.
- Regresi auxiliary
- Metode Klein 
- Variance Inflation Factor dan Tolerance 

Contoh Deteksi Multikolinieritas

Misal kita mempunyai model egresi berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 Time + \varepsilon$$

Hasil Estimasi Model Regresi



Equation: UNTITLED Workfile: TABLE10-7::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: Y
Method: Least Squares
Date: 11/25/07 Time: 10:11
Sample: 1947 1961
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	67271.28	23237.42	2.894954	0.0200
X1	-2.051082	8.709740	-0.235493	0.8197
X2	-0.027334	0.033175	-0.823945	0.4338
X3	-1.952293	0.476701	-4.095429	0.0035
X4	-0.958239	0.216227	-4.431634	0.0022
X5	0.051340	0.233968	0.219430	0.8318
TIME	1585.156	482.6832	3.284049	0.0111

R-squared	0.995512	Mean dependent var	64968.07
Adjusted R-squared	0.992146	S.D. dependent var	3335.820
S.E. of regression	295.6219	Akaike info criterion	14.52076
Sum squared resid	699138.2	Schwarz criterion	14.85119
Log likelihood	-101.9057	F-statistic	295.7710
Durbin-Watson stat	2.492491	Prob(F-statistic)	0.000000

Hasil Estimasi Model Regresi

Dari Hasil Estimasi, terlihat bahwa:

- **R² tinggi tapi,**
- **signifikansi parameter rendah (X1, X2, X5)**

Korelasi yang tinggi antar variabel regresor

	X1	X2	X3	X4	X5	TIME
X1	1.000000	0.993669	0.591734	0.468974	0.983316	0.990843
X2	0.993669	1.000000	0.575280	0.458778	0.989698	0.994789
X3	0.591734	0.575280	1.000000	-0.203285	0.674764	0.646567
X4	0.468974	0.458778	-0.203285	1.000000	0.371243	0.422210
X5	0.983316	0.989698	0.674764	0.371243	1.000000	0.995742
TIME	0.990843	0.994789	0.646567	0.422210	0.995742	1.000000

Korelasi tinggi :

X1 vs X2 ; X2 vs X5

X1 vs X5 ; X2 vs Time

X1 vs Time

Tolerance (TOL) dan variance inflation factor (VIF)

$$TOL = 1 - R_i^2$$

$$VIF(\hat{\beta}_i) = 1 / TOL = 1 / (1 - R_i^2)$$

- Bila mana $VIF > 10$ menunjukkan kolinearitas TINGGI.
- Bila mana $VIF 5 - 10$ menunjukkan kolinearitas SEDANG.
- Bila mana $VIF 1 - <5$ menunjukkan kolinearitas RENDAH.

Q baik.

Contoh Table 10_7

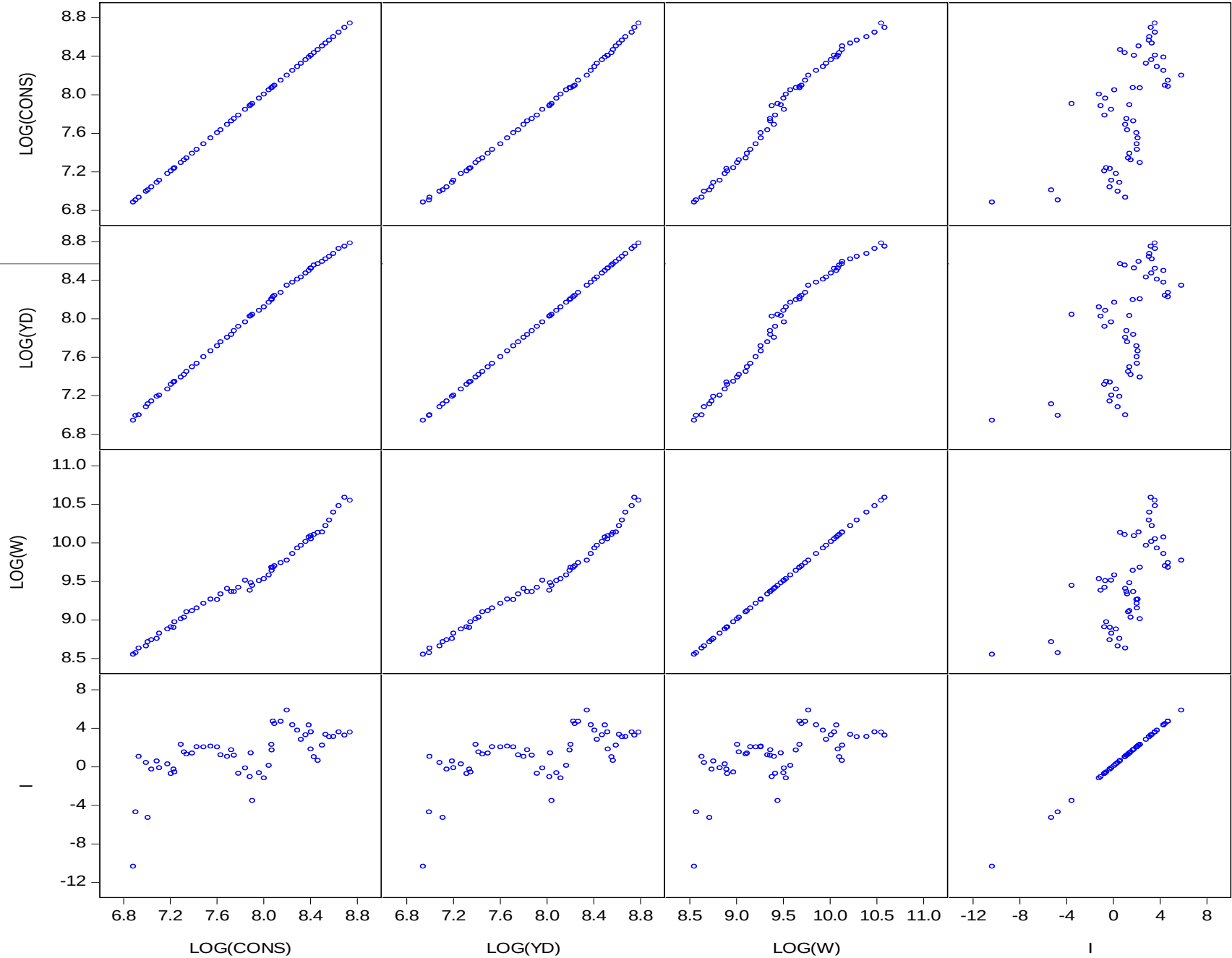
Masukkan data dalam Table 10_7

Estimasi dengan persamaan: $\log(\text{cons}) = c + \log(yd) + \log(w)$ |

Catatan log di sini adalah logaritma natural

Simpan persamaan estimasinya. Buatlah group baru dengan menekan:
Procs → make regressor group.

Scatter Plot



Bagaimana mengatasi Multikolinieritas?

➤ Do nothing

➤ Rule of Thumb Procedure

- A priori information
- Combining cross sectional and time series data
- Dropping a variable(s) and specification bias
- Transformation of variables
- Additional or new data

Do Nothing

- Multikolinieritas adalah masalah akibat ketidaksempurnaan data
- Untuk data ekonomi: tidak dapat dikontrol dan tidak ada pilihan
- Penduga secara keseluruhan tetap dapat dipakai walaupun penduga secara individu relatif kurang efisien dan tidak signifikan

A priori information

- ▶ Informasi dari penelitian sebelumnya mengenai hubungan fungsional antar parameter peubah yang berkorelasi

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

- ▶ Dengan X_2 yang berkorelasi tinggi dengan X_3
- ▶ Misal:
 - X_2 : pendapatan, X_3 : Kekayaan, Y : konsumsi
- ▶ Diketahui dari penelitian sebelumnya bahwa perubahan kekayaan terhadap perubahan konsumsi adalah 1/10 perubahan pendapatan terhadap perubahan konsumsi

$$\beta_3 = 0.1\beta_2$$

A priori information

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$\beta_3 = 0.1\beta_2$$

$$Y = \beta_1 + \beta_2 X_2 + 0.1\beta_2 X_3 + u$$

$$Y = \beta_1 + \beta_2 X + u$$

$$X = X_2 + 0.1X_3^*$$

- ▶ Lakukan transformasi terhadap kedua peubah eksogen dengan hubungan sesuai (*)
- ▶ Lakukan pendugaan menggunakan peubah yang sudah ditransformasi

Menggabungkan data cross section dan time series

▶ Misalkan:

Y : jumlah penjualan mobil

P : rata-rata harga mobil

I : pendapatan

Pada data time series, P dan I cenderung berkorelasi

$$\ln Y_t = \beta_1 + \beta_2 \ln P_t + \beta_3 \ln I_t + u_t$$

- ▶ β_2 : adalah elastisitas harga terhadap jumlah penjualan mobil
- ▶ β_3 : adalah elastisitas pendapatan terhadap jumlah penjualan mobil

Menggabungkan data cross section dan time series

- ▶ Jika terdapat data cross section (pada satu waktu) yang dapat dipakai untuk menduga koefisien elastisitas pendapatan β_3
 - ▶ Dengan asumsi bahwa pada satu waktu harga tidak terlalu bervariasi
- ▶ Gunakan penduga bagi β_3 untuk melakukan transformasi terhadap Y

$$\ln Y_t = \beta_1 + \beta_2 \ln P_t + \hat{\beta}_3 \ln I_t + u_t$$

$$\ln Y_t - \hat{\beta}_3 \ln I_t = \beta_1 + \beta_2 \ln P_t + u_t$$


$$Y_t^* = \beta_1 + \beta_2 \ln P_t + u_t$$

Dropping a variable(s) and specification bias

- ▶ Membuang salah satu dari peubah yang berkorelasi
- ▶ Masalah:
 - ▶ Jika semua peubah secara ekonomi harus ada di dalam model:
specification bias
 - ▶ Jika pendapatan dan kekayaan memang harus ada di dalam model konsumsi
- ▶ Tujuan perbaikan multikolinieritas dapat memunculkan masalah baru: specification bias
- ▶ Tetap gunakan dua-duanya

Transformation of variables

- ▶ Contoh pada data time series pada

X_2 : pendapatan, X_3 : Kekayaan, Y : konsumsi

Dengan X_2 yang berkorelasi tinggi dengan X_3 seiring dengan waktu

Pada waktu t berlaku:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

- ▶ Model yang sama dapat berlaku pada waktu $t-1$

$$Y_{t-1} = \beta_1 + \beta_2 X_{2,t-1} + \beta_3 X_{3,t-1} + u_{t-1}$$

Transformation of variables: First Difference

- ▶ Untuk meminimumkan multikolinieritas, dilakukan perbedaan dari model di waktu t dan waktu $t-1$

- ▶ First difference form

$$Y_t - Y_{t-1} = \beta_2(X_{2t} - X_{2t-1}) + \beta_3(X_{3t} - X_{3t-1}) + (u_t - u_{t-1})$$

$$\Delta Y_t = \beta_2 \Delta X_{2t} + \beta_3 \Delta X_{3t} + v_t$$

- ▶ Regresi dilakukan pada masing-masing peubah yang sudah dibedakan
 - ▶ Korelasi di antara peubah beda (ΔX_2 dan ΔX_3) tidak sebesar korelasi dari peubah aslinya

Additional or new data

- Jika multikolinieritas terjadi akibat pengambilan sampel
- Penambahan ukuran sampel dapat mengurangi efek dari multikolinieritas

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\left(\sum x_{2i}^2\right)(1 - r_{23}^2)}$$

Ragam lebih kecil/lebih efisien

Komponen ini diasumsikan tetap

Sampel bertambah akan memperbesar nilai komponen ini

Tugas

Kerjakan Latihan dari buku Gujarati & Porter Edisi Kelima Bab 10 No. 10.24 dan 10.33

Sumber Bacaan

Gujarati, Damodar N. and Dawn C. Porter (2009). *Basic Econometrics*, Fifth Edition. McGraw-Hill/Irwin. New York. Chapter 10.