

EKONOMETRIKA I

EKONOMI PEMBANGUNAN KELAS C

DOSEN PENGASUH:

IMAM AWALUDDIN, S.E., M.E.

FAKULTAS EKONOMI DAN BISNIS

UNIVERSITAS LAMPUNG

SEMESTER GENAP 2023/2024

Bahan Kuliah Ekonometrika I

Analisis Regresi Majemuk

IMAM AWALUDDIN

Review Asumsi-asumsi yang mendasari Metode OLS (CLRM)

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
4. Homoskedastisitas: ragam yang sama pada galat
5. Galat tidak saling berkorelasi (tidak ada otokorelasi)
6. Peubah bebas (eksogen) dan galat saling bebas

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

Review: Sifat-sifat Estimator Kuadrat Terkecil - Teori Gauss-Markov

Sebuah estimator, misalkan estimator OLS $\hat{\beta}$, yang dikatakan merupakan estimator terbaik, linier, dan tidak bias β_2 , jika memiliki sifat-sifat berikut:

1. Bersifat linier, di mana merupakan fungsi linier dari sebuah variabel acak seperti variabel dependen Y dalam sebuah model regresi.
2. Bersifat tidak bias, di mana rata-rata atau nilai ekspektasinya $E(\hat{\beta})$ sama dengan nilai sebenarnya β_2 .
3. Memiliki varians minimum dari semua kelompok estimator-estimator yang linier dan tidak bias; sebuah estimator tidak bias dengan varians terkecil dikenal sebagai estimator yang efisien (*efficient estimator*)

Review: Teorema Gauss-Markov

Teorema Gauss-Markov: mengacu pada asumsi-asumsi CLRM, estimator/penduga kuadrat terkecil (OLS), pada estimator-estimator yang linier dan tidak bias, memiliki varians minimum, sehingga estimatornya disebut *BLUE* (*best linear unbiased estimator*)

Review: Goodness of Fit dari garis regresi

- Sebagai alat untuk:
 - Menentukan apakah tidak ada alternatif garis lain yang dapat menjelaskan hubungan X dan Y
 - Mengukur seberapa baik model yang diperoleh menjelaskan Y
- Diperlukan penguraian nilai JK (jumlah kuadrat) Y di sekitar nilai tengahnya.

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i + u_i - \bar{y})^2$$

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$$

JK total JK Regresi JK Residual/Galat
 TSS ESS RSS

Review: Koefisien Determinasi (R^2)

Dari penguraian JK tersebut dapat diturunkan koefisien determinasi berikut:

$$R^2 = \frac{\text{JK Regresi}}{\text{JK Total}}$$

Koefisien determinasi: Sebagai ukuran seberapa besar (dalam proporsi/persen) keragaman total Y dapat dijelaskan oleh model regresi.

$$\text{JK Total} = \text{JK Regresi} + \text{JK Galat}$$

$$1 = \frac{\text{JK Regresi} + \text{JK Galat}}{\text{JK Total}} \quad \longrightarrow \quad 1 = R^2 + \frac{\text{JK Galat}}{\text{JK Total}}$$

Regresi Linier Berganda

Model Umum Regresi Linier Majemuk/Berganda:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + \varepsilon$$

- Ø Satu peubah respon (endogen)
- Ø Beberapa peubah penjelas (eksogen)

Regresi Linier Berganda

- Ø Satu peubah respon (endogen)
- Ø Beberapa peubah penjelas (eksogen)
- Ø Dinotasikan dalam matriks

$$Y = X\beta + u$$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \quad X = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1k} \\ X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nk} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

$n \times 1 \quad n \times k \quad k \times 1 \quad n \times 1$

Penduga OLS

Penduga yang meminimumkan Jumlah kuadrat galat (RSS), dalam notasi matriks:

$$\begin{aligned} RSS &= \hat{u}'\hat{u} \\ RSS &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) \\ &= (Y' - \hat{\beta}'X')(Y - X\hat{\beta}) \\ &= Y'Y - \hat{\beta}'X'Y + Y'X\hat{\beta} - \hat{\beta}'X'X\hat{\beta} \\ &= Y'Y - 2Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \end{aligned}$$

Penduga OLS

RSS akan minimum pada nilai penduga yang merupakan solusi dari turunan pertama RSS yang disamadengankan nol

$$\begin{aligned} \frac{\partial RSS}{\partial \hat{\beta}} &= -2X'Y + 2X'X\hat{\beta} = 0 \\ X'X\hat{\beta} &= X'Y \\ \hat{\beta} &= (X'X)^{-1}X'Y \end{aligned}$$

PENDUGA OLS

Model Regresi Linier 2 Variabel Bebas

Model:

$$Y_i = b_1 + b_2 X_{2i} + b_3 X_{3i} + u_i$$

b_1 adalah konstanta

b_2 dan b_3 adalah koefisien regresi parsial.

X_{2i} dan X_{3i} adalah variabel bebas

u_i adalah error term

Estimator OLS

min JKG
min RSS

$$Y_i = \hat{b}_1 + \hat{b}_2 X_{2i} + \hat{b}_3 X_{3i} + \hat{u}_i$$

$$\min \hat{u}_i^2 = \hat{a} (Y_i - \hat{b}_1 - \hat{b}_2 X_{2i} - \hat{b}_3 X_{3i})^2$$

Dengan menurunkan terhadap masing-masing parameter dan menyamadengankan nol maka diperoleh persamaan-persamaan normal:

$$\bar{Y}_i = \hat{b}_1 + \hat{b}_2 \bar{X}_{2i} + \hat{b}_3 \bar{X}_{3i} \quad \checkmark$$

$$\hat{a} Y_i X_{2i} = \hat{b}_1 \hat{a} X_{2i} + \hat{b}_2 \hat{a} X_{2i}^2 + \hat{b}_3 \hat{a} X_{2i} X_{3i} \quad \checkmark$$

$$\hat{a} Y_i X_{3i} = \hat{b}_1 \hat{a} X_{3i} + \hat{b}_2 \hat{a} X_{2i} X_{3i} + \hat{b}_3 \hat{a} X_{3i}^2 \quad \checkmark$$

Estimator OLS

Parameter-parameter regresinya diperoleh sebagai berikut:

$$\hat{b}_1 = \bar{Y} - \hat{b}_2 \bar{X}_{2i} - \hat{b}_3 \bar{X}_{3i} \quad \checkmark$$

$$\hat{b}_2 = \frac{(\hat{a} y_i x_{2i})(\hat{a} x_{3i}^2) - (\hat{a} y_i x_{3i})(\hat{a} x_{2i} x_{3i})}{(\hat{a} x_{2i}^2)(\hat{a} x_{3i}^2) - (\hat{a} x_{2i} x_{3i})^2} \quad \checkmark$$

$$\hat{b}_3 = \frac{(\hat{a} y_i x_{3i})(\hat{a} x_{2i}^2) - (\hat{a} y_i x_{2i})(\hat{a} x_{2i} x_{3i})}{(\hat{a} x_{2i}^2)(\hat{a} x_{3i}^2) - (\hat{a} x_{2i} x_{3i})^2} \quad \checkmark$$

Arti Koefisien Regresi Parsial

Model Regresi:

$$Y_i = b_0 + b_1 X_1 + b_2 X_2 + u_i$$

$$Y = (b_1) + (b_2)X_2 + (b_3)X_3 + u$$

b_1 = mengukur perubahan nilai rata – rata Y , untuk tiap perubahan dalam X_1 dengan X_2 konstan $\rightarrow b_1 = \frac{\Delta Y}{\Delta X_1}, X_2 \text{ konstan}$

b_2 = mengukur perubahan nilai rata – rata Y , untuk tiap perubahan dalam X_2 dengan X_1 konstan $\rightarrow b_2 = \frac{\Delta Y}{\Delta X_2}, X_1 \text{ konstan}$

Asumsi-asumsi pada regresi linier berganda

Sama dengan semua asumsi pada regresi linier sederhana, dengan tambahan:

- Tidak ada hubungan linier sempurna di antara dua atau lebih peubah penjelas (eksogen)

Dengan terpenuhinya asumsi maka penduga OLS akan bersifat:

- Linier: fungsi linier dari peubah respons (endogen)
- Tidak bias: nilai harapan penduga adalah nilai parameter
- Konsisten: untuk $n \rightarrow \infty$, penduga menuju nilai parameter yang sebenarnya, dan ragam penduga $\rightarrow 0$
- Ragam yang paling kecil di antara semua penduga yang mungkin
- BLUE: Best Linear Unbiased Estimators

Struktur Ragam Peragam dari Penduga

$$\text{var}(\hat{\beta}) = E(\hat{\beta} - b)(\hat{\beta} - b)' \rightarrow \text{matriks}$$

$$= s^2 (\mathbf{X}'\mathbf{X})^{-1}$$

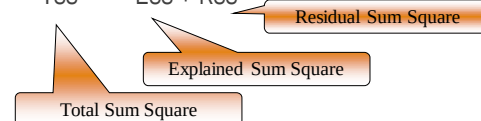
- Ø Matriks berukuran $k \times k$
- Ø Ragam (variance) pada diagonal utama
- Ø Peragam (covariance) selainnya

Koefisien determinasi majemuk (R^2)

Artinya proporsi variasi dalam Y yang dijelaskan oleh variabilitas variabel – variabel X secara bersama – sama.

$$100\% = R^2 + e$$

$$\text{TSS} = \text{ESS} + \text{RSS}$$



Koefisien determinasi majemuk (R^2)

Model Regresi:

$$Y_i = \hat{b}_1 + \hat{b}_2 X_{2i} + \hat{b}_3 X_{3i} + \hat{u}_i$$

$$= \hat{Y}_i + \hat{u}_i$$

Dengan mengkuadratkan masing-masing sisi dan menjumlahkan akan diperoleh:

$$\sum \hat{y}_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2 \sum \hat{y}_i \hat{u}_i$$

$$= \sum \hat{y}_i^2 + \sum \hat{u}_i^2$$

$$= \text{TSS (Total Sum of Squares)}$$

Koefisien determinasi majemuk (R^2)

$$\hat{y}_i^2 = \hat{b}_2 \hat{y}_{i2i} + \hat{b}_3 \hat{y}_{i3i}$$

= ESS

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\hat{b}_2 \hat{y}_{i2i} + \hat{b}_3 \hat{y}_{i3i}}{\hat{y}_i^2}$$

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\hat{u}_i^2}{\hat{y}_i^2}$$

ANOVA (Analysis of Variance)

Sumber Variasi	SS (Sum of square)	df	MSS (mean sum of square)	F
ESS	$\sum(\hat{Y}_i - \bar{Y})^2$	k - 1	ESS/k-1	
RSS	$\sum(Y_i - \hat{Y}_i)^2$	n - k	RSS/n-k	$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$
TSS	$\sum(Y_i - \bar{Y})^2$	n - 1	TSS/n-1	

Catatan: n adalah jumlah observasi, k adalah jumlah parameter estimasi termasuk konstanta

$n-k-1 \rightarrow$ k yg termasuk konstanta

F = Fisher
t = Student

Goodness of Fit dari garis Regresi Berganda

- ☐ R^2 pada regresi linier sederhana tidak dapat dipakai untuk membandingkan dua model dengan jumlah peubah eksogen yang berbeda.
- ☐ Ketika jumlah peubah X ditambah: $\text{EX} \uparrow \rightarrow R^2 \uparrow$
 - Proporsi keragaman Y yang dijelaskan oleh X akan selalu meningkat.
 - R^2 akan selalu meningkat seiring jumlah X, tanpa melihat penting tidaknya penambahan X dalam model.
- ☐ Digunakan $\text{adjusted } R^2$, $= \bar{R}^2$
 - Adjusted: disesuaikan terhadap jumlah peubah eksogen X yang digunakan

Adjusted $R^2 = \bar{R}^2$

$$R^2 = \frac{\text{JK Regresi}}{\text{JK Total}} = 1 - \frac{\text{JK Galat}}{\text{JK Total}}$$

Dengan penyesuaian terhadap jumlah peubah eksogen

$$\text{adjusted } R^2 = 1 - \frac{\text{JK Galat}/(n-k)}{\text{JK Total}/(n-1)} = 1 - \frac{\text{JK Galat}'(n-1)}{\text{JK Total}'(n-k)}$$

- } $\text{Adjusted } R^2$ dapat digunakan untuk memilih model mana yang terbaik berdasarkan jumlah peubah eksogen yang dipakai.
- } Terbaik: $\text{Adjusted } R^2 \rightarrow 1$

$$R^2 < \bar{R}^2$$

Kriteria lain untuk Pemilihan Model

Beberapa kriteria digunakan, AIC, FPE, SBC, HQC

Semua memberikan penalti terhadap JK Galat:

- Semakin banyak peubah eksogen semakin besar penaltinya

Model terbaik (berdasarkan jumlah peubah eksogen) dipilih dari nilai terkecil kriteria-kriteria tersebut.

Harapan: model terbaik mempunyai nilai terkecil untuk semua kriteria

- Tidak selalu terjadi akibat bobot yang berbeda

AIC: lebih banyak digunakan pada data deret waktu

Kriteria lain untuk Pemilihan Model

Akaike Information Criterion

$$AIC = \frac{\sum \text{JK Galat}}{n} e^{\frac{2k}{n}}$$

yg kecil

Finite Prediction Error

$$FPE = \frac{\sum \text{JK Galat}}{n} \frac{n+k}{n-k}$$

R² yg besar

Schwarz Bayesian Criterion

$$SBC = \frac{\sum \text{JK Galat}}{n} e^{\frac{k}{n}}$$

Hannah and Quinn Criterion

$$HQC = \frac{\sum \text{JK Galat}}{n} (\ln n)^{\frac{2k}{n}}$$

Beberapa Uji Hipotesis Pada Regresi Berganda

- Ø Uji keberartian koefisien secara individu → *uji hip. parsial/individual*
 - Uji t (sama dengan uji t pada kasus regresi linier sederhana)
- Ø Uji keberartian koefisien secara bersama-sama = *semipak = simultan*
 - Uji F → *join*
- Ø Uji linear restriction:
 - Uji hubungan linier antara dua atau lebih koefisien: uji F atau uji Wald (pengembangan uji t)
- Ø Uji untuk penambahan atau pengurangan peubah eksogen
 - Uji F atau Uji chi square dengan Likelihood Ratio
- Ø Semua uji merupakan perbandingan dari *unrestricted model* (menggunakan semua peubah eksogen) dan *restricted model*
 - Jika perbedaan tidak nyata maka restriction tidak berarti secara statistik.
 - Model unrestricted lebih baik digunakan.

Uji t

Model Regresi:

$$Y_i = b_0 + b_1 X_1 + b_2 X_2 + u_i$$

- Ø Uji keberartian koefisien secara individu
 - Uji t (sama dengan uji t pada kasus regresi linier sederhana)

Menambah data ln dari Excel

Workfile: LATIHAN REGRESI TABLE 7_3 ED 5 - (d:\unila\bahan pelati...)

Range: 1 51
Sample: 1 51

Group: GROUP01

Ln Y

Ln L

Ln K

33

Hasil Estimasi Model Double-log

Equation: EQ_LOG_LOG

Workfile: LATIHAN REGRESI TABLE ...

Dependent Variable: LOG(Y)
Method: Least Squares
Date: 01/23/17 Time: 10:25
Sample: 1 51
Included observations: 51

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(L)	0.468332	0.098926	4.734170	0.0000
LOG(K)	0.521279	0.096887	5.380274	0.0000
Constant	3.887600	0.396228	9.811514	0.0000

R-squared: 0.964175
Adjusted R-squared: 0.962683
S.E. of regression: 0.266752
Sum squared resid: 3.415520
Log likelihood: -3.426721
F-statistic: 645.9311
Prob(F-statistic): 0.000000

di EViews
log → ln

34

Hasil Estimasi dengan Excel

$t(\beta_i) = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)}$

$\hat{y} = \beta_0 + \beta_1 L + \beta_2 K + \epsilon$

k = jumlah parameter = 3

ESS

RSS

F = $\frac{ESS/(n-k)}{RSS/(n-k)}$

$\alpha = 0.05$

Regression Statistics		ANOVA				
Multiple R	0.981924345		df	SS	MS	F
R Square	0.964175419	Regression	2	91.92460691	45.9623	645.9311
Adjusted R Square	0.962682728	Residual	48	3.415520139	0.071157	1.99686E-35
Standard Error	0.266752075	Total	50	95.34012705		
Observations	51					

	Coefficients	Standard Error	t Stat	P-value
Intercept	3.887599524	0.396228315	9.811514	4.70478E-13
Ln L	0.468332205	0.098925936	4.73417	1.98088E-05
Ln K	0.521279114	0.09688709	5.380274	2.18316E-06

Interpretasi Hasil Estimasi

Nilai konstanta 3,887600. nilai konstanta ini dalam model double log tidak dapat secara langsung diartikan, tapi harus dikembalikan dalam bentuk anti ln karena $\theta_0 = \ln A$. Nilai A dalam model Cobb-Douglas menunjukkan technological progress. Dengan mengganti-ln-kan nilainya adalah sebesar 40,794.

Nilai koefisien lnL sebesar 0,468332 berarti bahwa setiap kenaikan 1 persen jam kerja akan meningkatkan output sebesar 0,468332 ribu dolar ceteris paribus. Angka ini juga menunjukkan besaran elastisitas parsial input L terhadap output.

36

Interpretasi Hasil Estimasi

Nilai koefisien $\ln K$ sebesar 0,521279 berarti bahwa setiap kenaikan 1 persen pengeluaran modal akan meningkatkan output sebesar 0,521279 ribu dolar ceteris paribus. Angka ini juga menunjukkan besaran elastisitas parsial input K terhadap output.

Jika kita jumlahkan koefisien $\ln L$ dengan koefisien $\ln K$ sebesar 0,989611 yang mendekati 1, menunjukkan bahwa skala hasil di industri manufaktur AS adalah mempunyai *constant return to scale*.

37

Contoh Uji t

mis. $\ln Y = \beta_0 + \beta_1 \ln L + \beta_2 \ln K + \varepsilon$

$$\beta_1 = \frac{\Delta \ln Y}{\Delta \ln L} = \frac{2 \Delta Y}{2 \Delta L} \quad \text{L & K konstan}$$

$$= 2 \cdot 0,468 \rightarrow 0,936 \rightarrow \text{L 1%} \rightarrow \text{Y 0,936\%}$$

β_2 ...

Contoh: Uji F

$$\ln Y = \beta_0 + \beta_1 \ln L + \beta_2 \ln K + \varepsilon$$

$H_0: \beta_1 = \beta_2 = 0 \rightarrow$ Sem var labor & modal sama dgn var teoretik. $F_{\text{hit}} = 695,931$

$H_a: \beta_1 \neq 0$ atau $\beta_2 \neq 0 \rightarrow$ pthg AL beda dgn koefisien $\beta_0 \neq 0$
 tdk hrs cul. 387 $\alpha = 0,05$ $\rightarrow$ kriteri besar sem parameter mng ex. teoretik.
 k-jud parameter $\rightarrow H_1 = 2$

df numerator $n-k-1 = 19$
 $F_{\text{hit}} = F_{\text{INV. RT}} = 3,191$
 Daerah tolak H_0 $\alpha = 0,05$ $\rightarrow$ tolak H_0
 Daerah terima H_0 $F_{\text{hit}} = 695,931$

Penggunaan uji F untuk Uji keberartian koefisien peubah X secara bersama-sama

Uji *goodness fit* secara keseluruhan

Pada dua model

- Unrestricted: menggunakan semua peubah eksogen
- Restricted: hanya menggunakan konstanta (*super restricted model*)

$$U: Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i$$

$$R: Y_i = \beta_1 + u_i$$

$$H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

H_1 : paling sedikit salah satu $\beta_i \neq 0, i = 2, 3, 4, 5$

Uji Chi-Square dengan Likelihood Ratio

Perbandingan likelihood dua model, restricted dan unrestricted

- *Unrestricted model*: menggunakan semua peubah eksogen (sejumlah k)
- *Restricted model*: terdapat beberapa peubah yang tidak digunakan atau ditambahkan (sejumlah m)

Hipotesis nol: beberapa parameter bernilai nol

Menggunakan statistik uji chi-square:

$$LR = -2(l_R - l_u) \sim \chi_m^2$$

Tugas

Kerjakan lagi contoh 4.2 buku Agus W.

Jawaban per individu ditulis tangan di buku Latihan.

Sumber Bacaan

Gujarati, Damodar N. and Dawn C. Porter (2009). *Basic Econometrics*, Fifth Edition. McGraw-Hill/Irwin. New York. Chapter 7.

Widarjono, Agus (2018). *Ekonometrika: Pengantar dan Aplikasinya*, Edisi Kelima. Penerbit UPP STIM YKPN. Yogyakarta. Bab 4.