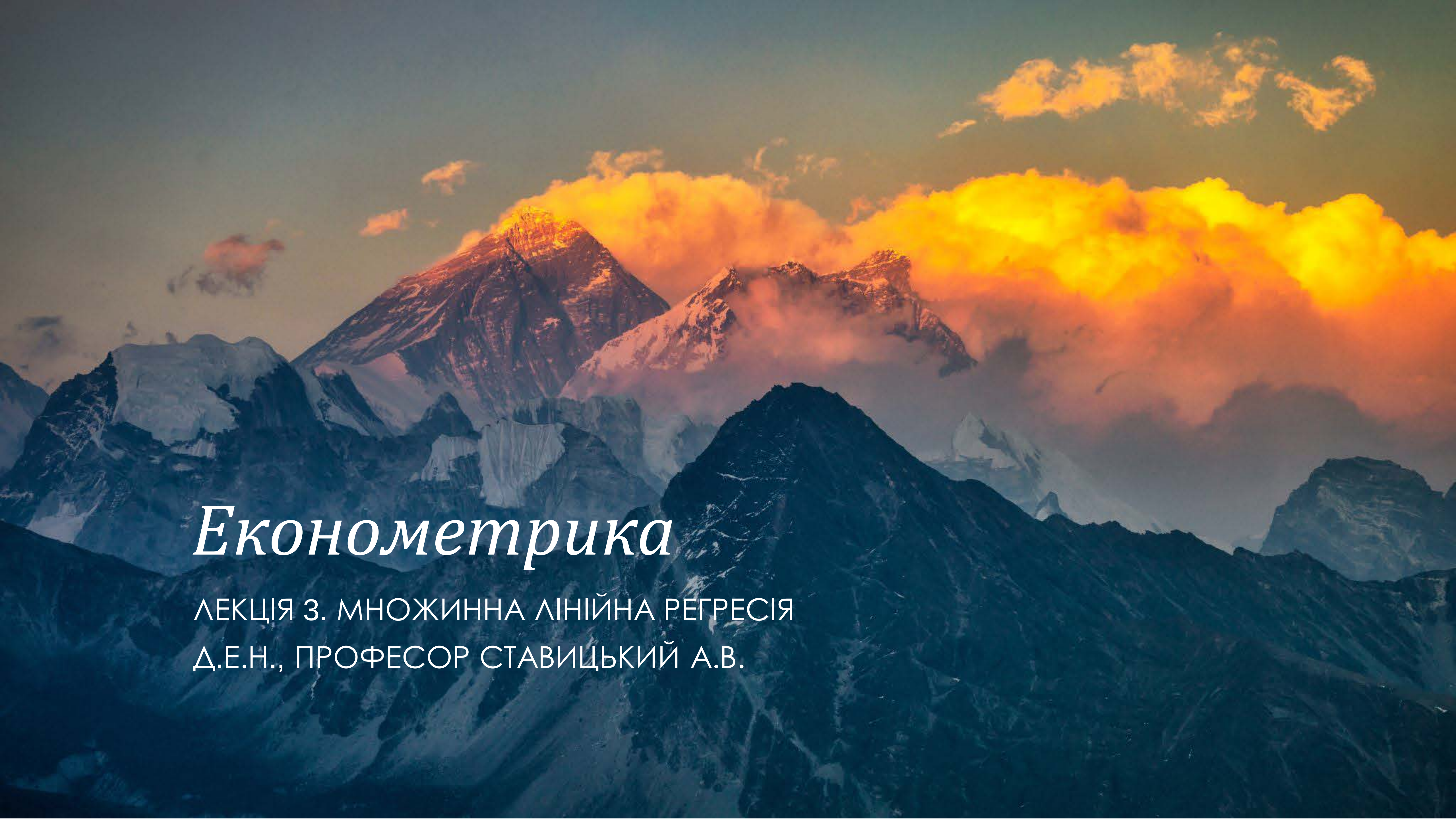


Економетрика

ЛЕКЦІЯ 3. МНОЖИННА ЛІНІЙНА РЕГРЕСІЯ

Д.Е.Н., ПРОФЕСОР СТАВИЦЬКИЙ А.В.

Опис моделі

- За допомогою моделі простої лінійної регресії визначається зв'язок між залежною змінною у та незалежною змінною х.
- Модель множинної лінійної регресії описує співвідношення між у та набором незалежних змінних x_1, x_2, x_3… .

Врахування впливу декількох факторів

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_{k-1} x_{k-1t} + \varepsilon_t, t = \overline{1,n}$$

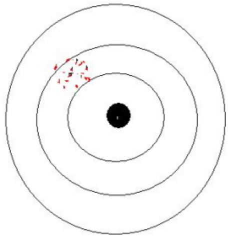
y_t - залежна змінна;

$x_{1t}, x_{2t}, \ldots, x_{k-1t}$ - незалежні змінні;

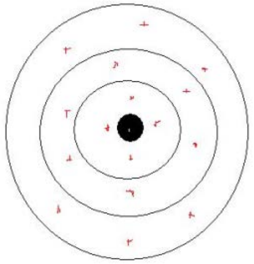
ε_t - помилки моделі.

Бажані характеристики

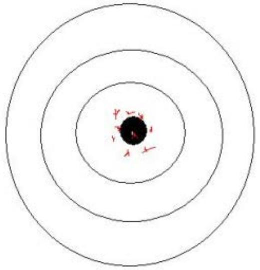
- **Відсутність зсуву**



- **Ефективність**
  - Стандартні помилки будуть мінімальними



- **Консистентність**
  - При збільшенні числа спостережень стандартна помилка зменшується.



Основна задача

- побудувати модель з великим числом факторів, визначивши при цьому вплив кожного з них окремо, а також їх сукупний вплив на заданий показник.

Припущення щодо збурень

- Нульове середнє: $M\varepsilon_t = 0, t = \overline{1,n}$
- Рівність дисперсій (гомоскедастичність): $D\varepsilon_t = M\varepsilon_t^2 = \sigma^2 = const, t = \overline{1,n}$
- Незалежність збурень: $\text{cov}(\varepsilon_t, \varepsilon_\tau) = M\varepsilon_t\varepsilon_\tau = 0, t \neq \tau$
- Незалежність збурень та регресорів: $\text{cov}(\varepsilon_t, x_{jt}) = 0, \forall t, j$
- Нормальність збурень: $\varepsilon_t \sim N(0, \sigma^2)\ \forall t$

Матричний вигляд

$$y = X\beta + \varepsilon$$

- вектор значень залежної змінної, $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}$
- матриця значень незалежних змінних, $\mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & . & . & . & x_{k-1,1} \\ 1 & x_{1,2} & . & . & . & x_{k-1,2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & x_{1,n} & . & . & . & x_{k-1,n} \end{pmatrix}$
- вектор збурень, $\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{pmatrix}$
- вектор параметрів (коефіцієнтів) регресії. $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_{k-1} \end{pmatrix}$

Коваріаційна матриця вектора збурень

$$\mathbf{D\varepsilon}=\begin{pmatrix} D\varepsilon_1 & \mathrm{Cov}(\varepsilon_1,\varepsilon_2) & \mathrm{Cov}(\varepsilon_1,\varepsilon_3) & \cdot & \cdot & \mathrm{Cov}(\varepsilon_1,\varepsilon_n) \\ \mathrm{Cov}(\varepsilon_2,\varepsilon_1) & D\varepsilon_2 & \mathrm{Cov}(\varepsilon_2,\varepsilon_3) & \cdot & \cdot & \mathrm{Cov}(\varepsilon_2,\varepsilon_n) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathrm{Cov}(\varepsilon_{\mathrm{n}},\varepsilon_1) & \mathrm{Cov}(\varepsilon_{\mathrm{n}},\varepsilon_2) & \cdot & \cdot & \cdot & D\varepsilon_n \end{pmatrix}=$$

$$=\begin{pmatrix} \sigma^2 & 0 & 0 & \cdot & \cdot & 0 \\ 0 & \sigma^2 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0\cdot & \cdot & \cdot & \sigma^2 \end{pmatrix}=\sigma^2\mathrm{E}_n$$

- Припущення 2 та 3 щодо збурень регресії зручно записувати у вигляді:

$D\varepsilon=\sigma^2 E_n$ $\qquad$ E_n - одинична матриця

Модель множинної лінійної регресії у матрично-векторних позначеннях

$$y = X\beta + \varepsilon,$$

$$M\varepsilon = 0,$$

$$D\varepsilon = \sigma^2 E,$$

ε не залежить від X

$$\varepsilon \sim N(0, \sigma^2 E)$$

Метод найменших квадратів

$$\varepsilon_t = y_t - \hat{y}_t$$

$$Q = Q(\hat{\beta}) = \sum_{t=1}^{n} \varepsilon_t^2 = \vec{\varepsilon}^{\,\mathrm{T}} \vec{\varepsilon}$$

$$\begin{cases} n\beta_0 + \beta_1 \sum x_{1,t} + \beta_2 \sum x_{2,t} + \ldots + \beta_{k-1} \sum x_{k-1,t} = \sum y_t, \\ \beta_0 \sum x_{1,t} + \beta_1 \sum x_{1,t}^2 + \beta_2 \sum x_{1,t} x_{2,t} + \ldots + \beta_{k-1} \sum x_{1,t} x_{k-1,t} = \sum x_{1,t} y_t, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \beta_0 \sum x_{k-1,t} + \beta_1 \sum x_1 x_{k-1,t} + \beta_2 \sum x_2 x_{k-1,t} + \ldots + \beta_{k-1} \sum x_{k-1,t}^2 = \sum x_{k-1,t} y_t. \end{cases}$$

$$\frac{\partial Q}{\partial \hat{\beta}} = \frac{\partial \left(y - X\hat{\beta} \right)^T \left(y - X\hat{\beta} \right)}{\partial \hat{\beta}} = -2X^T y + 2X^T X \hat{\beta} = 0$$

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_{k-1} x_{k-1} \qquad \hat{y} = X\hat{\beta}$$

Властивості залишків методу найменших квадратів

$$\sum_{t=1}^{n} \hat{\varepsilon}_t = 0$$

$$X^T \hat{\varepsilon} = 0$$

залишки методу найменших квадратів ортогональні до регресорів.

МНК-оцінка є незміщеною оцінкою для β

$$M\hat{\beta} = M\left(\left(X^T X\right)^{-1} X^T y\right) = \left(X^T X\right)^{-1} X^T M\left(y\right) = \left(X^T X\right)^{-1} X^T M\left(X\beta + \varepsilon\right) =$$

$$= \left(X^T X\right)^{-1} X^T X\beta + \left(X^T X\right)^{-1} X^T M\varepsilon = \beta$$

Коваріаційна матриця МНК-оцінки

$$V\left(\hat{\beta}\right) = V\left(\left(X^T X\right)^{-1} X^T y\right) = \left(X^T X\right)^{-1} X^T V\left(y\right)\left(\left(X^T X\right)^{-1} X^T\right)^T =$$

$$= \left(X^T X\right)^{-1} X^T \sigma^2 E\left(\left(X^T X\right)^{-1} X^T\right)^T = \sigma^2 \left(X^T X\right)^{-1} X^T X \left(X^T X\right)^{-1} =$$

$$= \sigma^2 \left(X^T X\right)^{-1}$$

$$\hat{\sigma}^2 = \frac{RSS}{n-k}$$

Теорема Гауса-Маркова

1. Нехай припущення про нормальність збурень не накладається. Тоді МНК–оцінки мають мінімальну коваріаційну матрицю в класі незміщених лінійних оцінок.
2. Припустимо, що збурення мають нормальний розподіл. МНК–оцінки мають мінімальну коваріаційну матрицю в класі усіх незміщених оцінок.

Розклад дисперсії залежної змінної

$$TSS = ESS + RSS$$

$$TSS = \sum_{t=1}^{n}(y_t - \overline{y})^2 = \sum_{t=1}^{n}(y_t - \hat{y}_t + \hat{y}_t - \overline{y})^2 = \sum_{t=1}^{n}(\hat{\varepsilon}_t + \hat{y}_t - \overline{y})^2 =$$

$$= \sum_{t=1}^{n}\hat{\varepsilon}_t^2 + 2\sum_{t=1}^{n}\hat{\varepsilon}_t(\hat{y}_t - \overline{y}) + \sum_{t=1}^{n}(\hat{y}_t - \overline{y})^2 = \sum_{t=1}^{n}\hat{\varepsilon}_t^2 + \sum_{t=1}^{n}(\hat{y}_t - \overline{\hat{y}})^2.$$

$$ESS = \sum_{t=1}^{n}(\hat{y}_t - \overline{\hat{y}})^2$$

$$RSS = \sum_{t=1}^{n}\hat{\varepsilon}_t^2$$

Коефіцієнт детермінації

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

- Не можна порівнювати моделі з різною кількістю факторів на основі R^2!!!
- Ні в якому разі не можна співвідносити моделі з константою і без константи на підставі порівняння коефіцієнтів детермінації. Якщо немає економічних підстав для вибору регресійної функції у вигляді без константи, то бажано розглядати модель з константою.

Скоригований коефіцієнт детермінації

$$R_{adj}^2 = 1 - \frac{\frac{RSS}{n-k}}{\frac{TSS}{n-1}}$$

$$R_{adj}^2 = 1 - \left(1 - R^2\right)\frac{n-1}{n-k}$$

$R^2 \geq R_{adj}^2, k \geq 1$

- скоригований коефіцієнт детермінації не може перевищувати 1.
- скоригований коефіцієнт детермінації може бути від'ємним.

Оцінка моделі Кобба-Дугласа – I

$$Y = \beta_0 K^{\beta_1} L^{\beta_2} + \varepsilon_t$$

де

- Y - індекс реального обсягу виробництва
- K - індекс реальних капітальних витрат
- L - індекс реальних витрат праці

Оцінка моделі Кобба-Дугласа – 2

$$Y = \beta_0 K^{\beta_1} L^{\beta_2} + \varepsilon_t$$

- Прологарифмувавши рівняння, маємо:

$$\ln Y_t = \ln \beta_0 + \beta_1 \ln K_t + \beta_2 \ln L_t + \varepsilon_t$$

- Введемо нові позначення:

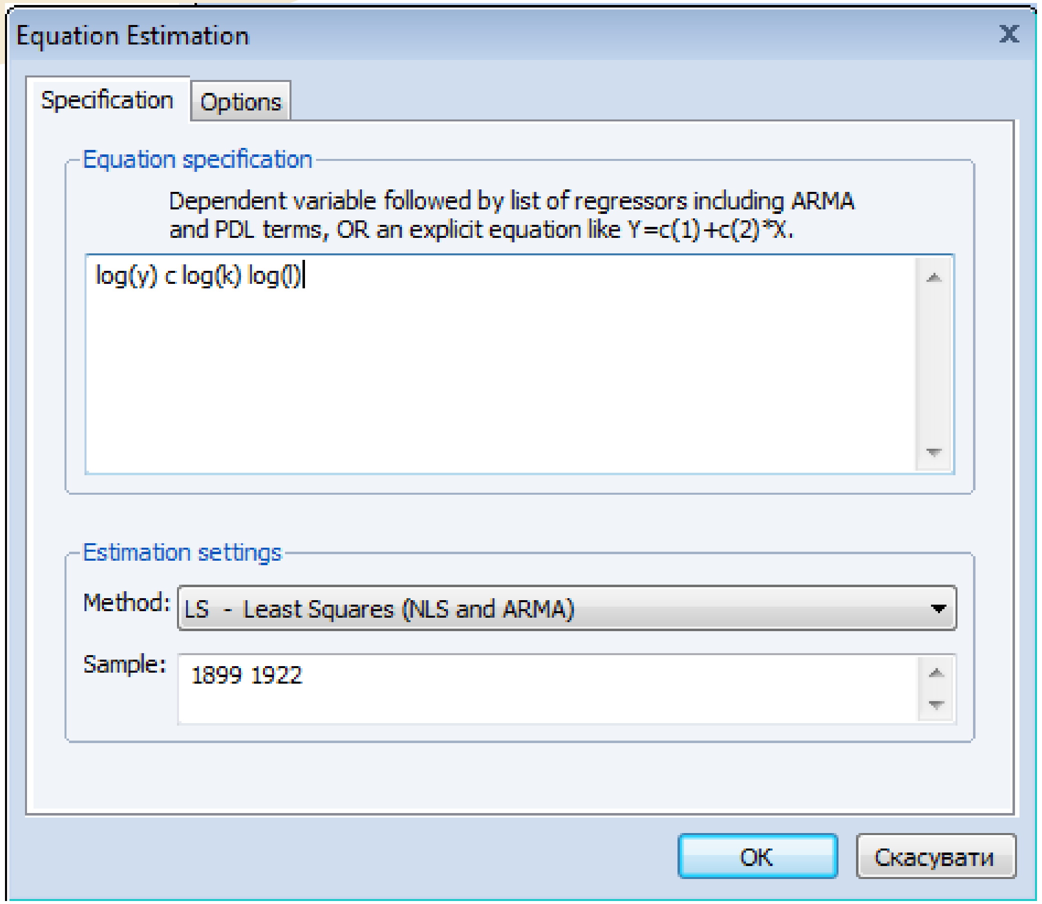
$$y_t^* = \ln Y_t \quad \beta_0^* = \ln \beta_0 \quad \beta_1^* = \beta_1 \quad \beta_2^* = \beta_2 \quad x_{1t} = \ln K_t \quad x_{2t} = \ln L_t$$

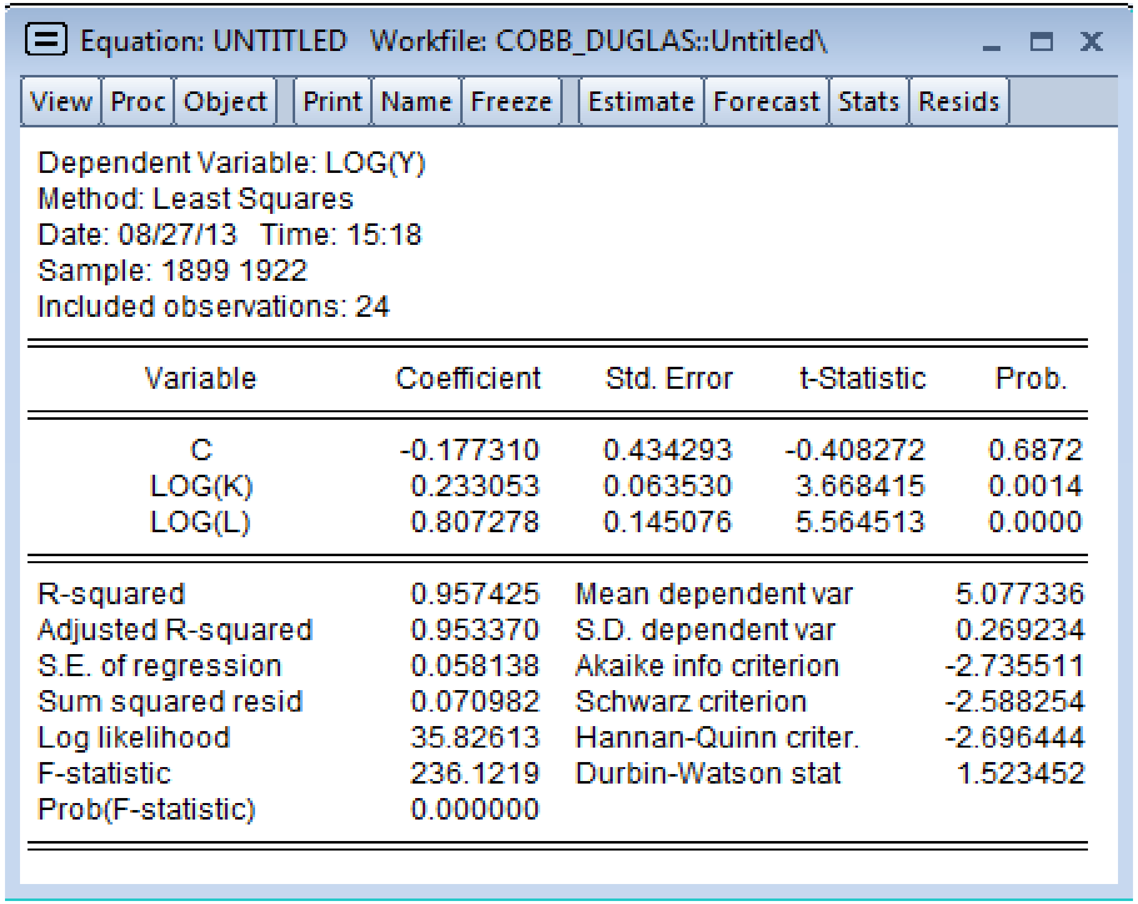
- Тоді модель можна записати у вигляді:

$$y_t^* = \beta_0^* + \beta_1^* x_{1t} + \beta_2^* x_{2t} + \varepsilon_t$$

- що відповідає моделі множинної лінійної регресії.

Оцінка моделі Кобба-Дугласа – 3





Dependent Variable: LOG(Y)
Method: Least Squares
Date: 08/27/13 Time: 15:18
Sample: 1899 1922
Included observations: 24

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.177310	0.434293	-0.408272	0.6872
LOG(K)	0.233053	0.063530	3.668415	0.0014
LOG(L)	0.807278	0.145076	5.564513	0.0000

R-squared	0.957425	Mean dependent var	5.077336
Adjusted R-squared	0.953370	S.D. dependent var	0.269234
S.E. of regression	0.058138	Akaike info criterion	-2.735511
Sum squared resid	0.070982	Schwarz criterion	-2.588254
Log likelihood	35.82613	Hannan-Quinn criter.	-2.696444
F-statistic	236.1219	Durbin-Watson stat	1.523452
Prob(F-statistic)	0.000000		

$$\beta_0 = e^{-0.177310}$$

$$\beta_1 = 0.233053$$

$$\beta_2 = 0.807278$$

Кейс: ціна морозива

- Каліфорнія
- Преміум-сегмент

- Визначити оптимальну ціну

	A	B	C	D	E
1	обсяг продажів	ціна за одиницю	сезон (1 -зима, 0 -літо)	дохід(сер\міс тис.дол.]	ціна товара конкурента
2	1400	40	1	12,50	5
3	3220	20	0	15,00	6
4	2240	20	1	25,00	3
5	3640	17	0	25,00	7
6	3010	17	0	10,00	8
7	2492	23	0	20,00	7
8	3920	10	1	20,00	6
9	3080	14	1	25,00	9
10	2310	35	1	15,00	5
11	1680	35	0	12,50	6
12	1820	32	0	15,00	7
13	3360	13	1	25,00	4
14	2212	30	0	25,00	5
15	3766	10	1	10,00	5
16	3080	13	1	20,00	6
17	2492	38	1	20,00	3
18	2352	37	0	25,00	8
19	2926	16	1	15,00	7
20	2772	19	1	12,50	5
21	2198	40	0	15,00	6
22	2842	29	1	25,00	4
23	3080	22	1	25,00	6
24	3444	16	1	10,00	7
25	2492	38	0	20,00	5
26	1960	42	0	20,00	8
27	2352	34	0	25,00	4
28	1806	32	0	15,00	5
29	2884	25	1	17,50	6
30	3430	22	1	12,50	8
31	3584	20	1	30,00	6

Підбір моделі

$$Q_t = \beta_0 + \beta_1 P_t + \varepsilon_t$$

$$\hat{Q}_t = 4103.2 - 54.35 P_t$$

Equation: UNTITLED   Workfile: PRICE_USA::Examples\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: SALES
Method: Least Squares
Date: 09/07/20   Time: 18:17
Sample: 1 30
Included observations: 30

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3347.150	585.5846	5.715911	0.0000
PR_COMP	45.56009	52.86038	0.861895	0.3969
INCOME	15.51769	12.82783	1.209689	0.2377
PRICE	-49.81841	8.623991	-5.776724	0.0000
SEASON	144.0503	174.4231	0.825867	0.4167

R-squared	0.701538	Mean dependent var	2728.133
Adjusted R-squared	0.653784	S.D. dependent var	663.7295
S.E. of regression	390.5390	Akaike info criterion	14.92394
Sum squared resid	3813018.	Schwarz criterion	15.15748
Log likelihood	-218.8592	Hannan-Quinn criter.	14.99865
F-statistic	14.69071	Durbin-Watson stat	2.411431
Prob(F-statistic)	0.000003		

Equation: UNTITLED   Workfile: PRICE_USA::Examples\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: SALES
Method: Least Squares
Date: 09/07/20   Time: 18:18
Sample: 1 30
Included observations: 30

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4103.161	193.4505	21.21039	0.0000
PRICE	-54.34894	7.123748	-7.629262	0.0000

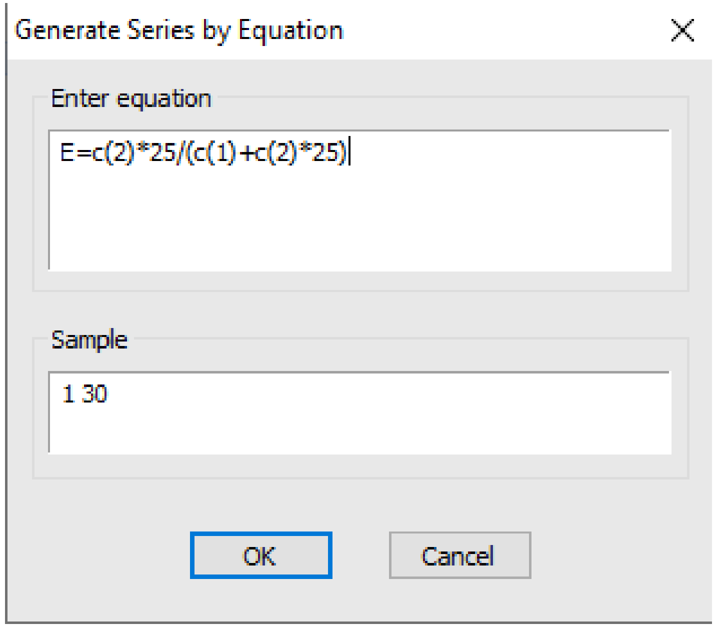
R-squared	0.675195	Mean dependent var	2728.133
Adjusted R-squared	0.663595	S.D. dependent var	663.7295
S.E. of regression	384.9659	Akaike info criterion	14.80853
Sum squared resid	4149565.	Schwarz criterion	14.90194
Log likelihood	-220.1279	Hannan-Quinn criter.	14.83841
F-statistic	58.20563	Durbin-Watson stat	1.982795
Prob(F-statistic)	0.000000		

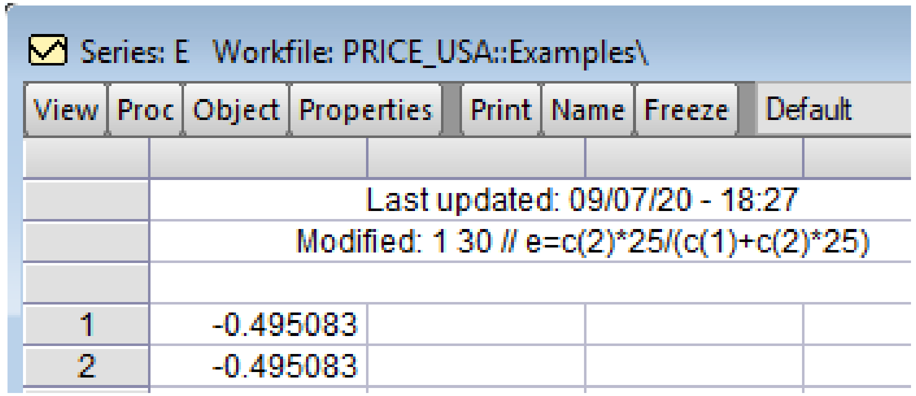
Еластичність

$$E_p = \frac{\partial Q}{\partial P} \cdot \frac{P}{Q}$$

$$E_p = \hat{\beta}_1 \cdot \frac{P}{\left(\hat{\beta}_0 + \hat{\beta}_1 P\right)}$$

- При ціні P=25:



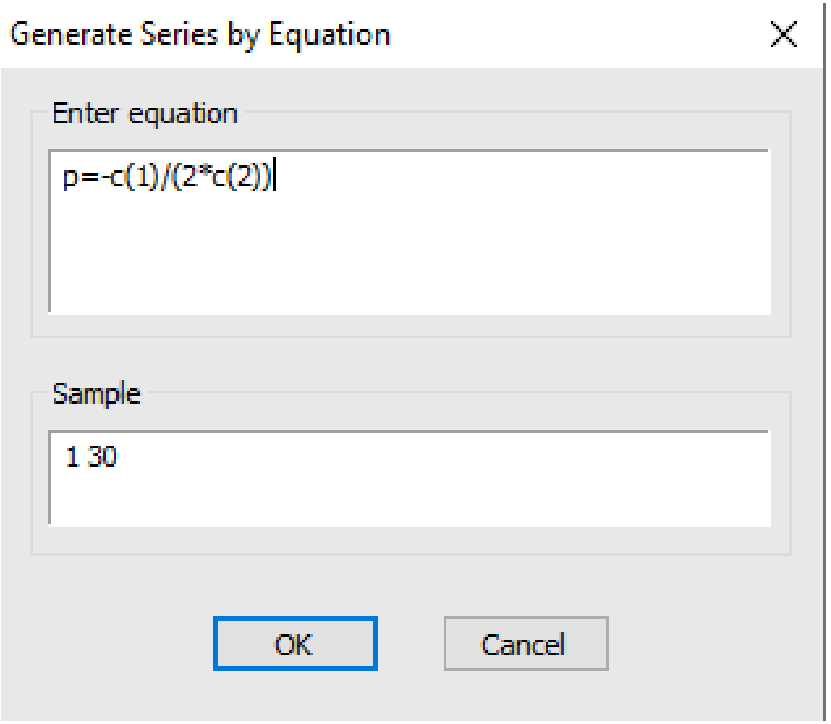


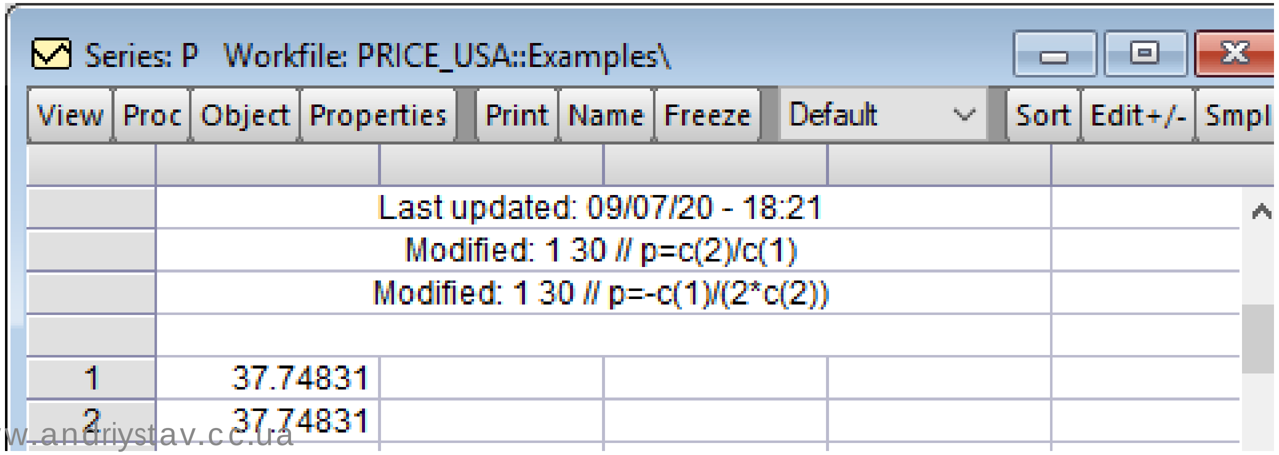
Визначення ціни

$$\Pi = Q \cdot P = (\hat{\beta}_0 + \hat{\beta}_1 P)P \to \max$$

$$\frac{\partial \Pi}{\partial P} = \hat{\beta}_0 + 2\hat{\beta}_1 P = 0$$

$$P = -\frac{\hat{\beta}_0}{2\hat{\beta}_1}$$





Питання?