Ekonometria – wykład do kursu DOiT
na kierunku LIT

W4/5

Temat: Szacowanie parametrów MEL Metodą Najmniejszych KWADRATÓW LAGRANGE'A

Przedstawione na W3 i W4 metody pozwalają ustalić zbiór $\{Y, X_1, \dots, X_m\}$, gdzie (MEL) ma postać

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_m X_m + \varepsilon,$$

oraz dane empiryczne z obserwacji tego zbioru SE następujące

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & & x_{nm} \end{bmatrix}$$

$$\begin{array}{ccc} X_1 & X_2 & X_m \end{array}$$

Gdyby jeszcze na przestrzeni przeszły szacowania parametrów strukturalnych: $\alpha_1, \alpha_2, \ldots, \alpha_m$ oraz parametrów losowych ε

Powszechnie stosowana jest MNK LAGRANGE'A

Cel: wyznaczenie takich wartości ocen $a_0, a_1, \ldots, a_m$ parametrów strukturalnych $\alpha_0, \alpha_1, \ldots \alpha_m$, aby suma kwadratów odchyleń zaobserwowanych wartości zmiennej objaśnianej od jej wartości teoretycznej (czyli zapisanej w MEL) była najmniejsza

Zatem, niech dla $j = 1, \ldots, n$

$$e_j \overset{df}{=} y_j - \hat{y}_j \text{, gdzie}$$

$$\hat{y}_j = a_0 + a_1 x_{j1} + a_2 x_{j2} + \ldots + a_{jm}$$

Wtedy musi być:

$$(*) \quad \sum_{j=1}^{n} e_j^2 \to \min$$

Przypadek jednej zmiennej objaśniającej

W tym przypadku

(MEL): $Y = \beta + \alpha X + \varepsilon$ $(\alpha_0 = \beta, \alpha_1 = \alpha, X_1 = X)$

Skorzystajmy z prawdy (*): MNK dla y, X:

Wtedy

$$(**) \quad S^{(a,b)} = \sum_{j=1}^{m} (y_j - (b + a x_j))^2 \to \min,$$

gdzie a, b są ocenami (oszacowaniami) α i β odpowiednio.

$(a,b) \longrightarrow S(a,b)$ traktujemy jako funkcję 2-zmiennych.

Warunek (**) dla funkcji S oznacza, iż rozważając parę funkcji:

$a \longrightarrow S(\cdot, b)$ (b-ustalone)

2

oraz

$$b \longrightarrow S(a, \cdot), \quad a\text{-ustalone},$$

obie z nich spełniają (**).

Zatem ich pochodne muszą być równe zero!

Rozważamy kolejno względem a (b stałe) i b (a stałe) dostaniemy:

$$\sum_{j=1}^{n} 2(y_j - (b+ax_j))(-2x_j) = 0,$$

co po przekształceniu daje

$$(1) \quad \boxed{b\sum_{j=1}^{n} x_j + a\sum_{j=1}^{n} x_j^2 = \sum_{j=1}^{n} y_j x_j}$$

Teraz po b (a stałe)

$$\sum_{j=1}^{n} 2(y_j - (b+ax_j)) = 0, \quad \text{co po przekształceniu}$$

$$\sum_{j=1}^{n}(b+ax_j) = \sum_{j=1}^{n} y_j \iff$$

(2) $mb + a\sum_{j=1}^{n} x_j' = \sum_{j=1}^{n} y_j'$

Daje to nam układ równań: (1) i (2)

$$(\ast\ast\ast)\begin{cases} b\sum_{j=1}^{n} x_j + a\sum_{j=1}^{n} x_j'^2 = \sum_{j=1}^{n} y_j' x_j' & /:n \\ b\cdot n + a\sum_{j=1}^{n} x_j' = \sum_{j=1}^{n} y_j' & /:n \end{cases}$$

Oznaczmy $\frac{1}{n}\sum_{j=1}^{n} x_j' = \bar{x}$, $\frac{1}{n}\sum_{j=1}^{n} y_j' = \bar{y}$

$(\ast\ast\ast)$ możemy zapisać

$$(\ast)\begin{cases} b\bar{x} + a\frac{1}{n}\sum_{j=1}^{n} x_j'^2 = \frac{1}{n}\sum_{j=1}^{n} y_j' x_j' \\ b + a\bar{x} = \bar{y} \end{cases}$$

Stąd $\boxed{b = \bar{y} - a\bar{x}}$ i po podstawieniu do r. pierwszego:

$$(\bar{y} - a\bar{x})\bar{x} + a\frac{1}{n}\sum_{j=1}^{n} x_j'^2 = \frac{1}{n}\sum_{j=1}^{n} y_j x_j' \Leftrightarrow$$

$$\bar{y}\bar{x} \underbrace{- a(\bar{x})^2 + a\frac{1}{n}\sum_{j=1}^{n} x_j'^2}_{} = \frac{1}{n}\sum_{j=1}^{n} y_j x_j' \Leftrightarrow$$

$$a\left(\frac{1}{n}\sum_{j=1}^{n} x_j'^2 - (\bar{x})^2\right) = \frac{1}{n}\sum_{j=1}^{n} y_j x_j' + \bar{x}\bar{y},$$

co daje

$$a = \frac{\frac{1}{n}\sum_{j=1}^{n} y_j x_j' + \bar{x}\bar{y}}{\frac{1}{n}\sum_{j=1}^{n} x_j'^2 - (\bar{x})^2} = \frac{\sum_{j=1}^{n} y_j x_j + n\bar{x}\bar{y}}{\sum_{j=1}^{n} x_j'^2 - n(\bar{x})^2}$$

Wtedy $(y_j - \bar{y})(x_j' - \bar{x}) = y_j x_j' - \bar{y}x_j' + \bar{x}\bar{y} + y_j\bar{x}$

Sumując od $j=1$ do n mamy

$$\sum_{j=1}^{n} (y_j - \bar{y})(x_j' - \bar{x}) = \sum_{j=1}^{n} y_j x_j' - \bar{y}\sum_{j=1}^{n} x_j' + n\bar{x}\bar{y} + \bar{x}\sum_{j=1}^{n} y_j$$

Ale $-\bar{y}\sum_{j=1}^{n} x_j' + \bar{x}\sum_{j=1}^{n} y_j = -n\bar{y}\bar{x} + n\bar{x}\bar{y} = 0$

Dlatego $\sum_{j=1}^{n}(y_j-\bar{y})(x_j-\bar{x})$ = Licznik dla a

Podobnie, weźmy

$$(x_j-\bar{x})^2 = x_j^2 - 2x_j\bar{x} + (\bar{x})^2$$ i sumujemy

$$\sum_{j=1}^{n}(x_j-\bar{x})^2 = \sum_{j=1}^{n}x_j^2 - 2\bar{x}\sum_{j=1}^{n}x_j + n(\bar{x})^2$$

$$= \sum_{j=1}^{n}x_j^2 - 2n(\bar{x})^2 + n(\bar{x})^2 =$$

$$= \sum_{j=1}^{n}x_j^2 - n(\bar{x})^2$$

Dlatego minimum układu (***) ma postać

MNK:

$u^* = $ (****)

$$\boxed{a = \frac{\sum_{j=1}^{n}(y_j-\bar{y})(x_j-\bar{x})}{\sum_{j=1}^{n}(x_j-\bar{x})^2} \qquad b = \bar{y} - a\bar{x}}$$

Ocena (wariancji) odchyleń losowych

Jeśli dla $j = 1, 2, \ldots n$

$$y_j = b + a x_j + \varepsilon_j \text{ , gdzie}$$

$$\varepsilon_j = (y_j - b - a x_j)$$

to ocena ε ma postać (kwadrat odchyleń S_ε^2)

(5*) $$\boxed{S_\varepsilon^2 = \frac{\sum_{j=1}^{n} e_j^2}{n-2}} \text{ , gdzie } S_\varepsilon = \sqrt{S_\varepsilon^2}$$

Ponadto, standardowe błędy $S(a)$, $S(b)$ oszacowania parametrów strukturalnych α i β wyrażamy ze wzorów:

(6*) $$\boxed{\begin{aligned} S(a) &= \frac{S_\varepsilon}{\sqrt{\sum_{j=1}^{n} (x_j - \bar{x})^2}} \\ S(b) &= S_\varepsilon \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n \sum_{j=1}^{n} (x_j - \bar{x})^2}} \end{aligned}}$$

~~Przykład~~

ZAD 1. Przygotuj arkusz w MS Excel pozwalający

na: - wprowadzenie danych dla y i X

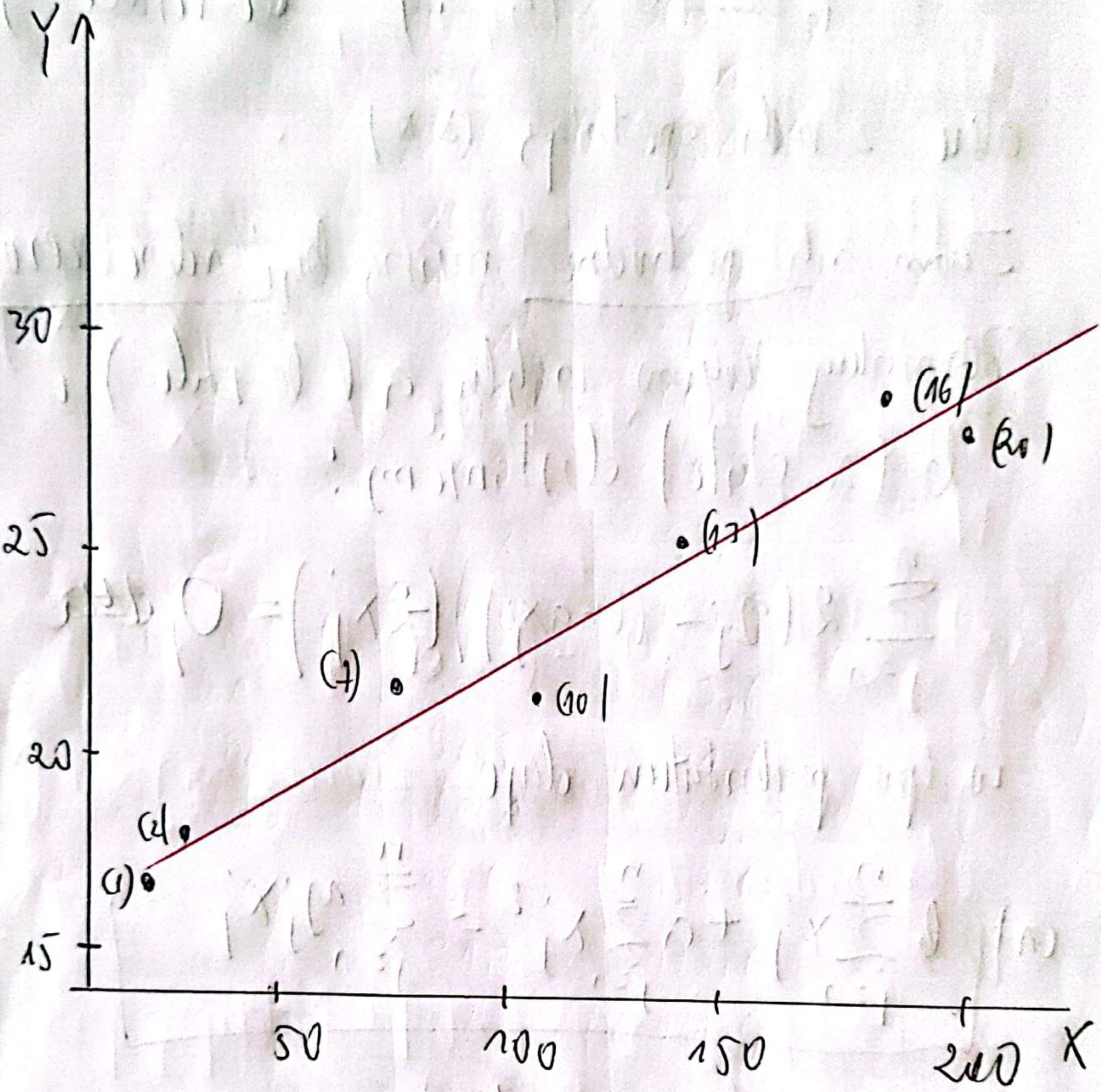
- obliczenie $h(x)$
- obliczenie $S(x)$
- obliczenie $C(x)$

Przykład

Należy zbadać zależność plonu w [q] z [1 ha] (Y) od zużycia nawozu na [1 ha] użytków w [kg] (X) na podstawie danych podanych w tabeli (n = 20)

Kolejny okres #n	Y	X
1	16,9	36,5
2	19,9	39,1
3	19,4	44,1
4	19,9	45,5
5	18,5	49,0
6	20,1	56,4
7	21,2	66,4
8	21,9	80,9
9	24,2	93,4
10	24,0	109,5
11	23,2	123,6
12	26,5	131,6
13	25,1	149,1
14	29,6	157,6
15	31,7	173,6
16	28,3	181,9
17	31,3	193,3
18	28,9	189,0
19	32,5	190,3
20	27,0	188,9
Σ	490,1	2299,7

Analiza wstępna



Rozrzut punktów empirycznych (x_j, y_j), $j = 1 \dots 20$ wskazuje, iż mamy do czynienia z modelem liniowym, zatem

$$(MEL) \quad Y = \beta_0 + \alpha X + \varepsilon$$

Szacujemy parametr α, czyli obliczamy a korzystając z (4*)

- obliczamy $\bar{y} = 24,505$, $\bar{x} = 114,985$
- dla $j = 1 \ldots 20$

$$y_j - \bar{y}, \quad x_j - \bar{x},$$

i podstawiamy do

$$a = \frac{\sum_{j=1}^{20} (y_j - \bar{y})(x_j - \bar{x})}{\sum_{j=1}^{20} (x_j - \bar{x})^2} =$$

$$= \frac{5118,1286}{67586,864} = \underline{\underline{0,076}}$$

Szacujemy parametr β, czyli obliczamy b z (4*)

$$b = \bar{y} - a\bar{x} = 24,505 - 0,076 \cdot 114,985$$

$$= \underline{\underline{15,767}}$$

Stąd, model (MEL) plonu względem zużycia nawozu ma postać

$$\hat{Y}_i = 15,767 + 0,076X.$$

Współczynnik oceny $a_1 = 0,076$ mówi, że przyrostowi zużycia nawozu na 1ha o 1kg odpowiada przyrost plonu z 1ha średnio o 7,6 [kg] (bo Y w q).

ZAD 2

Proszę wykonać te obliczenia na podstawie arkusza przygotowanego w ramach zad 1.

Szacujemy odchylenie składnika losowego ε – szacujemy wzór (S_ε)

$$S_\varepsilon^2 = \frac{\sum_{i=1}^{20} e_i^2}{20-2} = \frac{44,850}{18} = 2,4916,$$

stąd $S_\varepsilon = \sqrt{S_\varepsilon^2} = \sqrt{2,4916} = 1,578$

ZAD 2.

Korzystając z arkusza określ S_{ε}

Interpretacja

$S_{\varepsilon} = 1,578$ oznacza, że rzeczywiste płace różnią się średnio od płacy teoretycznej wynikającej z modelu ekon. o 1,578 tys.

Wyznaczanie standardowych błędów $S(a)$, $S(b)$

Sposób (ex):

$$S(a) = \frac{S_{\varepsilon}}{\sqrt{\sum_{j=1}^{20}(x_j - \bar{x})^2}} = \frac{1,578}{\sqrt{69169}} = 0,006$$

$$S(b) = 0,782$$

ZAD 3. Korzystając z arkusza określ $S(a)$, $S(b)$.

WNIOSEK.

Otrzymany model ma postać

$$\hat{Y} = 15{,}767(\pm 0{,}792) + 0{,}076(\pm 0{,}006)X,$$

gdzie $S_e = 1{,}578$