ECONOMETRIA - 6 CREDITI - Prof DEANA - 2018-2019

• ASSUNZIONI DEL MODELLO OLS •

A1 → Il modello e' lineare

$Y = B_1 + B_2 X + \varepsilon$

linearita' e stabilita' dei parametri

B_1 e B_2 sono costanti

Gli effetti marginali sono costanti

A2 → Non tutti i valori di X possono essere uguali

Dato $\{Y_i, X_i\}$ $i = 1, \dots, n$ $\exists i \neq j$ t.c. $X_i \neq X_j$

Ho bisogno di variabilita' nei miei dati

Il modello OLS spiega la differenza fra gli individui.

→ Non deve esserci multicollinearita' (non potrei ottenere gli stimatori).

A3 → Il campione dev'essere casuale

$\forall i \neq j$ $cov(Y_i, Y_j | X) = 0$, $cov(\varepsilon_i, \varepsilon_j | X) = 0$

Non deve esserci distorsione distorsione nella selezione

Non deve esserci correlazione tra gli errori.

A4 → Esogeneita'

• Forte (modello deterministico / sperimentale)

$E(\varepsilon) = 0 \Rightarrow E(Y) = B_1 + B_2 X$

• Debole (modello stocastico / realistico)

$E(\varepsilon | X) = 0 \Rightarrow E(Y|X) = B_1 + B_2 X \Rightarrow B_1 = E(Y|X)$ intercetta

$\Rightarrow Cov(\varepsilon, X) = 0$ $\qquad B_2 = \frac{\partial E(Y|X)}{\partial X}$ coeff ang

Vale anche per esogeneita' forte.

A5 → Omoschedasticita'

$V(\varepsilon | X) = E(\varepsilon^2 | X) = \sigma^2$

$V(Y|X) = E(B_1 + B_2 X + \varepsilon | X) = \sigma^2$

Varianza dell'errore dato nella distribuzione e' costante

$V(\varepsilon|X) = \sigma^2 = V(Y|X)$. Tenuto conto della distribuzione, l'unica cosa aleatoria, che ha una varianza, e' l'errore.

A6 → Casualita' degli errori.

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

A4, A5

$$Y|X \sim N(B_1 + B_2 X, \sigma^2)$$

Presuppone che il modello sia fatto talmente bene che l'errore, la spazzatura, sia casuale.

Se valgono queste assunzioni posso usare il modello OLS. Avra' determinate proprieta'.

Modello OLS.

Ordinary least squares: e' un metodo di stima che si basa su una funzione di perdita, da minimizzare.
Gli errori sono teorici, quindi devo minimizzare i residui al quadrato.

$\min_{B_1 B_2} \sum_{i=1}^{n} e_i^2$ → Minimizzo i residui al quadrato rispetto a B_1 e B_2

Per trovare gli stimatori:

$$\min_{B_0 B_1} \sum_{i=1}^{n} e_i^2 = \min_{B_0 B_1} \sum_{i=1}^{n} \underbrace{(y_i - B_0 - B_1 X_i)^2}_{\text{Funzione obiettivo} = D(B_0, B_1)}$$

$$\hat{B} = \begin{pmatrix} \hat{B}_0 \\ \hat{B}_1 \end{pmatrix} = \underset{B_0 B_1}{\text{argmin}}\, D(B_0, B_1)$$

↓

Stimatore OLS: stimatore che minimizza la funzione obiettivo, per minimizzare facciamo la derivata.

$\boxed{\text{FOC}_1:}$ (first order condition, condizione di primo ordine)

$$\frac{\partial D(B_0, B_1)}{\partial B_0} = 0 \qquad -2 \sum_i (y_i - \hat{B}_0 - \hat{B}_1 X_i) = 0$$

$$-\frac{2}{n} \sum_i (y_i - \hat{B}_0 - \hat{B}_1 X_i) = \frac{0}{n} \qquad \bar{Y} - \hat{B}_0 - \hat{B}_1 \bar{X} = 0 \qquad \hat{B}_0 = \bar{Y} - \hat{B}_1 \bar{X}$$

$\boxed{\text{FOC } 2:}$

$$\frac{\partial D(B_0, B_1)}{\partial B_1} = 0 \qquad -2 \sum_i (y_i - \hat{B}_0 - \hat{B}_1 X_i) X_i = 0$$

$$-\frac{2}{n} \sum (y_i - \hat{B}_0 - \hat{B}_1 X_i) X_i = \frac{0}{n}$$

$$\frac{1}{n}\sum_i (y_i x_i) - \frac{1}{n}\sum_i \hat{B}_0 x_i - \frac{1}{n}\sum_i \hat{B}_1 x_i^2 = 0$$

$$\frac{1}{n}\sum_i y_i x_i = \hat{B}_0 \bar{x} + \hat{B}_1 \frac{1}{n}\sum_i x_i^2$$

Sostituisco $\hat{B}_0 = \bar{y} - \hat{B}_1 \bar{x}$

$$\frac{1}{n}\sum_i y_i x_i = [\bar{y} - \hat{B}_1 \bar{x}]\bar{x} + \hat{B}_1 \frac{1}{n}\sum_i x_i^2$$

$$\frac{1}{n}\sum_i y_i x_i - \bar{y}\bar{x} = -\hat{B}_1 \left(\bar{x}^2 - \frac{1}{n}\sum_i x_i^2\right)$$

$$\frac{1}{n}\sum_i y_i x_i - \bar{y}\bar{x} = -\hat{B}_1 \left(\bar{x}\frac{1}{n}\sum_i x_i - \frac{1}{n}\sum_i x_i^2\right)$$

$$\frac{1}{n}\sum_i y_i x_i - \bar{y}\bar{x} = -\hat{B}_1 \frac{1}{n}\left(\bar{x}\sum_i x_i - \sum_i x_i^2\right)$$

$$\cancel{\frac{1}{n}}\sum_i y_i x_i - \bar{y}\cancel{\frac{1}{n}}\sum_i x_i = -\hat{B}_1 \cancel{\frac{1}{n}}\left(\bar{x}\sum_i x_i - \sum_i x_i^2\right)$$

$$\sum_i y_i x_i - \bar{y}\sum_i x_i = -\hat{B}_1 \left(\bar{x}\sum_i x_i - \sum_i x_i^2\right)$$

$$\sum_i x_i (y_i - \bar{y}) = -\hat{B}_1 \sum_i x_i (\bar{x} - x_i)$$

$$\sum_i x_i (y_i - \bar{y}) = \hat{B}_1 \sum_i x_i (x_i - \bar{x}) \Rightarrow \hat{B}_1 = \frac{\sum_i x_i (y_i - \bar{y})}{\sum_i x_i (x_i - \bar{x})}$$

Considerando che $\sum_i (x_i - \bar{x}) = 0$, quindi anche $\bar{x}\sum_i (x_i - \bar{x}) = 0$

$$\Rightarrow \sum_i x_i (x_i - \bar{x}) = \sum_i x_i (x_i - \bar{x}) + \underbrace{\bar{x}\sum_i (x_i - \bar{x})}_{=0} = \sum_i (x_i - \bar{x})(x_i - \bar{x}) = \sum_i (x_i - \bar{x})^2$$

Quindi: $\hat{B}_1 = \dfrac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \dfrac{\widehat{COV}(x, y)}{\widehat{VAR}(x)} \Rightarrow$

Infatti $VAR(x) \neq 0$, deve esserci variabilità

$\Rightarrow \hat{y}_i = \hat{B}_0 + \hat{B}_1 x_i$ ottenuto sostituendo i valori stimati

$e_i = y_i - \hat{y}$ differenza tra valore vero e valore fittato.

$y_i = \hat{y} + e_i$

$e_i \rightarrow$ componente residua.

$\hat{y} \rightarrow$ componente spiegata

$\hat{B}_1 = \dfrac{\widehat{COV}(X, Y)}{\widehat{VAR}(X)}$ $\qquad$ $B_1 = \dfrac{COV(X, Y)}{VAR(X)}$

$\downarrow$ stimato $\qquad$ $\downarrow$ teorico

Non è detto che il modello vero sia uguale a quello che ho stimato

· MODELLO MULTIVARIATO ·

$$\underset{n\times 1}{Y} = \underset{n\times k}{X}\ \underset{k\times 1}{B} + \underset{n\times 1}{\varepsilon}$$

n = osservazioni; K = regressori perche' il primo e' una costante

ε → errori

B → regressori che dovro' stimare

X → matrice dei regressori

Y → variabile dipendente

ASSUNZIONI MODELLO MULTIVARIATO.

A1 → Linearita'

$Y = XB + \varepsilon$

A2 → Variabilita'

Rango (X) = K, $K < n$ altrimenti piu' incognite e meno equazioni

Tutti i vettori linearmente indipendenti

A3 → Casualita'

$Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$

A4 → Esogeneita'

$E(\varepsilon | X) = 0 \Rightarrow E(Y|X) = XB$

$B_j = \frac{\partial E(Y|X)}{\partial X_j}$ effetto marginale

A5 → Omoschedasticita'

$V(\varepsilon_i) = \sigma^2 \quad \forall i$

Varianza: come si trova?

$$V(\varepsilon) = E\{(\varepsilon - E(\varepsilon))(\varepsilon - E(\varepsilon))'\} = E\left[\begin{pmatrix}\varepsilon_1\\ \vdots \\ \varepsilon_n\end{pmatrix}(\varepsilon_1 \cdots \varepsilon_n)\right] = E\begin{bmatrix}\varepsilon_1^2 & \cdots & \varepsilon_1\varepsilon_n \\ \vdots & \ddots & \\ \varepsilon_n\varepsilon_1 & & \varepsilon_n^2\end{bmatrix}$$

$0 = E(\varepsilon_1) = \dots = E(\varepsilon_n)$

$E(\varepsilon^2) = V(\varepsilon) = \sigma^2$

$\Rightarrow V(\varepsilon) = \sigma^2 I_n$

ε hanno varianza, tra di loro non c'e' correlazione

A6 → $\varepsilon \sim N(0, \sigma^2 I_n)$

Casualita' dell' errore

$Y \sim N(XB, \sigma^2 I_n)$

Per trovare gli stimatori dobbiamo minimizzare

$\underset{B}{Min} \sum_i \varepsilon_i^2 \qquad \sum_i \varepsilon_i^2 = \varepsilon'\varepsilon = (Y - XB)'(Y - XB)$

↳ prima il trasposto, così troviamo uno scalare.