Titolo nota 03/06/2019

# MICROECONOMETRIA - LEZIONE 13

## EFFETTI MARGINALI IN MODELLI LOGIT/PROBIT

$$p_i = E(y_i) = F(x_i\beta)$$

↳ CDF < LOGISTICA $\Lambda(x_i\beta)$ ; NORMALE STANDARD $\Phi(x_i\beta)$

$$\frac{\partial p_i}{\partial x_{ir}} = \frac{\partial E(y_i)}{\partial x_{ir}} = \frac{\partial F(x_i\beta)}{\partial x_i\beta} \cdot \frac{\partial x_i\beta}{\partial x_{ir}}$$

DIPENDE DAL $i = 1 \dots N$ ← $\frac{\partial F(x_i\beta)}{\partial x_i\beta}$ ; $\frac{\partial x_i\beta}{\partial x_{ir}} = \beta_r$

EFFETTI MARGINALI $\dfrac{\partial p_i}{\partial x_{iz}}$ VARIANO AL

VARIARE DI $i = 1 \ldots N$ _

NECESSITÀ DI SINTETIZZARE GLI EFFETTI MARGINALI

COME ?

1) EFFETTI MARGINALI MEDI : $\dfrac{1}{N} \sum_{i=1}^{N} \dfrac{\partial p_i}{\partial x_{iz}}$

2) EFFETTI MARGINALI VALUTATI NELLA MEDIA DEI REGRESSORI:

$$\frac{\partial p_i}{\partial x_{ir}} = \frac{\partial F(x_i\beta)}{\partial x_i\beta} \cdot \frac{\partial x_i\beta}{\partial x_{ir}} = \frac{\partial F(x_i\beta)}{\partial x_i\beta}\,\beta_r$$

$F(\bar{x}\beta)$

$\Downarrow$

MEDIA CAMPIONARIA DEI REGRESSORI

L'EFFETTO MARGINALE VALUTATO NELLA MEDIA DEI REGRESSORI

$$\frac{\partial F(\bar{x}\cdot\beta)}{\partial \bar{x}\cdot\beta} \cdot \beta_2$$ RISULTA

COSTANTE RISPETTO A $i = 1 \ldots N$

## EFFETTI MARGINALI CON VARIABILI ESPLICATIVE DI TIPO QUALITATIVA (DUMMY)

$$y_i = F(\beta_1 + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \delta D_i) + u_i$$

$$p_i = E(y_i) = F(\beta_1 + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \delta D_i)$$

dove $D_i = \begin{cases} 0 & \text{ALTRIMENTI} \\ 1 & \text{SE } i\text{-ESIMO INDIVIDUO APPARTIENE ALLA ``CATEGORIA'' } C \end{cases}$

$$\frac{\partial p_i}{\partial x_{i2}} = \frac{\partial F(\beta_1 + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \delta D_i)}{\partial(\dots)} \cdot \beta_2$$

$$\frac{\partial p_i}{\partial D_i}$$ ?? NON SENSO

$$D_i = 0 \longrightarrow p_i|_{D_i=0} = F(\beta_1 + \beta_2 x_{i2} + \dots + \beta_k x_{ik})$$

$$D_i = 1 \longrightarrow p_i|_{D_i=1} = F((\beta_1 + \delta) + \beta_2 x_{i2} + \dots + \beta_k x_{ik})$$

RAFFORZA INTERCETTA

<u>IMPATTO</u> DELLA "DUMMY" SULLA PROBABILITÀ $p_i$:

$$p_i|D_i=1 - p_i|D_i=0$$

## MISURA DI 'GOODNESS OF FIT' NEI MODELLI LOGIT/PROBIT

INCISO: LPM $y_i = x_i \beta + u_i \qquad i = 1 \dots N$

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

DOVE $TSS = ESS + RSS \qquad 0 \leq R^2 \leq 1$

LOGIT/PROBIT: PSEUDO-$R^2$

DOVE $$\text{PSEUDO-}R^2 = 1 - \frac{\log L_1}{\log L_0}$$

E "1" INDICA IL MODELLO DI INTERESSE

"0" INDICA IL MODELLO CON SOLA COSTANTE/INTERCETTA

$$y_i = F(\beta_1) + v_i$$

$$PSEUDO\_R^2 = 1 - \frac{\log L_1}{\log L_0}$$

DOVE $\log L_1$ = FUNZIONE DI LOG VEROSIMIGLIANZA MASSIMIZZATA (VALUTATA NELLE STIME DI ML) DEL MODELLO DI INTERESSE

$\log L_0$ = FUNZIONE DI LOG VEROSIMIGLIANZA MASSIMIZZATA (VALUTATA IN $\hat{\beta}_{1ML}$) DEL MODELLO CON SOLA INTERCETTA

SE $\log L_1 = \log L_0$, ALLORA PSEUDO $R^2 = 0$

INTERPRETAZIONE: IL MODELLO DI INTERESSE SI COMPORTA COME IL MODELLO CON SOLA COSTANTE, IL CUI POTERE ESPLICATIVO È <u>NULLO</u>

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

$R^2 = 0$ SE $ESS = 0$, O EQUIVALENTEMENTE, SE $RSS = TSS$

$\longrightarrow y_i = \beta_1 + u_i$

SE $\log L_1 = 0$, ALLORA PSEUDO-$R^2 = 1$

INTERPRETAZIONE: $L_1$ = FUNZIONE DI VEROSIMIGLIANZA / FUNZIONE DI PROBABILITÀ DEL MODELLO DI INTERESSE

"EVENTO CERTO"

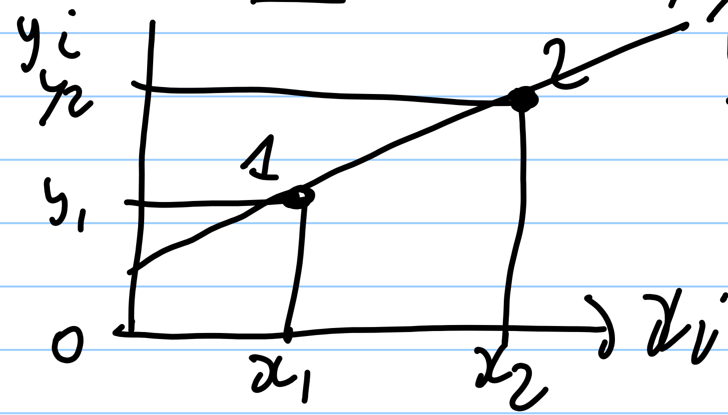
SE $L_1 = 1$, ALLORA $\log L_1 = 0$

FIT PERFETTO

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

$R^2 = 1$ SE $ESS = TSS$ o, EQUIVALENTEMENTE, $RSS = 0$

$\hookrightarrow$ $y_i = \underline{\beta_1} + \underline{\beta_2} x_i + u_i$ $\quad i = 1 \ldots N = \underline{\underline{2}}$




$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$

$\hat{u}_i = y_i - \hat{y}_i = 0$

$\forall i = 1 - N = \underline{\underline{2}}$

## CALCOLO DI $\log L_0$

$$\log L = \sum_{i=1}^{N} \left[ y_i \log p_i + (1-y_i) \log(1-p_i) \right]$$

DOVE $p_i \equiv F(x_i\beta)$

PER OTTENERE $\log L$ MASSIMIZZATA, DOBBIAMO STIMARE CON ML I PARAMETRI $\beta$, OTTENERE $\hat{\beta}_{ML}$ E VALUTARE $\log L$ IN $\hat{\beta}_{ML}$ —

RELATIVAMENTE AL MODELLO "0", ABBIAMO $F(\beta_1)$,

CIOÈ $F(\beta_1) = P_i$ → PROBABILITÀ DI OSSERVARE $y = 1$

$N$ INDIVIDUI
- $N_1$ = N° DI INDIVIDUI PER CUI $y_i = 1$
- $N_0$ = N° " " " $y_i = 0$

TALE CHE $N = N_1 + N_0$

PER STIMARE $p$ COSTANTE, SI UTILIZZA:

$$\hat{p} = \frac{N_1}{N} = \frac{\text{N° CASI FAVOREVOLI}}{\text{N° CASI POSSIBILI}}$$

$\Downarrow$

$$\log L_0 = \sum_{i=1}^{N} \left[ y_i \cdot \log \hat{p} + (1 - y_i) \log (1 - \hat{p}) \right] =$$

$$= \underbrace{\sum_i y_i}_{N_1} \cdot \log \left( \frac{N_1}{N} \right) + \underbrace{\sum_i (1 - y_i)}_{N_0} \log \left( \frac{N_0}{N} \right)$$

## MISURE DI BONTÀ DELLE PREVISIONI IN MODELLI LOGIT/PROBIT

INCISO: MODELLO LINEARE $y_i = x_i \beta + u_i$

$\hat{\beta}$ OLS

$\hat{u}_i = y_i - \hat{y}_i$

RESIDUI (IN SAMPLE)

ERRORI DI PREVISIONE (OUT OF SAMPLE)

FORECAST

$$MSFE = \frac{1}{N} \sum_i (y_i - \hat{y}_i)^2 = \frac{1}{H} \sum_i \hat{u}_i^2$$

MODELLI LOGIT / PROBIT : $y_i = F(x_i\beta) + u_i$

$0 \quad 1 \quad$ DUMMY

$\hat{\beta}_{ML}$

$$F(x_i\hat{\beta}_{ML}) = \underbrace{\hat{p}_i}_{\hat{y}_i} \quad \text{(VALORI FITTATI, PROBABILITÀ)}$$

$$\hat{u}_i = y_i - \hat{y}_i = \hat{p}_i \in (0,1)$$

$y_i \to 0,\ 1$

(ERRORI DI PREVISIONE)

$y_i$: VARIABILE DISCRETA

$\hat{y}_i$: VARIABILE CONTINUA

$\hat{u}_i$ CALCOLATO CONFRONTANDO UNA VARIABILE DISCRETA CON UNA VARIABILE CONTINUA <u>NON HA SENSO</u>

PER CONFRONTARE, ALL'INTERNO DELLA DEFINIZIONE DI ERRORE DI PREVISIONE, DUE VARIABILI DI NATURA OMOGENEA, INTRODUCIAMO <u>LA SEGUENTE REGOLA</u>:

"IL CASO $y_i = 1$ È PREVISTO CORRETTAMENTE DAL MODELLO SE $\hat{p}_i \geq 0.5$"

VALORE SCELTO "ARBITRARIO"

SULLA BASE DELLA REGOLA DI <u>CORRETTA PREVISIONE</u>

È POSSIBILE COSTRUIRE LA SEGUENTE VARIABILE:

$$\widehat{\widehat{y}}_i = \begin{cases} 1 & \text{SE } \hat{p}_i \geq 0.5 \\ 0 & \text{SE } \hat{p}_i < 0.5 \end{cases}$$

→ DUMMY

$\Downarrow$

ERRORE DI PREVISIONE $\widehat{u}_i = y_i - \widehat{y}_i$

$y_i$: 0 , 1

$\widehat{y}_i$: 0 , 1

$\Downarrow$

$$\boxed{MSFE = \frac{1}{N} \sum_{i=1}^{N} \widehat{u}_i^{\,2}}$$

LA CAPACITÀ PREVISIVA DEL MODELLO DI INTERESSE DEVE ESSERE SEMPRE VALUTATA RISPETTO A UN MODELLO DI RIFERIMENTO (ES. MODELLO CON SOLA COSTANTE)

$$MSFE_1 = \frac{1}{N}\sum_i \hat{u}_i^2 = \frac{1}{N}\sum_i (y_i - \hat{y}_i)^2$$

$$MSFE_0 = \frac{1}{N}\sum_i (y_i - \hat{y}_i)^2 \begin{cases} \xrightarrow{\hat{p} \geq 0.5} \frac{1}{N}\sum_i (y_i - 1)^2 \quad (*) \\ \xrightarrow{\hat{p} < 0.5} \frac{1}{N}\sum_i (y_i)^2 = \frac{1}{N}\sum_i y_i = \frac{M}{N} \end{cases}$$

(∗) $\frac{1}{N}\sum_i (y_i - 1)^2 = \frac{1}{N}\sum_i \left(y_i + 1 - 2y_i\right) =$

$$= \frac{N_1}{N} + \cancel{\frac{N}{N}} - 2\frac{N_1}{N} =$$

$$= 1 - \frac{N_1}{N} = \frac{N_0}{N} = 1 - \hat{p}$$

$$\Downarrow$$

$$\boxed{MSFE_0 = \begin{cases} \hat{p} & \text{se } \hat{p} < 0.5 \\ 1-\hat{p} & \text{se } \hat{p} \geq 0.5 \end{cases}}$$

ES. $N = 100$, $N_1 = 90$ $\Big\} \ \hat{p} = \frac{N_1}{N} = 0.9$

$MSFE_0 = 1 - \hat{p} = \underline{\underline{0.1}}$

$MSFE_1$ <u>DEVE</u> ESSERE INFERIORE AL 10% !!