Titolo nota ~~31/05/2018~~

MICROECONOMETRIA – LEZIONE 15a

MODELLI LOGIT/PROBIT PER SCELTE MULTIPLE

$j = 0, 1, 2, \dots, J$ (SCELTE)

$p_{ij} \equiv \Pr(y_i = j)$

RELAZIONE TRA p_{ij} E VARIABILI ESPLICATIVE (CARATTERISTICHE INDIVIDUALI E/O CARATTERISTICHE SCELTE)

E PARAMETRI

$$\max_{\text{PARAMETRI}} \log L \equiv \sum_{i=1}^{N} \ell_i$$

DOVE $\ell_i \equiv \sum_{j=0}^{J} I(y_i = j) \log p_{ij}$

FUNZIONE "INDICATORE" CHE SELEZIONA LA j-ESIMA SCELTA DELL'i-ESIMO INDIVIDUO

COMMENTI

1) p_{ij}, $j = 0 \ldots J$ E p_{it}, $j = 0, 1$, (V. MODELLI PER DATI PANEL CON VAR. DIP. <u>BINARIA</u>)

HANNO LA <u>STESSA</u> STRUTTURA

$\Downarrow$

I MODELLI PROBIT PER SCELTE MULTIPLE $(J+1)$

COINVOLGONO PROBABILITÀ DEFINITE IN TERMINI DI INTEGRALI MULTIPLI $(J+1)$ (K. MODELLI PANEL PROBIT EFFETTO RANDOM) –

DI CONSEGUENZA SONO PREFERIBILI **I MODELLI DI TIPO LOGISTICO**

2) I MODELLI DI TIPO LOGISTICO IN UN CONTESTO A SCELTE MULTIPLE SI FONDANO SU UNA GENERALIZZAZIONE DELLA DISTRIBUZIONE LOGISTICA, NOTA COME TYPE I EXTREME VALUE

ε VARIABILE CASUALE

$$F(\varepsilon) = \exp[-\exp(-\varepsilon)]$$

3) DUE CLASSI DI MODELLI LOGIT PER SCELTE MULTIPLE :

MULTINOMIAL LOGIT (TRATTANO REGRESSORI X_i)

CONDITIONAL LOGIT (TRATTANO REGRESSORI X_{ij})

DOVE X_i = VARIABILI CHE VARIANO SOLO RISPETTO AGLI INDIVIDUI (ES. REDDITO)

X_{ij} = VARIABILI CHE VARIANO RISPETTO AGLI INDIVIDUI E ALLE SCELTE (ES. COMMUTING TIME)

ii) È POSSIBILE DIMOSTRARE CHE I MODELLI MULTINOMIAL LOGIT (MLM) SONO CASI PARTICOLARI DEI MODELLI CONDITIONAL LOGIT (CLM)

MODELLI MULTINOMIAL LOGIT (MLM)

$$y_{ij}^* = x_i \beta_j + \varepsilon_{ij} \qquad , j = 0 \ldots J$$

VARIABILE LATENTE

PARAMETRI SONO SPECIFICI PER CIASCUNA SCELTA

VARIABILE OSSERVATA : $y_{ij} = j$ SE $y_{ij}^* = \max(y_{i0}^*, y_{i1}^*, \ldots \ldots y_{iJ}^*)$

È POSSIBILE DIMOSTRARE CHE :

$$p_{ij} = \frac{\exp(x_i' \beta_j)}{1 + \sum_{k=1}^{J} \exp(x_i' \beta_k)} \quad , \; j = 1 \ldots J$$

$$p_{i0} = 1 - \sum_{j=1}^{J} p_{ij} = \frac{1}{1 + \sum_{k=1}^{J} \exp(x_i' \beta_k)}$$

OSSERVAZIONI:

1) $$\frac{p_{ij}}{p_{i0}} = \frac{\exp(x_i \beta_j)}{\cancel{1+\sum_k \exp(x_i \beta_k)}} \cdot \cancel{\left(1+\sum_k \exp(x_i \beta_k)\right)}$$

$$\Downarrow$$

$$\log\left(\frac{p_{ij}}{p_{i0}}\right) = x_i \beta_j$$ (LOG ODDS RATIO LINEARE, V. LOGIT BINARIO)

2) $$\frac{p_{ij}}{p_{ih}} = \frac{\exp(x_i \beta_j)}{\exp(x_i \beta_h)} \qquad \begin{matrix} j \neq 0 \\ j \neq h \\ h \neq 0 \end{matrix}$$

TALE ODDS RATIO NON
DIPENDE DALLE ALTRE SCELTE
(CIOÈ DALLE SCELTE DIVERSE DA j E h)

3) $\max_{\beta_j} \log L(\cdot) = \sum_{i=1}^{N} \ell_i$,

DOVE $\ell_i \equiv \sum_{j=0}^{J} I(y_i = j) \log p_{ij}$

V. DEFINIZIONE DELLE PROBABILITÀ FORNITA IN PRECEDENZA

$\hat{\beta}_j$ È MLE (CONSISTENTE) PER β_j, $j = 0 \ldots J$.

$\boxed{\text{MODELLI CONDITIONAL LOGIT (CLM)}}$

$$y_{ij}^* = x_{ij}\beta + \varepsilon_{ij}$$

PARAMETRI COMUNI

VARIANO RISPETTO A i E j

$$y_{ij} = j \quad \text{SE} \quad y_{ij}^* = \max(y_{i0}^*, y_{i1}^*, \ldots, y_{iJ}^*)$$

È POSSIBILE DIMOSTRARE CHE :

$$p_{ij} = \frac{\exp(x_{ij}\beta)}{\sum_{k=0}^{J} \exp(x_{ik}\beta)} \qquad j, k = 0 \ldots J$$

$$\Downarrow$$

$$\frac{p_{ij}}{p_{ik}} = \frac{\exp(x_{ij}\beta)}{\exp(x_{ik}\beta)} = \exp\left[(x_{ij} - x_{ik})\beta\right] \quad j \neq k$$

↓ ODDS RATIO INDIPENDENTE DA INFO RIGUARDANTI TUTTE LE ALTRE SCELTE

SIA NEI MLM SIA NEI CLM GLI ODDS RATIO TRA SCELTA j E SCELTA h $(j \neq h)$ SONO INDIPENDENTI DA TUTTE LE ALTRE SCELTE

→ INDIPENDENZA DELLE ALTERNATIVE IRRILEVANTI (IIA, INDEPENDENCE OF IRRELEVANT ALTERNATIVES)

IIA PUÒ RISULTARE PARTICOLARMENTE RESTRITTIVA IN DATE SITUAZIONI, PER ESEMPIO QUANDO LE SCELTE SONO MOLTO SIMILI

ES. <u>PARADOSSO DELL'AUTOBUS "ROSSO"</u>

SCELTA 1: CAR (C)

SCELTA 2: RED BUS (RB)

$$Pr(C) = Pr(RB) = \frac{1}{2}$$

$$\Downarrow$$

$$\frac{Pr(C)}{Pr(RB)} = 1$$

LA METÀ DI RB VIENE DIPINTA DI "BLU".

SCELTA 3 : BLUE BUS (BB)

$$P(C) = \frac{1}{2}$$

$$P_R(RB) = P_R(BB) = \frac{1}{4}$$

$$\Downarrow$$

$$\frac{P_R(C)}{P_R(RB)} = \frac{1/2}{1/4} = 2$$

SE UTILIZZASSIMO MLM/CLM PER ANALIZZARE TALI SCELTE INDIVIDUALI DI TRASPORTO, ESSENDO MLM/CLM FONDATI SULLA IIA, IL RAPPORTO TRA $P_r(C)$ E $P_r(RB)$ DEVE ESSERE INDIPENDENTE DALLE ALTRE SCELTE (CIOÈ DALLA SCELTA BB) —

IN ALTRI TERMINI, $\dfrac{P_R(C)}{P_R(RB)} = 1$

ANCHE IN PRESENZA DI BB –

↳ MA IMPONE CHE $P_R(C)$ PASSI DA $\frac{1}{2}$ A $\frac{1}{3}$

$$\Downarrow \quad \frac{P(C)}{P(RB)} = \frac{1/3}{1/3} = \underline{\underline{1}}$$

LA RIDUZIONE DI $Pr(C)$ DA $\frac{1}{2}$ A $\frac{1}{3}$ <u>NON</u> È GIUSTIFICATA DA UN PUNTO DI VISTA LOGICO, VISTA LA NATURA DI "SOSTITUTI PERFETTI" DELLE SCELTE RB E BB

$\Downarrow$

MLM/CLM NON COSTITUISCONO MODELLI APPROPRIATI PER L'ANALISI DI <u>SCELTE MOLTO SIMILI TRA LORO</u>

MODELLI ALTERNATIVI, NON SOGGETTI A IIA

ES. NESTED LOGIT

IPOTESI: POSSIBILITÀ DI STRUTTURARE LE SCELTE INDIVIDUALI SU PIÙ LIVELLI

MEZZI DI TRASPORTO : AUTO, MOTO, BICI, TRAM, METRO, BUS

<u>LIVELLO 1</u> : PUBBLICO (TRAM, METRO, BUS)
PRIVATO (AUTO, MOTO, BICI)

<u>LIVELLO 2</u> : DATO <u>PUBBLICO</u>, SCELTA TRA TRAM, METRO, BUS
DATO <u>PRIVATO</u>, SCELTA TRA AUTO, MOTO, BICI

G_s = N° DI SCELTE A DISPOSIZIONE DELL'INDIVIDUO ALL'INTERNO DELL' S-ESIMO GRUPPO

DUE TIPI DI PROBABILITÀ: $p(y \in G_s)$ (1° LIVELLO)

$p(y=j \mid y \in G_s)$ (2° LIVELLO)

$$\Downarrow$$

$$\ell_i \equiv \sum_{s=1}^{S} \Big[I(y_i \in G_s) \log p(y \in G_s) + I(y_i = j \mid y \in G_s) \log p(y = j \mid y \in G_s) \Big]$$

$$\log L(\cdot) = \sum_{i=1}^{N} \ell_i$$

$$\max_{\text{PARAMETRI}} \log L(\cdot)$$

È POSSIBILE DIMOSTRARE CHE GLI ODDS RATIOS OTTENUTI SULLA BASE DELLE STIME <u>NESTED LOGIT</u> NON SOFFRONO DEL PROBLEMA DELLA IIA –

DALLA FORMA DELLA P_i SI EVINCE CHE GLI ODDS RATIOS TRA DUE QUALSIASI SCELTE TENGONO CONTO DELLA PRESENZA DI TUTTE LE ALTRE

(È LA NATURA A <u>DUE/PIÙ LIVELLI</u> DEL MODELLO NESTED LOGIT A GARANTIRE TALE RISULTATO) —