MICROECONOMETRIA - LEZIONE 17 (LAB 6a)

Titolo nota 07/06/2018

MODELLI PER VARIABILI DIPENDENTI LIMITATE (LDVM)

$$y_i^* = x_i\beta + \varepsilon_i \quad , \quad \varepsilon_i \sim N(0, \sigma^2)$$

CENSURA :

$$y_i = y_i^* \quad \text{SE } y_i^* > 0$$

$$y_i = 0 \quad \text{SE } y_i^* \leq 0$$

TRONCAMENTO : $y_i = y_i^*$ SE $y_i^* > 0$

NULLA SE $y_i^* \leq 0$

<u>TRONCAMENTO</u> : $E(y_i^* | y_i^* > 0) \neq x_i\beta$

<u>CENSURA</u> : $E(y_i^*) \neq x_i\beta$

OLS DISTORTO E INCONSISTENTE ($N \to \infty$)

CAUSA DEL PROBLEMA :

$$E(\varepsilon_i \mid \varepsilon_i > -x_i\beta) \neq 0$$

SOLUZIONI DEL PROBLEMA

1) OLS "CORRETTI"
2) MLE

STIMATORI CONSISTENTI PER I PARAMETRI DI INTERESSE

1) OLS "CORRETTI"

$$E(\varepsilon_i \mid \varepsilon_i > -x_i\beta)$$

RISULTATO : $z \sim N(0,1)$

$$E(z \mid z > k) = \frac{\phi(k)}{1 - \Phi(k)}$$

DENSITÀ DI UNA NORMALE STANDARD

CDF NORMALE STANDARD

$$\sigma E\left(\frac{\varepsilon_i}{\sigma} \,\middle|\, \frac{\varepsilon_i}{\sigma} > \overset{K}{-\frac{x_i\beta}{\sigma}}\right) =$$

$$= \sigma \frac{\phi(K)}{1-\Phi(K)} = \sigma \frac{\phi\left(-\frac{x_i\beta}{\sigma}\right)}{1-\Phi\left(-\frac{x_i\beta}{\sigma}\right)} =$$

$$= \sigma \frac{\phi\left(\frac{x_i\beta}{\sigma}\right)}{\Phi\left(\frac{x_i\beta}{\sigma}\right)}$$

IMPLICAZIONI DI TALE RISULTATO

TRONCAMENTO

$$E(y_i^* \mid y_i^* > 0) = E(y_i \mid y_i^* > 0) = x_i\beta + \underbrace{\boxed{E(\varepsilon_i \mid \varepsilon_i > -x_i\beta)}}_{\text{BIAS}}$$

$$= x_i\beta + \sigma \cdot \frac{\phi\left(\frac{x_i\beta}{\sigma}\right)}{\Phi\left(\frac{x_i\beta}{\sigma}\right)} =$$

$$= x_i\beta + \sigma\, \phi_i / \Phi_i$$, DOVE

$$\phi_i \equiv \phi\left(\frac{x_i\beta}{\sigma}\right)$$

$$\Phi_i = \Phi\left(\frac{x_i\beta}{\sigma}\right)$$

$$\Downarrow$$

$$\boxed{y_i^* = x_i\beta + \sigma\frac{\phi_i}{\Phi_i} + \eta_i} \quad E(\eta_i) = E(\eta_i \mid \dots) = 0$$

BIAS / MILLS' RATIO

VISTO CHE IL BIAS ORIGINARIO VIENE ORA INCORPORATO NELLA PARTE SISTEMATICA DEL MODELLO COME REGRESSORE AGGIUNTIVO, IL MODELLO <u>AUMENTATO</u> HA UN ERRORE IL CUI VALORE ATTESO CONDIZIONATO È <u>NULLO</u> $\Longrightarrow$

$\Longrightarrow$ OLS CONSISTENTE PER I PARAMETRI DEL <u>MODELLO AUMENTATO</u> (β E σ)

CENSURA :

$$E(y_i^*) = Pr(y_i^* > 0)\,\overset{\text{V. TRONCAMENTO}}{\boxed{E(y_i^* \mid y_i^* > 0)}} + Pr(y_i^* \leq 0)\,\underbrace{E(y_i^* \mid y_i^* \leq 0)}_{=0}$$

$$Pr(x_i\beta + \varepsilon_i > 0) = Pr(\varepsilon_i > -x_i\beta) = Pr(\varepsilon_i \leq x_i\beta) =$$

$$= Pr\left(\frac{\varepsilon_i}{\sigma} \leq \frac{x_i\beta}{\sigma}\right) = \Phi\left(\frac{x_i\beta}{\sigma}\right)$$

$$E(y_i^*) = \Phi\left(\frac{x_i\beta}{\sigma}\right)\left[x_i\beta + \sigma\frac{\phi\left(\frac{x_i\beta}{\sigma}\right)}{\Phi\left(\frac{x_i\beta}{\sigma}\right)}\right]$$

$$= \Phi_i x_i\beta + \sigma\phi_i$$

$$\Downarrow$$

$$\boxed{y_i^* = \Phi_i x_i\beta + \sigma\phi_i + \eta_i,} \quad \underbrace{E(\eta_i)}_{=0} = E(\eta_i | \cdots) =$$

IL MODELLO <u>AUMENTATO</u> HA UN ERRORE η_i: IL CUI VALORE ATTESO CONDIZIONATO È NULLO $\Rightarrow$ OLS CONSISTENTI PER β E σ NEL <u>MODELLO AUMENTATO</u>

PER STIMARE CON OLS I MODELLI "AUMENTATI"/"CORRETTI"

IN CASO DI TRONCAMENTO E CENSURA È NECESSARIO <u>OSSERVARE</u>

IL MILLS RATIO   <u>STIMARE</u>

<u>PROBLEMA</u> : IL MILLS RATIO DIPENDE DAI PARAMETRI

β E σ, PROPRIO QUEI PARAMETRI CHE VOGLIAMO STIMARE CONSISTENTEMENTE CON OLS APPLICATO AI <u>MODELLI AUMENTATI</u>

SOLUZIONE: NECESSITÀ DI STIMARE β E σ PRELIMINARMENTE (IN MODO APPROPRIATO)

PIÙ INFORMATIVO	MENO INFORMATIVO
$y_i = y_i^*$ SE $y_i^* > 0$	$D_i = 1$ SE $y_i^* > 0$
$y_i = 0$ SE $y_i^* \leq 0$	$= 0$ SE $y_i^* \leq 0$

PIÙ INFORMATIVO $\longrightarrow$ MENO INFORMATIVO

$\Downarrow$

PROBIT

PROBIT

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

DOVE $p_i \equiv \Phi\left(\frac{x_i \beta}{\sigma}\right)$

$$\log L = \sum_{i=1}^{N} \left[D_i \log \Phi\left(\frac{x_i \beta}{\sigma}\right) + (1 - D_i) \log\left(1 - \Phi\left(\frac{x_i \beta}{\sigma}\right)\right)\right]$$

MAX $\log L \Rightarrow \widehat{\frac{\beta}{\sigma}}$ MLE

$$\hat{\phi}_i \equiv \phi\left(x_i \widehat{\frac{\beta}{\sigma}}\right)$$

$$\hat{\Phi}_i \equiv \Phi\left(x_i \widehat{\frac{\beta}{\sigma}}\right)$$

MILLS RATIO ESTIMATE:

$$\frac{\hat{\phi}_i}{\hat{\Phi}_i}$$

TRONCAMENTO : $y_i^* = x_i\beta + \sigma \widehat{\dfrac{\phi_i}{\hat{\Phi}_i}} + \eta_i$

↳ OLS DI β E σ, CHE $\hat{\beta}$ E $\hat{\sigma}$, CONSISTENTI.

CENSURA : $y_i^* = \hat{\Phi}_i x_i\beta + \sigma \hat{\phi}_i + \eta_i$.

↳ OLS DI β E σ, CHE $\hat{\beta}$ E $\hat{\sigma}$, CONSISTENTI

COMMENTO

1) LO SCHEMA PROBIT (BINARIO : 0, 1) È MENO INFORMATIVO RISPETTO ALLO SCHEMA DI CENSURA $\Longrightarrow$ APPROCCIO PROBIT NON CONSENTE DI STIMARE SEPARATAMENTE β E σ — CONSENTE SOLO DI STIMARE IL RAPPORTO β/σ

2) PER IDENTIFICARE (STIMARE SEPARATAMENTE) β E σ

ABBIAMO BISOGNO DI MAGGIORI INFORMAZIONI, CIOÈ DI

SFRUTTARE I VERI VALORI DI y_i^* QUANDO $y_i^* > 0$ $\Rightarrow$

$\Rightarrow$ V. REGRESSIONI AUMENTATE, IN CUI LA VARIABILE

DIPENDENTE È OSSERVATA CON I SUOI VERI VALORI.

INFATTI, NEI MODELLI AUMENTATI È POSSIBILE

STIMARE SEPARATAMENTE β E σ.

2) MLE

CENSURA

LINEARE

$$y_i = y_i^* \quad \text{SE } y_i^* > 0 \quad \text{(SOLUZIONI INTERNE)}$$

PROBIT

$$y_i = 0 \quad \text{SE } y_i^* \leq 0 \quad \text{(SOLUZIONI D'ANGOLO)}$$

$$\log L = \underbrace{\sum_{i \in y>0} f(y_i)}_{\text{COMPONENTE LINEARE}} + \underbrace{\sum_{i \in y=0} \log(1-p_i)}_{\text{COMPONENTE PROBIT}}$$

DOVE $p_i = \Phi\left(\frac{x_i\beta}{\sigma}\right)$

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - x_i\beta)^2}{2\sigma^2}\right)$$

$$y_i^* = y_i \sim N(x_i\beta, \sigma^2) \quad [\varepsilon_i \sim N(0, \sigma^2)]$$

$$z_i = \frac{y_i - x_i\beta}{\sigma} \sim N(0,1)$$

$$\phi(z_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_i^2}{2}\right)$$

DENSITÀ DI NORMALE STANDARD

$$\Rightarrow f(y_i) = \frac{\phi(z_i)}{\sigma}$$

$$\log L = \sum_{i \in y>0} \log \frac{\phi(z_i)}{\sigma} + \sum_{i \in y=0} \log\left(1 - \Phi\left(\frac{x_i\beta}{\sigma}\right)\right)$$

$$\max_{\sigma,\, \frac{\beta}{\sigma}} \log L \Rightarrow \hat{\beta}_{MLE},\ \hat{\sigma}_{MLE} \text{ consistenti}$$

PER β E σ

TRONCAMENTO

$$\log L = \sum_{i \in y>0} \log f(y_i | y_i > 0)$$

N.B. $f(y_i) = Pr(y_i > 0)\, f(y_i | y_i > 0)$

$$\log L = \sum_{i \in y>0} \log \left(\frac{f(y_i)}{Pr(y_i > 0)} \right) =$$

$$\log L = \sum_{i \in y>0} \left[\log f(y_i) - \log \Pr(y_i > 0) \right]$$

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - x_i\beta)^2}{2\sigma^2}\right) \qquad \Phi\left(\frac{x_i\beta}{\sigma}\right)$$

$\max\limits_{\beta,\sigma} \log L \Rightarrow$ STIME DI MLE (CONSISTENTI) DI β <u>E</u> σ.

COMMENTO

1) OLS "CORRETTI" E MLE PRODUCONO STIME CONSISTENTI SEPARATE DI β E σ.

2) MLE PIÙ EFFICIENTE DI OLS "CORRETTI", IN QUANTO OLS "CORRETTI" SI BASA SU UNA PROCEDURA A 2 STADI (1° STADIO = MILLS RATIO ; 2° STADIO = REGRESSIONE AUMENTATA)