Principi di Econometria

lezione 15

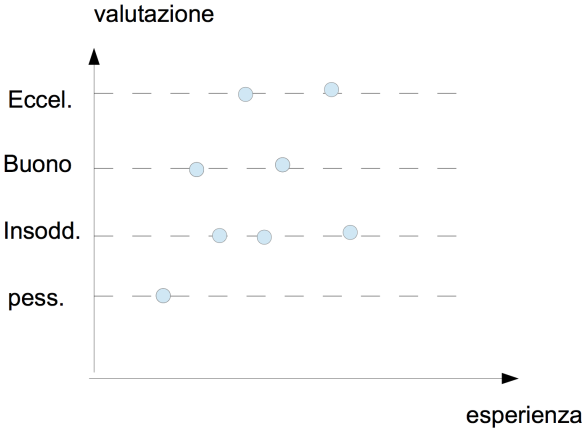
AA 2016-2017

Paolo Brunori

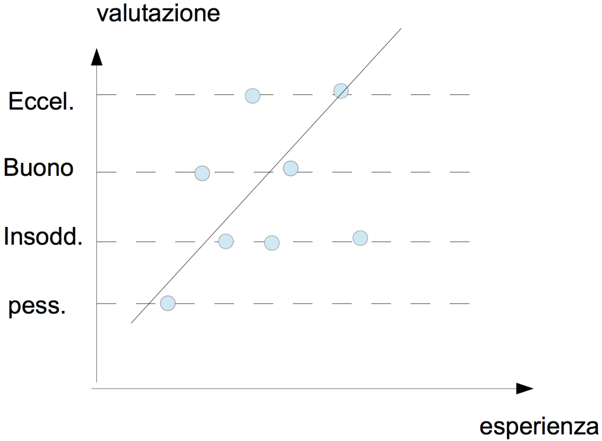
Variabili dipendenti ordinali

- quando la variabile dipendente è ordinale occorre trasformare la variabile
- immaginate di voler spiegare la variabile qualità del professore
- agli studenti è chiesto di dare un voto: pessimo, insoddisfacente, buono, eccellente.
- la nostra Y quindi assume 4 valori, la X consideriamo che siano gli anni di esperienza

variabile dipendente qualitativa



variabile dipendente qualitativa



Variabili dipendenti ordinali

- come si può interpretare il β_1?
- per farlo dobbiamo assegnare un valore cardinale ad una variabile ordinale
- ad esempio: Eccel.=4, Buono=3, Insodd.=2, pess.=1
- ma siamo sicuri che sia corretto assumere che la distanza in termini di qualità sia 1 fra tutte le categorie adiacenti?

Variabili dipendenti ordinali

- quando la variabile dipendente è ordinale occorre trasformare la variabile
- immaginate che le valutazioi possibile siano: pessimo, insoddisfacente,accettabile, buono, eccellente.
- in questo caso i valori sarebbero: Eccel.=5, Buono=4, accettabile=3 Insodd.=2, pess.=1
- la variazione fra insoddisfacente e buono è raddoppiata!

Variabili dipendenti ordinali

- come si risolve il problema?
- ipotesi 1 si impone arbitrariamente che la variabile abbia significato cardinale
- in molti casi è possibile che l'incertezza di β sia più elevata dell'errore dovuto a questa assunzione
- ipotesi 2 si 'dicotomizza' la variabile ovvero si crea una variabile con valore 0
- che assume valore 1 se la valutazione è superiore a una certa soglia
- in questo modo la regressione diventa un LPM che spiega la probabilità di ricevere una valutazione pari o superiore a una certa soglia

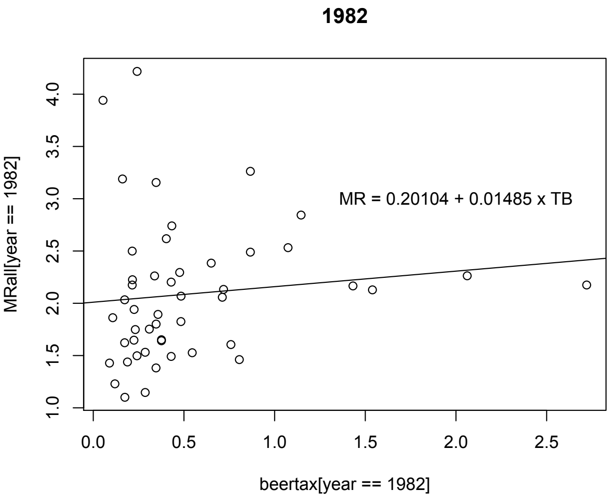
Dati longitudinali (panel)

- variabile dipendent: $Y_{i,t}$ $i = 1, ..., n,$ $t = 1, ..., T$
- variabili indipendenti: $X_{i,t}$
- variabili che non variano nel tempo ma fra osservazioni: Z_i
- variabili che non variano fra osservazioni ma variano nel tempo: S_t

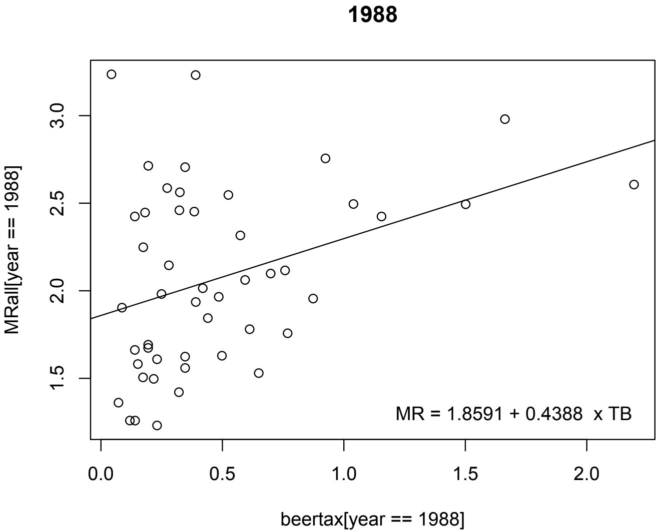
esempio del libro

- morti per il traffico e accise sull'alcool
- $y = morti/10.000$
- 48 stati Americani $n = 48$
- 7 anni (1982,..., 1988) $T = 7$
- panel bilanciato $7 \times 48 = 336$ ossevazioni

relazione di interesse: effetto causale delle imposte sulla mortalità 1982



relazione di interesse: effetto causale delle imposte sulla mortalità 1988



altri dati disponibili

- età degli automobilisti
- qualità delle strade
- densità automobilisti
- età minima per consumo alcol
- tasso di disoccupazione
- ...

distorsione da variabili omesse?

- è possibile che "intensità di traffico" sia una variabile omessa?
    1. determina Y?
    2. potrebbe essere correlato con X?
- sfruttando la struttura dei dati possiamo eliminare questa distorsione a patto che le variabili omesse non varino nel tempo

ci sono altre possibili variabili omesse?

- cosa determina la probabilità di un incidente mortale?

modello panel con 2 periodi

$$Y_{i,t} = \beta_0 + \beta_1 BT_{i,t} + \beta_2 Z_i + u_{i,t}$$

- $BT_{i,t}$ importo dell'accisa sulla birra
- Z_i attitudine culturale (non cambia nel tempo)

modello panel con 2 periodi

- ogni variazione dal 1988 al 1998 non può essere causata da Z_i
- se stimiamo un modello per ogni t

$$mort_{i,1982} = \beta_0 + \beta_1 BT_{i,1982} + \beta_2 Z_i + u_{i,1982}$$

$$mort_{i,1988} = \beta_0 + \beta_1 BT_{i,1988} + \beta_2 Z_i + u_{i,1988}$$

- assumiamo

$$E(u_{i,t}|BT_{i,t}, Z_i) = 0$$

modello panel con 2 periodi

- quindi possiamo riscrivere i due modelli:

$$MRT_{i,1988} - MRT_{i,1982} =$$

$$= \beta_0 + \beta_1 BT_{i,1988} + \beta_2 Z_i + u_{i,1988} - \beta_0 + \beta_1 BT_{i,1982} + \beta_2 Z_i + u_{i,1982}$$

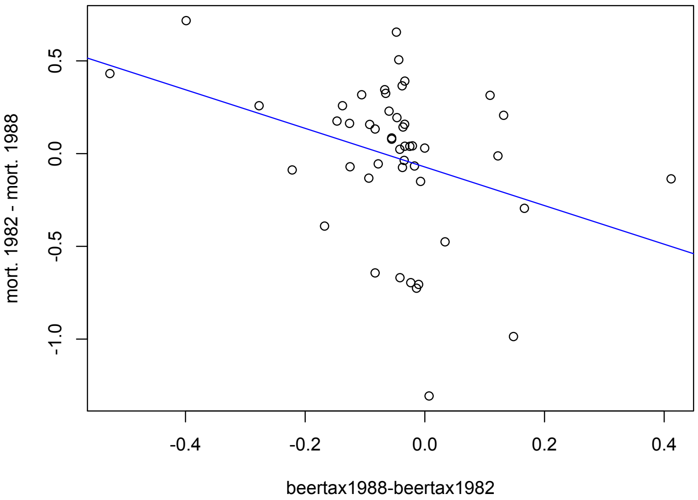
$$MRT_{i,1988} - MRT_{i,1982} = \beta_1 (BT_{i,1988} - BT_{i,1982}) + u_{i,1988} - u_{i,1982}$$

- $u_{i,t}$ non è correlato con i regressori $\rightarrow$ l'equazione può essere stimata con OLS

equazione in differenza

- $\Delta Y_i = \Delta X_i + u_i$
- anche se Z_i non è osservata non è una variabile omessa
- perchè?

OLS in differenza



regressione con effetti fissi

- quandto $T > 2$

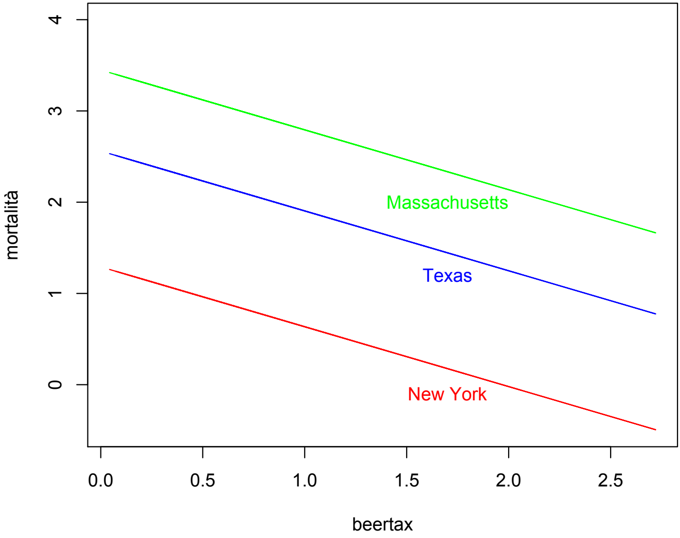
$$Y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \beta_2 Z_i + u_{i,t}$$

- dove Z_i è una variabile inosservata che varia fra stati ma non nel tempo
- vogliamo stimare β_1 tenendo costante Z

regressione con effetti fissi

- se definiamo $\alpha_i = \beta_0 + \beta_2 Z_i$ il modello da stimare diventa
$$Y_{i,t} = \beta_1 X_{i,t} + \alpha_i + u_{i,t}$$
- α_i è l'effetto fisso di stato: cambia fra stati ma non nel tempo
- α_i si introduce con una variabile binaria che assume valore 1 per lo stato $i - esimo$ e valore zero per gli altri

regressione con effetti fissi



ipotesi per l'inferenza

- 4+1 assunzioni per poter applicare l'inferenza OLS a questo modello
- gli errori devono essere incorrelati sia nel tempo sia tra le entità condizionatamente ai regressori:

$$cov(u_{i,t}, u_{i,s}|X_{i,t}.X_{i,s}.\alpha_i) = 0 \text{ per } t \neq s$$

- esempio: piogge intense in un anno aumentano gli incidenti in alcune zone
- ma questo non ha a che vedere con l'imposta sulla birra
- ed è incorrelato con le precipitazioni degli altri anni
- $cov(u_{i,t}, u_{i,s}|BT) = 0$ sono incorrelati
- la violazione dell'assunzione altera gli errori standard delle stime

effetti fissi come variazioni dalla media

- la stima OLS con effetti fissi pone problemi computazionali quando $n-1$ è grande
- generalmente i software procedono in due passi:

1. riscrivono i dati in termini di variazioni dalla media
2. stimano l'OLS

deviazioni dalle medie

- prendiamo la media da entrambe i lati del modello a effetti fissi

$$Y_{i,t} = \beta_1 X_{i,t} + \alpha_i + u_{i,t} \rightarrow \bar{Y}_i = \beta_1 \bar{X}_i + \alpha_i + \bar{u}_i$$

- dove

$$\bar{Y}_i = \frac{1}{T} \sum_{t=1}^{T} Y_{i,t}$$

$$\bar{X}_i = \frac{1}{T} \sum_{t=1}^{T} X_{i,t}$$

$$\bar{u}_i = \frac{1}{T} \sum_{t=1}^{T} u_{i,t}$$

deviazioni dalle medie

- sottraendo la media dal modello a effetti fissi

$$Y_{i,t} - \bar{Y}_i = \beta_1(X_{i,t} - \bar{X}_t) + (u_{i,t} - \bar{u}_i)$$

- se definiamo con la $\sim$ le deviazioni dalla media possiamo stimare:

$$\tilde{Y}_{i,t} = \beta_1 \tilde{X}_{i,t} + \tilde{u}_{i,t}$$

- dove β_1 stima esattamente lo stesso effetto stimato con $n-1$ variabili dicotomiche

stima con effetti fissi e effetti temporali omessi

- stimando il modello ad effetti fissi si trova $\beta_1 = -0.66$ nell'esempio del vostro libro
- come possiamo interpretare questo coefficiente?
- potrebbero esserci un problema di variabili omesse?
- se sia X_i è cresciuta nel tempo potrebbero essere cresciute anche altre variabili?

regressione ad effetti temporali

- se sia le accise sugli alcolici che gli standard di sicurezza sono cresciuti nel tempo β_1 sarà distorto
- il vero modello che spiega la mortalità è:

$$Y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \beta_2 Z_i + \beta_3 S_t + u_{i,t}$$

- dove S_t è una variabile che varia nel tempo ma è uniforme fra stati

regressione con effetti solo temporali

- la variabile S_t non è osservabile ma varia nel tempo uniformemente in tutti gli stati
- può quindi essere rimpiazzata applicando lo stesso metodo usato per variabili che variano nel tempo ma hanno un effetto uniforme in tutti gli stati
- siano $B1, ..., BT$ variabili dicotomiche che assumono valore 1 nel periodo $t - esimo$ e zero altrimenti

$$Y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \delta_2 B2_t +, ..., \delta_T BT_t + u_{i,t}$$

- come sono interpretabili i coefficienti δ_t?

regressione con effetti fissi e temporali

- nel modello ad effetti temporali sono stati esclusi gli effetti fissi di stato
- in realltà è possibile stimare un modello che comprenda entrambe le tipologie di effetti:

$$Y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \gamma_1 D2_i + ... + \gamma_n Dn_i + \delta_2 B2_t +, ..., \delta_T BT_t + u_{i,t}$$

- β_1 è stata stimata eliminando sia le distorsioni da variabili omesse fisse nel tempo che fisse in ogni stato
- quale variabilità è sfruttata?
- ci sono ancora possibilità di aver omesso qualcosa di rilevante?

l'effetto dell'accisa sugli alcolici sulla mortalità per incidente stradale

- la stima del modello sia a effetti fissi che temporali restituisce:

$$\hat{MR} = -0.64 + \text{effetti fissi di stato} + \text{effetti temporali}$$

Modelli per valutare l'eterogeneità dei coefficienti

	coefficiente	SE	t	$p - value$
β_{BT}	-0.6399	0.1973	-3.242	0.0013
$\beta_{ALABAMA}$	3.4595	0.3228	10.717	0.0000
$\beta_{ARIZONA}$	2.8210	0.1332	21.16	0.0000
...				
β_{1982}	0.0518	0.0396	1.307	0.1921
...				
β_{1987}	0.0009	0.0384	0.023	0.981

n.b. è possibile stimare tutti gli effetti fissi di paese e $k-1$ effetti fissi temporali (ma non tutti per via della multicollinearità perfetta)