

Gli effetti dell'attrito sulla stima della disuguaglianza in Italia

Federico Biagi, Anna Giraldo, Enrico Rettore

Università degli Studi di Padova

Modena, 30 gennaio 2009

Sommario

- 1 Motivazione
- 2 I dati
- 3 Prime evidenze empiriche
- 4 Metodologia
- 5 Risutati
- 6 Conclusioni

Motivazione

- L'indagine sui bilanci delle famiglie della Banca d'Italia (IBI) raccoglie informazioni su redditi e ricchezza delle famiglie italiane. È l'unica fonte di dati che consente di studiare l'andamento della disuguaglianza dei redditi nel tempo.
- La struttura dell'indagine è tale per cui ad ogni occasione di indagine vengono re-intervistate alcune delle famiglie già intervistate nelle occasioni precedenti e si aggiungono di volta in volta nuove famiglie.
- La componente longitudinale dell'indagine è affetta da un notevole attrito (progressiva perdita di unità campionarie dovuta a rifiuti o irreperibilità).

Motivazione

- L'attrito può essere un problema se le famiglie che permangono nel campione sono sistematicamente diverse da quelle che lo lasciano e se le famiglie nuove entranti non sono tali da rimpiazzare adeguatamente le famiglie uscite. Ciò può dare luogo ad una progressiva perdita di rappresentatività del campione.
- Accertato che la distribuzione del reddito varia con il numero di occasioni nell'indagine – le famiglie che rimangono sono tendenzialmente più ricche e meno disuguali di quelle che lasciano – studiamo se l'attrito che caratterizza IBI è tale da produrre distorsioni nella stima dell'andamento nel tempo della disuguaglianza.

L'indagine Banca d'Italia

- La componente longitudinale nell'IBI è stata introdotta nel 1989; la struttura è quella di uno split-panel: in ogni occasione d'indagine alcune famiglie entrate nel campione all'occasione precedente vengono reintervistate. Le famiglie reintervistate escono dal campione solo a causa dell'attrito.
- In ogni anno di indagine il campione è composto da: famiglie intervistate per la prima volta – la componente “fresca” – e che in parte verranno reintervistate nelle occasioni future, e famiglie intervistate più volte in passato – la componente longitudinale. Nella componente longitudinale i campioni panel sono disomogenei quanto ad anno di ingresso e durata della permanenza nell'indagine.

I campioni panel

Nel periodo 1989-2006, la componente longitudinale è formata da 36 panel bilanciati mutuamente esclusivi di diversa lunghezza che coinvolgono circa 11.500 famiglie e che vengono generati congiuntamente dal disegno split-panel dell'IBI e dall'attrito.

La numerosità campionaria

Tab. 1. Dimensione del campione dell'IBI secondo l'anno di ingresso nell'indagine e l'anno di riferimento dell'indagine, 1989-2006

Anno di ingresso nell'indagine	Anno di riferimento dell'indagine								
	1989	1991	1993	1995	1998	2000	2002	2004	2006
1989	8.274	2.187	1.050	827	544	404	307	230	189
1991		6.001	2.420	1.752	1.169	832	613	464	393
1993			4.619	1.066	583	399	270	199	157
1995				4.490	373	245	177	117	101
1998					4.478	1.993	1.224	845	636
2000						4.128	1.014	667	475
2002							4.406	1.082	672
2004								4.408	1.334
2006									3.811
N. famiglie	8.274	8.188	8.089	8.135	7.147	8.001	8.011	8.012	7.768
% di famiglie presenti nell'indagine in almeno 5 occasioni	6,5	16,8	20,3	21,7	33,6	32,4*	32,5*	33,9*	36,5*

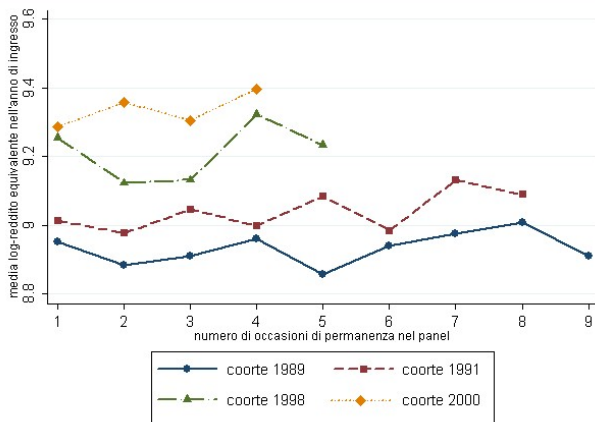
* Il numero di famiglie che rimarranno nell'indagine per almeno 5 occasioni tra quelle entrate nell'anno t=2000, 2002, 2004, 2006, non è ancora osservabile. E' stato stimato ipotizzando che i tassi di permanenza siano dello stesso ordine di grandezza di quelli osservati per le famiglie entrate negli anni precedenti

Fonte: Banca d'Italia [2008].

Prime evidenze empiriche

- Come evidenziato in tabella 1 aumenta nel tempo la percentuale di famiglie presenti nell'indagine in almeno in 5 occasioni. Dal 6,5% nel 1989 al 36,5% (dato stimato) nel 2006.
- La media del logaritmo del reddito familiare equivalente nell'anno di ingresso nel panel per durata della permanenza nel panel, a meno della variabilità campionaria tende a crescere al crescere della durata della permanenza.

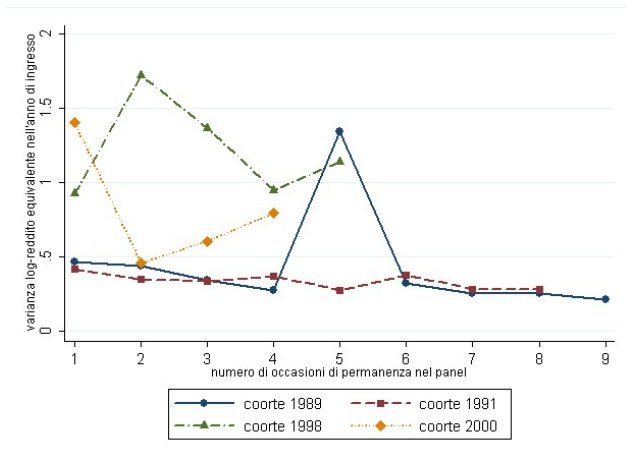
Grafico medie



Prime evidenze empiriche

- La varianza del logaritmo del reddito familiare equivalente degli stessi gruppi, sempre a meno della variabilità campionaria, tende a diminuire al crescere della durata della permanenza.

Grafico varianze



Prime evidenze empiriche

- Questi risultati evidenziano che tendono a rimanere più a lungo nel panel le famiglie con redditi relativamente più alti e meno diseguali tra loro.
- Questa evidenza è in linea con quanto trovato in Giraldo *et al.* (2007): considerando gli indici di diffusione della povertà il numero di occasioni nelle quali una famiglia permane nel panel è correlato negativamente con la probabilità di sperimentare un episodio di povertà.

Metodologia

- Per analizzare gli effetti dell'attrito sulla misura di disuguaglianza, consideriamo gruppi di unità campionarie definiti secondo l'anno di ingresso nel campione e il numero di occasioni di indagini alle quali l'unità campionaria prende parte. Le famiglie incluse nel campione in occasione dell'indagine relativa all'anno solare t differiscono infatti tra loro sia secondo il loro anno di ingresso nell'indagine, sia secondo il loro anno di uscita.

Metodologia

- Ai nostri fini risulta utile lavorare con indici di disuguaglianza perfettamente scomponibili. Consideriamo indici che appartengono alla classe Generalized Entropy, e in particolare sull'indice $GE(0)$, definito dalla seguente espressione:

$$GE(0) = \frac{1}{N} \sum_{i=1}^N \ln \frac{\mu}{y_i}$$

Metodologia

- Detto $GE(0)_{tcl}$ l'indice di disuguaglianza relativo all'anno solare t per il gruppo di soggetti entrati nell'indagine all'occasione c e rimasti nell'indagine per l occasioni successive, l'indice utilizzato può essere scomposto nel seguente modo:

$$GE(0)_t = \sum_{c,l} \left(\frac{N_{tcl}}{N_t} \right) GE(0)_{tcl} + \sum_{c,l} \left(\frac{N_{tcl}}{N_t} \right) \left(\ln \frac{\mu_t}{\mu_{tcl}} \right)$$

Metodologia

- Con riferimento ad una specifica occasione di indagine calcoliamo sia la disuguaglianza all'interno dei vari gruppi presenti in quella occasione – nel seguito componente within, sia quella tra gruppi - nel seguito componente between. Tale partizione risulta particolarmente utile nella fase di costruzione del profilo della disuguaglianza che si sarebbe osservata in assenza di distorsione da attrito, dal momento che viene alimentata solo dalla componente within.

Metodologia

- L'obiettivo della nostra analisi è stabilire se la disuguaglianza all'interno dei diversi gruppi così definiti varia con il numero di occasioni nell'indagine e, nel caso, se il mancato controllo di tale dipendenza distorce la stima dell'andamento nel tempo della disuguaglianza. Lo strumento al quale ricorriamo è la seguente regressione:

$$GE(0)_{tcl} = \vartheta_t + \gamma_c + \delta_l + \epsilon_{tcl}$$

- La sequenza dei coefficienti ϑ_t fornisce l'andamento nel tempo della disuguaglianza che si osserverebbe (a meno delle oscillazioni casuali dovute a ϵ_{tcl}) se il momento dell'ingresso e il numero delle occasioni di permanenza nell'indagine fossero irrilevanti.

Metodologia

- Accanto al modello di regressione visto sopra stimiamo anche il modello ridotto:

$$GE(0)_{tcl} = \vartheta_t + \nu_{tcl}$$

- Il confronto dell'andamento nel tempo delle stime di ϑ_t ottenute dai due modelli rivela se il mancato controllo per gli effetti della durata della permanenza nell'indagine e per gli effetti dell'anno di ingresso comporta una distorsione nella stima dell'andamento nel tempo della disuguaglianza.

Metodologia

Si noti infine che per tutti i soggetti ancora presenti nell'indagine all'ultima occasione, il 2006 nel nostro caso, la durata di permanenza nell'indagine è osservata in modo incompleto (non sappiamo quando usciranno dall'indagine). Risolviamo il problema attribuendo le probabilità (funzione di sopravvivenza) alla permanenze nell'indagine basate sulle osservazioni per le indagini precedenti. Ciò consente di utilizzare anche i gruppi che non sono osservati fino all'istante di uscita dal panel.

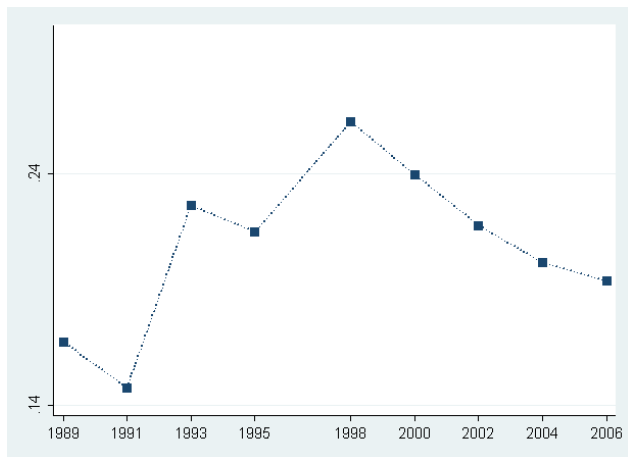
Risultati

La misura del reddito utilizzata in questo lavoro è il reddito familiare complessivo equivalente (= redditi da lavoro e assimilati, redditi da capitale e altri redditi in natura, percepiti dai membri dell'unità familiare). La scala di equivalenza utilizzata è la scala OCSE modificata che attribuisce un valore pari a 1 al primo adulto, 0,5 a ogni altro adulto (14 e più anni), 0,3 ai minori di 14 anni. Ad esempio una coppia senza figli avrà coefficiente 1,5 mentre una coppia con due figli minori 2,1.

Risultati

- Riportiamo per primo l'andamento dell'indice $GE(0)$ calcolato alle varie occasioni di indagine facendo uso dei pesi di riporto all'universo del campione sezionale forniti dalla Banca d'Italia. Il profilo che ne risulta presenta un brusco aumento negli anni 1993-1995, un ulteriore aumento nel 1998 seguito da una fase di graduale moderata riduzione.

Risultati



Risultati

- Riportiamo poi i risultati relativi alla specificazione in cui la disuguaglianza dei diversi gruppi è regredita sulle sole dummies temporali. In questo caso non controlliamo per gli effetti dell'occasione di ingresso nell'indagine e del numero di periodi trascorsi nell'indagine. Tale stima costituisce quindi un utile benchmark rispetto ai passaggi successivi.

Risultati



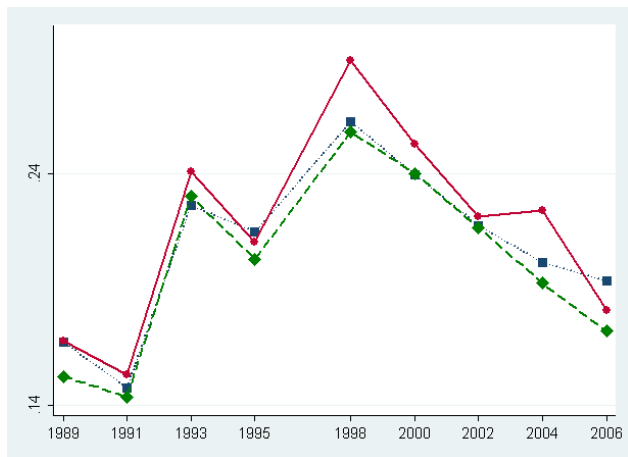
Risultati

- Se esaminiamo i risultati della regressione con le tre componenti troviamo che gli effetti dell'anno di ingresso nell'indagine sono trascurabili (a parte 1993). La stima degli effetti del numero di occasioni di indagine conferma la nostre ipotesi quanto agli effetti dell'attrito sulla misura della disuguaglianza: a parità di occasione di ingresso nell'indagine e di periodo di riferimento, le famiglie presenti nell'indagine almeno cinque occasioni sono meno diseguali tra loro delle famiglie presenti nell'indagine al più quattro occasioni. Tale differenza risulta statisticamente significativa in corrispondenza a $l=5, 7, 8$.

Risultati

- A partire da questa regressione siamo in grado di calcolare l'andamento nel tempo della disuguaglianza che si sarebbe osservato in assenza di attrito, vale a dire a parità di composizione dei campioni sezionali secondo il numero di occasioni di presenza delle varie unit campionarie.

Risultati



Conclusioni

I principali risultati della nostra analisi sono così riassumibili:

- 1** la disuguaglianza nei redditi personali equivalenti risulta marcatamente decrescente nel numero di occasioni di indagine alle quali le famiglie prendono parte. Ad esempio, i componenti delle famiglie rimaste nel campione 9 occasioni presentano un valore dell'indice inferiore di 0,0866 rispetto a quelle uscite dal campione dopo la prima occasione, per un indice che negli anni considerati prende valori nell'intervallo (0,15; 0,29).
- 2** con il trascorrere del tempo i campioni sezionali dell'indagine presentano una quota via via crescente di famiglie sopravvissute, o destinate a sopravvivere, più a lungo nell'indagine.

Conclusioni

Come conseguenza di [1] e [2], il campione dà luogo ad una sistematica sottostima della disuguaglianza. Il pattern temporale della disuguaglianza negli anni dal 1989 al 2006 risultante dal metodo che proponiamo risulta diverso da quello che si osserva facendo uso dei pesi di riporto all'universo dei campioni sezionali forniti dalla Banca d'Italia. La correzione apportata all'indice di disuguaglianza per tenere conto degli effetti dell'attrito dà luogo a valori dell'indice sistematicamente più alti (con l'eccezione del 2006), in alcuni anni in modo sensibile. La conclusione che ci sembra di poter trarre è che i pesi per il riporto all'universo dei campioni sezionali non correggono pienamente gli effetti dell'attrito.