

GLI EFFETTI DELL'ATTRITION NELL'INDAGINE 'BANCA D'ITALIA' SULLA
STIMA DELL'ANDAMENTO DELLA DISUGUAGLIANZA IN ITALIA NEGLI ANNI
DAL 1989 AL 2006

FEDERICO BIAGI, ANNA GIRALDO, ENRICO RETTORE

pubblicazione internet realizzata con contributo della

COMPAGNIA
di San Paolo

società italiana di economia pubblica

dipartimento di economia pubblica e territoriale – università di pavia

FEDERICO BIAGI, ANNA GIRALDO, ENRICO RETTORE

GLI EFFETTI DELL'ATTRITION NELL'INDAGINE 'BANCA
D'ITALIA' SULLA STIMA DELL'ANDAMENTO DELLA
DISUGUAGLIANZA IN ITALIA NEGLI ANNI DAL 1989 AL 2006

NOTA: IL TESTO È STATO SCRITTO NELL'AMBITO DELLA RICERCA
"OSSERVATORIO DISUGUAGLIANZE SOCIALI" E SARÀ A BREVE
PUBBLICATO DALLA CASA EDITRICE "IL MULINO" ALL'INTERNO DEL
VOLUME INTITOLATO "2° RAPPORTO SULLE DISUGUAGLIANZE"
(ATTUALMENTE IN FASE DI REVISIONE)

1. *Introduzione*

L'Indagine sui bilanci delle famiglie della Banca d'Italia (nel seguito IBI) raccoglie informazioni su redditi e ricchezza delle famiglie italiane a cadenza biennale. Si tratta dell'unica indagine che consente di ottenere l'andamento nel tempo della disuguaglianza nei redditi tra le famiglie italiane con riferimento ad un lungo arco di tempo, ad oggi dal 1989 al 2006. La struttura dell'indagine è tale per cui ad ogni occasione di indagine vengono re-intervistate alcune delle famiglie già intervistate nelle occasioni precedenti alle quali si aggiungono di volta in volta nuove famiglie.

Il processo che determina l'uscita dal campione delle famiglie è solo in parte sotto il controllo del gestore dell'indagine, per il resto essendo il risultato della libera scelta delle famiglie di continuare o meno la collaborazione. Tale fenomeno, che in letteratura prende il nome di *attrition*, può dare luogo ad una progressiva perdita della rappresentatività del campione. Ciò accade 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.

In questo lavoro studiamo se la distribuzione del reddito per le famiglie che rimangono nel campione è sistematicamente diversa dalla corrispondente distribuzione per le famiglie che lo lasciano. Accertato che la distribuzione del reddito varia con il numero di occasioni nell'indagine, - come vedremo, le famiglie che rimangono sono tendenzialmente più ricche e meno disuguali di quelle che lasciano - studiamo se il processo di *attrition* che caratterizza IBI è tale da produrre distorsioni nella stima dell'andamento nel tempo della disuguaglianza.

L'organizzazione del lavoro è la seguente. Nella sez. 2 presentiamo brevemente le caratteristiche dell'indagine, la consistenza dei flussi di unità campionarie in uscita e in ingresso alle varie occasioni di indagine e le prime evidenze descrittive degli effetti dell'*attrition* sulla stima della media e della

dispersione dei redditi familiari. Nella sez. 3 descriviamo l'indice di disuguaglianza utilizzato nello studio e le sue principali proprietà. Nella sez. 4 (e nell'Appendice per maggiori dettagli) presentiamo la strategia utilizzata per identificare gli effetti dell'*attrition* sulla stima della distribuzione dei redditi e del suo andamento nel tempo. La sez. 5 presenta i risultati dell'analisi. La sez. 6 conclude.

2. I termini del problema

L'Indagine sui bilanci delle famiglie¹ della Banca d'Italia raccoglie informazioni su redditi e ricchezza delle famiglie italiane a cadenza biennale². Dal 1989 nell'indagine è stata introdotta una componente longitudinale; la struttura è quella di uno *split-panel*: in ogni occasione d'indagine alcune famiglie entrate nel campione all'occasione precedente vengono reintervistate; una volta entrate a far parte del campione panel le famiglie vi escono solo a causa dell'attrito³. In ogni anno di indagine il campione è dunque 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.

Nello specifico, 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. Se consideriamo ad esempio l'indagine del 1989 essa comprende 8.274 famiglie intervistate per la prima volta⁴; nel 1991 solo 2.187 vengono ricontattate – il numero dei ricontatti dopo la prima occasione d'indagine dipende dal disegno dell'indagine e da considerazioni di efficienza del piano di campionamento. Nel 2006, dopo 17 anni e nove occasioni d'indagine, 189 famiglie sono ancora nell'indagine; queste 189 famiglie rappresentano il panel a nove occasioni. In totale, dall'indagine 1989 si generano 8 panel mutuamente esclusivi: il panel a due occasioni formato dalle famiglie che escono dal

¹ Per una sua accurata descrizione, vedi Brandolini [1999] e Banca d'Italia [anni vari].

² La cadenza dell'indagine è biennale dal 1987, con l'eccezione dell'indagine 1998 avvenuta a tre anni dalla precedente.

³ Con attrito si intende la progressiva perdita di unità campionarie di un'indagine longitudinale dovuta a rifiuti o irreperibilità.

⁴ La componente longitudinale della IBI parte nel 1987. In questo e in lavori precedenti si è scelto di non considerare la componente longitudinale iniziata nel 1987 e di considerare il campione 1989 come un campione “fresco”. Questo da un lato per la ridotta numerosità del campione longitudinale (la frazione di famiglie campionate nel 1987 e reintervistate nel 1989 sono circa il 15% del totale e il numero di famiglie della coorte 1987 che sopravvive fino al 2006 è pari a 30 su 189) e dall'altra perché il trattamento delle famiglie panel non sembrava, nelle fasi iniziali, condotto in maniera accurata.

campione dopo l'indagine del 1991; il panel a tre occasioni formato dalle famiglie che escono dal campione dopo l'indagine del 1993; e così via fino al panel a nove occasioni delle famiglie ancora presenti nel campione dell'indagine del 2006. In maniera del tutto analoga, dall'insieme di famiglie che entrano nell'IBI nel 1991 otteniamo sette panel (dal panel a due occasioni al panel a otto occasioni); e così via fino all'insieme di famiglie che entrano nell'indagine nel 2004, per le quali abbiamo un ultimo panel a due occasioni.

La progressiva riduzione della dimensione dell'insieme degli entrati in un dato anno, parte per disegno, parte a causa dell'attrito, è documentata in tabella 1 per i vari anni dal 1989 al 2006.

INSERIRE QUI TABELLA 1

La semplice riduzione di dimensione dei vari insiemi di entrati non sarebbe un problema se i soggetti che abbandonano l'indagine avessero le stesse caratteristiche dei soggetti che vi restano, ma diverse evidenze empiriche segnalano che il processo di attrition induce forte distorsione⁵. In figura 1 riportiamo l'andamento della media del logaritmo del reddito familiare per anno di ingresso nel panel e per durata della permanenza nel panel stesso. A meno della variabilità dovuta alla talvolta piccola numerosità campionaria, osserviamo che al crescere della durata della permanenza nel panel, il log-reddito medio aumenta. Di converso se andiamo a considerare in figura 2 la varianza del logaritmo del reddito degli stessi gruppi, sempre a meno della variabilità campionaria, questa tende a diminuire all'aumentare della permanenza nel panel.

INSERIRE QUI FIGURA 1 e FIGURA 2

Un comportamento del genere evidenzia il fatto che le famiglie che permangono nel panel più a lungo sono famiglie selezionate, con un reddito medio tendenzialmente più alto. Questi risultati di carattere descrittivo sono in linea con quanto trovato in Giraldo et al. [2007]: le famiglie meno esposte al rischio di povertà sopravvivono più a lungo nel campione. Infatti, considerando gli indici di diffusione della povertà (*head-count ratio*) in alcuni selezionati panel di diversa lunghezza gli autori notano come il numero di occasioni per le quali una famiglia permane nel panel è correlato con la probabilità di sperimentare un episodio di povertà. Per le famiglie che appartengono ai panel più lunghi, infatti, la probabilità di sperimentare un episodio di povertà è decisamente inferiore a quella delle famiglie che appartengono ai panel più corti.

È evidente che studiare la disuguaglianza dei redditi utilizzando dati

⁵ Per un'analisi del processo di attrito che affligge la componente panel dell'indagine relativa al periodo 1989-1995, vedi Giraldo, Rettore e Trivellato [2001]

sezionali senza considerare che il campione è formato da famiglie intervistate per la prima volta, per la seconda volta, ... per la nona volta e soprattutto senza tener conto che le famiglie che appartengono ai panel più lunghi sono quelle meno povere, fa sì che gli indici di disuguaglianza possano essere distorti.

Se poi si volesse studiare l'evoluzione nel tempo della disuguaglianza in gruppi omogenei, quali sono le famiglie con stesso anno di ingresso e di uscita dal campione, dovremmo tener conto che l'attrito seleziona fortemente il gruppo e quindi le misure di disuguaglianza possono risultare sottostimate.

Di qui la necessità di costruire misure di disuguaglianza che tengano conto degli effetti distorsivi della selezione non casuale delle famiglie causata dall'attrito.

3. Cenni sugli indici di disuguaglianza usati

Ai fini del nostro lavoro risulta particolarmente utile lavorare con indici di disuguaglianza che siano perfettamente scomponibili. Infatti, dato l'obiettivo di verificare l'effetto della selezione non casuale del campione sulle misure aggregate di disuguaglianza, è fondamentale poter distinguere i soggetti a seconda dell'anno in cui essi sono entrati nel campione e del numero di occasioni in cui sono osservati. Attraverso un indice esattamente scomponibile siamo in grado di cogliere in modo preciso la relazione tra la disuguaglianza complessiva e la disuguaglianza all'interno dei vari gruppi, definiti secondo l'anno di entrata nel campione e l'anno di uscita.

La nostra attenzione si è quindi concentrata sugli indici che appartengono alla classe Generalized Entropy, definiti dalla seguente espressione:

$$GE(\alpha) = \frac{1}{\alpha(\alpha-1)} \left[\sum_{i=1}^n w_i \left(\frac{y_i}{\mu} \right)^\alpha - 1 \right] \quad (1)$$

dove α è un parametro (posto tipicamente uguale a 0, 1 o 2), w_i è il peso dell'individuo i -esimo, y_i è il suo reddito equivalente e μ è il reddito medio della popolazione considerata (l'indice è calcolato in relazione ad un dato istante di tempo, nel nostro caso l'anno). Si noti che, come mostrato da Cowell and Mercader-Prats (1999), Coulter *et al.* (1992a) e Biagi e Casalone (2008), la relazione tra il valore degli indici che appartengono a questa classe e le scale di equivalenza non è lineare (dipende dal valore di α e dalla

distribuzione del reddito e delle caratteristiche utilizzate per calcolare la scala di equivalenza⁶).

Nel nostro lavoro abbiamo scelto di applicare la scala di equivalenza OECD modificata (che varia in modo non lineare al variare del numero di componenti della famiglia) e di concentrarci sull'indice $GE(0)$, quello cioè ottenuto ponendo il parametro α pari a zero, definito dalla seguente espressione:

$$GE(0) = \frac{1}{N} \sum_{i=1}^N \ln \frac{\mu}{y_i} \quad . \quad (2)$$

Questo indice, sensibile a variazioni concentrate nella parte destra della distribuzione, è di fatto proporzionale al coefficiente di variazione del reddito equivalente. Infatti, sviluppando $\ln(\mu) - \ln(y)$ in serie di Taylor fino al secondo ordine, rispetto a y in un intorno di μ , si ottiene:

$$\ln \mu - \ln y_i \approx \frac{\mu - y_i}{\mu} + 0,5 \left(\frac{\mu - y_i}{\mu} \right)^2 \quad (3)$$

da cui risulta:

$$GE(0) \approx 0,5 \sum_{i=1}^N \frac{1}{N} \left(\frac{\mu - y_i}{\mu} \right)^2 \quad (4)$$

L'indice $GE(0)$ possiede la caratteristica fondamentale di essere esattamente scomponibile e i risultati della scomposizione sono interpretabili in modo molto chiaro. In particolare, esso è scomponibile nelle due componenti *within* e *between*. Infatti, se dividiamo la popolazione considerata al tempo t in G gruppi tra loro mutuamente esclusivi, possiamo scrivere:

$$GE(0)_t = \sum_{g=1}^G \left(\frac{N_{gt}}{N_t} \right) GE(0)_{gt} + \sum_{g=1}^G \left(\frac{N_{gt}}{N_t} \right) \left(\ln \frac{\mu_t}{\mu_{gt}} \right) = GE(0)_{Wt} + GE(0)_{Bt} \quad (5)$$

⁶ Coulter *et al.* (1992) dimostrano come, per valori di α nell'intervallo (-1,2) esista tipicamente una relazione a U tra il valore dell'indice di disuguaglianza e il valore del parametro utilizzato per determinare la scala di equivalenza.

dove N_{gt} è il numero di soggetti che appartengono al gruppo g al tempo t , $GE(0)_{gt}$ rappresenta il valore dell'indice di disuguaglianza $GE(0)$ per il gruppo g al tempo t , e μ_t e μ_{gt} sono – rispettivamente – il valore del reddito (equivalente) medio e quello medio in riferimento al solo gruppo g ($g=1\dots G$) al tempo t .

4. L'identificazione congiunta degli effetti dell'attrito e dell'andamento temporale della disuguaglianza

I soggetti inclusi nel campione in occasione dell'intervista sui redditi dell'anno solare t differiscono tra loro sia secondo il loro anno di ingresso nell'indagine, sia secondo il loro anno di uscita. Ad esempio, tra gli intervistati per l'anno 1995 ve ne sono alcuni entrati nell'indagine in occasione dell'intervista per l'anno 1989, altri per l'anno 1991, altri ancora per l'anno 1993 e infine alcuni alla loro prima intervista. Allo stesso modo, alcuni lasciano l'indagine dopo la prima intervista, altri dopo l'intervista relativa al 1998 e così via fino al 2006.

Alla luce del potenziale ruolo dell'attrito sulla misura di disuguaglianza, nel nostro lavoro 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. Con riferimento ad una specifica occasione di indagine siamo quindi in grado di calcolare sia la disuguaglianza all'interno dei vari gruppi presenti in quella occasione (che – attraverso gli opportuni pesi – genera la componente *within* dell'indice aggregato) sia quella tra gruppi (che ne determina invece la componente *between*). Tale partizione risulta particolarmente utile nella fase di costruzione del profilo della disuguaglianza che si sarebbe osservata in assenza di distorsione da *attrition*, dal momento che – come dimostriamo nel paragrafo 5 – quest'ultima viene alimentata solo dalla componente *within*. Proprio per questa ragione, nel nostro lavoro empirico ci concentriamo sugli elementi che caratterizzano la disuguaglianza all'interno dei vari gruppi.

Quindi, data la metodologia proposta, il pedice g con cui vengono identificati i gruppi è composto da due riferimenti che identificano in modo univoco i diversi soggetti secondo l'anno di entrata e la durata della permanenza nel campione. Ciò implica che l'indice che misura la disuguaglianza all'interno di un dato gruppo g al tempo t – $GE(0)_{gt}$ – può essere riscritto come $GE(0)_{ict}$. Questo è 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'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:

$$GE(0)_{icl} = \gamma_c + \mathcal{G}_t + \delta_l + \varepsilon_{icl} \quad (6)$$

La sequenza dei coefficienti \mathcal{G}_t fornisce l'andamento nel tempo della disuguaglianza che si osserverebbe (a meno delle oscillazioni casuali dovute a ε_{icl}) se il momento dell'ingresso e il numero delle occasioni di permanenza nell'indagine fossero irrilevanti. A questo andamento si sovrappone l'effetto della durata della permanenza nell'indagine: se, come le prime evidenze mostrate nel paragrafo 2 lasciano pensare, il processo di selezione dà progressivamente luogo a gruppi di sopravvissuti nell'indagine sempre più omogenei, la sequenza degli effetti δ_l dovrebbe risultare tale da ridurre progressivamente la disuguaglianza al crescere di l . Nel modello di regressione (6) sono infine inclusi degli effetti di anno di ingresso, γ_c . Tali effetti servono a tenere conto del fatto che, ad esempio, i responsabili dell'indagine potrebbero selezionare il campione dei nuovi entranti in modo tale da compensare le distorsioni causate dalla selezione non casuale indotta dal processo di attrito.

Il fatto che gli effetti della durata della permanenza nell'indagine, i coefficienti δ_l , e gli effetti dell'anno di ingresso, i coefficienti γ_c , risultino diversi da zero di per sé non implica che l'andamento nel tempo osservato per la disuguaglianza risulti deformato rispetto al 'vero' andamento. Può infatti darsi il caso che il campione dei soggetti intervistati sui loro redditi nell'anno t sebbene selezionato non casualmente ad opera del processo di attrito e del processo di selezione dei nuovi entranti riproduca correttamente la disuguaglianza nella popolazione. Ciò accade ad esempio, se il campione dei nuovi entranti viene selezionato in modo tale da compensare le distorsioni generate dal processo di attrito. Per verificare se questo è il caso, accanto al modello di regressione (6) stimiamo anche il modello ridotto:

$$GE(0)_{icl} = \mathcal{G}_t + u_{icl}. \quad (7)$$

Il confronto dell'andamento nel tempo delle stime di \mathcal{G}_t , rispettivamente, dal modello (6) e dal modello (7) 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.

L'identificazione degli effetti della durata di permanenza nell'indagine è garantita dal fatto che a parità di anno di riferimento t e di anno di ingresso

c sono presenti nel campione soggetti che presentano diversa durata nell'indagine. Analogamente il fatto che a parità di anno di riferimento t e di durata di permanenza nell'indagine l siano presenti nel campione soggetti che presentano diverso anno di ingresso nell'indagine garantisce l'identificabilità degli effetti di anno di ingresso.

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. Ad esempio, il gruppo costituito dagli entrati nel 1998 ancora presenti nel 2006 include soggetti che permangono nell'indagine esattamente 5 occasioni (sono quelli che lasciano l'indagine subito dopo l'intervista del 2006), esattamente 6 occasioni (sono quelli che lasciano l'indagine subito dopo l'intervista del 2008), esattamente 7 occasioni (sono quelli che lasciano l'indagine subito dopo l'intervista del 2010), Tali gruppi sono tra loro indistinguibili. Nell'Appendice mostriamo nei dettagli come teniamo conto di questo problema nella stima dei parametri delle regressioni (6) e (7).

5. I risultati dell'analisi

Le analisi sono state condotte utilizzando i dati dell'Indagine Banca d'Italia sui bilanci delle famiglie italiane, dal 1989 al 2006; oltre a nove campioni sezionali, composti da famiglie intervistate una volta sola tra il 1989 e il 2006, come detto in sezione 2, si distinguono diversi sottocampioni mutuamente esclusivi: un campione longitudinale a nove occasioni (1989-1991-...-2006), due campioni longitudinali a otto occasioni, tre campioni longitudinali a sette occasioni e così via fino a otto campioni longitudinali a 2 occasioni (il primo è il panel 1989-1991 e l'ultimo il panel 2004-2006) per un totale di 36 campioni panel mutuamente esclusivi.

La misura del reddito utilizzata in questo lavoro è il reddito familiare complessivo equivalente. Il reddito familiare complessivo comprende redditi da lavoro e assimilati, redditi da capitale e altri redditi in natura, percepiti dai membri dell'unità familiare. La scala di equivalenza utilizzata (necessaria per rendere confrontabili tra loro i redditi complessivi di cui dispongono famiglie di diversa dimensione) è quella 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).

La prima cosa che possiamo notare è l'andamento dell'indice di disuguaglianza aggregato $GE(0)$. L'andamento dell'indice segnala un profilo crescente sino al 1998, cui segue una fase di moderata riduzione (si veda la tab. 3, colonna 1).

Se, seguiamo l'equazione 5), che scompone l'indice nelle sue componenti *between* e *within*, notiamo che i fattori che influenzano l'andamento della componente *within* – oltre alla numerosità relativa dei vari gruppi – sono da

ricercare nei valori di $GE(0)_{ict}$, cioè la disuguaglianza all'interno di ciascun gruppo. Per quanto riguarda invece la componente *between*, a parte anche in questo caso la numerosità relativa dei gruppi, ciò che rileva sono i valori di μ_{ct} , cioè il reddito medio equivalente calcolato in riferimento ad un dato gruppo, relativamente al reddito medio μ_t . La nostra attenzione si concentra quindi su $GE(0)_{ict}$ e μ_{ict} . Per entrambe le variabili (che abbiamo calcolato in relazione a ciascun gruppo) proponiamo una scomposizione del tipo anno, coorte, durata.

Per quanto concerne la misura di disuguaglianza, riprendendo la (6) notiamo come γ_c catturi gli elementi della disuguaglianza comuni a tutti i gruppi che appartengono alla stessa coorte, nel senso che sono entrati nel campione nello stesso anno⁷, \mathcal{A}_t rappresenti invece gli elementi comuni a tutti coloro che sono osservati nello stesso anno⁸ t (ed è quindi un effetto tempo), mentre δ_t rappresenti gli elementi comuni a tutti coloro che permangono nell'indagine per lo stesso numero di anni (cioè presentano la stessa lunghezza di osservazione; assimilabile ad un effetto età). Poiché vogliamo depurare la misura della disuguaglianza dagli effetti di coorte e di età, l'informazione su cui concentriamo la nostra attenzione è quella relativa all'andamento nel tempo dei redditi (medi per gruppo), cioè l'andamento nel tempo di \mathcal{A}_t . Una analoga scomposizione è proposta per i redditi medi, che rilevano nella fase di costruzione della disuguaglianza controfattuale (si veda l'Appendice).

I primi risultati che è utile riportare sono quelli relativi alla specificazione in cui la disuguaglianza dei diversi gruppi è regredita sulle sole *dummies* temporali, come in (7). In questo caso non controlliamo quindi per gli effetti coorte ed età e tale stima costituisce quindi un utile *benchmark* rispetto ai passaggi successivi.

I risultati di tale regressione sono riportati nella Tab.2, Col.1.

INSERIRE QUI TABELLA 2

Come si può notare, gli effetti di periodo evidenziano un profilo della disuguaglianza piuttosto variabile nel tempo: ad una prima fase in cui la disuguaglianza è crescente – con un picco nel 1998 - segue una fase in cui essa decresce in modo significativo, restando tuttavia più elevata (anche se non in modo statisticamente significativo) nel 2006 rispetto al 1989 (che costituisce il nostro termine di paragone).

⁷ Tra gli elementi comuni esistono certamente le metodologie del disegno campionario e le altre metodologie utilizzate nello svolgimento dell'indagine.

⁸ Tra di essi vi sono tutti gli shock di tipo macroeconomico che possono influenzare i salari in un dato anno.

Quando controlliamo anche per gli effetti di coorte e di età (quando cioè stimiamo l'equazione (6); si veda Tab.2, Col.2) troviamo immediatamente una chiara conferma delle nostre ipotesi quanto agli effetti dell'attrito sulla misura della disuguaglianza⁹. Infatti, gli effetti età mostrano un profilo di segno negativo e crescente in valore assoluto a partire dai gruppi presenti 5 volte nell'indagine sino ai gruppi presenti (almeno) 9 volte. Questo risultato evidenzia in modo chiaro come – anche controllando per gli effetti coorte relativi all'anno di entrata nell'indagine - i gruppi che restano più a lungo nel campione tendono ad avere al loro interno una minore disuguaglianza, a causa del fatto che con il passare del tempo la coda sinistra della distribuzione viene perduta nelle reinterviste. Naturalmente questo esercizio modifica anche la misura del profilo temporale della disuguaglianza, cioè l'andamento degli effetti di periodo. Rispetto al caso in cui non controlliamo per effetti di coorte ed età, possiamo notare che ora il profilo temporale è sostanzialmente più piatto: i coefficienti relativi a tutti gli anni sono infatti ridotti. Resta comunque l'immagine di una disuguaglianza più elevata nel corso degli anni 90' che poi decresce nel corso dei primi anni del secolo successivo.

Prima di considerare le implicazioni di tale esercizio sulla misura della disuguaglianza aggregata, come già segnalato nel paragrafo 4, dobbiamo correggere le nostre stime in modo da tenere conto della presenza all'interno del campione analizzato di gruppi tra loro non omogenei. Infatti, mentre per i gruppi che entrano in un dato anno nel campione ed escono in un determinato anno siamo certi della numerosità delle occasioni in cui sono stati osservati, per tutti i gruppi che sono osservati nell'ultimo anno disponibile (il 2006) noi non sappiamo con esattezza il numero di periodi per cui essi saranno osservati¹⁰. La soluzione proposta, descritta diffusamente nell'Appendice, prevede di assegnare delle probabilità di sopravvivenza ai diversi gruppi presenti nel 2006.

Le stime che otteniamo dopo aver introdotto questa correzione si differenziano da quelle prima descritte in ordine a due aspetti: il profilo temporale e la rilevanza del numero delle occasioni (gli effetti di coorte sono infatti – con una sola eccezione – non significativi sia nel caso in cui si controlli per la censura della durata nell'indagine sia nel caso in cui non si tenga conto di questo problema).

Per quanto riguarda il profilo temporale della disuguaglianza, i risultati ottenuti (Tab. 2; Col. 3) ci sembra evidenzino il venir meno di un chiaro profilo temporale declinante nel corso dei primi anni 2000. Infatti, il coefficiente per il 2004 (0,049) – che è ora significativo – risulta

⁹ Si noti che in questo caso il termine di paragone è costituito dal valore dell'indice di disuguaglianza del gruppo di coloro che sono entrati nel 1989, sono osservati nel 1989 e restano nel campione solo per una volta.

¹⁰ Se ad esempio, un gruppo è entrato nel 2000 ed è ancora presente nel 2006, noi possiamo solo dire che questo gruppo è presente almeno 4 volte, ma non possiamo concludere che esso è presente esattamente 4 volte.

sostanzialmente identico a quello osservato nel 1993 (0,05). Solo per il 2006 si evidenzia una riduzione della disuguaglianza, che è però vicina ai valori osservati nel 1995 (oltre che nel 2002). Per quanto riguarda invece l'effetto età, se da un lato possiamo notare come resti confermata la relazione negativa tra numero di occasioni e disuguaglianza, possiamo notare come quest'ultima sia particolarmente significativa dal punto di vista quantitativo per i gruppi osservati 8 e 9 volte.

A questo punto è forse utile ricordare che l'obiettivo del nostro lavoro è duplice. In primo luogo abbiamo inteso stimare l'effetto dei tre diversi effetti sulla disuguaglianza all'interno dei gruppi (quella cioè responsabile della disuguaglianza *within*). I risultati appena descritti ci confermano che la lunghezza della permanenza nel campione tende ad essere associata in modo sistematico ad una minore disuguaglianza. Il secondo obiettivo del nostro lavoro consiste nella quantificazione del ruolo dell'attrito sulla misura della disuguaglianza complessiva. Per fare ciò dobbiamo costruire una misura della disuguaglianza che sia depurata degli effetti dell'attrito e dell'anno di ingresso e che si concentri invece sui coefficienti che descrivono l'andamento nel tempo della disuguaglianza di ciascun gruppo.

Data la scomposizione della disuguaglianza aggregata proposta nella (5) e date le ipotesi contenute nella (6) e nella analoga scomposizione proposta¹¹ per $\ln \mu_{ict}$, possiamo notare che l'andamento complessivo della disuguaglianza depurata degli effetti di coorte ed età dipenda solo dall'andamento nel tempo degli effetti di periodo. Questo perché la componente della disuguaglianza *between* non appena si siano eliminati gli effetti di età e coorte risulta per definizione pari a zero:

Quindi la nostra misura 'controfattuale' della disuguaglianza aggregata è alimentata solo dalla componente *within*, e in particolare dalla misura di quest'ultima depurata degli effetti di età e di coorte. Ne consegue che possiamo scrivere – per la misura delle disuguaglianza controfattuale –:

$$GE_t = GE_{wt} = \sum_{c,l} \left(\frac{N_{clt}}{N_t} \right) GE_{ict} = \mathcal{G}_t. \quad (9)$$

Per rendere la misura della disuguaglianza 'controfattuale' direttamente confrontabile con quella calcolata applicando la (2), abbiamo imposto un valore identico nell'anno base (il 1989). Se confrontiamo questa serie con quella non corretta per gli effetti di coorte e di età (si veda Tab.2) possiamo notare come, a partire dal 1993, la misura della disuguaglianza controfattuale risulti sempre più elevata. In particolare, lo scostamento risulta particolarmente significativo in termini assoluti nel 1993, nel 1998 e nel 2004.

¹¹ Si veda l'Appendice.

INSERIRE QUI TABELLA 3

6. Conclusioni

In questo capitolo abbiamo analizzato le conseguenze dell'*attrition* che caratterizza l'indagine Banca d'Italia sui redditi e la ricchezza delle famiglie per la stima dell'andamento nel tempo della disuguaglianza nel reddito.

L'analisi muove da una duplice evidenza: i) nelle occasioni di indagine dal 1989 al 2006 la distribuzione del campione secondo il numero di occasioni di indagine alle quali le unità campionarie prendono parte si sposta progressivamente verso destra; ii) le unità campionarie che rimangono più a lungo nel campione sono comparativamente più ricche e meno diseguali. Le due evidenze messe assieme suggeriscono che ne può risultare una sistematica sottostima della disuguaglianza.

La disuguaglianza nei redditi familiari equivalenti viene misurata mediante l'indice *Generalized Entropy* di ordine zero. Tale indice viene calcolato per ogni occasione di indagine distintamente per i gruppi di unità campionarie definiti secondo l'anno di ingresso nell'indagine e il numero di occasioni di permanenza nell'indagine. L'indice così calcolato viene regredito congiuntamente su effetti specifici di periodo, di anno di ingresso nell'indagine e di numero di occasioni di permanenza nell'indagine.

I principali risultati della nostra analisi sono riassumibili in due affermazioni:

- 1) la disuguaglianza nei redditi equivalenti tra le famiglie incluse nel campione risulta marcatamente decrescente nel numero di occasioni di indagine alle quali le famiglie prendono parte. Ad esempio, le famiglie rimaste nel campione 9 occasioni presentano un valore dell'indice inferiore di 0.06 rispetto a quelle uscite dal campione dopo la prima occasione, per un indice che negli anni considerati prende valori nell'intervallo (.15, .20).
- 2) tenuto conto dell'effetto al punto precedente, il *pattern* temporale della disuguaglianza negli anni dal 1989 al 2006 risulta diverso da quello che si osserva ignorando gli effetti dell'*attrition*. Innanzitutto, il livello della disuguaglianza risulta sistematicamente più alto a partire dal 1993. In particolare, negli anni in cui la disuguaglianza presenta dei picchi – 1993, 1998 e 2004 – la correzione degli effetti dell'*attrition* amplifica la dimensione di tali picchi. Ad esempio, nel 1998 la correzione degli effetti dell'*attrition* sposta il valore dell'indice da 0.22 a 0.25.

Nel complesso, questi risultati confermano le evidenze già mostrate da Giraldo, Rettore e Trivellato [2007]: le determinanti dell'*attrition* nell'indagine della Banca d'Italia sono strettamente associate al reddito familiare.

Appendice: censura dei tempi di permanenza nell'indagine

Sia $GE(0)_{tc(l)}$ l'indice di disuguaglianza al tempo t per il gruppo costituito dagli entrati all'occasione di indagine c che rimangono nell'indagine *almeno* l occasioni. Ad esempio, per gli entrati nel 1995 ancora presenti nel 2006 $c=1995$ e $l=6$. Ricordando la definizione di $GE(0)$, è immediato esprimere la disuguaglianza al tempo t per il gruppo costituito dagli entrati all'occasione di indagine c che rimangono nell'indagine *almeno* l occasioni in funzione del reddito medio e dell'indice di disuguaglianza al tempo t per i gruppi costituiti dagli entrati all'occasione di indagine c che rimangono nell'indagine, rispettivamente, l occasioni, $l+1$ occasione, $l+2$ occasioni,....:

$$(A1) \quad GE(0)_{tc(l)} = \sum_j w_{cj} \ln\left(\frac{\mu_{tc(l)}}{\mu_{tcj}}\right) + \sum_j w_{cj} GE(0)_{tcj}$$

w_{cj} essendo il peso dei soggetti che rimangono nell'indagine esattamente j occasioni tra quelli entrati all'occasione c . La soluzione che proponiamo per risolvere il problema della censura dei tempi di permanenza nell'indagine si articola in tre passi:

- 1) si derivano i pesi w_{cj} dalla *funzione di rischio* relativa ai tempi di permanenza nell'indagine.
- 2) Si stima la quantità $\sum_j w_{cj} \ln(\mu_{tc(l)}/\mu_{tcj})$ ricorrendo ad una regressione dei redditi sugli effetti di periodo, di anno di ingresso e di durata della permanenza nell'indagine analoga alla (6).
- 3) Sostituendo a $GE(0)_{tcj}$ in (A1) il modello in (6) si ottiene:

$$(A2) \quad GE(0)_{tc(l)} = \gamma_c + \mathcal{G}_t + \sum_{j=l} w_{cj} \delta_j + \sum_{j=l} \varepsilon_{tcj}$$

Tale espressione consente di utilizzare le disuguaglianze osservate per i gruppi che presentano durata di permanenza nell'indagine incompleta, $GE(0)_{tc(l)}$, accanto alle disuguaglianze osservate per i gruppi con durata completa, $GE(0)_{tcj}$.

Quanto al punto 1), sia $b(l) = Pr(D=l | D \geq l)$, $l = 1, 2, 3, \dots$, la *funzione di rischio* associata al tempo di permanenza nell'indagine. Tale funzione è agevolmente stimabile rapportando il numero di soggetti che lasciano l'indagine dopo avervi trascorso esattamente l occasioni al numero di soggetti che vi trascorrono almeno l occasioni. Vista la sostanziale stabilità nel tempo dei tassi di abbandono dell'indagine, nell'analisi svolta abbiamo assunto che i pesi w_{cj} che compaiono nella (A1) siano gli stessi per i gruppi di entranti nell'indagine in anni diversi. Tali pesi corrispondono alla *funzione di sopravvivenza* associata alla funzione di rischio $b(l)$. Si ottengono dalla funzione di rischio mediante le seguenti relazioni:

$$\begin{aligned}
& w_1 = h(1) \\
\text{(A3)} \quad & w_2 = (1-h(1))h(2) \\
& w_3 = (1-h(1))(1-h(2))h(3) \\
& \dots\dots\dots
\end{aligned}$$

Quanto al punto 2), in analogia all'equazione di regressione (6), specifichiamo per il reddito medio nel periodo t relativo al gruppo di soggetti entrati nell'indagine nel periodo c e rimastivi per l occasioni la seguente regressione:

$$\text{(A4)} \quad \ln \mu_{tcl} = \nu_c + \tau_t + \varphi_l + u_{tcl}$$

dove ν_c cattura gli elementi del logaritmo del reddito comuni a tutti i gruppi che appartengono alla stessa coorte, τ_t rappresenta invece gli elementi comuni a tutti coloro che sono osservati nello stesso anno¹² t , mentre φ_l rappresenta gli elementi comuni a tutti coloro che permangono nell'indagine per lo stesso numero di occasioni di indagine .

I gruppi la cui durata di permanenza è osservata incompleta contribuiscono alla stima di tale equazione mediante la seguente relazione:

$$\text{(A5)} \quad \ln \mu_{tcl} = \nu_c + \tau_t + \sum_j w_{cj} \varphi_j + \sum_j w_{cj} u_{tcj}$$

Ottenute le stime degli effetti di periodo, di anno di ingresso e di durata della permanenza, è possibile stimare la quantità μ_{tcj} che compare al denominatore del primo termine al secondo membro della (A1) mediante la regressione (A4):

$$\ln \mu_{tcl} = \nu_c + \tau_t + \varphi_l .$$

Ottenuta tale quantità, si procede con il passo 3) della procedura di stima.

Indicazioni bibliografiche

Banca d'Italia
(anni vari) *I bilanci delle famiglie italiane nell'anno* [da 1989 a 2006], *Supplemento al Bollettino Statistico* (nuova serie), Banca d'Italia, Roma, anni e numeri vari.

¹² Tra di essi vi sono tutti gli shock di tipo macroeconomico che possono influenzare i redditi in un dato anno.

Biagi, F. e Canalone, G.

2008 *The evolution of inequality and poverty in Italy: 1987-2006*, mimeo

Brandolini, A.

1999 *The distribution of personal income in post-war Italy: source description, data quality, and the time pattern of income inequality*, in «Giornale degli Economisti e Annali di Economia», vol. 58, pp. 183-239.

2001 *Disuguaglianza e povertà*, in *Manuale di economia del lavoro*, a cura di Brucchi Luchino, Bologna, Il Mulino, 411-432.

2008 *Income inequality in Italy: facts and measurement*, in «Atti della XLIV riunione scientifica della Società Italiana di Statistica», Università della Calabria, 25-27 giugno 2008, Cleup, pp. 55-77.

Coulter, F. A. E., Cowell, F. A. e Jenkins, S. P.

1992 *Differences in needs and assessment of income distributions*, *Bulletin of economic Research*, **44**, 77-124

Cowell, F. A. e Mercader-Prats, M.

1999 *Equivalence Scales and Inequality*, Discussion paper N. 27 DARP, LSE.

Giraldo, A., Rettore, E. e Trivellato, U.

2001 *Attrition bias in the Bank of Italy's Survey of Household Income and Wealth*, in «Proceedings of the International Conference on Quality in Official Statistics», Stockholm, May 14-15, 2001.

2007 *Gli episodi di povertà causano ulteriori episodi di povertà? Evidenze dal panel sui bilanci delle famiglie della Banca d'Italia*, in *Povertà e benessere. Una geografia delle disuguaglianze in Italia*, a cura di A. Brandolini e C. Saraceno, Bologna, Il Mulino, pp. 237-257.

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

Fonte: Banca d'Italia [2008].

Tab.2 Stima degli effetti di periodo, di coorte e di età sulla componente within dell'indice di disuguaglianza

	Solo effetti di periodo	Effetti di periodo, di coorte e di età	Effetti di periodo, di coorte e di età (Con correzione per la censura dei tempi di permanenza nell'indagine)
Anno 1991	-0,001 (0,009)	-0,004 (0,007)	-0,004 (0,007)
Anno 1993	0,060** (0,012)	0,052** (0,010)	0,050** (0,010)
Anno 1995	0,046** (0,010)	0,036** (0,008)	0,033** (0,009)
Anno 1998	0,070** (0,012)	0,064** (0,011)	0,069** (0,011)
Anno 2000	0,042** (0,010)	0,035** (0,009)	0,037** (0,010)
Anno 2002	0,038** (0,010)	0,034** (0,010)	0,033** (0,011)
Anno 2004	0,026* (0,013)	0,022 (0,014)	0,049** (0,014)
Anno 2006	0,014 (0,011)	0,009 (0,011)	0,027** (0,012)
2 Occasioni	N.A.	0,016* (0,008)	0,017* (0,010)
3 Occasioni	N.A.	0,004 (0,008)	0,007 (0,008)
4 Occasioni	N.A.	0,003 (0,008)	0,011 (0,010)
5 Occasioni	N.A.	-0,021** (0,008)	-0,017** (0,008)
6 Occasioni	N.A.	-0,023** (0,008)	-0,017* (0,009)
7 Occasioni	N.A.	-0,037** (0,008)	-0,022** (0,008)
8 Occasioni	N.A.	-0,045** (0,009)	-0,054** (0,011)
9 Occasioni	N.A.	-0,047** (0,009)	-0,061** (0,010)
Ingresso 1991	N.A.	0,004 (0,006)	-0,005 (0,006)
Ingresso 1993	N.A.	0,015 (0,010)	0,017* (0,009)
Ingresso 1995	N.A.	0,015 (0,011)	-0,001 (0,012)
Ingresso 1998	N.A.	-0,011 (0,011)	-0,013 (0,012)
Ingresso 2000	N.A.	0,008 (0,015)	0,019 (0,013)
Ingresso 2002	N.A.	-0,008 (0,012)	-0,002 (0,012)
Ingresso 2004	N.A.	0,001 (0,014)	-0,007 (0,015)
Ingresso 2006	N.A.	0,019 (0,013)	0,003 (0,013)
Intercetta	0,143** (0,008)	0,159** (0,006)	0,158 (0,007)
R2	0,28	0,56	0,58
N. osservazioni	165	165	165

Nota: Errori standard robusti in parentesi.

** 95% di significatività; * 90% di significatività.

Tab.3 Andamento nel tempo dell'indice di disuguaglianza corretto e non corretto per gli effetti di coorte e di età

	GE(0) non corretto	GE(0) corretto
Anno 1989	0,162699997	0,162699997
Anno 1991	0,153500006	0,153601274
Anno 1993	0,209600002	0,225422785
Anno 1995	0,194700003	0,201592401
Anno 1998	0,217099994	0,247232899
Anno 2000	0,198200002	0,206864566
Anno 2002	0,183699995	0,190592393
Anno 2004	0,185800001	0,200996205
Anno 2006	0,176200002	0,180813923

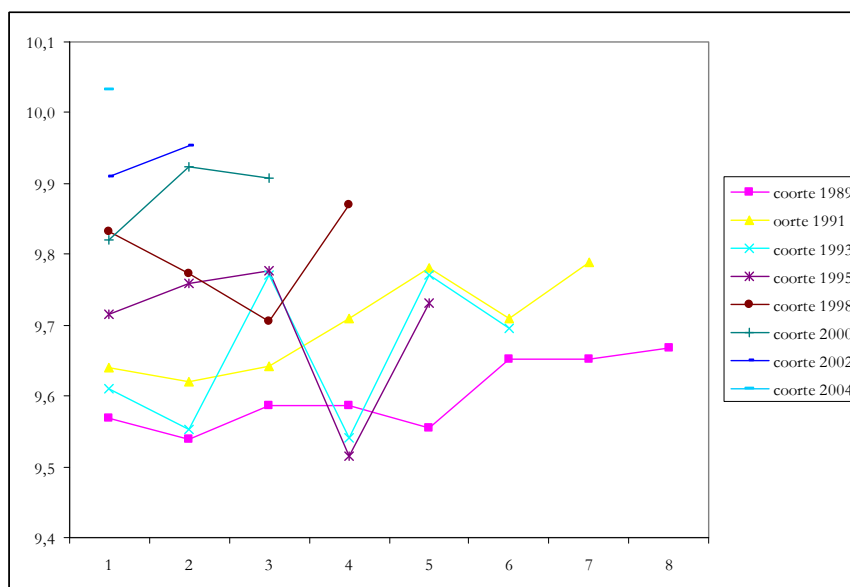


Fig. 1. Media del logaritmo dei redditi familiari per secondo il numero di occasioni di permanenza nel panel.

Fonte: nostre elaborazioni su dati Banca d'Italia.

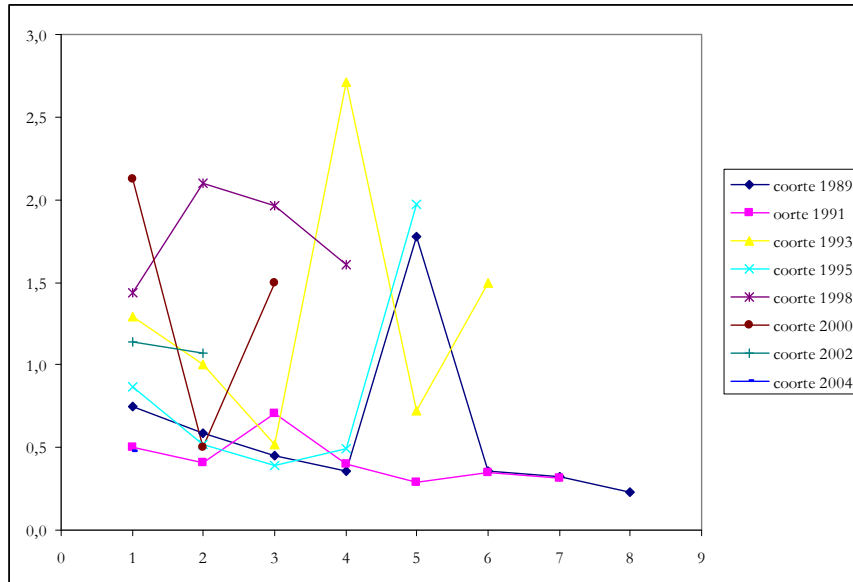


Fig. 2. Varianza del logaritmo dei redditi medi familiari per secondo il numero di occasioni di permanenza nel panel.

Fonte: nostre elaborazioni su dati Banca d'Italia.