

GLI EFFETTI DELL'ATTRITO SULLA STIMA DELLA DISUGUAGLIANZA IN ITALIA

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 per l'Italia 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 attrito, 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 capitolo riprendiamo il filone di ricerche sulla disuguaglianza e sulla povertà già avviato nel precedente volume e studiamo se la distribuzione del reddito tra i componenti delle famiglie che rimangono nel campione è sistematicamente diversa dalla corrispondente distribuzione per le famiglie che lo lasciano. Dopo aver 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 attrito che caratterizza IBI è tale da produrre distorsioni nella stima dell'andamento nel tempo della disuguaglianza.

Il principale risultato che otteniamo è che il processo di attrito dà luogo ad una distorsione della stima della disuguaglianza. Il metodo che proponiamo consente di correggere tale distorsione in modo semplice.

L'organizzazione del capitolo 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'attrito sulla stima della media e della dispersione dei redditi familiari. Nella sez. 3 (e nelle Appendici per maggiori dettagli) presentiamo la strategia utilizzata per identificare gli effetti dell'attrito sulla stima della distribuzione dei redditi e del suo andamento nel tempo. La sez. 4 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

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.

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 complessivamente 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 tra queste 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 tra quelle entrate nell'indagine nel 1989 sono ancora presenti; 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 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.

Dato un certo insieme di entrati nell'anno t tale insieme, nel tempo, progressivamente si riduce parte per disegno, parte a causa dell'attrito. Tale processo, per gli anni dal 1989 al 2006 è documentato in tabella 1.

Si noti come per effetto dell'attrito la composizione del campione sezionale secondo il numero di occasioni di indagine alle quali le varie famiglie prendono parte cambia con il passare del tempo. Ad esempio, le famiglie che prendono parte ad *almeno 5* occasioni di indagine rappresentano il 6,5% del campione nell'indagine del 1989 (544 famiglie ancora presenti nel 1998 delle 8.274 entrate nell'indagine del 1989); rappresentano il 16,8% del campione nel 1991 (544 famiglie entrate nel 1989 e ancora presenti nel 1998 alle quali vanno aggiunte le 832 famiglie entrate nel 1991 e ancora presenti nel 2000, il totale rapportato a 8.188, il numero di famiglie coinvolte nell'indagine del 1991). Nelle indagini successive a quella del 1991 tale percentuale è pari a, rispettivamente, 20,3, 21,7, 33,6, 32,4, 32,5, 33,9 e 36,5⁵.

³ Con il termine attrito si denota 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.

⁵ Per le famiglie entrate nell'indagine dal 2000 in poi il calcolo delle percentuali richiede la stima di quante tra loro rimarranno nell'indagine almeno 5 occasioni, un evento che al 2006, ultima occasione di indagine disponibile, nel loro caso non si è ancora manifestato. La stima è stata ottenuta ipotizzando che i tassi di permanenza nell'indagine per queste famiglie siano gli stessi osservati per le famiglie entrate nell'indagine negli anni precedenti al 2000.

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].

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 attrito induce distorsione, in alcuni casi tutt'altro che trascurabile⁶. In figura 1 riportiamo la media del logaritmo del reddito familiare equivalente⁷ nell'anno di ingresso nell'indagine secondo la durata della permanenza nell'indagine stessa (da 1 a 9 sull'asse orizzontale) e secondo l'anno di ingresso nel panel (le linee corrispondenti alle varie "coorti"). A meno dell'elevata variabilità dovuta alla talvolta piccola numerosità campionaria, sembra emergere una tendenza del logaritmo del reddito ad aumentare al crescere della durata della permanenza nel panel. Di converso se andiamo a considerare in figura 2 la varianza del logaritmo del reddito per gli stessi gruppi, sempre a meno della variabilità campionaria, questa tende a diminuire all'aumentare della permanenza nel panel.

Questi risultati evidenziano che tendono a rimanere più a lungo nel panel le famiglie con redditi relativamente più alti e meno diseguali tra loro. Si tratta di risultati in linea con quanto trovato in Giraldo et al. [2007]: 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 nelle quali una famiglia permane nel panel è correlato negativamente con la probabilità di sperimentare un episodio di povertà.

⁶ 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]

⁷ Per i dettagli del calcolo del reddito equivalente si veda la successiva sez. 4.

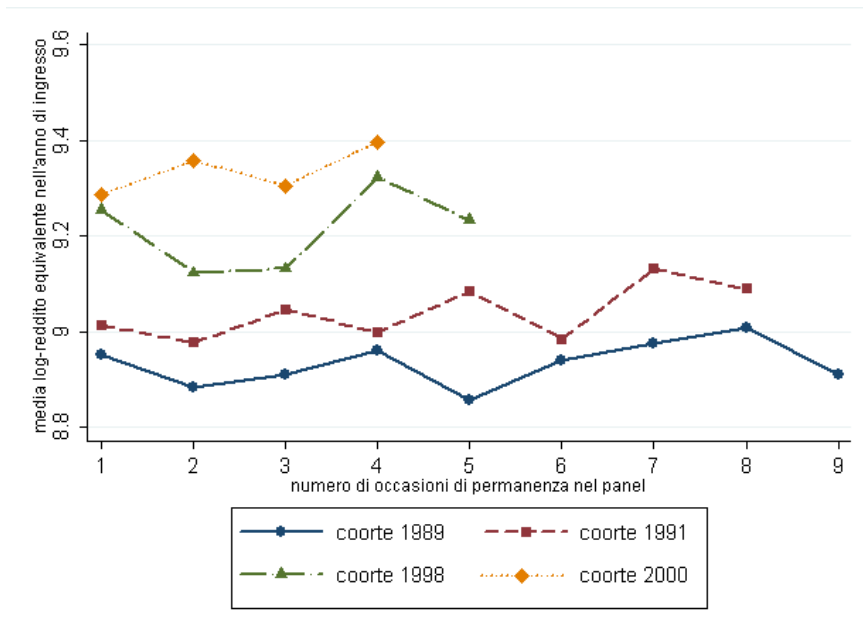


Fig. 1. Media del logaritmo dei redditi equivalenti nell'anno di ingresso nell'indagine secondo il numero di occasioni di permanenza nel panel per alcune coorti di entranti.

Fonte: nostre elaborazioni su dati Banca d'Italia.

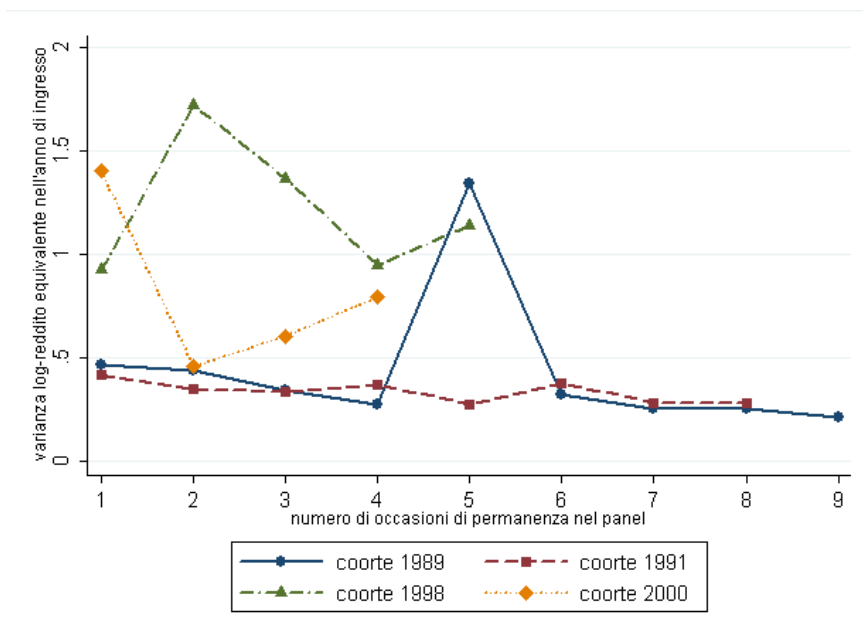


Fig. 2. Varianza del logaritmo dei redditi equivalenti nell'anno di ingresso nell'indagine secondo il numero di occasioni di permanenza nel panel per alcune coorti di entranti.

Fonte: nostre elaborazioni su dati Banca d'Italia.

E' possibile che tale selezione non casuale delle famiglie che permangono nel campione IBI induca una distorsione nella stima della disuguaglianza dei redditi. Più specificamente, il fatto che aumenti nel tempo il peso nel campione di famiglie sopravvissute, o destinate a sopravvivere, a lungo nell'indagine, famiglie tendenzialmente più ricche e tra loro meno diseguali, può indurre una deformazione della stima del trend della disuguaglianza.

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

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

Nella nostra analisi facciamo uso dell'indice di disuguaglianza *Generalized Entropy* di ordine zero definito dalla seguente espressione:

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

nella quale y_i è il reddito equivalente dell' i -esimo soggetto e μ è il reddito equivalente medio nella popolazione considerata. Nell'Appendice 1 illustriamo con qualche dettaglio le proprietà della classe di indici di disuguaglianza della quale fa parte $GE(0)$. Per l'interpretazione dei valori assunti dall'indice – e dei risultati della nostra analisi sugli effetti dell'attrito – si tenga presente che la quantità $[2*GE(0)]^{1/2}$ è circa pari al rapporto tra la deviazione standard dei redditi e la loro media (si veda l'Appendice 1).

Per analizzare gli effetti 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. 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. Ad esempio, nel campione per l'anno 1995 sono comprese famiglie entrate nel campione nell'anno 1989, altre nell'anno 1991, altre ancora nell'anno 1993 e infine alcune alla loro prima intervista. Allo stesso modo, alcune lasciano l'indagine dopo la prima intervista, altre dopo l'intervista relativa al 1998 e così via fino al 2006.

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 – 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 – come mostriamo nel paragrafo 4 – 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.

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:

$$\begin{aligned} GE(0)_t &= \sum_{c,l} \left(\frac{N_{tcl}}{N_t} \right) GE(0)_{tcl} + \sum_{c,l} \left(\frac{N_{tcl}}{N_t} \right) \left(\ln \frac{\mu_t}{\mu_{tcl}} \right) = \\ &= GE(0)_{wt} + GE(0)_{Bt} \end{aligned} \quad (2)$$

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:

$$GE(0)_{ict} = \gamma_c + \mathcal{G}_t + \delta_t + \varepsilon_{ict}. \quad (3)$$

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_{ict}) se il momento dell'ingresso e il numero delle occasioni di permanenza nell'indagine fossero irrilevanti. Infatti, l'andamento complessivo della disuguaglianza depurata degli effetti dell'anno di ingresso e del numero di occasioni di indagine dipende solo dall'andamento nel tempo degli effetti di periodo \mathcal{G}_t , dato che la componente *between* della disuguaglianza eliminati gli effetti dell'occasione di ingresso e del numero di occasioni di indagine risulta per definizione pari a zero.

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 sopravvivenenti nell'indagine sempre più omogenei, la sequenza degli effetti δ_t dovrebbe risultare tale da ridurre progressivamente la disuguaglianza al crescere di l .

Nel modello di regressione (3) 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.

Muovendo dalle stime di \mathcal{G}_t , nel prossimo paragrafo proporremo una misura controfattuale della disuguaglianza, vale a dire la disuguaglianza che si sarebbe osservata in assenza degli effetti dell'occasione di ingresso, γ_c , e degli effetti del numero di occasioni nell'indagine, δ_t .

Si noti che il fatto che gli effetti della durata della permanenza nell'indagine, i coefficienti δ_t , 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 accadrebbe 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 (3) stimiamo anche il modello ridotto:

$$GE(0)_{ict} = \mathcal{G}_t + u_{ict}. \quad (4)$$

Il confronto dell'andamento nel tempo delle stime di \mathcal{G}_t , rispettivamente, dal modello (3) e dal modello (4) rivela se il mancato controllo degli effetti della durata della permanenza nell'indagine e dell'anno di ingresso comporta una distorsione nella stima dell'andamento nel tempo della disuguaglianza.

L'identificazione degli effetti della durata della 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 dell'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 (3) e (4).

4. 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 sola volta alle varie occasioni 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, l'ultimo il panel 2004-2006) per un totale di 36 campioni panel mutuamente esclusivi.

La misura del reddito personale 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). Il reddito familiare equivalente così calcolato viene attribuito ad ognuno dei componenti la famiglia. Infine, l'indice di disuguaglianza relativo al periodo t per il gruppo di famiglie entrate nell'indagine all'occasione c e rimastovi per l occasioni si ottiene applicando l'espressione (1) ai redditi equivalenti dei componenti di tali famiglie.

La linea a tratteggio fine in Fig. 3 (simbolo il quadrato) presenta 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.

Con riferimento all'equazione (2), 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_{tel} . Per entrambe le variabili proponiamo una scomposizione che metta in luce gli effetti, rispettivamente, del periodo di riferimento, dell'occasione di ingresso nell'indagine e del numero di periodi trascorsi nell'indagine.



Fig. 3. Andamento nel tempo dell'indice di disuguaglianza GE(0) ottenuto, rispettivamente, facendo uso dei pesi di riporto all'universo del campione sezionale (—■—), correggendo (—●—) e non correggendo (—◆—) per gli effetti dell'attrito.

Fonte: nostre elaborazioni su dati Banca d'Italia.

Per quanto concerne la misura di disuguaglianza, riprendendo la (3) notiamo che γ_c cattura gli elementi della disuguaglianza comuni a tutti i gruppi entrati nell'indagine alla stessa occasione⁸, \mathcal{G}_t rappresenta invece gli elementi comuni a tutti coloro che sono osservati nello stesso anno⁹ t , mentre δ_t rappresenti gli elementi comuni a tutti coloro che permangono nell'indagine per lo stesso numero di occasioni. Poiché vogliamo depurare la misura della disuguaglianza dagli effetti dell'occasione di ingresso e del numero di occasioni di permanenza nell'indagine, l'informazione su cui concentriamo la nostra attenzione è quella relativa all'andamento nel tempo della disuguaglianza, cioè l'andamento nel tempo di \mathcal{G}_t . Una analoga scomposizione è proposta per i redditi medi, rilevanti per la 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 (4). 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.

I risultati di tale regressione sono riportati in Fig. 3, linea a tratteggio lungo (simbolo il

⁸ Tra gli elementi comuni esistono certamente il disegno campionario e le altre metodologie utilizzate nello svolgimento dell'indagine.

⁹ Tra di essi vi sono tutti gli shock di tipo macroeconomico che possono influenzare i redditi in un dato anno.

rombo; vedi anche la Tab.2, col.1).

Tab.2: *Stima degli effetti di periodo, di anno di ingresso e di numero di occasioni di indagine sull'indice di disuguaglianza GE(0) (n=165).*

	Solo effetti di periodo	Effetti di periodo, di anno di ingresso e di n. di occasioni di indagine	Effetti di periodo e di n. di occasioni di indagine
Anno 1991	-0,0087 (0,0124)	-0,0167 (0,0172)	-0,0146 (0,0130)
Anno 1993	0,0775** (0,0187)	0,0587** (0,0205)	0,0732** (0,0192)
Anno 1995	0,0506** (0,0136)	0,0282 (0,0192)	0,0427** (0,0136)
Anno 1998	0,1233** (0,0201)	0,1116** (0,0262)	0,1210** (0,0216)
Anno 2000	0,0875** (0,0230)	0,0740** (0,0246)	0,0850** (0,0235)
Anno 2002	0,0604** (0,0177)	0,0461* (0,0246)	0,0538** (0,0184)
Anno 2004	0,0402** (0,0190)	0,0575** (0,0266)	0,0561** (0,0183)
Anno 2006	0,0195 (0,0170)	0,0313 (0,0217)	0,0131 (0,0136)
2 Occasioni	-	0,0027 (0,0125)	-
3 Occasioni	-	-0,0048 (0,0125)	-
4 Occasioni	-	-0,0088 (0,0144)	-
5 Occasioni	-	-0,0457** (0,0148)	-
6 Occasioni	-	0,0342 (0,0246)	-
7 Occasioni	-	-0,0324* (0,0189)	-0,0311** (0,0097)
8 Occasioni	-	-0,0293 (0,0336)	
9 Occasioni	-	-0,0866** (0,0175)	
Ingresso 1991	-	0,0045 (0,0115)	-
Ingresso 1993	-	0,0373** (0,0184)	-
Ingresso 1995	-	0,0202 (0,0196)	-
Ingresso 1998	-	-0,0076 (0,0169)	-
Ingresso 2000	-	0,0250 (0,0210)	-
Ingresso 2002	-	0,0159 (0,0196)	-
Ingresso 2004	-	-0,0129 (0,0206)	-
Ingresso 2006	-	-0,0017 (0,0186)	-
Intercetta	0,1526** (0,0115)	0,1716** (0,0157)	0,1699** (0,0124)

Nota: Tra parentesi sono riportati gli errori standard delle stime robuste rispetto all'eteroschedasticità. ** 95% di significatività; * 90% di significatività.

Il profilo della disuguaglianza che ne risulta ricalca nella sostanza quello ottenuto calcolando l'indice sulle singole cross-sections dell'indagine facendo uso dei pesi di riporto all'universo.

Infine, in Tab. 2, col. 2, presentiamo i risultati della stima dell'equazione (3). Si nota innanzitutto che gli effetti dell'anno di ingresso nell'indagine sono trascurabili (fa eccezione

il 1993 che risulta statisticamente significativo). La stima degli effetti del numero di occasioni di indagine conferma la nostre ipotesi quanto agli effetti dell'attrito sulla misura della disuguaglianza¹⁰: 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 $l=5, 7, 8$.

Per agevolare l'interpretazione dei risultati abbiamo stimato una versione semplificata dell'equazione (3), eliminando gli effetti dell'anno di ingresso - come abbiamo visto, sostanzialmente irrilevanti - e cogliendo gli effetti del numero di occasioni di indagine mediante una dummy che vale 1 per i gruppi presenti nell'indagine almeno in 5 occasioni e 0 per gli altri gruppi. I risultati sono in Tab.2, colonna 3. A parità di periodo di riferimento, i gruppi che rimangono nell'indagine per almeno 5 occasioni sono meno diseguali dei rimanenti gruppi per un ammontare pari a 0,0311 nella scala dell'indice $GE(0)$, statisticamente significativo.

Gli effetti di questo risultato per la stima dell'andamento nel tempo della disuguaglianza sono presto detti. Tenendo presente l'equazione (2), la semplificazione qui considerata dell'equazione (3) e i risultati in Tab. 2, col. 3, l'andamento nel tempo della componente *within* della disuguaglianza è dato da:

$$\mathcal{G}_t - 0,0311 \cdot p(l > 4)_t \quad (5)$$

dove \mathcal{G}_t sono gli effetti di periodo riportati in Tab. 2, col. 3 e $p(l > 4)_t$ è la proporzione di unità campionarie presenti nell'indagine per almeno 5 occasioni tra quelle incluse nel campione al tempo $t=1989 \dots 2006$, riportate nell'ultima riga della Tab. 1. Il termine $-0,0311 \cdot p(l > 4)_t$ rappresenta la distorsione nella stima della disuguaglianza al tempo t causata dall'attrito. Dato che la proporzione $p(l > 4)_t$ è crescente nel tempo e che la stima del suo coefficiente è statisticamente significativa, l'effetto distorsivo del numero di occasioni di presenza nell'indagine si manifesta in modo a sua volta crescente nel tempo.

Facendo uso della (5), è immediato 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 unità campionarie. Scegliendo come distribuzione di riferimento quella relativa al 1989 - percentuale di unità presenti in almeno 5 occasioni pari a 6,6 - si ottengono i valori dell'indice $GE(0)$ riportati in Tab. 3, col.3 e rappresentati in Fig. 3, linea continua (simbolo il pallino)¹¹.

La stima dell'indice di disuguaglianza che ne risulta va confrontata con la stima ottenuta facendo uso dei pesi di riporto all'universo dei campioni sezionali forniti dalla Banca d'Italia (in col. 1 di Tab. 3). L'indice corretto dagli effetti dell'attrito secondo la nostra procedura risulta tendenzialmente più elevato dell'indice ottenuto facendo uso dei pesi di riporto all'universo, nel 1998 e nel 2004 in modo marcato, con l'eccezione del 2006 (e, marginalmente, del 1995) nel quale l'ordinamento tra i due indici si rovescia. Ricordando che il coefficiente di variazione (il rapporto tra deviazione standard e media dei redditi) è circa pari alla radice quadrata del doppio di $GE(0)$, le differenze osservate tra le due stime

¹⁰ Si noti che in questo caso il termine di paragone è costituito dal valore dell'indice di disuguaglianza osservato nel 1989 relativo al gruppo di coloro che sono entrati nell'indagine nel 1989 e vi sono rimasti per una sola occasione.

¹¹ I valori in col.3 di Tab. 3 si ottengono nel seguente modo. Per il 1989 il valore è dato dall'effetto di periodo relativo al 1989 in col. 3 di Tab. 2 al quale va sottratta la quantità $0,0311 \cdot 0,066$ per tenere conto dell'attrito. Per gli anni dal 1991 al 2006 al valore così ottenuto è sufficiente aggiungere la stima degli altri effetti di periodo riportati in col. 3 di Tab. 2.

di $GE(0)$ relative al 1998 corrispondono approssimativamente ad una differenza pari a 0,035 nella scala propria del coefficiente di variazione (la differenza è di poco inferiore per il 2004).

Tab. 3. *Andamento nel tempo dell'indice di disuguaglianza $GE(0)$ ottenuto facendo uso dei pesi di riporto all'universo del campione sezionale, corretto e non corretto per gli effetti dell'attrito*

	Facendo uso dei pesi di riporto all'universo	non corretto per gli effetti dell'attrito	corretto per gli effetti dell'attrito
Anno 1989	0,1675	0,1526	0,1678
Anno 1991	0,1477	0,1439	0,1532
Anno 1993	0,2262	0,2301	0,2410
Anno 1995	0,2149	0,2032	0,2105
Anno 1998	0,2623	0,2759	0,2888
Anno 2000	0,2396	0,2401	0,2528
Anno 2002	0,2177	0,2166	0,2216
Anno 2004	0,2016	0,1928	0,2239
Anno 2006	0,1938	0,1721	0,1809

5. Conclusioni

In questo capitolo abbiamo analizzato le conseguenze dell'attrito 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 personale equivalente.

L'analisi muove da una duplice evidenza: i) per effetto del processo di attrito, nelle occasioni di indagine dal 1989 al 2006 i campioni sezionali presentano una frazione via via crescente di famiglie che permangono a lungo nell'indagine; ii) le unità campionarie che rimangono più a lungo nell'indagine sono comparativamente più ricche e tra loro meno diseguali. Le due evidenze messe assieme suggeriscono che ne può risultare una sistematica sottostima della disuguaglianza.

La disuguaglianza nei redditi personali equivalenti viene misurata mediante l'indice *Generalized Entropy* di ordine zero. Tale indice viene calcolato per ogni occasione di indagine distintamente per le unità campionarie appartenenti a gruppi 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 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.

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.

Nel complesso, questi risultati confermano le evidenze già mostrate da Giraldo, Rettore e Trivellato [2007] con riferimento alla dinamica della povertà, le generalizzano al caso della misura della disuguaglianza, ne forniscono una giustificazione analitica ed infine indicano una strategia operativa utile alla correzione degli effetti distorsivi dell'attrito.

Appendice 1: L'indice di disuguaglianza Generalized Entropy di ordine zero

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:

$$(A1) \quad GE(\alpha) = \frac{1}{\alpha(\alpha-1)} \left[\sum_{i=1}^n w_i \left(\frac{y_i}{\mu} \right)^\alpha - 1 \right]$$

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 lineare¹².

Nel nostro lavoro abbiamo scelto di applicare la scala di equivalenza OCSE 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:

$$(A2) \quad GE(0) = \frac{1}{N} \sum_{i=1}^N \ln \frac{\mu}{y_i} .$$

Quanto all'interpretazione dei valori osservati dell'indice, si noti che 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:

$$(A3) \quad \ln \mu - \ln y_i \approx \frac{\mu - y_i}{\mu} + 0,5 \left(\frac{\mu - y_i}{\mu} \right)^2$$

da cui risulta:

$$(A4) \quad GE(0) \approx 0,5 \sum_{i=1}^N \frac{1}{N} \left(\frac{\mu - y_i}{\mu} \right)^2 .$$

¹² 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.

Pertanto, la quantità $[2*GE(0)]^{1/2}$ è circa pari al rapporto tra la deviazione standard dei redditi e la loro media. Ad esempio, un valore dell'indice pari a 0,25 sta a significare che la deviazione standard dei redditi è pari a circa il 70% del reddito medio.

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:

$$(A5) \quad GE(0)_t = \sum_{g=1}^G \left(\frac{N_{gt}}{N_t} \right) GE(0)_{gt} + \sum_{g=1}^G \left(\frac{N_{gt}}{N_t} \right) \left(\ln \frac{\mu_t}{\mu_{gt}} \right) = \\ = GE(0)_{wt} + GE(0)_{Bt}$$

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 .

Appendice 2: censura dei tempi di permanenza nell'indagine

Sia $GE(0)_{ic(l)}$ 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,....:

$$(A6) \quad GE(0)_{ic(l)} = \sum_j w_{cj} \ln\left(\frac{\mu_{ic(l)}}{\mu_{icj}}\right) + \sum_j w_{cj} GE(0)_{icj}$$

w_j 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_j dalla *funzione di rischio* relativa ai tempi di permanenza nell'indagine.
- 2) Si stima la quantità $\sum_j w_{cj} \ln(\mu_{ic(l)}/\mu_{icj})$ ricorrendo ad una regressione dei redditi sugli effetti di periodo, di anno di ingresso e di durata della permanenza nell'indagine analoga alla (3).
- 3) Sostituendo a $GE(0)_{icj}$ in (A6) il modello in (3) si ottiene:

$$(A7) \quad GE(0)_{ic(l)} - \sum_j w_{cj} \ln\left(\frac{\mu_{ic(l)}}{\mu_{icj}}\right) = \gamma_c + \mathcal{G}_t + \sum_{j=l} w_{cj} \delta_j + \sum_{j=l} \varepsilon_{icj}$$

Tale espressione consente di utilizzare le disuguaglianze osservate per i gruppi che presentano durata di permanenza nell'indagine incompleta, $GE(0)_{ic(l)}$, (corrette secondo la quantità $\sum_j w_{cj} \ln\left(\frac{\mu_{ic(l)}}{\mu_{icj}}\right)$) accanto alle disuguaglianze osservate per i gruppi con durata completa, $GE(0)_{icj}$.

Quanto al punto 1), sia $h(l) = Pr(D=l | D \geq l)$, $l = 1, 2, 3, \dots$, 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_j che compaiono nella (A6) 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 $h(l)$. Si ottengono dalla funzione di rischio mediante le seguenti relazioni:

$$(A8) \quad \begin{aligned} w_1 &= h(1) \\ w_2 &= (1-h(1))h(2) \\ w_3 &= (1-h(1))(1-h(2))h(3) \\ &\dots\dots \end{aligned}$$

Quanto al punto 2), in analogia all'equazione di regressione (3), 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:

$$(A9) \quad \ln \mu_{tcl} = \nu_c + \tau_t + \varphi_l + u_{tcl}$$

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 anno¹³ 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:

$$(A10) \quad \ln \mu_{tc(l)} = \nu_c + \tau_t + \sum_j w_{ej} \varphi_j + \sum_j w_{ej} u_{tcj}$$

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 (A6) mediante la regressione (A9).

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

¹³ Tra di essi vi sono tutti gli shock di tipo macroeconomico che possono influenzare i redditi in un dato anno.

RIFERIMENTI BIBLIOGRAFICI

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.

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.