



BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Stime campionarie del reddito e della ricchezza familiare
coerenti con le stime aggregate: alcuni esperimenti

di Giovanni D'Alessio e Andrea Neri

Giugno 2015

Numero

272



BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional papers)

Stime campionarie del reddito e della ricchezza familiare
coerenti con le stime aggregate: alcuni esperimenti

di Giovanni D'Alessio e Andrea Neri

Numero 272 – Giugno 2015

La serie Questioni di economia e finanza ha la finalità di presentare studi e documentazione su aspetti rilevanti per i compiti istituzionali della Banca d'Italia e dell'Eurosistema. Le Questioni di economia e finanza si affiancano ai Temi di discussione volti a fornire contributi originali per la ricerca economica.

La serie comprende lavori realizzati all'interno della Banca, talvolta in collaborazione con l'Eurosistema o con altre Istituzioni. I lavori pubblicati riflettono esclusivamente le opinioni degli autori, senza impegnare la responsabilità delle Istituzioni di appartenenza.

La serie è disponibile online sul sito www.bancaditalia.it.

ISSN 1972-6627 (stampa)

ISSN 1972-6643 (online)

Stampa a cura della Divisione Editoria e stampa della Banca d'Italia

STIME CAMPIONARIE DEL REDDITO E DELLA RICCHEZZA FAMILIARE COERENTI CON LE STIME AGGREGATE: ALCUNI ESPERIMENTI

di Giovanni D'Alessio* e Andrea Neri*

Sommario

L'indagine sui bilanci delle famiglie Italiane (IBF) è una fonte di informazione largamente utilizzata nel dibattito economico e nella ricerca accademica. Le stime di alcuni aggregati importanti, come il reddito e la ricchezza, risentono tuttavia della minore propensione delle famiglie più abbienti a partecipare alla rilevazione e della reticenza degli intervistati a riportare con veridicità il proprio bilancio familiare. Questo lavoro, dopo aver passato in rassegna i principali contributi condotti nel corso degli anni, volti a correggere queste distorsioni ne valuta l'applicazione con riferimento ai dati del periodo 1995-2012. Sono inoltre sperimentati stimatori campionari che, utilizzando le tecniche di calibrazione, risultano coerenti con le informazioni macroeconomiche disponibili da altre fonti.

Classificazione JEL: D10, D31.

Parole chiave: reddito, ricchezza, famiglia, calibrazione.

Indice

1. Introduzione	5
2. Una breve rassegna delle metodologie di correzione	5
3. Precedenti aggiustamenti sui dati IBF	8
4. La correzione dell' <i>under-reporting</i> e della mancata risposta	12
4.1 La correzione proporzionale - C_1	12
4.2 La correzione sulla base del giudizio dell'intervistatore - C_2	12
4.3 Correzioni su singoli fenomeni - C_3	14
4.3.1 Peso di non risposta - C_{3A}	14
4.3.2 La correzione del reddito degli autonomi - C_{3B}	15
4.3.3 La correzione degli immobili non di residenza - C_{3C}	16
4.3.4 La correzione delle attività finanziarie - C_{3D}	18
4.4 Le calibrazioni - C_4 / C_8	19
5. Valutazione delle stime al 2012	21
6. Conclusioni	26
Appendice A – Tavole statistiche	28
Bibliografia	35

* Banca d'Italia, Dipartimento di Economia e Statistica.

1. Introduzione¹

L'indagine sui bilanci delle famiglie Italiane è una fonte di informazione largamente utilizzata nel dibattito economico e nella ricerca accademica. I dati di questa indagine tuttavia risentono, come quelli della gran parte delle indagini di questo tipo², di alcuni problemi di qualità dovuti alla minor propensione delle famiglie più abbienti a partecipare alla rilevazione e alla reticenza degli intervistati a riportare con veridicità il proprio bilancio familiare.

Nel corso degli anni numerose sono state le analisi volte ad approfondire questi fenomeni che danno luogo a stime campionarie distorte, come si evince dalle ampie differenze rispetto alle altre fonti informative sulle famiglie (sia macroeconomiche, come la contabilità nazionale, sia di tipo amministrativo, come le segnalazioni di vigilanza o i censimenti).

Il presente studio, dopo aver passato in rassegna le metodologie di correzione e aver illustrato i principali esperimenti condotti nel corso degli anni sui dati IBF, ne riesamina l'applicazione con riferimento ai dati del periodo 1995-2012. Si intende in tal modo valutare la possibilità di svolgere analisi a livello micro sulle condizioni di vita delle famiglie italiane lungo le diverse direttrici generalmente esaminate (reddito, ricchezza e debiti), con delle stime che siano coerenti con le altre informazioni macroeconomiche. Sebbene l'errore statistico caratterizzi anche le fonti non campionarie, si cerca in tal modo di valorizzare l'uso congiunto delle informazioni disponibili, sfruttando i punti di forza di ciascuna fonte. Il lavoro discute infine in che misura questi archivi di dati micro possono essere usati come base dati di input di modelli di microsimulazione.

2. Una breve rassegna delle metodologie di correzione

Le indagini campionarie sulle famiglie sono inevitabilmente afflitte da problemi legati agli errori di misurazione e alla mancata partecipazione di alcune unità selezionate. Nonostante gli sforzi utilizzati per ridurre e controllare tali errori durante le fasi di rilevazione, l'uso di metodi di aggiustamento nella fase di stima è in generale necessario.

I metodi di aggiustamento per dati campionari presenti in letteratura possono essere ricondotti a due approcci (per una rassegna recente si veda Nicolini *et. al.*, 2013).

L'approccio basato sul piano di campionamento (*design-based approach*), utilizzato prevalentemente per la correzione del problema della non risposta, prevede l'aggiustamento dei pesi campionari. L'idea è quella di considerare il campione osservato come un campione a due fasi. Il campione selezionato è quello ottenuto nella prima fase, mentre il campione effettivamente intervistato (i rispondenti) è considerato come una seconda fase di campionamento. Ciascuna unità della popolazione ha una certa probabilità di partecipare a questa seconda fase. Tale probabilità può essere

¹ Gli autori desiderano ringraziare Giovanna Ranalli, Luigi Cannari, Romina Gambacorta, Stefano Iezzi e Giuseppe Ilardi per i numerosi commenti ricevuti durante la stesura del lavoro. Si ringraziano anche i partecipanti al seminario "L'indagine sui bilanci delle famiglie italiane. Metodi, problemi e linee evolutive" tenutosi a Roma l'11 Dicembre 2014.

² Si veda al riguardo BCE (2013) che riporta i confronti delle stime provenienti dalle indagini dell'HFCS con quelle di fonte macro.

stimata attraverso vari approcci e successivamente può essere utilizzata per costruire degli stimatori con proprietà asintotiche migliori^{3 4}.

Nell'approccio *model-based* i metodi di aggiustamento sono caratterizzati da due requisiti: 1) un modello (con le relative assunzioni) relativo all'errore di misura, 2) informazioni ausiliarie di varia natura che vengono utilizzate per stimare i parametri di tale modello. Fra i vari modelli che esistono in letteratura, quelli che più si avvicinano ai nostri scopi sono sicuramente i metodi di imputazione. Per una descrizione dettagliata si rimanda ai lavori seminali di Rubin (1978, 1987). Questi metodi sono principalmente impiegati nelle correzioni per le mancate risposte parziali, ma sono facilmente generalizzabili anche al problema dell'errore di misura. La variabile affetta da errore potrebbe essere per alcuni rispondenti considerata non attendibile e dunque imputata con un valore plausibile⁵.

I due approcci hanno comunque dei tratti in comune che rendono talvolta difficile una netta separazione. Ad esempio, la correzione per la non risposta attraverso la ponderazione può anche essere vista come un metodo di imputazione che consiste nel compensare le risposte mancanti con quelle dei rispondenti (che hanno caratteristiche simili). D'altra parte, l'imputazione di valori plausibili in luogo di valori dichiarati dagli intervistati può essere rappresentato come una riponderazione delle unità.

Inoltre, nell'ambito dell'approccio *design-based* è stato recentemente sviluppato un approccio che è anche assistito da modello (*model-assisted approach*), in cui il modello viene usato per descrivere la relazione fra la variabile di interesse e una o più variabili di cui sono noti i totali, alla ricerca di stimatori con migliori proprietà asintotiche (che comunque sono valutate sempre in un contesto *design-based*).

Ciò premesso è comunque possibile sintetizzare alcuni dei pro e contro dei due approcci. Discussioni più approfondite possono trovarsi ad esempio nei lavori di Gelman (2007) e Brick (2013) e le relative discussioni.

Un aspetto che accomuna molti dei metodi presenti in letteratura è l'ipotesi di *missing at random*, secondo la quale le variabili ausiliarie a disposizione sia per la riponderazione sia per l'imputazione contengono tutta l'informazione necessaria per realizzare l'aggiustamento.

³ Deville e Särndal (1992) estendono le tecniche di calibrazione ai totali di variabili quantitative. Fuller et. al (1994) per primi notano che la calibrazione lineare consente implicitamente di aggiustare per la mancata risposta se il modello per la non risposta è lineare. Sulla base di questa intuizione altri lavori hanno introdotto estensioni. Folsom e Singh (2000) trovano una formulazione generale che permettere di includere nella calibrazione anche funzioni diverse da quella lineare. Deville (2000) introduce il concetto di calibrazione generalizzata che permette di utilizzare nella calibrazione anche variabili che spiegano la non risposta ma per le quali non sono disponibili totali a livello di popolazione (come per esempio le informazioni raccolte dagli intervistatori). Kott e Chang (2010), riprendendo un'idea di Deville (2000), propongono di includere la stessa variabile di interesse nella calibrazione generalizzata per correggere la distorsione legata alla non risposta non ignorabile.

⁴ Per una descrizione più approfondita dell'approccio si veda ad esempio Oh e Scheuren (1983). Le proprietà statistiche di questi stimatori sono analizzate in vari lavori. Ad esempio, Little e Vartivarian (2005) mostrano come nel caso le variabili utilizzate sono associate sia alla mancata partecipazione sia alla variabile di interesse, consentono ridurre la distorsione e la varianza degli stimatori. Più di recente Kott e Liao (2012) presentano stimatori assistiti da modello che consentono una duplice protezione contro la distorsione legata alla non risposta.

⁵ Per un recente esempio di utilizzo di questi modelli di imputazione si veda il lavoro di Peytchev (2012) che utilizza questa tecnica per correggere congiuntamente sia la non risposta che l'errore di misura.

La differenza tra i metodi emerge chiaramente in presenza di trattamenti che riguardano più variabili. Se, come risulta ad esempio da Neri e Zizza (2010), le risposte sulle forme di investimento finanziario e sui redditi da lavoro autonomo sono soggette a forme di *under-reporting* differenziate, l'aggiustamento basato su modelli permette di usare modelli diversi, realizzando una più efficace correzione della distorsione.

Nel caso di indagini complesse, l'imputazione di una variabile comporta la necessità di ricalcolare le variabili derivate, come ad esempio può accadere nel caso dell'imputazione di una certa forma di ricchezza che richiede, oltre alla modifica dell'aggregato relativo al patrimonio, anche una revisione coerente dei redditi da capitale. L'utilizzo di modelli per l'imputazione di più variabili pone in generale il problema della correlazione con gli altri fenomeni rilevati e tra gli stessi fenomeni imputati⁶.

Nel caso delle correzioni basate sui pesi, le coerenze interne tra le variabili sono preservate per definizione, ciò che costituisce un indubbio vantaggio soprattutto per le analisi micro. La correzione di una distribuzione tramite i pesi comporta però in generale la modifica delle distribuzioni di altri caratteri, che vanno pertanto tenute sotto controllo.

L'approccio basato su modelli, lavorando a livello di singola osservazione consente in generale di ottenere stime con minore varianza di quelle che si avrebbero modificando i pesi. Si pensi, ad esempio, al caso di attività finanziarie fortemente concentrate nelle mani di un numero limitato di famiglie e oggetto di *under-reporting*. Nel campione le unità che possono essere oggetto di riponderazione possono essere poche e le relative stime divenire eccessivamente variabili. In casi di questo tipo, non è raro che le procedure iterative di modifica dei pesi non convergano, in quanto non riescono ad allineare la variabile di interesse, condizionatamente ad altre variabili rilevanti, come ad esempio le caratteristiche socio-demografiche.

I metodi basati su modello forniscono probabilmente maggiori opportunità di presentare aggiustamenti che si allontanano dall'assunzione di *missing at random* in quanto occorre specificare e stimare un modello esplicito e dunque ci sono maggiori possibilità di rimuovere alcune ipotesi. Gli stimatori basati su modello possono avere problemi di robustezza dei risultati di fronte a violazione delle ipotesi del modello, ipotesi che normalmente non sono testabili. Holt e Smith (1979) illustrano come invece la robustezza, ossia la protezione da erronee specificazioni del modello, sia una dei punti di forza della post-stratificazione.

Secondo Lohr (2007) l'uso di stimatori basati su modello sarebbe meno raccomandabile per i produttori ufficiali in quanto richiedono un numero maggiore di scelte da difendere rispetto al caso degli stimatori da disegno. Le procedure da disegno, al contrario, sono ritenute più semplici da usare e accessibili per una grande varietà di utilizzatori.

L'uso dei pesi permette inoltre un facile allineamento dei risultati dell'indagine a quelli di fonti esterne quali il censimento attraverso tecniche di calibrazione. Gli stimatori ottenute tramite queste tecniche hanno in generale proprietà statistiche desiderabili: permettono nella maggior parte dei casi di aumentare la precisione degli stimatori e se le variabili usate per la calibrazione sono correlate anche con la non risposta, consentono di ridurre la distorsione ad essa collegata (Little, Vartivarian 2005).

⁶ Una soluzione possibile è quella di ricorrere a imputazioni secondo uno schema sequenziale, in modo da assicurare la coerenza nell'archivio finale.

È tuttavia utile ricordare che la scelta del metodo di aggiustamento è in pratica guidata dalle informazioni disponibili e dall'obiettivo della correzione. Se ad esempio sono disponibili come informazioni ausiliarie solo dei totali noti ad esempio di fonte censuaria, l'uso di metodi da disegno rappresenta la scelta più ragionevole. Se invece sono disponibili anche informazioni ausiliarie a livello di singola osservazione, o di celle, allora anche i metodi basati su modello potrebbero essere impiegati.

I due approcci non devono comunque essere considerati alternativi. L'applicazione alla base del presente lavoro rappresenta un potenziale esempio di come i due approcci possano essere utilizzati congiuntamente per allineare l'indagine IBF a una molteplicità di fonti esterne.

3. Precedenti aggiustamenti sui dati IBF

La studio della discrepanza tra le stime del campione IBF e le corrispondenti stime macro è oggetto di studio da molti decenni. Nel Bollettino del 1970 (Banca d'Italia, 1970) Adalberto Ulizzi, illustrando i risultati dell'indagine sul 1968, osservava che *“Tra i vari errori menzionati (gli errori non campionari n.d.r.), un particolare rilievo assumono quelli riconducibili alla reticenza degli intervistati sul possesso di attività finanziarie. L'esperienza acquisita in numerose indagini, alcune delle quali specifiche sull'argomento, ha permesso di rilevare una notevole reticenza da parte delle famiglie a fornire informazioni sul possesso di attività finanziarie (...). Per il risparmio e il reddito la collaborazione degli intervistati è generalmente migliore, essendo minore l'avversione a fornire dati sui flussi che non sulle consistenze.”*

In quegli anni, Ulizzi aveva condotto una ricerca sui comportamenti di *under-reporting* delle attività finanziarie utilizzando tecniche di *exact matching*. Ulizzi aveva cercato di intervistare circa 900 soggetti di cui era noto per altra fonte il valore dei titoli detenuti. Di questi soggetti, il 30 per cento non aveva aderito all'indagine. Il valore medio delle attività finanziarie detenute dai soggetti non intervistati era di poco più elevato di quello vero dei soggetti intervistati. Questo risultato è di grande interesse, in quanto suggerisce che l'impatto della mancata partecipazione sulle stime complessive sia relativamente contenuto. Il valore medio dei titoli dichiarati dagli intervistati era tuttavia assai più basso di quello realmente detenuto dagli stessi soggetti (la rilevazione coglieva poco più del 15 per cento del valore medio dei titoli). Il grosso della discrepanza tra i valori veri e quelli dichiarati derivava dal fatto che oltre il 60 per cento degli intervistati negava il possesso dei titoli (*non-reporting*). Tra i restanti intervistati, buona parte dichiarava un valore inferiore a quello vero (*under-reporting*) o si rifiutava di rispondere. I fenomeni di *non-reporting* e *under-reporting* risultavano più marcati tra le famiglie più ricche.

Dopo quella prima esperienza, numerosi altri studi sono stati condotti su questi fenomeni⁷, e in particolare sulla mancata partecipazione all'indagine e sulla sottostima nella dichiarazione degli intervistati.

L'indagine si è tradizionalmente posta l'obiettivo di rappresentare la popolazione residente in Italia. Poiché la selezione delle unità da intervistare avviene dalle liste anagrafiche che non sempre presentano una totale accuratezza, è possibile che vi sia una sottorappresentazione nel campione di alcune fasce di popolazione, come ad

⁷ Diversi studi hanno sono stati condotti sul confronto delle stime dell'indagine con quelle desunte da altre fonti (si veda, ad esempio, Brandolini, 1999 e Bonci Marchese e Neri, 2005). Nel seguito faremo riferimento ai soli lavori che propongono forme di aggiustamento delle stime campionarie. Per una rassegna si veda D'Alessio e Ilardi, 2013.

esempio i cittadini di recente immigrazione, che non sempre rispettano l'obbligo di comunicare cambi di residenza in Italia o il ritorno in patria.

È tuttavia assai più probabile che i maggiori effetti sulla correttezza delle stime che coinvolgono la composizione del campione derivino dalle mancate interviste⁸.

Le mancate interviste, siano esse dovute a rifiuto esplicito o a irreperibilità, rappresentano un problema nelle indagini statistiche in quanto possono condurre a campioni nei quali i segmenti di popolazione meno disposti a collaborare (o meno raggiungibili) possono essere sottorappresentati, producendo stime distorte (*selectivity bias*). Poiché nella stima si utilizzano solo le informazioni di coloro che hanno partecipato alla rilevazione, la distorsione cresce all'aumentare della mancata risposta e del divario tra il valore medio della variabile oggetto di studio per i rispondenti e per i non rispondenti.

Per limitare gli effetti della mancata partecipazione nell'IBF sono adottate varie misure (Banca d'Italia, 2014). In primo luogo, le famiglie che non è possibile intervistare sono sostituite con altre estratte con criteri casuali negli stessi comuni. Ciò consente di tenere sotto controllo una potenziale fonte di distorsione, cioè quella che origina dalla relazione tra territorio e caratteristiche familiari. In secondo luogo, al termine della rilevazione, viene effettuata una stratificazione a posteriori sulla base di alcune caratteristiche individuali dei soggetti intervistati, che consente di riequilibrare, all'interno del campione, il peso dei diversi segmenti della popolazione. Essa viene condotta imponendo al campione finale, tramite tecniche di *raking*, l'allineamento alle caratteristiche della popolazione per quanto riguarda il sesso, la classe di età, l'area geografica e la dimensione del comune di residenza.

È tuttavia possibile che permangano elementi di distorsione, cioè che a parità di quelle caratteristiche che sono tenute sotto controllo si continuino a selezionare con minore probabilità gruppi particolari di famiglie (ad esempio le più ricche). Questo aspetto non è facilmente valutabile poiché le informazioni relative ai non rispondenti non sono disponibili nell'indagine.

In un esperimento condotto negli anni novanta da Cannari e D'Alessio (1992), per l'insieme di famiglie panel, che aveva cioè già partecipato a una o più indagini, sono state valutate le caratteristiche di coloro che interrompevano la loro collaborazione rispetto al gruppo di quelle che la proseguivano (*attrition*). Il comportamento di non risposta, osservato solo sulla parte panel del campione, veniva poi estrapolato a tutte le famiglie, consentendo di valutare la sottostima del reddito dovuta alle mancate interviste, pari a circa il 5 per cento.

Diversi altri metodi sono stati sperimentati nel corso del tempo per approfondire questo aspetto, e in particolare per cercare di avere informazioni anche sulle famiglie che non sono mai state intervistate (non panel), per le quali non è possibile condurre studi come quello appena citato.

L'esame dello sforzo necessario a ottenere l'intervista (ad esempio il numero di visite o il numero di contatti telefonici volti a convincere le famiglie a partecipare) può aiutare a capire quali sono le famiglie più difficili da intervistare e correggere i coefficienti di ponderazione individuando, tramite dei modelli, la effettiva probabilità di partecipazione di ogni famiglia intervistata. Elaborando informazioni di questo tipo, D'Alessio e Faiella (2002) mostrano che le stime di reddito e ricchezza sono più elevate quando si tiene conto di questi aspetti. Si osserva, infatti, che i valori medi di reddito e

⁸ D'Alessio e Faiella (2002).

ricchezza delle famiglie differiscono in funzione del livello di facilità con cui sono intervistate. Posto pari a 100 il valore medio del totale campione, coloro che collaborano dopo aver ricevuto la telefonata della società che ha convertito un rifiuto in partecipazione presentano un valore medio di circa 120 in termini di reddito e di circa 130 in termini di ricchezza netta; coloro che non sono stati trovati in casa hanno invece redditi e ricchezza solo di pochi punti percentuali o inferiore alla media.

Sempre nello stesso lavoro gli autori, basandosi sulle informazioni raccolte presso circa 2000 famiglie selezionate per l'intervista IBF, e a cui erano state abbinare in modo anonimo e nel rispetto della riservatezza alcune informazioni di natura bancaria, mostrano che la non risposta non è casuale ma caratterizza più frequentemente le famiglie più ricche. La distorsione osservata dagli autori è maggiore per le attività finanziarie (le stime corrette sono tra il 15 e il 30 per cento più elevate di quelle non corrette per tale fenomeno) che per il reddito (la cui sottostima oscilla tra il 5 e il 14 per cento), probabilmente a causa della maggiore asimmetria che caratterizza la distribuzione della ricchezza.

Neri e Ranalli (2011), utilizzando i risultati di una rilevazione telefonica supplementare condotta sui non rispondenti dell'IBF, segnalano una maggiore difficoltà di ottenere le interviste dalle famiglie più agiate, e propongono una corrispondente correzione dei pesi campionari. Il risultato è confermato anche dal recente studio di D'Alessio e Iezzi (2014).

L'altro filone di interesse per la correzione delle stime campionarie è quello che riguarda l'*under-reporting*, cioè la mancata dichiarazione o la sottovalutazione di beni immobili, attività finanziarie o forme di reddito.

Con riferimento alle componenti della ricchezza, il primo lavoro che propone forme di correzione delle stime campionarie dell'IBF è quello di Cannari e D'Alessio (1990). Il lavoro mostra che, mentre le abitazioni utilizzate come residenza era stimate bene nell'indagine, una sottostima caratterizzava le seconde case date in affitto, che non risultavano coerenti con le informazioni desunte, sempre nell'indagine, da parte degli affittuari. Inoltre, dal confronto delle stime totali del censimento con quelle desunte dall'indagine, risultava che anche le abitazioni utilizzate per le vacanze erano sottostimate. Gli autori proponevano pertanto un metodo per correggere il numero di abitazioni dichiarate dai proprietari nell'indagine, imputando alle famiglie del campione ulteriori abitazioni sulla base di probabilità stimate di averne⁹.

Cannari, D'Alessio Raimondi e Rinaldi (1990) conducono un esperimento di *matching* statistico tra le risposte fornite dagli intervistati sulle attività finanziarie in IBF e quelle fornite da clienti bancari, intervistate nel corso di un'indagine della loro stessa banca). Ipotizzando l'assenza di *under-reporting* nei dati di fonte bancaria, gli autori stimano, tramite dei modelli statistici, sia la probabilità di possedere i vari strumenti finanziari sia gli ammontari che le varie tipologie familiari dovrebbero mediamente possedere. Ponendo a confronto queste stime con quanto risulta in IBF lo studio mostra come per alcune tipologie familiari (i più poveri e i meno istruiti), risulta rilevante la mancata dichiarazione del possesso degli strumenti finanziari (*non-reporting*), mentre

⁹ La distribuzione del numero di abitazioni (ad esclusione di quella di residenza) viene modellata attraverso la distribuzione di Poisson il cui valor medio dipende da un vettore di caratteristiche osservabili (come ad esempio l'età, il titolo di studio e il genere del capofamiglia, il reddito della famiglia, il comune di residenza, ecc.). Con i dati dell'indagine vengono stimate le probabilità che le famiglie detengano un certo numero di seconde case. Queste probabilità vengono poi utilizzate per imputare le abitazioni mancanti (ossia la differenza fra quelle stimate dall'indagine e un qualche dato amministrativo ritenuto più affidabile come i risultanti dal censimento).

per altre tipologie sia più rilevante la sottovalutazione degli ammontari posseduti. Quest'ultima forma risulta, nel complesso, la maggiore responsabile della sottostima rispetto ai dati aggregati. Le stime corrette per l'*under-reporting*, ottenute con approccio *design-based*, risultano circa il doppio di quelle non corrette, ma una certa distanza rispetto ai dati macro rimane. Nonostante le revisioni in alcuni casi anche profonde, i rapporti relativi tra le varie categorie di famiglie non vengono modificati in maniera consistente. Cannari e D'Alessio (1993), adottando una metodologia più articolata rispetto al lavoro precedente, di tipo *model-based*, mostrano, inoltre, che l'indice di concentrazione di Gini non è influenzato in modo rilevante dalla correzione dei dati, che passa da 0,644 a 0,635 con riferimento al 1991.

Lo studio di Brandolini et al. (2004) propone uno studio della ricchezza degli italiani, dopo aver provveduto a correggere la sottostima delle poste reali e finanziarie.

L'esperimento di *matching* statistico tra i dati IBF e quelli di una banca commerciale è stata replicata in anni più recenti (D'Aurizio et al., 2006). In media le stime corrette delle attività finanziarie risultano più del doppio di quelle originarie, raggiungendo l'85 per cento degli ammontari macroeconomici. L'aggiustamento è più consistente per le famiglie la cui persona di riferimento è poco istruita o anziana. Il lavoro effettua una correzione anche delle passività finanziarie, rivalutandole in media di circa il 40 per cento.

Il lavoro di Neri e Monteduro (2013) propone un esperimento di correzione della ricchezza immobiliare basato sia sui dati aggregati sia sulle distribuzioni delle quote di proprietà disponibili dalla fonte fiscale. I risultati mostrano come l'indagine tenda a sottostimare il numero dei contribuenti che possiedono o un numero di quote di abitazioni molto piccolo o un numero molto elevato (oltre 5). Dopo aver corretto i dati campionari mediante l'utilizzo di modelli, allineando i dati alle informazioni amministrative, la ricchezza immobiliare delle famiglie risulta complessivamente superiore di circa un quarto a quella dichiarata nell'indagine. I risultati non sembrano invece mostrare che la sottostima comporti significative distorsioni nella misura della concentrazione della ricchezza e nell'associazione fra ammontare posseduto e alcune caratteristiche socio demografiche delle famiglie.

Per quanto riguarda l'*under-reporting* del reddito, il lavoro di Cannari e Violi (1995), traendo spunto dal lavoro condotto sui dati inglesi da Pissarides e Weber (1989), si basa su un metodo «indiretto» di ricostruzione del reddito «vero». Il metodo si fonda sull'ipotesi che il reddito sia rilevato correttamente per alcuni gruppi della popolazione e che alcune componenti della spesa per consumi siano rilevate senza errore (sistematico) per tutti i gruppi della popolazione. In accordo con tale ipotesi, la relazione tra la spesa per consumi (alimentari) e il reddito può essere stimata utilizzando il sottoinsieme del campione che dichiara accuratamente il proprio reddito. Per i restanti sottoinsiemi del campione, la relazione può essere invertita, consentendo la ricostruzione del reddito «vero» a partire dalla spesa per consumi¹⁰.

Questo approccio è stato ripreso da Neri e Zizza (2010) che utilizzano come indicatore il valore dell'abitazione di residenza (ipotizzato non affetto da *under-reporting*) in luogo della spesa per consumi. Il rapporto tra valore dell'abitazione e reddito viene dapprima stimato sui dipendenti pubblici e poi applicato ai lavoratori indipendenti, per derivarne un reddito da lavoro coerente: la correzione delle stime che ne consegue è rilevante (circa il 36 per cento del reddito). Gli autori sviluppano poi

¹⁰ Una procedura simile si trova in Hurst, Li, e Pugsley (2010).

delle correzioni per altre componenti del reddito, in buona parte basate su rielaborazioni dei lavori prima esposti.

Cifaldi e Neri (2013), applicando forme di correzione ai dati di reddito e consumo rilevati in IBF discutono le distorsioni che l'*under-reporting* differenziale tra i due fenomeni implica nell'esame della propensione al risparmio.

4.La correzione dell'*under-reporting* e della mancata risposta

Come si è visto, nell'IBF le stime campionarie di alcuni fenomeni (ad esempio il reddito o la ricchezza) quando vengono comparate alle stime macro, mostrano la presenza di ampie sottostime. Questi divari sono imputabili ai problemi di mancata risposta ma soprattutto a quelli di *under-reporting*.

In questo paragrafo si mostreranno alcuni metodi che possono essere impiegati per correggere i dati campionari da queste fonti di distorsione. Le correzioni si fondano talvolta sulla disponibilità di dati micro ritenuti esenti da *under-reporting*; in altri casi, i metodi assumono che i totali nazionali dei fenomeni oggetto di studio siano disponibili ed esenti da errore, e individuano le stime campionarie compatibili con le grandezze date, minimizzando una funzione di distanza applicata ai pesi campionari.

Di seguito si mostreranno diversi metodi di aggiustamento che è possibile applicare ai dati di reddito e ricchezza rilevati in IBF, lasciando al paragrafo successivo l'analisi comparativa dei risultati che ne scaturiscono e il giudizio sui metodi impiegati.

4.1 La correzione proporzionale - C₁

Il caso più elementare di correzione, e che qui utilizzeremo come benchmark e indicheremo con C₁, è quello per cui i valori delle grandezze campionarie y_i vengono semplicemente riproporzionate sulla base del coefficiente $k = Y^T / y^T$ che rapporta il valore totale noto dalla popolazione a quello stimato dal campione.

Si tratta di un modello di *under-reporting* molto semplice, nel quale si ipotizza che per l'individuo i -esimo l'ammontare dichiarato y_{id} sia una frazione costante dell'ammontare vero y_i , più una componente di errore e_i che si ipotizza a media nulla:

$$y_{id} = y_i/k + e_i \quad (1)$$

Nonostante la sua semplicità, l'applicazione di questo modello può risultare utile, in particolare se si adotta il riproporzionamento a livello delle singole componenti di reddito e di ricchezza. Gli aggregati di reddito e ricchezza che si ottengono come somma delle singole componenti riproporzionate possono fornire delle indicazioni utili sul possibile impatto dell'*under-reporting* sulle medie o sugli indici di concentrazione. Per il reddito, ad esempio, è possibile correggere disgiuntamente nell'indagine il reddito da lavoro dipendente (YL), da pensione e altri trasferimenti (YT), da attività autonoma (YM) e da capitale (YC); per la ricchezza è possibile tenere conto delle diverse componenti (attività reali AR, attività finanziarie AF e passività finanziarie PF), facendo emergere una maggiore sottostima delle componenti finanziarie rispetto a quelle reali.

Questo stimatore non corregge in nessun modo per il *non-reporting*, ovvero la mancata dichiarazione del possesso di un certo strumento o fonte di reddito, visto che solo gli importi dichiarati sono riproporzionati.

4.2 La correzione sulla base del giudizio dell'intervistatore - C₂

Nell'IBF per disporre di elementi di giudizio sull'*under-reporting*, fenomeno che per sua natura risulta difficilmente investigabile a livello di singola famiglia, viene

richiesto agli intervistatori di formulare, immediatamente dopo l'intervista, un giudizio sulla presunta attendibilità delle risposte, basando tale giudizio sulla rispondenza tra i dati forniti e gli elementi oggettivi a loro disposizione (zona e tipologia dell'immobile di residenza, tenore di vita desumibile dagli arredi, ecc.). Nel 1993 e nel 1995 si trattava di domande che ammettevano risposte qualitative (per niente, poco, abbastanza, molto); dal 1998 in poi i giudizi degli intervistatori sono stati espressi con un punteggio compreso tra 1 (per nulla attendibile) e 10 (totalmente attendibile).

Il livello di attendibilità così misurato risulta, nel complesso, soddisfacente per tutti gli anni esaminati (tavola 1): nel 1993 e nel 1995 la quota di interviste le cui informazioni sul reddito e la ricchezza sono giudicate abbastanza o molto soddisfacenti sono comprese tra l'85 e il 90 per cento; per gli anni successivi, considerando le frequenze di tutti i punteggi uguali o superiori al 6, si ottiene una quota analoga. Il livello medio appare crescente negli ultimi due anni.

Ciò nonostante, il livello di attendibilità non è omogeneo all'interno del campione. Giudizi più elevati caratterizzano stabilmente le famiglie con capofamiglia non solo lavoratore dipendente, ma anche quelle con elevato titolo di studio e residente al Centro e al Nord. Si tratta di evidenze che sembrano integrare quelle emerse in precedenza e che possono essere utilizzate per correggere le stime campionarie.

Tavola 1

Veridicità delle risposte sul reddito e la ricchezza, 1993-2012

(valori percentuali, punteggi in decimi)

Anno	Giudizio qualitativo sulla veridicità delle risposte (secondo l'intervistatore)				
	Per niente	Poco	Abbastanza	Molto	Totale
1993.....	0,9	9,4	50,5	39,2	100,0
1995.....	1,0	11,7	53,3	34,1	100,0

Punteggio in decimi sulla veridicità delle risposte (secondo l'intervistatore)												
Anno	1	2	3	4	5	6	7	8	9	10	Totale	Media
1998.....	1,5	1,3	1,7	2,7	6,5	12,3	16,5	22,0	17,5	18,1	100,0	7,6
2000.....	0,6	0,7	1,3	3,1	6,7	11,8	16,6	20,0	19,7	19,5	100,0	7,7
2002.....	0,7	1,2	1,3	2,2	6,3	12,3	17,2	21,1	18,2	19,6	100,0	7,7
2004.....	1,0	1,4	1,2	2,6	7,0	12,1	17,8	22,0	16,9	18,0	100,0	7,6
2006.....	0,3	0,7	1,1	2,4	6,3	13,1	18,7	23,5	17,8	16,1	100,0	7,6
2008.....	0,7	0,8	1,0	2,3	6,1	13,4	18,8	23,8	19,7	13,5	100,0	7,6
2010.....	0,6	0,5	0,7	1,6	4,0	8,6	15,7	22,8	26,0	19,6	100,0	8,0
2012.....	0,3	0,4	0,6	1,0	3,7	8,8	13,3	22,0	25,6	24,3	100,0	8,2

Si può dunque procedere nella stima di un modello di questo tipo:

$$\log(y_{id}) = x_i \beta + \gamma v_i + e_i \quad (2)$$

dove x_i sono variabili di controllo e v_i è il punteggio assegnato dall'intervistatore alla veridicità delle risposte fornite sul reddito e la ricchezza. Una volta individuato il contributo attribuibile alla componente V , è possibile stimare il reddito che la famiglia avrebbe dovuto dichiarare in presenza di un indice di veridicità v_i pari al suo massimo.

Secondo questo modello, il giudizio degli intervistatori sembra effettivamente cogliere alcuni elementi importanti alla base dei comportamenti di *under-reporting*. In generale la rivalutazione del reddito è più sensibile per gli autonomi rispetto ai pensionati e ai dipendenti. I valori medi aggiustati del reddito pur avvicinandosi ai valori noti dalle fonti aggregate, rimangono però piuttosto lontani da essi.

Uno stimatore alternativo (di seguito indicato come C_2) che, tenendo conto del giudizio degli intervistatori, colmi l'intero divario rispetto alle stime macro, può essere definito come:

$$y_{id} = y_i / k_i + e_i \quad (3)$$

dove k è una funzione inversa della giudizio dell'intervistatore v_i

$$k_i = 1 + (10 - v_i) \alpha \quad (4)$$

Quando v_i assume il suo massimo (10) non vi è correzione; quando scende sotto questo livello vi è una correzione dell'importo dichiarato che è funzione della distanza del punteggio dal suo massimo. Il coefficiente α viene calibrato in modo tale che sia soddisfatta la condizione per cui la stima campionaria del totale y^T coincida con il totale noto dalle fonti macro Y^T .

Come nel caso precedente, lo stimatore non corregge per il *non-reporting*.

4.3 Correzioni su singoli fenomeni – C_3

Alcune volte sono disponibili informazioni esterne che consentono di migliorare la stima di singoli fenomeni. Nel seguito si presentano le correzioni per la mancata risposta, per i redditi dei lavoratori autonomi, per le abitazioni diverse da quelle di residenza e per le attività finanziarie. Queste correzioni sono rispettivamente indicate come C_{3A} , C_{3B} , C_{3C} , C_{3D} , e considerate complessivamente con C_3 .

4.3.1 Peso di non risposta – C_{3A}

L'aggiustamento per la non risposta è basato sul lavoro Neri, Ranalli (2011). La metodologia prevede una correzione dei pesi campionari del tipo:

$$w_c^{(NR)} = w_c^{(DES)} \alpha_c \quad (5)$$

dove $w_c^{(NR)}$ rappresenta il peso aggiustato per mancata risposta per le famiglie nella classe c , $w_c^{(DES)}$ è il peso da disegno e α_c è il fattore correttivo ossia l'inverso della probabilità stimata per quelle famiglie di partecipare all'indagine.

Per le famiglie panel vengono utilizzate le informazioni disponibili dall'indagine precedente unite a quelle sui tentativi di contatto effettuati dagli intervistatori e sul loro esito finale (raccolte nel foglio contatti). La probabilità di partecipazione è stata stimata attraverso un modello logistico utilizzando come covariate l'area geografica e l'ampiezza del comune in cui la famiglia risiede, la classe di ricchezza, la classe di reddito e il clima dell'intervista precedente (basato sul giudizio dell'intervistatore). Le probabilità stimate dal modello sono state poi raggruppate in decili e a ciascuna famiglia è stato attribuita la probabilità media del decile di appartenenza.

Per le famiglie non panel sono state invece utilizzate delle informazioni raccolte su un campione di non rispondenti¹¹. Nelle rilevazioni sul 2008, sul 2010 e sul 2012, l'indagine principale è stata seguita da una indagine telefonica su un campione di circa 500 non rispondenti. Il campione è stato selezionato fra coloro per cui è stato possibile individuare un recapito telefonico e che si sono dimostrati disponibili a rispondere ad un

¹¹ Queste informazioni non sono al momento utilizzate nella costruzione dei pesi ufficiali rilasciati per l'indagine. Viene invece adottata una correzione per le famiglie panel simile a quella descritta nel paragrafo. Sul punto si veda l'appendice metodologica del supplemento I bilanci delle famiglie italiane nel 2012.

breve intervista telefonica. In totale, cumulando tutte le rilevazioni, si tratta di 863 famiglie (non panel). Per ciascuna indagine, comprese quelle sugli anni precedenti il 2008, questo campione è stato utilizzato accodandolo a quello dei rispondenti non panel intervistati. È stato poi stimato un modello logistico per la probabilità di appartenere al gruppo dei non rispondenti. Le covariate utilizzate sono l'area geografica e l'ampiezza del comune, l'età, la condizione professionale, il titolo di studio del capofamiglia, il possesso dell'abitazione di residenza, il numero di componenti e di percettori. In fase di stima al campione dei rifiuti è stato attribuito un peso in funzione del numero totale dei rifiuti risultanti dal foglio contatti dell'anno in questione.

Il metodo di correzione adottato si basa necessariamente su una serie di ipotesi semplificatrici. In primo luogo si è ipotizzato che il meccanismo di mancata risposta sia funzione delle sole caratteristiche osservabili (*missing at random*). In secondo luogo è stato ipotizzato che la mancata risposta e gli errori di misurazione descritti in questo paragrafo siano fenomeni indipendenti. Di conseguenza, l'aggiustamento descritto in questo paragrafo è realizzato al momento in modo indipendente dagli altri aggiustamenti.

4.3.2 La correzione del reddito degli autonomi – C_{3B}

Come si è già detto nel paragrafo precedente, per valutare il livello di *under-reporting* è possibile assumere come *benchmark* le dichiarazioni sul reddito di un gruppo di intervistati (ad esempio i lavoratori dipendenti), ipotizzando che queste dichiarazioni siano esenti da *under-reporting*. Se per l'intero campione si dispone di altri indicatori, legati al reddito ma non soggetti ad *under-reporting*, è possibile risalire indirettamente a una stima del reddito che tali indicatori suggeriscono per i soggetti che invece tendono a sottostimarli.

Nel seguito considereremo come variabile in grado di segnalare l'*under-reporting* degli autonomi il valore dell'abitazione di residenza che, essendo le interviste svolte di persona, non può facilmente essere nascosto nei confronti dell'intervistatore; è quindi ragionevole assumere che il suo valore non sia sistematicamente sottostimato, o almeno sia sottostimato in misura inferiore rispetto al reddito.

La misura dell'*under-reporting* da parte delle famiglie di lavoratori autonomi può essere stimata per confronto rispetto alle altre famiglie (per le quali si ipotizza l'assenza di *under-reporting*), stimando un modello del tipo:

$$\log(V) = \alpha + \beta \log(Y_d) + \gamma A + \theta X \quad (4)$$

dove cioè si ipotizza che il logaritmo dell'indicatore V utilizzato sia funzione di una costante α , del logaritmo del reddito dichiarato Y_d (che nel caso del gruppo di controllo coincide con il reddito effettivo Y), di altre caratteristiche di controllo (sesso, età, ecc.) raccolte nella matrice X e di una *dummy* A riferita alle famiglie di lavoratori autonomi. Nell'ipotesi che i due gruppi di famiglie abbiano gli stessi comportamenti in rapporto all'indicatore V , la quota di reddito dichiarata dagli autonomi π può essere stimata dall'equazione (4) come:

$$\pi = Y_d/Y = \exp(-\gamma/\beta) \quad (5)$$

Il coefficiente π non è per costruzione limitato all'intervallo 0-1, anche se nelle stime effettuate questo si verifica pressoché costantemente.

La tavola 2 riporta, nella prima colonna, i coefficienti di π stimati per le 3 aree geografiche e per il complesso del campione. Il coefficiente π segnala un livello di *under-reporting* di circa il 35 per cento, leggermente più elevato nel Mezzogiorno.

Per tenere conto di possibili errori di misura nelle variabili indipendenti, si è proceduto a una stima tramite variabili strumentali; secondo queste nuove stime l'*under-reporting* del reddito da parte degli autonomi sarebbe tra il 10 e il 20 per cento. Il maggiore *under-reporting* nelle regioni meridionali non è confermato da queste stime. Nel seguito si considererà una rivalutazione intermedia tra questi valori, del 20 per cento, uniforme a livello nazionale¹².

Tavola 2

Coefficienti di reporting

(valori percentuali)

Valore dell'abitazione	Logaritmo	Log (IV)
Nord	0,7369	0,8438
Centro	0,7873	0,8717
Sud e Isole	0,6276	0,9087
Totale Italia	0,6761	0,8709

4.3.3 *La correzione degli immobili non di residenza – C_{3C}*

Le famiglie italiane detengono una parte rilevante della loro ricchezza sotto forma di immobili. Una quota rilevante di questi immobili è costituita dalle abitazioni di residenza, che l'indagine IBF coglie in misura piuttosto accurata, presentando divergenze moderate sia con le stime di altre indagini campionarie come EU-SILC, sia con i dati di censimento.

Gli immobili non di residenza posseduti dalle famiglie sono invece sottostimati nell'IBF. Una prima evidenza di ciò si ritrova svolgendo una verifica di coerenza delle risposte degli intervistati dell'IBF (Cannari e D'Alessio, 1990). È infatti possibile confrontare i dati forniti dalle famiglie proprietarie di abitazioni concesse in affitto con quelle delle famiglie affittuarie presso abitazioni di proprietà di altre famiglie. In assenza di *under-reporting*, le due stime dovrebbero essere vicine, a meno di fluttuazioni campionarie.

In realtà, le abitazioni dichiarate dai proprietari sono largamente sottostimate; nel corso dei vari anni ne risultano tra un milione e un milione e mezzo, a fronte di un numero di famiglie affittuarie di abitazioni di proprietà di altre famiglie pari a circa 3

¹² Neri e Zizza (2010), con un metodo leggermente diverso, ottenevano una rivalutazione dei redditi da lavoro autonomo di circa il 36 per cento; Cannari e Violi (1995) stimavano invece incrementi di circa il 25 per cento.

milioni. Secondo questa stima, le abitazioni date in affitto sarebbero dichiarate nell'indagine nella misura di circa il 30-40 per cento (tavola 3)¹³.

Tavola 3

Abitazioni dichiarate dai proprietari e dagli affittuari, 1991-2012

(valori percentuali, punteggi in decimi)

Anno	Famiglie in affitto presso abitazioni di proprietà di altre famiglie (a)	Abitazioni dichiarate dai proprietari in affitto da altre famiglie (b)	Quota percentuale (b) / (a)
1991.....	3.291.258	983.777	29,9
1993.....	3.220.253	1.391.772	43,2
1995.....	3.360.512	1.533.344	45,6
1998.....	3.255.218	1.112.374	34,2
2000.....	3.182.180	1304.149	41,0
2002.....	2.970.913	978.709	32,9
2004.....	3.304.629	967.758	29,3
2006.....	3.360.706	861.826	25,6
2008.....	3.320.834	1.529.607	46,1
2010.....	3.646.078	1.205.595	33,1
2012.....	3.683.863	1.210.284	32,9
Media	-	-	35,8

Il confronto dei dati relativi alle abitazioni dichiarate dalle famiglie con i dati di censimento segnala simili misure di *under-reporting* (tavola 4). Secondo i dati dell'indagine sul 1991, ci sarebbero 15,3 milioni di abitazioni di proprietà delle famiglie; il censimento ne stima 22,9 milioni¹⁴. Se consideriamo che in quell'anno 12,4 milioni sono le abitazioni di residenza in proprietà, se ne deduce che la quota di case dichiarate, escludendo quelle di residenza che presumibilmente non sono affette da questo fenomeno, è di poco meno del 30 per cento. La stessa comparazione nel 2001 (con i dati dell'indagine del 2002) conduce a una quota dichiarata di case diverse da quelle di residenza pari al 35 per cento. Simili livelli di *under-reporting* richiedono trattamenti adeguati¹⁵.

Sulla base di evidenze di questo tipo Cannari e D'Alessio (1990) hanno sviluppato un metodo che consente di imputare gli immobili che risultano mancanti sul totale alle famiglie che con maggiore probabilità si può presumere possano esserne proprietarie¹⁶.

¹³ La ripartizione di questo indicatore per area geografica mostra valori più elevati per il Nord e il Centro rispetto al Sud e Isole. Poiché secondo i dati dell'indagine circa il 90 per cento degli immobili posseduti dalle famiglie è collocato nella stessa area geografica di residenza (la quota sale al 98 per cento per le abitazioni date in affitto a famiglie), è presumibile che il divario osservato sia da attribuire al più elevato livello di *under-reporting* che caratterizza le famiglie meridionali.

¹⁴ È possibile che una parte del divario sia da attribuire alla presenza di abitazioni in usufrutto o in uso gratuito.

¹⁵ Si veda ad esempio Cannari e D'Alessio (1990) e Brandolini, Cannari, D'Alessio e Faiella (2004).

¹⁶ Il metodo ipotizza che la distribuzione del numero di immobili (escluse le abitazioni di residenza) segua la distribuzione di Poisson.

Abitazioni dichiarate nell'indagine e dati di censimento, 1991-2012

Anno	Stima su dati indagine				Stima su dati di censimento (*)	Quota percentuale di immobili dichiarati (c) / (d)
	Residenza principale in proprietà (a)	Altre case in proprietà (b)	Totale case in proprietà (c) = (a) + (b)	Per memoria: usufrutto o uso gratuito	Abitazioni di proprietà delle famiglie (d)	
1991	12.791.339	3.181.017	15.972.357	2.020.510	22.958.865	69,6
2002	14.825.485	3.823.484	18.648.969	2.151.803	25.257.775	73,8

(*) Negli anni di censimento, la stima è stata effettuata attribuendo alle famiglie, per le case non occupate, la stessa quota che risulta per le case occupate.

Il metodo (C₃) viene qui applicato imputando la differenza tra il numero degli immobili dichiarati in IBF e quelli risultanti dai censimenti, opportunamente interpolati per gli anni intermedi e aggiornati con altre informazioni per gli anni successivi all'ultimo censimento disponibile (Banca d'Italia, 2012). Ovviamente nelle imputazioni si tiene conto delle diverse caratteristiche e del diverso valore medio delle abitazioni di residenza e delle altre.

Quanto alla valutazione degli immobili, nella correzione C₃ si tiene conto della tendenza degli intervistati a sovrastimare il loro effettivo valore di mercato, non considerando la normale discrepanza tra il prezzo richiesto dal venditore e quello effettivamente pagato dall'acquirente. Questo divario, che secondo il sondaggio congiunturale sul mercato delle abitazioni (Banca d'Italia, 2013) si aggira mediamente tra il 10 e il 15 per cento, è stato posto pari al valore intermedio del 12 per cento¹⁷.

4.3.4 La correzione delle attività finanziarie – C_{3D}

Per la ricchezza finanziaria un confronto puntuale tra la Contabilità finanziaria e IBF è stato effettuato da Bonci, Marchese e Neri (2005), che quantificano in dettaglio le discrepanze tra le due fonti attribuendole ai vari possibili fattori: differenze di definizione, errori di misura, errori campionari e non campionari. Secondo un recente confronto (Banca d'Italia, 2012) la stima campionaria delle attività e passività finanziarie è pari a circa il 30-40 per cento della corrispondente stima aggregata.

La correzione delle attività finanziarie che qui si propone si basa su un'estensione del metodo descritto in D'Aurizio et al. (2006). In quel lavoro i dati dell'IBF sul 2004 furono messi a confronto con quelli di un'indagine condotta nel 2003 da una banca commerciale sui propri clienti e corretti conseguentemente. Per rendere il confronto efficace le modalità operative di questa indagine erano state rese più possibile simili a quelle dell'IBF.

Il campione, stratificato secondo la classe di ricchezza finanziaria, l'area geografica e l'ampiezza del comune di residenza, era composto di 1.834 famiglie. Prima di procedere all'esperimento di *matching*, si operò una post-stratificazione in modo da riprodurre le principali caratteristiche socio-demografiche dell'intera popolazione di clienti bancari in Italia.

¹⁷ Il confronto tra i dati dell'indagine e i dati dell'OMI sulle compravendite conferma che i rispondenti tendono a sovrastimare il valore di mercato delle abitazioni possedute.

L'aggiustamento dei dati IBF avveniva in due stadi successivi. Nel primo si misurava la reticenza, confrontando le dichiarazioni dei clienti intervistati con i dati sulle consistenze effettivamente detenute, in funzione degli importi dichiarati e delle caratteristiche socio-economiche delle famiglie. Nel secondo stadio, le relazioni stimate allo stadio precedente venivano estese al campione IBF, ottenendo valori aggiustati della ricchezza finanziaria per l'intera popolazione italiana di clienti bancari.

La metodologia che qui si propone modifica quella appena descritta per consentire l'estrapolazione delle stime corrette agli anni successivi. Per gli anni precedenti al 2010, si è invece adottata la correzione di Cannari e D'Alessio (1993).

4.4 Le calibrazioni – C_4 / C_8

Nelle indagini campionarie è piuttosto usuale incorporare alcune informazioni ausiliarie note da fonte esterna nei pesi, in modo da ottenere stime più accurate e coerenti con l'informazione a priori. Un utilizzo tipico di tale approccio è quello della post-stratificazione o del raking, che sono tecniche utilizzate anche in IBF. Si impone, ad esempio, l'allineamento della composizione socio-demografica del campione ad alcuni distribuzioni note dai censimenti o dalle fonti anagrafiche in modo da ridurre tendenzialmente gli errori standard delle stime delle variabili che sono legate alla composizione socio-demografica (ad esempio, il reddito), oltre a fornire campioni che per le caratteristiche note (ad esempio, la composizione per sesso o per età della popolazione) riproducono esattamente quanto noto da altre fonti.

A partire dal lavoro di Deville e Särndal (1992) le tecniche di calibrazione sono state generalizzate per consentire di includere nel set di informazioni note a priori non solo distribuzioni di caratteri qualitativi o ordinali ma anche totali di variabili quantitative. Il metodo, tramite algoritmi numerici, ricerca i pesi di calibrazione che siano quanto più prossimi ai pesi da disegno (secondo un criterio di distanza), e in grado di soddisfare i vincoli sulla composizione del campione (come nel raking tradizionale) e sugli ammontari totali di certe grandezze (ad esempio i redditi totali o di un certo tipo). Nel seguito faremo riferimento alle tecniche di calibrazione implementate nella macro SAS Calmar (Sautory, 1993)¹⁸.

La strategia seguita è stata quella di prevedere l'allineamento, oltre che alle caratteristiche socio-demografiche del capofamiglia risultanti dall'IBF, anche ai totali delle fonti di reddito o delle forme di ricchezza, come descritto nella tavola 5.

L'allineamento del campione ai totali delle 4 forme di reddito considerato (da lavoro dipendente YL, da pensioni e altri trasferimenti YT, da lavoro autonomo YM e da capitale YC) e alla ricchezza netta nel suo complesso W viene ottenuta con un incremento della deviazione standard dei pesi che, nella media degli anni considerati, passa da 1,01 a 1,87¹⁹.

¹⁸ La macro Calmar prevede 4 criteri di ricerca delle soluzioni: lineare, raking, logistico, lineare troncata. Nel seguito è stato utilizzato quest'ultimo criterio, che consente nella maggior parte dei casi di pervenire a una soluzione, e consente di evitare pesi negativi, di difficile interpretazione.

¹⁹ Secondo alcuni calcoli effettuati sul reddito e la ricchezza nel 2010, un incremento della deviazione standard dei pesi dovuto alla calibrazione si riflette sugli errori standard delle stime in misura pressoché corrispondente. Se, ad esempio, l'errore standard del reddito medio è di circa 500 euro nel 2010, con una media di circa 35.000 euro, con dei pesi che presentano una variabilità circa doppia si ottengono errori standard approssimativamente di 1.000 euro. Si tratta ovviamente di un'approssimazione che consente però di valutare in prima battuta, senza complessi calcoli, l'impatto della calibrazione sulla variabilità delle stime.

L'allineamento delle stime dei totali ai valori noti delle diverse forme di ricchezza presenta difficoltà maggiori. Le calibrazioni che prendono in considerazione solo i totali delle macrocategorie di attività reali AR, attività finanziarie AF e passività finanziarie PF, oltre a quello del reddito Y, convergono solo in alcuni anni e con un incremento consistente della variabilità dei pesi, la cui deviazione standard in media supera la soglia di 3. Se si inseriscono ulteriori vincoli, come ad esempio quello sul totale delle attività finanziarie rischiose AF3 o la distribuzione delle abitazioni diverse da quelle di residenza (ALTRIM), non si raggiunge la convergenza. A maggior ragione non appare praticabile procedere con l'inserimento di vincoli riguardanti congiuntamente il reddito e la ricchezza.

In sintesi, da questo primo blocco di calibrazioni sembra di poter concludere che, mentre per il reddito si raggiunge la convergenza con un set di pesi che presentano una variabilità non troppo elevata rispetto ai pesi da disegno, nel caso della ricchezza la convergenza viene raggiunta con pesi molto più variabili e con un set di variabili limitato. Ciò è dovuto sia al maggiore livello di *under-reporting* sia alla maggiore concentrazione che caratterizza la ricchezza rispetto al reddito; potrebbe contribuire inoltre, una certa incoerenza tra i dati IBF e i vincoli usati nella calibrazione.

La calibrazione sui totali di ricchezza è stata replicata, limitatamente al 2010, con un campione che integra le famiglie IBF con 198 famiglie rilevate dall'Associazione Italiana Private Banking (AIPB) con una procedura e un questionario analoghi ad IBF. Si tratta di famiglie che, essendo selezionate tra la clientela dell'AIPB, detengono oltre 500.000 euro di attività finanziarie, anche se – come in IBF – non necessariamente dichiarano nell'intervista quanto effettivamente posseduto.

L'integrazione tra i due campioni è stata effettuata tramite una post-stratificazione, calcolando nel campione IBF il peso delle famiglie per le classi di ricchezza più elevate e riproducendo lo stesso peso per il complesso delle famiglie IBF e AIPB.

La maggiore numerosità campionaria di famiglie ricche che si ottiene nel campione integrato consente di ottenere deviazioni standard dei pesi di calibrazione leggermente più contenuti, quando si impone il controllo dei totali delle forme di ricchezza (2,60). L'allineamento del campione rimane però problematico quando si richiede l'allineamento anche con il numero di immobili (diversi dall'abitazione di residenza) posseduti.

I risultati fin qui mostrati sembrano suggerire la difficoltà di applicazione dei metodi di calibrazione in presenza di un *under-reporting* molto consistente. Gli esperimenti di calibrazione sono stati pertanto ripetuti sugli archivi IBF i cui pesi tengono conto della non risposta e in cui i dati sugli immobili, quelli sulle attività finanziarie e sui redditi dei lavoratori autonomi sono stati preventivamente aggiustati secondo le procedure C₃ descritte in precedenza.

Le calibrazioni sui dati IBF aggiustati che prendono in considerazione le fonti di reddito e il totale della ricchezza (C₆) presentano pesi con una variabilità relativamente contenuta (la deviazione standard dei pesi finali è, in media tra i vari anni, pari a 1,91). Considerando invece i totali delle attività reali AR, delle attività finanziarie AF e delle passività finanziarie PF, oltre il totale del reddito Y (C₇), i pesi presentano una variabilità dei pesi (1,35) solo di poco superiore a quelli da disegno (tavola 5).

Il rispetto dei totali delle componenti di reddito e di ricchezza congiuntamente (C₈), applicato ai dati già corretti, comporta una deviazione standard dei pesi sensibilmente superiore ai casi precedenti (2,77).

Fermo restando che diverse altre ipotesi potrebbero essere valutate, inserendo o eliminando vincoli, si ritiene di avere materiale sufficiente per procedere a una valutazione comparativa dei risultati che scaturiscono dalle correzioni dei dati appena mostrate.

Tavola 5

Risultati della calibrazione

(Deviazione standard dei pesi di calibrazione*)

Anno	IBF (C ₀)	Peso non risposta (C ₃)	Controlli sui totali**					
			YL YM YT YC W (C ₄)	AR AF PF AF3 Y (C ₅)	YL YM YT YC AR AF PF	YL YM YT YC W (C ₆)	AR AF PF Y (C ₇)	YL YM YT YC AR AF PF (C ₈)
			Dati IBF originari		Dati IBF originari + AIPB	Dati IBF corretti ***		
1995	0,94	1,04	1,85	Non converge	-	1,99	1,47	2,74
1998	0,98	1,16	1,97	2,76	-	2,10	1,01	3,10
2000	0,94	1,19	1,98	Non converge	-	1,98	1,10	2,84
2002	1,04	1,48	2,12	Non converge	-	2,02	1,57	3,18
2004	1,05	1,34	1,70	Non converge	-	1,75	1,47	2,61
2006	1,04	1,34	1,50	Non converge	-	1,61	1,36	2,48
2008	1,03	1,12	1,64	Non converge	-	1,61	1,46	2,89
2010	1,06	1,21	1,96	2,96	2,60	1,98	1,38	2,26
2012	1,07	1,14	2,08	2,82	-	2,13	1,33	2,83
Media	1,01	1,24	1,84	2,85	2,60	1,91	1,35	2,77

(*) La deviazione standard dei pesi delle correzioni C₁ e C₂ è uguale a quella di C₀. (**) Oltre ai controlli su sesso, età, qualifica professionale, numero di componenti della famiglia, ampiezza comunale e area geografica di residenza. (***) Correzione per la non risposta, per il numero di case non di residenza e per il valore delle abitazioni, per le attività finanziarie e per il reddito dei lavoratori autonomi.

5. Valutazione delle stime al 2012

Le tavole A1, A2, A3 e A4 in Appendice riportano i valori medi del reddito e della ricchezza netta, calcolati sui dati IBF e per le varie correzioni considerate per le principali caratteristiche familiari.

Nella correzione proporzionale (C₁), la maggiore rivalutazione che caratterizza i redditi da lavoro autonomo rispetto ai redditi da lavoro dipendente modifica la posizione relativa delle famiglie di imprenditori rispetto a quelle dei dirigenti, i cui redditi sono solo marginalmente rivalutati. Anche gli altri lavoratori autonomi hanno una rivalutazione maggiore della media, mentre invece l'opposto si osserva per i dipendenti. I profili medi per le altre caratteristiche non sono però profondamente modificati da questa correzione; i rapporti tra i valori delle famiglie residenti nelle aree geografiche, ad esempio, sono pressoché identici.

Per la ricchezza netta, la più ampia rivalutazione delle componenti finanziarie tende a modificare maggiormente gli ammontari di ricchezza delle famiglie del Nord rispetto a quelle del Centro e del Sud e Isole. È inoltre maggiormente rivalutata la ricchezza delle famiglie con capofamiglia anziano e laureato.

L'aggiustamento C₂, che tiene conto del giudizio degli intervistatori, non si discosta in misura importante dal C₁; rivaluta i redditi e la ricchezza delle famiglie di imprenditori e laureati un po' di meno e quelle di altri autonomi e residenti nel Sud e Isole un po' di più.

Nell'ambito delle correzioni indicate come C_3 , l'aggiustamento per la non risposta (C_{3A}) implica una rivalutazione di circa il 9 per cento nel valore medio per il reddito e del 15 per cento per la ricchezza, La rivalutazione è relativamente maggiore per gli imprenditori e gli altri autonomi e minore per i dirigenti e i quadri.

La correzione del reddito da lavoro autonomo, che viene incrementato per tutti del 25 per cento, comporta una rivalutazione complessiva del reddito pari al 3,9 per cento, ovviamente maggiore per coloro che svolgono quel tipo di attività.

La correzione degli immobili diversi da quelli di residenza (che implica un aumento degli immobili posseduti ma anche una valutazione più contenuta del valore di mercato) implica un aumento del reddito del 3,8 per cento e della ricchezza del 3,1 per cento.

La correzione delle attività finanziarie comporta un aumento dei valori medi della ricchezza del 18 per cento; il reddito viene invece rivalutato indirettamente tramite l'imputazione dei redditi corrispondenti del 3,7 per cento. Contrariamente alle due precedenti correzioni, in questo caso la rivalutazione della ricchezza dei lavoratori autonomi e dei dirigenti è minore della media.

Complessivamente, le 4 correzioni C_{3A} - C_{3D} comportano una rivalutazione del reddito medio del 19 per cento e della ricchezza del 37,7 per cento, valori che sono comunque ancora inferiori ai corrispondenti totali noti dalla Contabilità Nazionale. Il reddito degli autonomi risulta rivalutato in modo particolare (per la correzione specifica in C_{3B}) mentre la ricchezza di queste famiglie viene rivalutata meno della media.

La calibrazione sulle fonti di reddito (C_4) comporta rivalutazioni sensibili sia del reddito (30 per cento) sia della ricchezza netta (23 per cento), allineando le medie a quelle desunte dalla contabilità nazionale. Le rivalutazioni sono più elevate per le famiglie di autonomi, con maggior numero di componenti, residenti nei comuni fino a 40.000 abitanti e nelle regioni del Sud e Isole. Rivalutazioni modeste si osservano invece per i lavoratori dipendenti (in particolare operai e insegnanti). Sulla ricchezza, che il metodo tiene sotto controllo solo nel suo aggregato totale, il valore delle famiglie di lavoratori autonomi e imprenditori sarebbe di molto rivalutato; le famiglie di quadri, operai e pensionati avrebbero però una ricchezza inferiore a quella dichiarata.

La calibrazione che opera sulle forme di ricchezza e controlla il reddito nel suo complesso (C_5) presenta una notevole instabilità. Nel complesso conferma le tendenze emerse con la correzione C_4 , rivalutando con maggiore evidenza sia il reddito sia la ricchezza delle famiglie residenti al Nord.

Le calibrazioni C_6 , C_7 e C_8 (applicate ai dati già corretti con le correzioni C_{3A} - C_{3D}) forniscono rivalutazioni non sempre pienamente concordanti. Nel complesso rimane confermata la maggiore rivalutazione dei redditi degli autonomi e dei laureati. Per la ricchezza, invece, i risultati di queste tipologie familiari sono contrastanti, in crescita oltre la media in alcuni casi e sotto la media in altri.

Le figure 1 e 2 forniscono un quadro di insieme delle diverse correzioni (la linea più spessa indica il dato IBF). I profili di reddito per sesso, età e titolo di studio mostrano una certa stabilità. I risultati del Nord e del Centro presentano una certa variabilità che implica in alcuni casi un'inversione nell'ordine mentre il Sud e Isole rimane stabilmente staccato, con una distanza dal Nord che oscilla intorno ai 30 punti percentuali. Per ampiezza demografica, la maggior parte delle stime conferma valori un po' più alti per le città maggiori.

Nel complesso, i profili di reddito che si registrano con la correzione C_3 presentano la maggiore correlazione con quelli desunti dai dati IBF, sia per il reddito sia per la ricchezza. La correzione C_3 è dunque una correzione che tende a conservare piuttosto fedelmente l'immagine che si presenta nei dati originari.

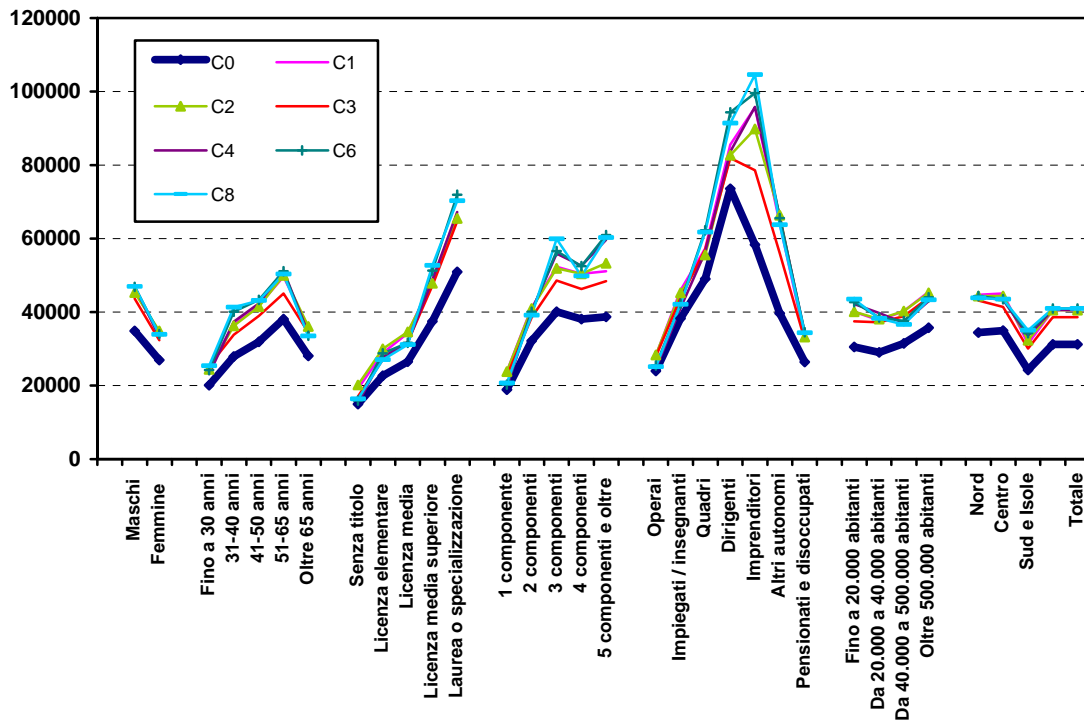
Tra i metodi utilizzati, la maggiore variabilità nelle stime dei redditi la si riscontra però nell'analisi per qualifica professionale. Le stime dei redditi delle famiglie di dirigenti, imprenditori e altri autonomi mostrano un'ampia variabilità, in parte attribuibile anche al loro ridotto peso campionario.

Nel complesso le stime della ricchezza netta confermano quanto detto ora, con un'accentuazione della variabilità che caratterizza le stime di questo aggregato rispetto al reddito. È visibile, inoltre, che alcune stime corrette della ricchezza netta siano inferiori rispetto alle stime IBF, in buona parte a causa della sottovalutazione della attività reali operata nelle correzioni, per tener conto dei problemi di valutazione degli immobili nei dati di indagine.

Le correzioni implicano con una certa frequenza un indebolimento delle correlazioni tra i fenomeni oggetto di aggiustamento (tavola A5 e A6). La correlazione semplice tra reddito e ricchezza, che nei dati non corretti è pari a 0,57, si riduce a 0,50 e 0,44 con le correzioni C_1 e C_2 ; aumenta invece con la correzione C_3 (0,62), che imputa case e attività finanziarie e congiuntamente i redditi prodotti da queste attività. Le calibrazioni non presentano una tendenza comune, in alcuni casi incrementano la correlazione (C_5 e C_7) e in altri la riducono. Gli altri indici che misurano il grado di concordanza tra le variabili forniscono indicazioni analoghe.

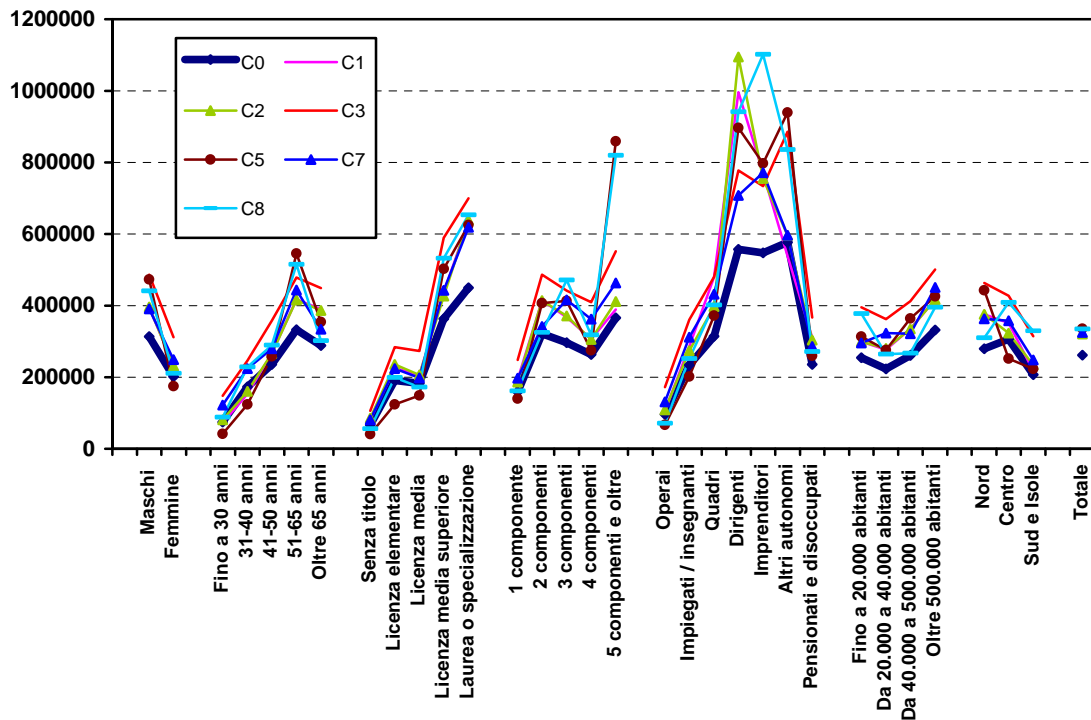
L'indice di concentrazione dei redditi, sia equivalenti sia familiari, risulta sempre maggiore di quello misurato sui dati di partenza, in particolare nel caso delle calibrazioni. Per la ricchezza l'indice di concentrazione della correzione C_3 risulta leggermente inferiore a quello di partenza; le correzioni che operano sui dati aggiustati con C_3 riportano però valori nuovamente maggiori di IBF (tavole A7 e A8). Nel complesso, queste evidenze sembrano suggerire che l'indagine potrebbe sottostimare i livelli di concentrazione degli aggregati esaminati.

Figura 1 – Profili di reddito familiare medio secondo le caratteristiche del capofamiglia, confronto tra le correzioni^(*)
(euro)



(*) La figura riporta i valori medi del reddito sui dati non corretti (C₀, linea in grassetto) e su quelli aggiustati (C₁, C₂, C₃, C₄, C₆ and C₈) (vedi tav. A1).

Figura 2 – Profili di ricchezza netta familiare media secondo le caratteristiche del capofamiglia, confronto tra le correzioni^(*)
(euro)



(*) La figura riporta i valori medi della ricchezza netta sui dati non corretti (C₀, linea in grassetto) e su quelli aggiustati (C₁, C₂, C₃, C₅, C₇ and C₈) (vedi tav. A3).

La tavola A9 mostra la variabilità e una stima della distanza di questi stimatori dai valori della Contabilità Nazionale, che per semplicità indichiamo con il termine di distorsione, anche se è noto che le stime aggregate sono esse stesse soggette a errori di varia natura.

La stima IBF presenta un ridotto errore standard sia per il reddito sia per la ricchezza ma un'elevata distorsione. Gli altri stimatori presentano tutti una minore (o nulla) distorsione, anche se ottenuta con un incremento della varianza.

L'errore quadratico medio consente di valutare congiuntamente questi due aspetti. Nel complesso, escludendo gli stimatori che procedono con un semplice riproporzionamento dei valori (C_1 e C_2), lo stimatore C_7 è quello che sembra offrire la migliore performance sia per il reddito sia per la ricchezza.

È interessante osservare il confronto della distribuzione del numero di immobili (diversi dall'abitazione di residenza) posseduti, come può essere stimata a partire dai dati fiscali²⁰, e quanto risulta nelle varie correzioni (Tavola 6). I dati originari di IBF presentano un'ampia sottostima di immobili posseduti, che si riflette nella più elevata frequenza di famiglie che segnalano di non possedere altri immobili (85 contro 68,2 per cento). Tutte le correzioni riducono questo divario, ad eccezione della C_4 , che è rivolta principalmente alla correzione dei redditi e di quella che fa riferimento ai dati AIPB (C_9)²¹.

Le correzioni più soddisfacenti sono quelle che fanno riferimento ai dati aggiustati (C_3) e in particolare le correzioni C_6 , C_7 e C_8 che sono anche calibrate.

Tavola 6

**Distribuzione degli immobili non di residenza nei dati fiscali
e in quelli ottenuti nelle correzioni IBF**

	Immobili diversi dall'abitazione di residenza							Totale
	0	1	2	3	4	5	6 e oltre	
Dati fiscali	68,2	23,0	6,8	0,5	0,5	0,4	0,6	100,0
C_0	85,0	11,7	2,3	0,6	0,2	0,1	0,0	100,0
C_3	66,0	22,9	7,0	2,7	0,8	0,4	0,1	100,0
C_4	87,7	9,5	2,0	0,4	0,2	0,0	0,1	100,0
C_5	82,5	12,6	3,3	0,9	0,6	0,1	0,0	100,0
C_6	72,3	18,9	6,0	2,1	0,3	0,4	0,1	100,0
C_7	68,4	21,5	6,5	2,4	0,6	0,5	0,0	100,0
C_8	73,8	17,9	5,6	1,9	0,3	0,4	0,1	100,0
C_9	90,9	7,2	1,4	0,3	0,1	0,0	0,1	100,0

Concludiamo questo paragrafo mostrando come l'uso di queste tecniche di correzione illustrate nei paragrafi precedenti possa essere esteso nelle fasi di stima e di analisi dei dati.

Si è mostrato che le calibrazioni, nel fornire risultati vincolati a totali noti da fonti macro, possono aumentare la variabilità dei coefficienti di ponderazione e produrre

²⁰ Si veda Neri e Monteduro (2013).

²¹ Le famiglie AIPB sono particolarmente rivolte alla ricchezza finanziaria, ciò che consente di allineare i totali di ricchezza senza incrementare il numero di immobili.

stimatori instabili. Per ottenere stime più robuste è possibile utilizzare dati provenienti da indagini contigue, sotto l'ipotesi che queste rappresentino sufficientemente bene le caratteristiche strutturali della popolazione, utilizzando poi le tecniche di calibrazione per riportare le stime all'anno di interesse. Ad esempio, per ottenere stime più robuste riferite al 2012 si possono considerare i dati provenienti dalle indagini sul 2008, 2010 e 2012, applicando poi la calibrazione con i vincoli riferiti a quell'anno. La stessa procedura può essere ripetuta con riferimento agli altri due anni considerati²². Per tenere conto dell'inflazione nel periodo, si possono riproporzionare le variabili monetarie sul livello dei singoli anni oggetto di stima. Poiché le stime relative ai 3 anni esaminati vengono applicati allo stesso dataset e differiscono solo per i vincoli utilizzati, questo metodo permette di ottenere informazioni sulle modifiche dei profili indotte dalle variazioni dei vincoli stessi.

La tavola A4 mostra le correzioni C_6 per il periodo 2008-2012 a confronto con questi stimatori robusti (C_6^R). Gli stimatori C_6^R appaiono assai meno variabili da un anno all'altro rispetto agli stimatori C_6 , in quanto, come detto, esprimono solo l'effetto sulle stime dato dal cambiamento dei vincoli. Non si registrano per questi stimatori variazioni esagerate e poco credibili, come quella che si riscontra, ad esempio, per lo stimatore C_6 per le famiglie con capofamiglia dotato di licenza media.

Il metodo qui mostrato può essere anche impiegato per la valutazione di scenari previsivi, utilizzando vincoli riferiti ad anni successivi a quelli a cui si riferiscono i dati micro. Per una valutazione di questa pratica è possibile applicare i vincoli della correzione C_6 relativi al 2012 al dataset formato dalle famiglie rilevate nell'indagine sul 2010, dopo aver riproporzionato i valori medi del reddito su quelli del 2012.

Le variazioni percentuali tra queste stime previste al 2012 e le stime C_6 del 2010 mostrano una certa coerenza (correlazione 0,54) con quelle che si ottengono confrontando le stime robuste C_6^R del 2010 e del 2012, indice che una parte significativa dell'informazione contenuta nei vincoli viene trasferita nelle stime.

6. Conclusioni

Il lavoro ha esaminato diversi metodi di correzione che possono essere adottati per correggere gli errori non campionari (principalmente *selectivity bias* e *under-reporting*) presenti nei dati IBF.

Le correzioni basate sulla conoscenza specifica dei fenomeni, danno conto di *under-reporting* differenziali dei singoli fenomeni ma sono costose, richiedono molte assunzioni e non colmano totalmente i divari tra le stime e i totali noti.

Le calibrazioni sembrano essere uno strumento interessante ma, con riferimento ai dati IBF, il loro impiego appare davvero utile soltanto congiuntamente ai metodi di correzione basata sui modelli. Quando i dati di base sono molto distanti dai vincoli,

²² Come illustrato in Cannari D'Alessio (2003), in presenza di famiglie panel e di correlazione non nulla tra i fenomeni nel corso del tempo, come nel caso del reddito e della ricchezza IBF, ai fini dell'efficienza dello stimatore della media è preferibile adottare un sistema di ponderazione che tenga conto di questo aspetto. In particolare, le osservazioni riferite alle famiglie presenti più volte nel periodo considerato dovrebbero essere ponderate in funzione inversa della media dei coefficienti di correlazione del reddito e della ricchezza nel tempo: il peso delle osservazioni presenti in due indagini dovrebbe essere diviso per $(1+\rho)$; quello delle famiglie presenti in tutte e tre le indagini invece per $[1 + (4/3)\rho + (2/3)\rho^2]$. Nel presente lavoro, per semplicità, non si è operata tale trasformazione.

infatti, le calibrazioni non convergono e, quando lo fanno, restituiscono stimatori instabili.

La pratica applicazione della calibrazione ai dati IBF ha mostrato come sia difficile immaginare una correzione unica per tutti gli scopi. Correggere i tipi di reddito può implicare una maggiore difficoltà nell'ottenere stime adeguate con riferimento alle componenti di ricchezza e viceversa.

Le correzioni danno nel complesso risultati che, in termini di profili, sono piuttosto simili tra di loro e rispetto ai dati di partenza (la correzione dei singoli fenomeni C_3 è la più conservativa rispetto a IBF). In altri termini, spesso i rapporti relativi tra le diverse categorie di famiglie non sono alterati in misura importante nei dati corretti. L'analisi degli errori quadratici medi evidenzia tuttavia che, tra le calibrazioni esaminate, le migliori sembrano essere quelle che correggono le singole componenti delle variabili di interesse (la correzione C_4 ma anche la C_7 per il reddito e la C_7 per la ricchezza).

Diversi indizi lasciano immaginare che le stime IBF sottostimino la concentrazione del reddito e della ricchezza.

I metodi di correzione basati sulla calibrazione sembrano interessanti sia per la fase interpretativa sia per la costruzione di scenari previsivi. Nei casi in cui la varianza degli stimatori calibrati sia ritenuta troppo elevata, si può ipotizzare di procedere aggregando alcune indagini contigue.

Appendice A – Tavole statistiche

Tavola A1

Reddito medio nei dati IBF (C₀) e nelle varie correzioni (C₁-C₈), 2012 (euro)

	Stime IBF	Correzioni			Calibrazioni				
	C ₀	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈
Sesso									
Maschi	34.896	45.656	45.312	43.619	46.918	49.027	46.893	46.014	46.935
Femmine	26.982	34.478	34.878	32.348	33.196	30.910	34.043	34.224	33.918
Classe di età									
Fino a 30 anni	20.058	24.634	24.413	25.397	23.123	22.826	24.219	25.387	25.372
31-40 anni	27.917	36.885	36.232	33.820	37.433	33.047	40.101	36.316	41.319
41-50 anni	31.912	41.830	41.511	38.957	42.387	42.267	43.323	43.801	43.140
51-65 anni	38.118	49.683	50.010	45.072	51.294	51.421	51.109	50.744	50.331
Oltre 65 anni	28.129	35.761	36.051	34.428	34.242	36.565	33.537	33.870	33.477
Titolo di studio									
Senza titolo	14.962	19.160	20.179	17.011	15.252	13.333	15.930	16.434	16.339
Licenza elementare	22.658	28.901	29.965	27.079	27.796	21.678	28.752	26.211	27.057
Licenza media	26.488	34.282	34.595	31.895	31.644	29.624	31.636	32.126	31.121
Licenza media superiore	37.439	48.698	47.866	47.203	48.781	48.560	51.305	48.738	52.678
Laurea o specializzazione	50.947	66.634	65.482	64.726	67.188	69.758	71.924	71.872	70.268
Numero di componenti									
1 componente	18.888	23.812	23.851	22.676	19.456	21.492	20.236	22.159	20.687
2 componenti	32.131	41.227	40.966	38.483	40.309	42.899	40.069	39.175	39.108
3 componenti	40.082	52.371	51.940	48.585	55.952	60.509	56.639	56.198	59.881
4 componenti	38.129	50.399	50.456	46.261	52.618	42.326	52.549	52.937	49.867
5 e più componenti	38.686	51.154	53.339	48.425	59.622	53.815	60.974	50.283	60.274
Condizione professionale									
Dipendente – Operai	24.039	28.195	28.386	27.874	24.618	25.090	25.506	27.764	25.165
Impiegati/insegnanti	38.275	45.901	45.244	43.832	41.005	42.636	42.246	45.490	42.083
Quadri	49.085	57.328	55.542	57.071	56.378	58.963	62.260	60.127	61.721
Dirigenti	73.602	85.523	82.673	81.945	83.670	95.055	94.368	92.317	91.395
Autonomi – Imprenditori	58.320	95.637	89.847	78.566	95.841	84.069	99.524	97.865	104.620
Altri autonomi	39.675	64.108	66.348	56.038	65.911	61.905	65.531	59.788	63.718
Pensionati e disoccupati	26.455	33.058	33.214	32.423	34.272	31.692	34.538	31.951	34.329
Ampiezza comunale									
Fino a 20.000 abitanti	30.554	40.087	40.063	37.456	42.148	39.810	42.723	38.559	43.471
Da 20.000 a 40.000 abitanti	29.033	37.779	38.059	37.244	39.748	35.976	38.562	40.091	38.281
Da 40.000 a 500.000 abitanti	31.506	40.284	40.301	38.842	37.026	41.816	37.583	40.760	36.577
Oltre 500.000 abitanti	35.760	45.488	45.232	43.886	42.994	46.756	44.038	48.396	43.318
Area geografica									
Nord	34.400	44.704	44.375	43.138	44.109	51.591	44.353	46.597	43.814
Centro	34.971	45.021	44.377	41.379	43.336	38.440	43.500	43.529	43.456
Sud e Isole	24.247	31.430	32.307	30.079	33.577	25.368	34.272	29.674	35.029
TOTALE	31.236	40.487	40.487	38.602	40.579	40.670	40.966	40.562	40.929

Reddito medio nelle singole correzioni di C₃ (C_{3A}-C_{3D}), 2012
(euro)

	C _{3A}	C _{3B}	C _{3C}	C _{3D}	C ₃
Sesso					
Maschi	38.300	36.391	36.318	37.497	43.619
Femmine	28.691	27.883	27.920	28.852	32.348
Classe di età					
Fino a 30 anni	23.357	20.501	20.345	21.250	25.397
31-40 anni	30.186	29.346	28.553	29.372	33.820
41-50 anni	34.342	33.564	32.897	33.828	38.957
51-65 anni	39.822	39.760	39.547	40.695	45.072
Oltre 65 anni	30.016	28.702	29.680	30.905	34.428
Titolo di studio					
Senza titolo	15.735	15.095	15.188	15.800	17.011
Licenza elementare	24.210	23.108	23.437	24.721	27.079
Licenza media	28.739	27.433	27.263	27.895	31.895
Licenza media superiore	40.729	39.153	39.083	40.494	47.203
Laurea o specializzazione	56.365	53.503	53.418	54.799	64.726
Numero di componenti					
1 componente	19.838	19.433	19.569	20.641	22.676
2 componenti	33.520	33.124	33.949	34.781	38.483
3 componenti	43.362	41.794	41.271	42.575	48.585
4 componenti	40.727	40.054	39.263	40.569	46.261
5 e più componenti	43.451	40.611	39.770	40.405	48.425
Condizione professionale					
Dipendente – Operai	25.852	24.203	24.635	25.287	27.874
Impiegati/insegnanti	39.833	39.031	39.516	40.486	43.832
Quadri	51.991	49.765	50.828	51.987	57.071
Dirigenti	75.530	74.443	77.338	76.836	81.945
Autonomi – Imprenditori	65.930	66.848	60.356	60.956	78.566
Altri autonomi	44.584	45.235	41.162	43.386	56.038
Pensionati e disoccupati	28.836	26.855	27.662	28.718	32.423
Ampiezza comunale					
Fino a 20.000 abitanti	33.068	31.799	31.518	32.805	37.456
Da 20.000 a 40.000 abitanti	32.673	30.183	30.024	31.363	37.244
Da 40.000 a 500.000 abitanti	34.138	32.643	33.053	33.681	38.842
Oltre 500.000 abitanti	38.834	37.143	37.353	38.177	43.886
Area geografica					
Nord	37.557	35.781	35.821	37.291	43.138
Centro	37.011	36.366	35.997	37.205	41.379
Sud e Isole	26.892	25.121	25.210	25.574	30.079
TOTALE	34.022	32.457	32.435	33.499	38.602

Tavola A3

Ricchezza netta media nei dati IBF (C₀) e nelle varie correzioni (C₁-C₈), 2012
(euro)

	Stime IBF	Correzioni			Calibrazioni				
	C ₀	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈
Sesso									
Maschi	313.142	395.878	396.845	486.200	445.503	473.194	448.956	390.987	441.410
Femmine	201.529	232.328	231.204	311.990	179.725	175.300	206.172	248.830	211.089
Classe di età									
Fino a 30 anni	75.172	72.694	81.676	148.506	61.199	41.519	85.984	122.667	88.101
31-40 anni	173.029	153.789	160.344	246.223	191.810	124.269	236.198	223.753	228.159
41-50 anni	235.283	266.314	266.435	358.123	276.497	257.611	276.446	279.533	289.000
51-65 anni	332.956	417.878	416.406	478.319	510.884	545.910	509.721	444.195	515.026
Oltre 65 anni	288.292	389.963	386.753	449.206	295.853	354.337	317.286	334.255	301.408
Titolo di studio									
Senza titolo	67.374	82.588	86.006	106.114	51.395	41.459	57.299	80.877	56.332
Licenza elementare	191.477	226.259	236.060	284.143	240.564	123.981	247.818	223.867	199.091
Licenza media	182.845	205.294	205.946	273.627	191.134	149.274	202.408	197.937	172.134
Licenza media superiore	363.070	436.489	426.455	589.566	479.788	502.293	529.836	443.280	532.087
Laurea o specializzazione	449.685	635.338	638.474	700.396	502.158	624.849	558.879	618.546	653.575
Numero di componenti									
1 componente	158.900	193.970	186.534	248.368	122.662	140.015	146.083	197.689	161.316
2 componenti	321.158	413.888	414.602	486.939	341.410	407.296	388.373	342.177	324.831
3 componenti	296.399	367.437	371.548	442.330	391.125	413.467	427.359	416.388	471.684
4 componenti	263.033	306.232	305.178	410.031	377.197	275.267	309.128	362.298	318.104
5 e più componenti	366.466	389.185	411.193	551.597	816.919	858.753	797.726	463.451	819.943
Condizione professionale									
Dipendente – Operai	96.471	105.360	108.990	172.406	58.245	66.607	80.337	131.448	71.346
Impiegati/insegnanti	233.888	285.337	274.269	360.461	199.001	201.998	236.076	312.005	252.125
Quadri	314.046	480.278	397.120	480.560	249.325	372.752	372.788	431.893	401.562
Dirigenti	557.100	996.609	1.094.885	777.241	579.566	896.363	865.000	707.902	941.373
Autonomi – Imprenditori	547.060	770.382	754.938	733.743	920.047	797.505	895.092	771.686	1.101.767
Altri autonomi	576.900	538.551	595.709	885.983	1.111.378	939.265	962.957	596.365	836.056
Pensionati e disoccupati	235.546	310.730	303.212	367.175	240.904	259.984	271.728	285.550	271.818
Ampiezza comunale									
Fino a 20.000 abitanti	254.598	298.791	298.746	395.293	382.342	313.777	399.465	295.160	377.494
Da 20.000 a 40.000 abitanti	223.650	277.055	280.066	362.644	286.318	275.730	237.060	323.531	263.906
Da 40.000 a 500.000 abitanti	260.980	336.078	335.093	412.585	229.683	364.600	249.394	321.753	267.044
Oltre 500.000 abitanti	331.793	417.698	416.598	500.789	331.258	425.961	393.176	450.637	395.708
Area geografica									
Nord	279.878	375.508	375.954	463.468	318.996	442.720	339.946	363.410	310.114
Centro	306.401	351.747	323.652	428.075	353.066	251.597	369.384	358.188	409.338
Sud e Isole	207.352	217.977	233.808	314.043	310.465	222.778	313.157	248.073	329.539
TOTALE	261.529	320.248	320.248	408.649	322.718	335.778	336.959	325.250	335.137

Ricchezza netta media nelle singole correzioni di C₃ (C_{3A}-C_{3D}), 2012

(euro)

	C _{3A}	C _{3B}	C _{3C}	C _{3D}	C ₃
Sesso					
Maschi	366,284	313,142	315,556	420,410	486.200
Femmine	222,716	201,529	216,537	272,850	311.990
Classe di età					
Fino a 30 anni	97,007	75,172	75,087	116,572	148.506
31-40 anni	186,304	173,029	173,554	229,249	246.223
41-50 anni	273,881	235,283	238,080	311,590	358.123
51-65 anni	363,390	332,956	336,312	441,410	478.319
Oltre 65 anni	316,253	288,292	309,303	397,335	449.206
Titolo di studio					
Senza titolo	75,513	67,374	67,858	95,869	106.114
Licenza elementare	212,423	191,477	186,786	269,666	284.143
Licenza media	210,510	182,845	183,575	239,976	273.627
Licenza media superiore	428,140	363,070	389,557	484,851	589.566
Laurea o specializzazione	510,320	449,685	463,376	611,922	700.396
Numero di componenti					
1 componente	174,855	158,900	162,895	224,737	248.368
2 componenti	345,679	321,158	348,223	430,200	486.939
3 componenti	331,435	296,399	296,618	399,839	442.330
4 componenti	307,291	263,033	263,323	358,271	410.031
5 e più componenti	479,719	366,466	357,804	439,106	551.597
Condizione professionale					
Dipendente – Operai	118,037	96,471	103,454	140,784	172.406
Impiegati/insegnanti	260,801	233,888	240,228	323,669	360.461
Quadri	338,169	314,046	324,740	435,040	480.560
Dirigenti	578,530	557,100	589,716	730,398	777.241
Autonomi – Imprenditori	621,936	547,060	540,922	667,788	733.743
Altri autonomi	714,447	576,900	566,758	728,387	885.983
Pensionati e disoccupati	260,338	235,546	248,913	325,152	367.175
Ampiezza comunale					
Fino a 20.000 abitanti	299,057	254,598	253,454	345,539	395.293
Da 20.000 a 40.000 abitanti	262,940	223,650	222,438	313,001	362.644
Da 40.000 a 500.000 abitanti	291,127	260,980	287,812	349,124	412.585
Oltre 500.000 abitanti	380,728	331,793	347,308	428,050	500.789
Area geografica					
Nord	333,286	279,878	284,053	394,186	463.468
Centro	331,216	306,401	307,657	396,673	428.075
Sud e Isole	238,414	207,352	225,852	262,347	314.043
TOTALE	302,374	261,529	269,767	352,174	408.649

Tavola A5

Livello di coerenza tra le variabili di reddito e di ricchezza nei dati IBF (C₀) e nelle varie correzioni (C₁-C₈)

(medie 1995-2012)

	Stime IBF	Correzioni			Calibrazioni				
		C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈
Correlazione reddito/ricchezza	0,573	0,499	0,438	0,619	0,462	0,729	0,499	0,616	0,530
Alpha di Cronbach (*).....	0,768	0,760	0,721	0,786	0,779	0,828	0,760	0,799	0,746
Varianza spiegata dalla prima componente principale (*).....	0,420	0,415	0,373	0,454	0,436	0,504	0,417	0,467	0,415

(*) Variabili considerate: Y, YL, YT, YM, YC, AR, AF, PF, W.

Tavola A6

Livello di coerenza tra le variabili di reddito e di ricchezza nei dati IBF (C₀) e nelle singole correzioni di C₃ (C_{3A}-C_{3D})

(medie 1995-2012)

	Calibrazioni				
	C _{3A}	C _{3B}	C _{3C}	C _{3D}	C ₃
Correlazione reddito/ricchezza	0,593	0,584	0,640	0,565	0,438
Alpha di Cronbach (*).....	0,777	0,770	0,774	0,760	0,721
Varianza spiegata dalla prima componente principale (*).....	0,432	0,427	0,433	0,413	0,373

(*) Variabili considerate: Y, YL, YT, YM, YC, AR, AF, PF, W.

Tavola A7

Indice di concentrazione nei dati IBF (C₀) e nelle varie correzioni (C₁-C₈)
(indice di Gini, medie 1995-2012)^(*)

	Stime IBF	Correzioni			Calibrazioni				
	C ₀	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈
Reddito familiare	0,350	0,383	0,389	0,383	0,403	0,439	0,402	0,419	0,405
Reddito equivalente	0,319	0,351	0,357	0,359	0,367	0,423	0,365	0,392	0,378
Ricchezza netta	0,596	0,632	0,637	0,577	0,680	0,694	0,652	0,626	0,696

^(*) Stime winsorizzate al 1° e 99° percentile.

Tavola A8

Indice di concentrazione nei dati IBF (C₀) e nelle varie correzioni (C₁-C₈)
(indice di Gini, medie 1995-2012)^(*)

	Calibrazioni				
	C _{3A}	C _{3B}	C _{3C}	C _{3D}	C ₃
Reddito familiare	0,356	0,360	0,367	0,361	0,383
Reddito equivalente	0,325	0,329	0,319	0,319	0,359
Ricchezza netta	0,576	0,596	0,595	0,587	0,577

^(*) Stime winsorizzate al 1° e 99° percentile.

Tavola A9

Variabilità degli stimatori di reddito e di ricchezza nei dati IBF (C₀) e nelle varie correzioni (C₁-C₈)

	Stime IBF	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈
Reddito									
Media	31.236	40.487	40.487	38.602	40.579	40.670	40.966	40.562	40.929
Std.err.	364	581	608	552	1.291	1.617	1.327	1.118	1.563
Distorsione....	9.251	0	0	3.103	92	185	458	77	279
EQM	9.258	581	608	3.152	1.295	1.628	1.404	1.121	1.587
Ricchezza netta									
Media	261.529	320.248	320.248	408.649	322.718	335.778	336.959	325.250	335.137
Std.err.	10.087	10.519	11.727	17.811	36.587	37.902	38.903	15.412	35.552
Distorsione....	58.719	0	0	39.859	2.501	15.400	21.020	5.186	8.852
EQM	59.579	10.519	11.727	43.657	36.672	40.911	44.219	16.261	36.637

Reddito medio – Correzione C₆ e simulazione, 2008 - 2012

(euro, valori percentuali)

	Correzione C ₆			Simulazione ^(*)			Previsione ^(**)
	2008	2010	2012	2008	2010	2012	2012
Sesso							
Maschi	48.820	46.868	46.893	47.746	47.588	47.692	47.668
Femmine	33.316	34.816	34.043	35.195	33.900	33.082	33.003
Classe di età							
Fino a 30 anni	23.932	26.889	24.219	24.837	25.648	23.716	28.046
31-40 anni	36.839	38.620	40.101	38.149	39.245	37.748	33.105
41-50 anni	44.651	46.048	43.323	46.314	44.995	43.521	43.269
51-65 anni	55.256	52.756	51.109	55.099	51.341	50.742	52.957
Oltre 65 anni	38.296	32.749	33.537	36.598	34.522	34.876	34.476
Titolo di studio							
Senza titolo	18.160	16.929	15.930	17.083	17.388	16.219	15.940
Licenza elementare	29.396	24.586	28.752	28.543	28.254	27.330	22.933
Licenza media	42.632	39.776	31.636	38.904	38.389	37.205	38.565
Licenza media superiore	48.774	46.360	51.305	50.286	48.778	46.905	46.827
Laurea o specializzazione	71.673	72.270	71.924	72.832	72.231	69.996	68.427
Numero di componenti							
1 componente	23.661	23.000	20.236	23.272	22.216	21.329	21.082
2 componenti	43.377	40.165	40.069	42.361	39.729	41.138	43.295
3 componenti	56.156	50.358	56.639	55.645	53.551	53.488	49.229
4 componenti	54.449	56.953	52.549	55.822	54.597	54.838	55.755
5 e più componenti	54.968	51.832	60.974	59.317	54.140	54.691	55.743
Condizione professionale							
Dipendente – Operai	26.441	27.261	25.506	26.777	27.199	24.964	25.506
Impiegati/insegnanti	41.964	46.722	42.246	44.590	45.071	42.093	43.899
Quadri	66.172	53.491	62.260	61.112	63.035	57.987	48.261
Dirigenti	98.933	82.508	94.368	93.087	93.716	89.551	78.889
Autonomi – Imprenditori	82.033	85.744	99.524	90.127	85.231	85.911	87.241
Altri autonomi	75.390	71.755	65.531	73.978	69.262	70.000	66.239
Pensionati e disoccupati	36.290	33.375	34.538	35.717	34.075	34.689	35.035
Ampiezza comunale							
Fino a 20.000 abitanti	46.783	39.641	42.723	44.605	42.983	42.224	40.656
Da 20.000 a 40.000 abitanti	36.145	39.799	38.562	37.981	38.130	37.236	36.476
Da 40.000 a 500.000 abitanti	41.604	42.572	37.583	42.082	40.422	40.022	40.893
Oltre 500.000 abitanti	41.130	47.994	44.038	46.009	42.050	42.252	46.947
Area geografica							
Nord	49.148	44.274	44.353	47.289	45.409	44.818	44.633
Centro	44.352	49.353	43.500	46.995	45.721	45.616	48.605
Sud e Isole	32.953	32.271	34.272	34.074	32.858	32.303	30.679
TOTALE	43.149	41.520	40.966	43.149	41.520	40.963	40.963

(*) La simulazione è stata ottenuta applicando i vincoli della correzione C₆ relativi a ciascun anno tra il 2008 e il 2012 al dataset formato dall'insieme delle famiglie rilevate nelle 3 indagini, dopo aver riproporzionato i valori medi del reddito su quelli del 2012.

(**) La previsione è stata ottenuta applicando i vincoli della correzione C₆ relativi al 2012 al dataset formato dalle famiglie rilevate nell'indagine sul 2010, dopo aver riproporzionato i valori medi del reddito su quelli del 2012.

Bibliografia

- Banca d'Italia (1970), *Risparmio e struttura della ricchezza delle famiglie italiane nel 1968*, in A. Ulizzi, (a cura di), *Bollettino*, Banca d'Italia, n. 1, gennaio-febbraio, pp. 103-167.
- Banca d'Italia (2012), *La ricchezza delle famiglie italiane - 2011*, Supplementi al Bollettino statistico n. 65, dicembre.
- Banca d'Italia (2013), *Sondaggio congiunturale sul mercato delle abitazioni in Italia*, Luglio 201, Supplementi al Bollettino statistico n. 41, agosto.
- Banca d'Italia (2014), *I bilanci delle famiglie italiane nell'anno 2012*, a cura di F. Carta, R. Gambacorta, G. Ilardi, A. Neri, C. Rondinelli, Supplementi al Bollettino Statistico (nuova serie), Banca d'Italia, n. 5, Gennaio
- BCE (2013), *The Eurosystem Household Finance and Consumption Survey - Methodological Report for the First Wave*. Statistics Paper Series, N.1, April.
- Bonci, R., G. Marchese, A. Neri (2005), *La ricchezza finanziaria nei conti finanziari e nell'indagine sui bilanci delle famiglie italiane*, Temi di Discussione n. 565 - Novembre.
- Brandolini A. (1999), *The Distribution of Personal Income in Post-War Italy: Source Description, Data Quality, and the Time Pattern of Income Inequality*, *Giornale degli Economisti e Annali di Economia*, vol. 58, n. 2, pp. 183-239.
- Brandolini A. L. Cannari, G. D'Alessio, I. Faiella, (2004), *Household wealth distribution in Italy in the 1990s*, Working papers, The Levy Economics Institute, No. 414.
- Brick, J. M. (2013), *Unit Nonresponse and Weighting Adjustments: A Critical Review*, *Journal of Official Statistics*, n. 29(3): 329-469.
- Cannari L., G. D'Alessio (1990), *Housing Assets in the Bank of Italy's Survey of Household Income and Wealth*, in Dagum e Zenga (a cura di), "Income and Wealth Distribution, Inequality and Poverty", Springer Verlag, Berlino, p. 326-334.
- Cannari L., G. D'Alessio (1992), *Mancate interviste e distorsione degli stimatori*, Banca d'Italia, Temi di discussione, n.172.
- Cannari L., G. D'Alessio (1993), *Non-reporting and Under-reporting Behavior in the Bank of Italy's Survey of Household Income and Wealth*, in "Bulletin of the International Statistical Institute", vol. LV, n. 3, Pavia, p. 395-412.
- Cannari L., G. D'Alessio (2003), *La distribuzione del reddito e della ricchezza nelle regioni italiane*, Temi di Discussione n. 482, Banca d'Italia, Roma, Giugno.
- Cannari L., G. D'Alessio, G. Raimondi, A.I. Rinaldi (1990), *Le attività finanziarie delle famiglie italiane*, Banca d'Italia, Temi di discussione, n. 136.
- Cannari L., R. Violi (1995), *Reporting Behaviour in the Bank of Italy's Survey of Italian Household Income and Wealth*, *Research on Economic Inequality*, vol. 6, JAI Press Inc., pp. 117-130
- Cifaldi G., A. Neri (2013), *Asking income and consumption questions in the same survey: what are the risks?*, Temi di discussione, n. 908 - Aprile.

- D'Alessio G., I. Faiella (2002), *Nonresponse behaviour in the Bank of Italy's Survey of Household Income and Wealth*, Banca d'Italia, Temi di discussione, n. 462.
- D'Alessio G., S. Iezzi (2014), *L'effetto del momento delle interviste sulla stima del reddito e della ricchezza*, mimeo, Banca d'Italia.
- D'Alessio G., G. Iardi (2012), *Non sampling errors in sample surveys: the Bank of Italy's experience*, in C. Davino e L. Fabbri (Eds), *Survey data collection and integration*, Springer-Verlag.
- D'Aurizio L., I. Faiella, S. Iezzi, A. Neri (2006), *L'under-reporting della ricchezza finanziaria nell'indagine sui bilanci delle famiglie*, Temi di discussione n. 610.
- Deville, Jean-Claude (2000), *Generalized calibration and application to weighting for non-response*, COMPSTAT, Physica-Verlag HD: 65-76.
- Deville, J., C. Särndal, (1992), *Calibration estimators in survey sampling*, Jour. Amer. Statist. Assoc. n. 87, 376-382.
- Folsom R.E, Singh, A.C. (2000). *The Generalized Exponential Model for Sampling Weight Calibration for Extreme Values, Nonresponse, and Poststratification*, ASA Proceedings of the Section on Survey Research Methods, 598-603.
- Fuller, W.A., Loughin, M.M., Baker, H.D, (1994), *Regression Weighting for the 1987-88 National Food Consumption Survey*, Survey Methodology, n. 20: 75-85.
- Gelman, A., (2007) *Struggles with Survey Weighting and Regression Modeling*, Statistical Science 22, n. 2: 153--164.
- Hurst E., G. Li, B. Pugsley (2010), *Are Household Surveys Like Tax Forms: Evidence from Income Underreporting of the Self-Employed*, NBER WP 16527, Cambridge MA.
- Kott, P.S, Chang, T, (2010) *Using calibration weighting to adjust for non-ignorable unit nonresponse*, Journal of the American Statistical Association n. 105(491):1265–1275.
- Kott P.S, Liao D, (2012), *Providing double protection for unit nonresponse with a nonlinear calibration-weighting routine*, Survey Research Methods, 6 n. 2: 105-111.
- Little, R J., Vattavarian S. (2005), *Does Weighting for Nonresponse Increase the Variance of Survey Means?*, Survey Methodology n. 31(2):161–68.
- Lohr, S. (2007) Comment: *Struggles with Survey Weighting and Regression Modeling*, Statistical Science 22, no. 2: 175--178.
- Sautory O. (1993), *Le macro CALMAR, Redressement d'un échantillon par calage sur marges*, Document n. F 9310, INSEE.
- Neri A., T. Monteduro (2013), *La ricchezza immobiliare delle famiglie italiane: un confronto fra dati campionari e censuari*, Questioni di Economia e Finanza, n. 146 - Gennaio.
- Neri A., M.G. Ranalli (2011), *To misreport or not to report? The measurement of household financial wealth*, Statistics in transition new series, 12, 2, 281-300.
- Neri A., R. Zizza (2010), *Income reporting behaviour in sample surveys*, Banca d'Italia, Temi di discussione, n. 777.

- Oh, H. L., Scheuren, F. J. (1983). Weighting adjustments for unit non-response. In W. G. Madow, I. Olkin, and D. B. Rubin (Eds.), *Incomplete data in sample surveys* (Vol. 2): Theory and bibliographies, pp. 143-184. Academic Press (New York; London).
- Nicolini, G., Marasini, D., Montanari, G.E., Pratesi, M., Ranalli, M.G., Rocco, E, (2013), *Metodi di stima in presenza di errori non campionari*, UNITEXT, Collana di Statistica e Probabilità Applicata, Springer Science & Business.
- Pissarides C.A., G. Weber (1989), *An Expenditure Based Estimate of Britain's Black Economy*, *Journal of Public Economics*, 39(1), Giugno.
- Rubin, D B, (1978) *Multiple Imputations in Sample Surveys—A Phenomenological Bayesian Approach to Nonresponse*, *Proceedings of the Survey Research Methods Section*, American Statistical Association.
- Rubin, D.B, (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley.
- Statistics Canada, (2009), *Quality Guidelines*, Fifth Edition.