



BANCA D'ITALIA
EUROSISTEMA

L'indagine sui bilanci delle famiglie italiane: aspetti metodologici e utilizzo dei dati

Dicembre 2016

L'indagine sui bilanci delle famiglie italiane: aspetti metodologici e utilizzo dei dati

L'indagine sui bilanci delle famiglie italiane (IBF) è un'indagine campionaria condotta dalla Banca d'Italia sin dagli anni '60 con l'obiettivo di raccogliere informazioni sui redditi e i risparmi delle famiglie italiane.

Le motivazioni dell'indagine sono esplicitamente ricordate in uno dei primi rapporti:

“L'importanza economica che rivestono le famiglie nel nostro sistema, così come nella maggioranza di quelli ad economia di mercato, appare evidente ove si consideri che esse possiedono direttamente o indirettamente la quasi totalità della ricchezza nazionale, percepiscono quasi tutto il reddito nazionale e da esse provengono, attualmente in Italia, circa i tre quarti della domanda globale interna. Anche dal punto di vista finanziario il peso delle famiglie è notevole, dando esse origine a una parte sostanziale dei flussi finanziari e possedendo una quota notevole della ricchezza mobiliare. [...] Allo scopo ultimo di acquisire tale conoscenza e, in via preliminare, di stimare alcune grandezze concernenti il settore delle famiglie, la Banca d'Italia ha iniziato una nuova serie di ricerche campionarie sul reddito, il consumo e il risparmio delle famiglie italiane.”¹.

Nel corso degli anni l'oggetto della rilevazione si è andato estendendo per includere anche la ricchezza e altri aspetti inerenti i comportamenti economici e finanziari delle famiglie, come ad esempio l'uso dei mezzi di pagamento.

In connessione con lo svolgimento dei propri compiti istituzionali, la Banca d'Italia effettua anche attività di raccolta, produzione e pubblicazione di informazioni statistiche. I microdati delle indagini campionarie forniscono informazioni complementari a quelle derivanti dalle fonti macro (ad es., Contabilità Nazionale) in quanto consentono le analisi economiche su particolari sottopopolazioni di interesse come sulle famiglie benestanti o povere, le famiglie a basso o alto reddito, i nuclei (altamente) indebitati o le famiglie finanziariamente vulnerabili.

Questo documento fornisce una sintesi delle principali caratteristiche metodologiche dell'indagine, con particolare enfasi su quelle riguardanti le procedure statistiche e il trattamento dei dati. Esso è destinato a fornire un solido *background* per i potenziali utenti dei microdati dell'indagine.

1. Il disegno campionario

L'indagine ha da sempre l'obiettivo di rappresentare la popolazione (ufficialmente) residente in Italia. Sono esclusi dalla popolazione obiettivo coloro che vivono in istituti (conventi, ospedali, prigionieri...) o che sono illegalmente presenti sul territorio nazionale.

Il disegno di campionamento dell'indagine prevede una procedura di selezione a due stadi. Le unità di primo stadio sono i comuni²; le unità di secondo stadio sono le famiglie. Prima di procedere all'estrazione delle unità di primo stadio, queste vengono suddivise in base alla regione e alla classe di ampiezza demografica³ (*stratificazione* delle unità di primo stadio).

All'interno di ogni strato i comuni nei quali effettuare le interviste sono selezionati includendo tutti quelli con popolazione superiore ai 40.000 abitanti (comuni *auto-rappresentativi*) ed estraendo i comuni rimanenti con una modalità di selezione che assegna ai comuni di maggiore dimensione una probabilità più elevata di essere inclusi nel campione (*probability proportional to size - PPS*)⁴. In una seconda fase, per ciascun comune selezionato, le famiglie da intervistare vengono estratte casualmente dalle liste anagrafiche.

¹ Banca d'Italia (1966), Reddito, risparmio e alcuni consumi delle famiglie italiane nel 1965, Bollettino n. 4.

² Nelle ultime edizioni dell'indagine il numero di unità di primo stadio è variato tra i 350 e i 390 comuni.

³ Le classi utilizzate sono: fino a 20.000 abitanti, fra 20.000 e 40.000 abitanti, oltre 40.000 abitanti.

⁴ Questa tecnica consente di avere un campione a due stadi autoponderante quando la dimensione campionaria è costante negli strati. Infatti, fissando il numero di famiglie da intervistare in un dato comune, la più alta probabilità di includere un comune di maggiore dimensione nel primo stadio è esattamente compensata dalla minor probabilità di estrarre le unità di quel comune nel secondo stadio.

L'indagine è stata effettuata fino al 1987 sulla base di rilevazioni indipendenti a cadenza pressoché annuale, mentre successivamente si è passati a rilevazioni biennali (con eccezione del 1998). Nel corso degli anni l'ampiezza del campione obiettivo è cresciuta fino a raggiungere le 8.000 famiglie.

Dall'indagine sul 1989, per favorire l'analisi dell'evoluzione dei fenomeni oggetto di indagine, è stato introdotto uno schema che prevede la presenza nel campione di una quota di unità già intervistate in occasione di precedenti edizioni dell'indagine (famiglie *panel*). La tavola 1 mostra la numerosità campionaria delle indagini dal 1987 al 2014, evidenziando il numero di famiglie intervistate più di una volta nel corso del tempo. Per la parte *panel* del campione, sono state selezionate le famiglie presenti da almeno due rilevazioni e una parte di quelle rilevate solo nella precedente indagine⁵. Le famiglie non *panel* sono state invece estratte casualmente dalle liste anagrafiche.

2. Il questionario e la fase di rilevazione

I questionari utilizzati nelle rilevazioni, integralmente disponibili alla pagina web “Documentazione per l'utilizzo dei microdati” della sezione dedicata all'indagine del sito internet della Banca d'Italia⁶, sono predisposti seguendo una struttura modulare. Si compongono di una parte di base, nella quale sono rilevati i fenomeni ai quali tutte le famiglie sono interessate, e di diversi allegati, in cui sono raccolte informazioni che riguardano soltanto specifici sottoinsiemi di famiglie.

A partire dall'indagine sul 1998 la rilevazione dei dati è stata effettuata in prevalenza (oltre il 90 per cento nelle ultime rilevazioni) con l'aiuto del computer (CAPI, *Computer-Assisted Personal Interviewing*). Con questa metodologia i dati vengono rilevati presso le famiglie mediante un questionario elettronico, che, oltre a memorizzare le informazioni, aziona una serie di controlli consentendo di risolvere, in presenza della famiglia, le eventuali incoerenze⁷.

Le restanti interviste sono state realizzate con un questionario cartaceo (PAPI, *Paper-And-pencil Personal Interviewing*) e successivamente trasferite su supporto elettronico dalla società di rilevazione, utilizzando il programma CAPI come procedura di immissione dei dati.

Nelle ultime rilevazioni la durata media dell'intervista si è attestata in media a circa 50 minuti. Tuttavia questa è abbastanza variabile all'interno del campione, risultando positivamente correlata al reddito, alla ricchezza e al numero dei componenti (e in particolare dei percettori) della famiglia (tav. 2).

L'esecuzione delle interviste è preceduta da un ciclo di incontri in cui funzionari della Banca d'Italia e rappresentanti della società di rilevazione impartiscono le istruzioni direttamente agli intervistatori e da una indagine pilota in cui vengono sottoposte a test le nuove domande presenti nel questionario.

3. Il trattamento dei dati personali

Le famiglie che vengono contattate per partecipare all'indagine sono selezionate dall'Istat a partire dalle liste anagrafiche. I nominativi sono estratti attraverso un criterio di casualità e sono poi comunicati alla Banca d'Italia al fine di consentire lo svolgimento dell'indagine.

La partecipazione delle famiglie all'indagine è su base volontaria e pertanto il conferimento dei dati è facoltativo⁸. Le interviste raccolte sono sottoposte in primo luogo a controlli di qualità volti a selezionare quelle che possono essere considerate valide. In particolare, in presenza di un numero elevato di domande non compilate, o nel caso in cui l'intervista sia compilata da una famiglia diversa da quella che ha ricevuto l'invito a partecipare, l'intervista è annullata⁹. Ai sensi dell'articolo 7 del d.lgs. n.196/2003, a ogni famiglia

⁵ Come in passato, per disporre di informazioni utili per le analisi intergenerazionali sono state inoltre contattate per l'intervista le famiglie che si sono formate dalle originarie unità *panel* incluse nel campione (tipicamente quelle formate dai figli che costituiscono una nuova famiglia).

⁶ <https://www.bancaditalia.it/statistiche/tematiche/indagini-famiglie-imprese/bilanci-famiglie/documentazione/index.html>

⁷ Numerose sono le cause alla base delle incoerenze che caratterizzano i dati rilevati: da parte dell'intervistato possono esserci problemi di comprensione della domanda, di memoria nel fornire una risposta adeguata o anche reticenza nel fornire informazioni percepite come riservate; da parte dell'intervistatore gli errori più frequenti riguardano la digitazione dei codici di risposta o la scrittura dei valori in un'unità di misura diversa da quella prevista nel questionario.

⁸ Le famiglie contattate per l'intervista ricevono un opuscolo che illustra le finalità dell'indagine e fornisce alcuni esempi di utilizzo dei dati raccolti. Le famiglie non ricevono alcun compenso per l'intervista rilasciata; in occasione della pubblicazione dei risultati dell'indagine ad esse viene inviata una lettera di ringraziamento alla quale vengono allegati alcuni articoli di giornale riportanti i commenti della stampa.

⁹ La sezione “Il controllo dei dati e l'imputazione dei dati mancanti” offre maggiori informazioni sulle procedure di controllo dei dati.

intervistata è riconosciuto tra l'altro il diritto di: (i) accedere ai propri dati personali; (ii) chiederne la rettifica, l'aggiornamento o la cancellazione; (iii) opporsi al loro trattamento per motivi legittimi. Tali diritti potranno essere esercitati facendone richiesta al Responsabile designato per il riscontro in caso di esercizio dei menzionati diritti.

Il Titolare del trattamento è la Banca d'Italia, Istituto di diritto pubblico con sede legale in Roma, via Nazionale 91. Il Responsabile del trattamento dei dati, per le operazioni di trattamento svolte dalla Banca d'Italia, è il Capo del Servizio Analisi Statistiche, Via Nazionale 91, 00184 Roma.

Le informazioni relative alle famiglie selezionate a partecipare all'indagine sono trattate nel rispetto della vigente normativa in materia di protezione dei dati personali (d.lgs. n. 196/2003 e s.m.i.) e sono utilizzate dalla Banca d'Italia esclusivamente per le finalità dichiarate. Tali informazioni potrebbero essere inoltre elaborate congiuntamente a dati di fonte amministrativa (banca dati reddituale, banca dati integrata della proprietà immobiliare e archivio dei rapporti con operatori finanziari dell'Agenzia delle Entrate, Sistema Informativo Integrato di Acquirente Unico S.p.A, archivi di Infocamere, Centrale dei Rischi, COVIP e INPS). Per l'effettuazione del *record linkage*, i nominativi selezionati per partecipare all'indagine verrebbero condivisi con le Istituzioni proprietarie di questi dati amministrativi. Il trattamento dei nominativi sarebbe in ogni caso effettuato unicamente da personale di queste Istituzioni espressamente autorizzato ai sensi dell'art. 30 del decreto legislativo n. 196/2003, nel rispetto dei principi di necessità, di pertinenza e di non eccedenza (artt. 3 e 11). Nessuna informazione economica rilevata durante l'intervista sarà conferita insieme ai nominativi a tali Istituzioni. La banca dati integrata (che unisce tutte le informazioni rilevate con l'indagine con quelle amministrative) sarebbe gestita esclusivamente dagli addetti della Divisione Indagini campionarie della Banca d'Italia. Ove si rendesse necessaria una condivisione di parte di tali dati integrati con le Istituzioni ricordate in precedenza, ad esempio ai fini del miglioramento della loro qualità, i dati condivisi sarebbero preliminarmente resi anonimi e potrebbero essere usati solo per finalità di ricerca.

La comunicazione di dati personali da parte dell'Istat alla Banca d'Italia è realizzata secondo le condizioni e le modalità stabilite dall'articolo 21, comma 2, del Regolamento (CE) n. 223/2009, il quale autorizza esplicitamente la "trasmissione di dati riservati" da un'autorità del Sistema statistico europeo a un membro del Sistema europeo delle Banche Centrali a condizione che tale trasmissione, come nel caso di specie, "sia necessaria ai fini dell'efficienza dello sviluppo, della produzione e della diffusione di statistiche europee o del miglioramento della loro qualità".

I microdati diffusi pubblicamente sul sito internet della Banca d'Italia contengono solo informazioni campionarie e sono resi totalmente anonimi in modo tale che non sia possibile risalire ai soggetti ai quali si riferiscono.

4. Il controllo dei dati e l'imputazione dei dati mancanti

La rilevazione mediante la metodologia CAPI ha fortemente limitato la necessità di verifiche di coerenza effettuate a posteriori sui dati acquisiti. Le tradizionali fasi di controllo sono però state condotte per le interviste svolte con il questionario cartaceo (meno di un decimo) per le quali, come si è detto, il programma CAPI è stato utilizzato come maschera di immissione dei dati per sfruttare la sua capacità di segnalare le incoerenze.

Conclusa la fase di controllo, è stata effettuata l'imputazione delle mancate risposte, che possono dipendere sia dalla reticenza dell'intervistato, sia dalla difficoltà da questi incontrata nel fornire risposta a un particolare quesito. L'imputazione si rende necessaria per tutte le variabili elementari che costituiscono le componenti di variabili aggregate, in quanto anche la mancanza di una sola componente non consentirebbe il calcolo della variabile aggregata (è ad esempio necessario imputare le integrazioni non monetarie come i buoni pasto per calcolare il reddito da lavoro dipendente).

Il numero dei dati imputati è in generale modesto. Tra le variabili rilevate che presentano valori mancanti vi sono le integrazioni non monetarie dei lavoratori dipendenti, i proventi dei lavoratori autonomi e il valore delle aziende.

La metodologia utilizzata per l'imputazione fa riferimento a modelli di regressione, mediante i quali si stimano i valori da attribuire alle mancate risposte sulla base di altre informazioni disponibili. Al fine di evitare una eccessiva concentrazione intorno ai valori medi, al dato così stimato viene aggiunta una componente casuale, estratta da una variabile di tipo normale, con media zero e varianza pari a quella dei

residui del modello di regressione utilizzato. In tal modo si intendono preservare la media e la varianza che caratterizzano i dati effettivamente rilevati¹⁰.

5. Le mancate interviste

Il campione di famiglie effettivamente intervistate è composto da circa 8.000 famiglie, di cui circa la metà *panel*. Per ottenere queste interviste, è stato necessario contattare nelle ultime rilevazioni circa 15.000 famiglie. Il tasso di partecipazione all'indagine è poco più del 50 per cento, storicamente più elevato per le famiglie *panel* (circa l'80 per cento) rispetto a quelle non *panel* (circa il 35 per cento; tav. 3).

Il motivo più frequente all'origine della mancata intervista è l'indisponibilità. In circa un quarto dei casi le famiglie sono risultate irreperibili, ovvero non è stato possibile contattare telefonicamente la famiglia e non è stato trovato nessuno in casa in tre visite effettuate dagli intervistatori in giorni e orari diversi.

Le mancate interviste costituiscono un problema nelle indagini statistiche in quanto possono condurre a campioni nei quali i segmenti di popolazione meno disposti a collaborare sono sottorappresentati, producendo stime distorte. Gli stimatori utilizzati si basano sulle informazioni di coloro che hanno partecipato alla rilevazione e possono presentare una distorsione che cresce all'aumentare della mancata risposta e del divario tra il valore atteso della variabile oggetto di studio per i rispondenti e per i non rispondenti¹¹.

Nel corso degli anni sono state messe a punto diverse strategie volte a limitare il fenomeno della mancata risposta¹². Prima della rilevazione le famiglie ricevono una lettera volta a sensibilizzarle sull'importanza dell'indagine e a rassicurarle sulla riservatezza delle informazioni fornite; le famiglie hanno a disposizione un numero verde e i contatti della Banca d'Italia che si occupano dell'indagine per eventuali chiarimenti; l'intervistatore consegna alle famiglie un opuscolo illustrativo sulle finalità dell'indagine; gli intervistatori sono selezionati con sempre maggiore accuratezza tra quelli di maggiore esperienza; il numero totale di nominativi a disposizione di ciascun intervistatore per raggiungere il proprio obiettivo di famiglie intervistate è stato progressivamente ridotto.

Per limitare a posteriori i possibili effetti negativi della mancata partecipazione sono state inoltre adottate alcune misure. In primo luogo, le famiglie che non è stato possibile intervistare sono state sostituite con altre estratte con criteri casuali negli stessi comuni. La sostituzione avviene attraverso una rigida procedura sotto il controllo della società di rilevazione che stabilisce sia quando un nominativo può essere sostituito sia il nuovo nominativo da contattare. In secondo luogo, al termine della rilevazione, è stata 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 (si veda sezione 8).

Nelle ultime rilevazioni per intervistare le circa 8000 famiglie del campione, è stato necessario porre in atto nei loro confronti circa 13.000 azioni di contatto. La difficoltà a ottenere l'intervista è crescente al crescere del reddito, della ricchezza e del titolo di studio del capofamiglia (tav. 2).

6. La valutazione della qualità delle risposte

Sebbene per un'indagine sul reddito e la ricchezza la mancata risposta costituisca un fenomeno rilevante nel determinare problemi di qualità delle stime, numerosi altri fattori sono indicati in letteratura come potenziali elementi di disturbo.

Un aspetto che può influire sulla qualità delle stime riguarda la reticenza delle famiglie a dichiarare le proprie fonti di reddito o le forme di attività finanziarie o reali possedute. Sebbene la partecipazione all'indagine sia volontaria e il contenuto della rilevazione noto all'intervistato sin dall'inizio, è possibile che

¹⁰ Si veda il paragrafo 10.6 in *Survey Methodology*, a cura di Groves R. M., Fowler F. J., Couper M. P., Lepkowski J. M., Singer E. e Tourangeau R., New York, Wiley, 2004.

¹¹ Si veda G. D'Alessio e I. Faiella, *Nonresponse behaviour in the Bank of Italy's Survey of Household Income and Wealth*, Banca d'Italia, Temi di discussione, n. 462, 2002.

¹² Il fenomeno della riduzione della partecipazione alle indagini campionarie è comune ad altri paesi. Si veda E. de Leeuw e W. de Heer, *Trends in Household Survey Nonresponse: A Longitudinal and International Comparison*, in *Survey Nonresponse*, a cura di Groves R. M., Dillman D. A., Eltinge J. L. e Little R. J. A., New York, Wiley, 2002, pp. 41-54; si veda inoltre il paragrafo 6.5 in *Survey Methodology*, a cura di Groves R. M., Fowler F. J., Couper M. P., Lepkowski J. M., Singer E. e Tourangeau R., New York, Wiley, 2004.

talvolta questi non risponda con totale sincerità alle domande più delicate, quali quelle riguardanti il reddito o la ricchezza.

Per disporre di elementi di giudizio su tale fenomeno, che per sua natura risulta difficilmente investigabile, è stato richiesto agli intervistatori di formulare, immediatamente dopo l'intervista, un giudizio sintetico sulla presunta attendibilità delle risposte su reddito e ricchezza, in base alla 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.)¹³.

Storicamente si rileva che, sebbene il livello di attendibilità sia risultato in media soddisfacente, esso non è omogeneo all'interno del campione. Si riscontrano giudizi migliori per le famiglie con livelli di reddito e ricchezza bassi, residenti al Centro o con capofamiglia con elevato titolo di studio o dirigente. Un'attendibilità leggermente inferiore si riscontra per le famiglie con capofamiglia lavoratore indipendente (imprenditore o lavoratore autonomo) (tav. 2).

Ulteriori elementi per valutare l'attendibilità delle risposte fornite dagli intervistati possono essere ottenuti comparando le stime desunte dall'indagine con i dati della Contabilità Nazionale. Queste comparazioni vanno effettuate con cautela, tenendo conto che, almeno in parte, le differenze riscontrate possono essere dovute alle diverse definizioni adottate¹⁴.

Il reddito medio familiare stimato dall'indagine è inferiore a quello della Contabilità Nazionale di circa il 25 per cento. Per il totale della ricchezza reale, dal confronto con i dati diffusi dalla Banca d'Italia nel supplemento *La ricchezza delle famiglie italiane*, si evidenzia come nel complesso le due fonti siano sostanzialmente allineate. Le componenti della ricchezza finanziaria appaiono invece soggette a sottostima¹⁵.

7. L'aggregazione delle variabili

I principali aggregati economici, costruiti a partire dalle voci del questionario, sono riportati negli schemi qui presentati.

¹³ I giudizi degli intervistatori sono espressi con un punteggio compreso tra 1 (per nulla attendibile) e 10 (totalmente attendibile). Gli intervistatori forniscono informazioni anche sul livello di comprensione delle domande da parte dell'intervistato e sul clima complessivo nel quale si è svolta l'intervista.

¹⁴ Per un'analisi delle incoerenze tra le due fonti e per possibili esercizi di riconciliazione, si veda G. D'Alessio e A. Neri, *Income and wealth sample estimates consistent with macro aggregates: some experiments*, in Banca d'Italia, *Questioni di Economia e Finanza* n. 272, 2015. Le stime desunte dall'indagine sono state, in passato, comparate anche con quelle derivanti dalle dichiarazioni fiscali, segnalando un sostanziale allineamento nei redditi da lavoro dipendente e una sottostima consistente dei redditi da lavoro autonomo dichiarati al fisco rispetto a quelli rilevati nell'indagine. Al riguardo, si veda L. Cannari, V. Ceriani e G. D'Alessio, *Il recupero degli imponibili sottratti a tassazione*, in Banca d'Italia, *Ricerche quantitative per la politica economica - 1995*, Roma, 1997.

¹⁵ Oltre che ai noti fenomeni di mancata risposta e *under-reporting*, parte di questa discrepanza è spiegabile da come le poste sono valutate nelle diverse fonti nonché da alcune voci, come le azioni, che per il settore delle famiglie sono stimate residualmente nei conti finanziari. Per un'analisi particolareggiata delle differenze di cui tener conto nel confrontare le stime micro dell'indagine con quelle macro della contabilità finanziaria si veda R. Bonci, G. Marchese e A. Neri, *La ricchezza finanziaria nei conti finanziari e nell'indagine sui bilanci delle famiglie italiane*, Banca d'Italia, Temi di discussione, n. 565, Banca d'Italia, 2005. Per l'analisi dell'*under-reporting* si vedano: L. D'aurizio, I. Faiella, S. Iezzi e A. Neri, *L'under-reporting della ricchezza finanziaria nell'indagine sui bilanci delle famiglie*, Temi di discussione, n. 610, 2006; L. Cannari e G. D'Alessio, *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, 1993, p. 395-412. Per quella sulla mancata risposta G. D'Alessio e I. Faiella, *Nonresponse behaviour in the Bank of Italy's Survey of Household Income and Wealth*, Banca d'Italia, Temi di discussione, n. 462, 2002. Per un lavoro che considera entrambi gli aspetti si veda A. Neri e M.G.: Ranalli, *To misreport or not to report? The case of the Italian Survey on Household Income and Wealth*,. *Statistics in Transition*,12, 281-300 (2011).

Piano di aggregazione delle variabili: conto del reddito

Nome	Descrizione ⁽¹⁾
Y	Reddito disponibile netto
YL	Reddito da lavoro dipendente
YL1	Retribuzioni nette
YL2	Integrazioni non monetarie
YT	Pensioni e trasferimenti netti
YTP	Pensioni e arretrati
YTP1	Pensioni
YTP2	Arretrati
YTA	Altri trasferimenti
YTA1	Assistenza economica (CIG, ecc.)
YTA2	Borse di studio
YTA3	Assegni alimenti e regali/contributi ricorrenti
YTA31	in entrata
YTA32	in uscita (-)
YM	Reddito netto da lavoro autonomo
YMA1	Reddito netto da lavoro autonomo
YMA2	Utili e altri redditi da partecipazione
YC	Reddito da capitale
YCA	Reddito da fabbricati
YCA1	Affitti effettivi
YCA2	Affitti imputati ⁽²⁾
YCF	Reddito da capitale finanziario ⁽³⁾
YCF1	Interessi attivi su depositi
YCF2	Interessi attivi su titoli di Stato
YCF3	Rendimenti altri titoli
YCF4	Interessi passivi (-)

$$Y = YL + YT + YM + YC$$

(1) Il simbolo (-) indica che la posta entra con il segno negativo nel calcolo dell'aggregato gerarchicamente superiore.

(2) Ad eccezione degli immobili utilizzati per l'attività di lavoro indipendente.

(3) Tasso di interesse * stock di capitale.

Piano di aggregazione delle variabili: conto dell'utilizzazione reddito

Nome	Descrizione
Y	Reddito disponibile netto
C	Consumi
CD	Consumi durevoli
CD1	Spese per mezzi di trasporto
CD2	Spese per mobili, ecc.
CN	Consumi non durevoli
S	Risparmio ⁽¹⁾

$$Y = C + S$$

(1) Determinato a saldo.

Piano di aggregazione delle variabili: conto del patrimonio

Nome	Descrizione ⁽¹⁾
W	Ricchezza netta
AR	Attività reali
AR1	Immobili
AR2	Aziende
AR3	Oggetti di valore
AF	Attività finanziarie
AF1	Depositi
AF2	Titoli di Stato
AF3	Altri titoli
AF4	Crediti commerciali e da altre fam.
PF	Passività finanziarie (-)
PF1	Debiti verso banche e soc. finanz.
PF2	Debiti commerciali
PF3	Debiti verso altre famiglie
Per memoria:	
BD	Beni di consumo durevoli
BD1	Mezzi di trasporto
BD2	Mobili, ecc.
W = AR + AF - PF	
(1) Il simbolo - indica che la posta entra con il segno negativo nel calcolo dell'aggregato gerarchicamente superiore.	

8. Le stime campionarie

Data la generica variabile oggetto di studio y , uno stimatore corretto¹⁶ della media della popolazione è lo stimatore di Horwitz-Thompson-Narain dato da:

$$(1) \quad \bar{y} = \frac{\sum_h \sum_i \sum_j y_{hij} w_{hij}}{\sum_h \sum_i \sum_j w_{hij}} \quad j = 1, \dots, n_{hi} \quad i = 1, \dots, a_h \quad h = 1, \dots, H$$

dove y_{hij} rappresenta il valore della variabile y osservato per la j -esima famiglia intervistata nel i -esimo comune dell' h -esimo strato, mentre w_{hij} costituisce il suo peso campionario.

Il peso da attribuire alle unità campionarie è ottenuto per mezzo di una procedura articolata nelle seguenti fasi¹⁷:

- a) inizialmente viene calcolato il *peso da disegno* come reciproco della probabilità di inclusione delle unità ($w^{(0)}$);
- b) si ottiene poi un peso corretto per la mancata risposta totale ($w^{(1)}$) moltiplicando $w^{(0)}$ per l'inverso del tasso di risposta del comune cui ciascuna unità appartiene;

¹⁶ Uno stimatore è "corretto" (*unbiased*) quando il suo valore medio coincide con il parametro della popolazione. Lo stimatore (1) è in realtà "approssimativamente" corretto, ossia presenta una distorsione di entità limitata che tende a zero al crescere della numerosità campionaria. Tale stimatore, presenta una distorsione massima pari al coefficiente di variazione campionario dei pesi (normalmente intorno all'1 per cento per le stime familiari e allo 0,7 per cento per quelle individuali). Si veda L. Kish, *Survey Sampling*, New York, Wiley, 1995.

¹⁷ Per una descrizione dettagliata della struttura di ponderazione adottata nell'indagine e degli effetti dei pesi sulla varianza campionaria degli stimatori si veda I. Faiella e R. Gambacorta, *The weighting process in the SHIW*, Banca d'Italia, Temi di discussione, n. 636, 2007.

- c) il peso $w^{(1)}$ viene inoltre modificato per tener conto della presenza della componente *panel* (ottenendo il peso $w^{(2)}$);
- d) infine, $w^{(2)}$ viene aggiustato sulla base di informazioni ausiliarie (*post-stratificazione*), giungendo al peso finale $w^{(3)}$.

Inizialmente, in accordo con il piano di campionamento, a ciascun individuo della famiglia viene attribuito un peso, definito come l'inverso della probabilità di inclusione nel campione (*peso da disegno*). Il coefficiente di ponderazione risulta costante a livello di comune e assume la forma:

$$(2) \quad w_{hi}^{(0)} = \frac{1}{m_h} \frac{P_h}{P_{hi}} \frac{P_{hi}}{n'_{hi}} = \frac{1}{m_h} \frac{P_h}{n'_{hi}}$$

dove P_h e m_h sono rispettivamente la popolazione residente e il numero di comuni campione nell'*h-esimo* strato, e P_{hi} e n'_{hi} rispettivamente la popolazione e il numero di componenti da intervistare (teorici) nell'*i-esimo* comune dell'*h-esimo* strato¹⁸.

Il peso da disegno viene aggiustato per tener conto del rifiuto di alcune famiglie di partecipare all'indagine. La correzione per la mancata risposta consiste nell'inflazionare i pesi per il rapporto tra dimensione del campione teorico e dimensione del campione effettivo (*Weighting Class Adjustment*):

$$(3) \quad w_{hi}^{(1)} = w_{hi}^{(0)} \frac{n_{hi}}{n'_{hi}}$$

dove n_{hi} rappresenta il numero di componenti effettivamente intervistati nell'*i-esimo* comune dell'*h-esimo* strato¹⁹.

Come si è visto nel paragrafo 1, il disegno dell'indagine prevede che una parte del campione sia costituita da famiglie intervistate nelle rilevazioni precedenti (famiglie *panel*). Le famiglie *panel* possono presentare caratteristiche socio-demografiche in parte diverse rispetto a quelle dell'intero campione, sostanzialmente a causa del processo di deterioramento della componente *panel* legato alla non partecipazione in successive *wave* (*attrition*). Per correggere tale possibile fonte di distorsione delle stime, si procede a una stratificazione a posteriori della parte *panel* del campione sulla base di alcune caratteristiche relative alla precedente indagine. In particolare l'aggiustamento è basato sull'incrocio fra classi di reddito (a 4 modalità) e condizione professionale del capofamiglia (atre modalità)²⁰:

$$(4) \quad w_c^{(2)} = w_c^{(1)} \alpha_c$$

dove α_c rappresenta il fattore di aggiustamento per la cella c (con $c = 1, \dots, 12$).

Inoltre, il peso viene modificato per riprodurre la quota ottimale della componente *panel* (stimata in circa il 50 per cento del campione).

Infine, i pesi vengono calibrati utilizzando informazioni esterne che, correlate con le variabili oggetto di indagine, hanno lo scopo di migliorare l'accuratezza delle stime e limitare ulteriormente le distorsioni legate alla mancata risposta. La correzione viene condotta imponendo al campione finale l'allineamento alle caratteristiche della popolazione per quanto riguarda il sesso, la classe di età (meno di 26 anni, da 26 a 45, da

¹⁸ Ciascun comune *auto-rappresentativo* è inserito in uno strato a se stante ed incluso con certezza nel campione. In tale caso la probabilità di estrazione di una famiglia è approssimativamente pari a n_{hi}/P_{hi} . I comuni *non auto-rappresentativi* sono selezionati con probabilità proporzionale alla dimensione (*probability proportional to size - PPS*); la probabilità di selezione dell'*i-esimo* comune nell'*h-esimo* strato è pertanto pari a $m_h P_{hi}/P_h$. La probabilità di inclusione di una famiglia nel campione può quindi essere scritta come: $m_h n_{hi}/P_{hi}$.

¹⁹ La correzione per la mancata partecipazione permette di correggere anche per i casi in cui le famiglie non sono state trovate dall'intervistatore all'indirizzo anagrafico a causa di trasferimenti, decessi o di indirizzi errati.

²⁰ Per tenere conto dell'*attrition* si sarebbero potuti utilizzare modelli statistici per la mancata partecipazione, come proposto da A. Giraldo, E. Rettore e U. Trivellato, *Attrition bias in the Bank of Italy's Survey of Households' Income and Wealth*, International Conference on Quality in Official Statistics, Stockholm, 14-15 Maggio 2001. Va comunque notato che, l'utilizzo di modelli o di tecniche di riponderazione nell'aggiustamento per la mancata risposta producono un impatto sostanzialmente simile, e normalmente contenuto, sulle stime campionarie. Si veda G. Kalton e I. Flores Cervantes, *Weighting Methods*, in *Journal of Official Statistics*, Vol.19, No.2, 2003, pp. 81-97.

46 a 65, oltre 65 anni), l'area geografica (Nord, Centro, Sud e Isole) e la dimensione del comune di residenza (fino a 20.000 abitanti, da 20.000 a 40.000, da 40.000 a 500.000, oltre 500.000 abitanti)²¹. I pesi finali sono dunque ottenuti come:

$$(5) \quad w_c^{(3)} = w_c^{(2)} \beta_c$$

dove β_c rappresenta il fattore di aggiustamento per la cella c .

9. La distribuzione dei dati

I microdati anonimi delle indagini sui bilanci delle famiglie condotte dal 1977 al 2014 sono disponibili sul sito Internet della Banca d'Italia all'indirizzo <http://www.bancaditalia.it/statistiche/tematiche/indagini-famiglie-imprese/bilanci-famiglie>²². Le informazioni necessarie per lo sfruttamento dei dati (questionari delle indagini più recenti, nomi delle variabili, note per l'utilizzo dei dati, ecc.) sono disponibili nella sezione documentazione in formato PDF. I dati sono disponibili in formato SAS, STATA e in formato ASCII separato da virgola (CSV, *comma separated value*).

È prevista la distribuzione di due tipologie di archivi. Gli **archivi annuali** contengono i dati delle indagini dal 1989 al 2014 in versione pressoché integrale (incluse le sezioni monografiche delle singole indagini). L'**archivio storico** riporta le informazioni riferite al periodo 1977-2014, in formato omogeneo, per il sottoinsieme delle variabili ritenute utili per le analisi longitudinali.

I pesi degli archivi annuali possono essere diversi dai corrispondenti dell'archivio storico. I pesi degli archivi annuali infatti non subiscono alcun processo di revisione, mentre i pesi dell'archivio storico vengono allineati alle statistiche demografiche sulla popolazione italiana rilasciate dall'ISTAT (ad es. ricostruzioni intercensuarie), quando queste sono oggetto di revisione. La revisione dei pesi campionari può causare delle discrepanze tra le statistiche desumibili dall'attuale versione dell'archivio storico e quelle di versioni precedentemente rilasciate.

10. L'uso dei dati

L'indagine rappresenta una importante fonte di informazione sulle condizioni economiche delle famiglie e contribuisce in modo rilevante al dibattito socio-politico su tale tema. I principali risultati vengono diffusi tramite i Supplementi al Bollettino Statistico e hanno un ampio impatto mediatico: vengono infatti ripresi dai principali quotidiani nazionali e citati nei dibattiti politici televisivi.

I micro dati dell'indagine vengono inoltre impiegati per la realizzazione di progetti di ricerca nell'ambito delle attività istituzionali della Banca d'Italia. Esempi di studi riguardano gli effetti delle manovre di politica monetaria sui consumi delle famiglie, la distribuzione del reddito e della ricchezza, il funzionamento dei mercati finanziari e aspetti inerenti la tassazione o le riforme del sistema pensionistico²³. Un impiego di particolare rilevanza è la valutazione dell'impatto di particolari misure di politica economica attraverso un modello di micro-simulazione.

I dati dell'indagine, dopo essere resi totalmente anonimi, sono messi a disposizione anche di utenti esterni appartenenti ad istituzioni di ricerca pubbliche e private. Le principali pubblicazioni scientifiche basate sull'indagine si possono quantificare in circa 800 lavori.

11. L'indagine nel contesto internazionale

Negli anni più recenti l'indagine è entrata a far parte di progetti internazionali che consentono il confronto della condizione economica delle famiglie tra i principali paesi occidentali.

²¹ La tecnica utilizzata consente di allineare in modo simultaneo i pesi campionari alla distribuzione di alcune caratteristiche note da fonti esterne considerando esclusivamente le probabilità marginali. Tale metodo è noto come *Iterative Proportional Fitting* (o *Raking*). Si veda G. Kalton e I. Flores Cervantes, *Weighting Methods*, in *Journal of Official Statistics*, Vol.19, No.2, 2003, pp. 81-97.

²² La documentazione e le pubblicazioni sono disponibili sia in italiano sia in inglese.

²³ Si veda la lista dei riferimenti bibliografici dei principali lavori di ricerca che sfruttano la base dati dell'indagine all'indirizzo <http://www.bancaditalia.it/statistiche/tematiche/indagini-famiglie-imprese/bilanci-famiglie/documentazione/Bibliografia-30112015.pdf>.

I primi due progetti sono il *Luxembourg Income Study* ed il *Luxembourg Wealth Study*, il cui scopo è quello di produrre archivi di dati armonizzati sul reddito e ricchezza delle famiglie (www.lisdatacenter.org/). L'armonizzazione avviene "a posteriori" individuando le grandezze rilevate in modo comparabile e creando una documentazione unica per le varie indagini in modo da rendere più facile il loro uso. I microdati sono messi a disposizione tramite un sistema di elaborazione a distanza e un *web tabulator*.

A partire dal 2008 l'indagine è anche parte integrante di un progetto della Banca Centrale Europea che ha lo scopo di realizzare un'indagine armonizzata sui bilanci familiari nell'area dell'euro (*Household Finance and Consumption Survey - HFCS*). L'indagine rileva le variabili necessarie a ricostruire i bilanci delle famiglie sia in termini di consistenze sia di flussi, con particolare attenzione alle componenti della ricchezza, e fornisce informazioni sul comportamento delle famiglie utili per una migliore comprensione dei meccanismi di trasmissione della politica monetaria e per la valutazione di profili di stabilità finanziaria. I micro dati anonimizzati sono a disposizione degli studiosi per finalità di ricerca (www.ecb.europa.eu/pub/economic-research/research-networks/html/researcher_hfcn.en.html). Il principale vantaggio di questo progetto è l'armonizzazione delle metodologie e delle definizioni usate dai vari paesi per condurre le varie indagini che ne fanno parte²⁴.

²⁴ Gambacorta R., Ilardi G., Locatelli A. Pico R., Rampazzi C. (2013), Principali risultati dell'Household Finance and Consumption Survey: l'Italia nel confronto internazionale, *Questioni di Economia e Finanza*, Banca d'Italia, n. 161.

Tavola 1

Famiglie intervistate nel corso del tempo, indagini 1987-2014

Anno della prima partecipazione all'indagine	Anno di rilevazione													
	1987	1989	1991	1993	1995	1998	2000	2002	2004	2006	2008	2010	2012	2014
1987.....	8.027	1.206	350	173	126	85	61	44	33	30	28	23	21	13
1989.....		7.068	1.837	877	701	459	343	263	197	159	146	123	102	64
1991.....			6.001	2.420	1.752	1.169	832	613	464	393	347	293	244	166
1993.....				4.619	1.066	583	399	270	199	157	141	124	106	78
1995.....					4.490	373	245	177	117	101	84	75	62	46
1998.....						4.478	1.993	1.224	845	636	538	450	380	267
2000.....							4.128	1.014	667	475	398	330	256	170
2002.....								4.406	1.082	672	525	416	340	221
2004.....									4.408	1.334	995	786	631	395
2006.....										3.811	1.143	856	648	414
2008.....											3.632	1.145	806	481
2010.....												3.330	1.015	579
2012.....													3.540	1565
2014.....														3697
Numerosità campionaria	8.027	8.274	8.188	8.089	8.135	7.147	8.001	8.011	8.012	7.768	7.977	7.951	8.151	8.156
<i>Quota delle famiglie panel</i>		14,6	26,7	42,9	44,8	37,3	48,4	45,0	45,0	50,9	54,4	58,1	56,6	54,7

Contatti, durata media dell'intervista e attendibilità delle risposte, indagine 2014
(numero, minuti, punteggi in decimi, valori percentuali)

Modalità (*)	Contatti	Famiglie intervistate	Contatti ogni 100 famiglie	Durata media dell'intervista	Attendibilità delle risposte	Intervista CAPI
Sesso						
maschi.....	8.603	5.267	163,3	51,9	8,1	92,9
femmine	4.552	2.889	157,6	47,0	8,0	92,9
Età						
fino a 34 anni	809	475	170,3	52,8	8,1	86,7
da 35 a 44 anni	1.638	967	169,3	52,1	8,2	91,9
da 45 a 54 anni	2.715	1.566	173,4	54,2	8,2	93,3
da 55 a 64 anni	2.629	1.599	164,4	53,3	8,1	93,4
oltre 64 anni	5.365	3.549	151,2	46,1	8,0	93,5
Titolo di studio						
senza titolo	422	326	129,4	42,2	7,4	93,9
licenza elementare	2.614	1.804	144,9	43,4	7,9	94,2
media inferiore	4.684	2.909	161,0	49,1	8,0	93,2
media superiore	3.598	2.086	172,5	54,5	8,3	91,9
laurea	1.838	1.031	178,2	58,6	8,5	91,3
Condizione professionale						
Lavoratore dipendente						
operaio	2.442	1.507	162,1	51,9	8,1	92,8
impiegato	2.197	1.237	177,6	54,7	8,4	91,8
dirigente, direttivo	633	330	191,8	63,2	8,6	91,5
totale	5.272	3.074	171,5	54,2	8,3	92,3
Lavoratore indipendente						
imprenditore, libero professionista	638	363	175,7	59,0	7,8	91,5
altro autonomo	666	403	165,2	56,1	7,8	91,6
totale	1.303	766	170,2	57,5	7,8	91,5
Condizione non professionale						
pensionati	5.963	3.926	151,9	45,9	8,0	93,7
altri non occupati	617	390	158,1	46,3	7,4	91,8
totale	6.580	4.316	152,4	46,0	8,0	93,6
Numero di componenti						
1 componente	3.700	2.394	154,6	42,1	7,9	92,5
2 componenti	4.105	2.588	158,6	49,7	8,1	92,5
3 componenti	2.509	1.500	167,2	53,9	8,2	92,7
4 componenti	2.079	1.222	170,1	58,2	8,3	94,3
5 e più componenti	763	452	168,8	61,2	7,9	94,0
Numero di percettori						
1 percettore	6.744	4.336	155,5	45,1	7,9	92,9
2 percettori	5.162	3.085	167,3	54,4	8,3	92,6
3 percettori	1.006	604	166,6	61,1	8,2	94,7
4 e più percettori	243	131	185,5	67,9	8,1	91,6
Quinti di ricchezza netta familiare⁽¹⁾						
1° quinto.....	2.232	1.415	157,7	45,6	7,9	93,3
2° quinto.....	2.185	1.412	154,7	47,3	8,0	92,4
3° quinto.....	2.727	1.747	156,1	47,7	8,0	94,2
4° quinto.....	2.934	1.813	161,8	50,7	8,1	91,5
5° quinto.....	3.078	1.769	174,0	58,1	8,4	93,2
Quinti di reddito familiare⁽²⁾						
1° quinto.....	2.368	1.570	150,8	42,2	7,6	93,6
2° quinto.....	2.374	1.583	150,0	44,7	7,9	93,2
3° quinto.....	2.571	1.634	157,3	48,9	8,2	92,8
4° quinto.....	2.809	1.693	165,9	53,2	8,3	92,3
5° quinto.....	3.034	1.676	181,0	60,9	8,4	92,6
Ampiezza comunale						
fino a 20.000 abitanti	3.318	2.145	154,7	48,3	8,2	94,3
da 20.000 a 40.000 abitanti	2.416	1.477	163,6	50,5	8,2	92,3
da 40.000 a 500.000 abitanti	6.226	3.827	162,7	51,7	7,9	93,8
oltre 500.000 abitanti	1.195	707	169,0	47,1	8,2	85,1
Area geografica						
Nord	6.245	3.687	169,4	50,5	8,2	91,6
Centro	2.648	1.679	157,7	49,4	8,7	91,1
Sud e Isole	4.262	2.790	152,8	50,1	7,6	95,6
Totale	13.155	8.156	161,3	50,2	8,1	92,9

(*) Le caratteristiche individuali sono riferite al capofamiglia, inteso come il maggior percettore di reddito all'interno della famiglia; (1) 1° quinto: fino a 6.500€, 2° quinto: da 6.500€ a 92.000€, 3° quinto: da 92.000€ a 183.500 €, 4° quinto: da 183.500€ a 313.466€, 5° quinto: oltre 313.466€; (2) 1° quinto: fino a 14.982 €, 2° quinto: da 14.982€ a 21.600€, 3° quinto: da 21.600€ a 30.065€, 4° quinto: da 30.065€ a 43.748€, 5° quinto: oltre 43.748€

Tavola 3

Famiglie contattate e motivi della mancata intervista, indagine 2014

Famiglie:	<i>Panel</i>		<i>Non panel</i>		Totale	
	<i>(unità)</i>	<i>(percentuali)</i>	<i>(unità)</i>	<i>(percentuali)</i>	<i>(unità)</i>	<i>(percentuali)</i>
Intervistate	4.459	84,9	3.697	36,8	8.156	53,3
Indisponibili	626	11,9	3.975	39,6	4.601	30,1
Irreperibili	169	3,2	2.376	23,6	2.545	16,6
Totale	5.254	100,0	10.048	100,0	15.302	100,0
Inleggibili (*).....	180	3,4	618	6,2	798	5,2

(*) Famiglie non esistenti all'indirizzo anagrafico (indirizzi errati, decessi, trasferimenti).