



BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Cinquant'anni di indagini sui bilanci delle famiglie italiane:
storia, metodi, prospettive

di Alberto Baffigi, Luigi Cannari e Giovanni D'Alessio

Dicembre 2016

Numero

368



BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional papers)

Cinquant'anni di indagini sui bilanci delle famiglie italiane:
storia, metodi, prospettive

di Alberto Baffigi, Luigi Cannari e Giovanni D'Alessio

Numero 368 – Dicembre 2016

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

CINQUANT'ANNI DI INDAGINI SUI BILANCI DELLE FAMIGLIE ITALIANE: STORIA, METODI, PROSPETTIVE

di Alberto Baffigi*, Luigi Cannari* e Giovanni D'Alessio*

Sommario

Il lavoro descrive l'evoluzione dell'indagine sui bilanci delle famiglie dalla sua origine nei primi anni sessanta fino ad oggi, illustrando in che modo le innovazioni che si sono susseguite nel corso del tempo hanno consentito di migliorare la qualità dei dati raccolti e ampliare le possibilità di analisi. Il lavoro esamina inoltre estesamente in che modo i dati dell'indagine si rapportano a quelli desumibili da altre fonti (conti nazionali, dati fiscali, censimenti, altre indagini campionarie e così via), riassumendo i principali risultati dei numerosi lavori svolti su questo aspetto. Nelle conclusioni si rammentano le principali linee evolutive dell'indagine, che mettono in evidenza la necessità di perseguire una sempre maggiore integrazione dell'indagine con altre rilevazioni sul piano internazionale, con altre fonti campionarie e amministrative nel nostro Paese e con le statistiche aggregate.

Classificazione JEL: C81, C83, D31.

Parole chiave: indagini campionarie, microdati economici, distribuzione del reddito e della ricchezza delle famiglie.

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

Indice

1. Introduzione.....	9
2. La storia dell'indagine	9
Le indagini sui bilanci delle famiglie nella storia della statistica italiana	9
I primi passi delle indagini sulle famiglie in Italia	14
L'indagine del 1951	17
“Effetti di ipotetiche variazioni del reddito sulle uscite di un “campione” di famiglie italiane”	20
Le prime indagini della Banca d'Italia.....	22
Le indagini più recenti	26
L'utilizzo dell'indagine per la politica economica e la ricerca.....	30
3. Le principali caratteristiche dell'indagine	32
La composizione del campione.....	32
Il panel	34
Il questionario e la fase di rilevazione	37
Campionamento e pesi.....	40
Pesi individuali e pesi familiari.....	42
L'editing dei dati.....	45
La stima degli aggregati.....	47
Gli archivi annuali e l'archivio storico	52
Errori standard delle stime	53
Sovracampionamento.....	54
L'IBF e l'HFCS	56
4. Mancate risposte, under-reporting e confronto con altre fonti	59
Mancate risposte parziali	59
Mancate risposte totali e selection bias.....	61
Il confronto tra stime micro e stime macro e l'under-reporting.....	67
I redditi: confronto con i dati della Contabilità Nazionale	67
I redditi: confronto con i dati fiscali	70
I redditi: confronto con i dati dell'indagine Eu-Silc.....	75
I consumi: confronto con dati dell'indagine Istat	79
Le attività reali: confronto con dati di CN.....	80
Le abitazioni: coerenza delle risposte dell'indagine e confronto con dati di Censimento	80
Le abitazioni: confronto con i dati fiscali.....	82
Le attività finanziarie e la ricchezza nel suo complesso.....	84
Lo studio dei comportamenti di reticenza sulle attività finanziarie.....	88
Le attività finanziarie: confronto con le segnalazioni di Vigilanza.....	91
Il contante e le monete.....	94
Il giudizio degli intervistatori sulla veridicità delle risposte	94
Stime micro coerenti con stime macro.....	97
Altri errori di misura	97
Il ruolo dell'intervistatore	101
5. Problemi di qualità legati a particolari utilizzi dei dati	103
Dati pooled, stime regionali o per piccoli domini.....	103
L'utilizzo ai fini della microsimulazione.....	108
6. Conclusioni.....	110
Appendice A: Cronologia dei documenti utilizzati nel testo	113
Riferimenti bibliografici.....	115

1. Introduzione¹

A distanza di cinquant'anni dalla prima indagine sui bilanci delle famiglie questo lavoro ne ripercorre la storia, l'evoluzione della metodologia e i principali problemi affrontati nel corso del tempo.

In particolare viene effettuata una rassegna delle analisi effettuate sulla qualità delle informazioni rilevate con l'IBF e si comparano i risultati con quelli di altre fonti (CN e altri aggregati macroeconomici; dati amministrativi). Da questa rassegna emergono spunti di approfondimento per la ricerca futura e suggerimenti in merito alle possibili linee evolutive dell'indagine.

2. La storia dell'indagine

Le indagini sui bilanci delle famiglie nella storia della statistica italiana

Vi è un argomento unificante, una chiave di lettura, che consente di accedere sinteticamente, e allo stesso tempo con la dovuta attenzione ai dettagli, alla storia della statistica mondiale nel Ventesimo secolo: si tratta del graduale sviluppo delle tecniche di campionamento a partire dalla fine dell'Ottocento fino al secondo dopoguerra. Come osserva Bellhouse (1988), si tratta di un fondamentale cambiamento di paradigma che comporta il superamento di una concezione della statistica, tipicamente ottocentesca, che considera irrinunciabile la rilevazione esaustiva della popolazione; si passa a una concezione che consente via via di affinare metodi volti a ottenere informazioni sulla popolazione, sulla base di un sottoinsieme della stessa². Se si studia lo sviluppo della teoria dei campioni e i vari percorsi che hanno portato alla sua affermazione nei diversi paesi, si assiste – come a uno spettacolo, in prima fila – a una drammatizzazione, all'intreccio instabile fra metodologia, teoria della probabilità e impatto politico-istituzionale che ha portato all'affermazione di quello che i primi fautori della rilevazione campionaria chiamavano “metodo rappresentativo”.

Come osservava Stephan (1948), «modern sampling practice is the result of combining statistical theory with scientific knowledge about the material that is being sampled and with the experience that has been gained through the use of various sampling techniques in surveys and experimental work» (Stephan, 1948, p. 12).

¹ Si ringraziano i numerosi colleghi che a vario titolo hanno contribuito nel corso del tempo alla realizzazione dell'indagine e degli approfondimenti metodologici qui riportati. Nella stesura della storia dell'indagine dobbiamo molto alla consulenza storico-archivistica di Elisabetta Loche che ha consentito il reperimento dei documenti dell'Archivio Storico della Banca d'Italia utilizzati nell'articolo. Le opinioni espresse sono quelle degli autori e non impegnano la responsabilità della Banca d'Italia.

² Sul graduale affermarsi della rilevazione campionaria si veda, oltre al citato Bellhouse (1988), Desrosières (1998, capitolo 7), Seng (1951), Kruskal e Moesteller (1979a, 1979b, 1979c, 1980), Henneck (1987), Stephan (1948), SIS (1996), per citare solo alcuni importanti contributi. Sulla resistenza degli statistici italiani al “metodo rappresentativo” si veda anche Baffigi (2007, cap. 4.1).

È così che il metodo si è fatto strada in modo diverso nei diversi paesi, a seconda delle tradizioni metodologiche nazionali, dell'assetto politico e dell'importanza e del grado di autonomia del settore privato nel produrre statistiche: si trattava di combinare progressi teorici ed esigenze conoscitive eminentemente pratiche, non di rado connesse alle necessità di governo tipiche di società sempre più complesse³. In quest'ottica, non può sfuggire l'importanza di identificare e comprendere i processi politici, istituzionali e interpersonali che hanno con fatica consentito l'accettazione del nuovo paradigma scientifico in un paese, come l'Italia, che è riuscito a colmare il divario economico con i paesi più avanzati proprio nell'arco di tempo in cui la teoria e la pratica della rilevazione campionaria coronavano il loro successo. Quali furono gli ostacoli che vi si opposero? Quali furono le idee, i personaggi le istituzioni che ne favorirono la diffusione?

Questo è il tema che si affronta in queste prime pagine. L'obiettivo è circoscritto, concentrato su una catena di episodi specifici e documentati, quelli che portarono la Banca d'Italia a familiarizzare sempre più con le rilevazioni campionarie, al fine di conoscere un'economia in rapido mutamento strutturale, nella quale i comportamenti dei consumatori e degli operatori finanziari mutavano rapidamente. L'azione di personaggi in vario modo legati alla Banca d'Italia, come Einaudi, Menichella e Baffi, insieme ad altre figure accademiche e istituzionali come Del Vecchio e Vanoni consentì di aprire un varco a questo tipo di ricerche, facendole uscire dallo stato di minorità di fatto in cui si trovavano, e preparando un fecondo terreno di coltura per il loro ulteriore sviluppo.

Quando, nell'estate del 1966, la Banca d'Italia pubblicò i risultati della prima indagine ufficiale sui bilanci delle famiglie (Banca d'Italia, 1966), l'evento – sicuramente simbolico – non era il frutto di una decisione improvvisa ma la conseguenza di una lunga maturazione che aveva potuto beneficiare dell'humus culturale e politico nel quale, almeno a partire dalla istituzione del Servizio studi nel 1936, si era sviluppata la tecnostruttura della Banca. In realtà, infatti, non si trattava della prima indagine campionaria sulle famiglie italiane condotta dalla Banca, anche se essa fu sicuramente la prima in cui l'Istituto di emissione figurava ufficialmente come committente e referente scientifico dell'iniziativa. Altre indagini analoghe erano state svolte: tra il 1961 e il 1962 ben sei indagini sul reddito, sul consumo e sul risparmio delle famiglie italiane i cui risultati non furono pubblicati se non in maniera sintetica, con commenti, nel Bollettino di luglio-agosto 1966, che riportava i dati della prima indagine condotta nel marzo

³ Si veda, a titolo di esempio, l'articolo di Hennock (1987) il quale mostra come lo sviluppo dei metodi campionari, dovuto allo statistico Arthur Bowley, consentì di superare l'ambito puramente locale delle inchieste sulla povertà urbana e di costruire una conoscenza territorialmente più ampia di tale fenomeno. Nonostante gli sforzi prodotti in inchieste molto dettagliate, notava Bowley nel 1906, «we have nothing to go on but guesses as to the real extent of poverty»; e proseguiva osservando che «we are not taking any steps as yet to learn our existing condition in any complete way, though the work done in intensive inquiries would have been sufficient, if directed over the whole field, to have given us an adequate sample» (citato in Hennock, 1987, p. 220). I vantaggi del campionamento, venivano ricondotti al risparmio in termini di lavoro e di tempo che esso determina, grazie alla teoria della probabilità che consente di scegliere il grado di accuratezza delle misurazioni di fenomeni riferiti a una popolazione basate su un sottoinsieme della stessa, dato il budget a nostra disposizione.

del 1966. Ma in quelle occasioni la Banca appariva già pronta e organizzata a interagire con la società incaricata alla rilevazione.

Più interessante appare la meno nota indagine del 1951, condotta dalla Doxa di Pier Paolo Luzzatto Fegiz: rilevante perché la prima e anche perché ai più sconosciuta; una indagine corredata da una pubblicazione nella quale però non compariva la Banca tra i committenti; interessante, soprattutto perché l'iniziativa, con tutte le difficoltà che dovette affrontare per essere realizzata, emblematicamente ci ricorda i motivi, le idee e le persone che favorirono la coraggiosa, e per niente scontata, decisione di adeguare gli strumenti di ricerca del Servizio studi alle tecniche più moderne.

Dunque, fu una scelta coraggiosa: negli anni in cui la Banca d'Italia cominciò a svolgere indagini sui bilanci delle famiglie, tra l'inizio degli anni Cinquanta e gli anni Sessanta, la statistica italiana, proseguendo una tradizione solida e pluridecennale, non vedeva di buon occhio le indagini campionarie; d'altra parte, le principali famiglie politiche nutrivano una acuta ostilità nei confronti di questa pratica statistica, ormai affermatasi da anni negli Stati Uniti.

Gli statistici italiani, in primo luogo Luigi Bodio, capo della statistica ufficiale dal 1872 al 1898, si erano opposti al metodo campionario fin da quando esso aveva cominciato a delinearsi, abbozzato nelle proposte del norvegese Anders Nicolai Kiaer, nelle riunioni dell'Istituto internazionale di statistica, a partire da quella tenuta a Berna nel 1895⁴. La tradizione statistica italiana ha mantenuto questa posizione per molto ancora e in questo ebbe un ruolo di grande rilevanza il primo presidente dell'Istat, Corrado Gini il quale riuscì a dare dignità scientifica alle critiche opposte al cosiddetto "metodo rappresentativo". La prospettiva metodologica di Gini, di fronte alla rapida diffusione del metodo rappresentativo, puntò sul cosiddetto campionamento a scelta ragionata o giudiziosa (*purposive*, in inglese), nel quale le unità statistiche che compongono il campione vengono selezionate in quanto per qualche motivo ritenute, nel loro insieme, tipiche rispetto alla popolazione di riferimento⁵. Il contributo di Gini e Galvani (1929) fa parte della storia della statistica: non si può disconoscerne il valore scientifico, ma la sua impostazione si contrapponeva a quella del campionamento casuale che si sarebbe di lì a poco affermata, per le sue utili proprietà e per l'efficacia e la flessibilità con cui rendeva possibile studiare una popolazione sulla base di un suo sottoinsieme opportunamente selezionato. Data la statura scientifica di Gini, comunque, non meraviglia il fatto che la moderna teoria dei campioni abbia origine da una sconfitta sul campo dello statistico italiano: l'articolo di Jerzy Neyman (1934), cui si fa storicamente riferimento se si

⁴ Seng (1951), p. 218. Favero e Trivellato (2000) ricordano come con la sua posizione di rifiuto «di fronte al progetto presentato da Anders Kiaer nel 1895 per un utilizzo del "metodo rappresentativo nei censimenti"», Bodio contribuì ampiamente a «far cadere, dopo cinque riunioni successive, l'interesse per l'argomento, che verrà ripreso all'interno dell'Istituto soltanto nel 1925, ritardando notevolmente l'applicazione estensiva dei metodi propri della statistica campionaria» (Favero e Trivellato, 2000, pp. 247-248).

⁵ Più in dettaglio, la selezione avviene per gruppi o distretti territoriali e si prefigge di fare in modo che il campione fornisca risultati (medie, varianze, ecc.) uguali a quelli della popolazione, rispetto ad alcune variabili di controllo di cui si conosce un elevato grado di correlazione con quelle oggetto dell'indagine (Gini e Galvani, 1929, pp. 6 ss.; Neyman, 1934, p. 571).

cerca il seme da cui è nata la moderna teoria dei campioni, si snoda come critica approfondita, analitica ed empirica, del lavoro in cui Gini e Galvani (1929) proponevano la via del campionamento giudizioso; utilizzando i loro stessi dati, Neyman dimostra la superiorità del campionamento casuale rispetto a quello giudizioso proposto, peraltro con moderato scetticismo, dai due autori italiani.

L'atmosfera in cui si trovò a operare un pioniere delle rilevazioni campionarie in Italia, come il fondatore della Doxa, Pier Paolo Luzzatto Fegiz, non era quindi favorevole e non lo sarà per molto tempo ancora. Del resto, l'opposizione giniana ai metodi campionari, e all'uso cui si prestavano, conteneva anche aspetti di tipo politico che Gini argomenta in una importante polemica che nel 1942 lancia contro Luzzatto Fegiz dalle pagine dell'*Archivio di Studi corporativi*. In questo caso la polemica di Gini si scagliava contro i sondaggi di opinione resi possibili dal metodo campionario. La discussione di Gini si concentra sui riflessi politici ed ideologici della tecnica dei campioni rappresentativi⁶.

Sul piano ideologico, la critica di Gini a quello che definiva il “referendum Gallup”, dal nome della nota società di sondaggi americana, è legata alla sua visione della società come un tutto organico del quale il potere statale deve conservare l'equilibrio spontaneo stratificatosi nel corso della storia; in questa concezione, gli individui, i loro comportamenti e le loro idee sono considerati come meri elementi incidentali di un entità più elevata, la nazione⁷. In quest'ottica, Gini mette in evidenza il rischio che la pubblica diffusione dei risultati dei sondaggi, inevitabile in un contesto democratico come quello statunitense, possa far «rilevare ai maneggioni dell'opinione pubblica i punti deboli delle tendenze popolari contrarie ai loro interessi, verso i quali essi potrebbero così dirigere la loro azione propagandistica e talvolta corruttrice per deviare a loro favore l'opinione della maggioranza» (Gini, 1942, p. 469). In questo senso egli non nega l'utilità dei sondaggi di opinione, ma scorge forti rischi nei feedback che essi possono generare nel comportamento sociale: la società che conosce se stessa può rischiosamente spostarsi dal suo equilibrio naturale che le autorità hanno la responsabilità di conservare. La concezione politica che ispira l'uso della statistica suggerito da Gini è di chiara marca paternalistico-autoritaria: annulla la responsabilità individuale e la necessità di prendere posizione sulle scelte di natura collettiva; la libera circolazione delle opinioni nuoce a coloro che di tale libertà si avvalgono. Per questo motivo Gini ritiene che i sondaggi di opinione esplichino tutti i loro benefici negli stati autoritari, dove la conoscenza che permettono di ottenere è prerogativa di chi detiene le leve del comando.

La posizione di Gini, costituisce una testimonianza di un clima generale che, sia dal punto di vista scientifico sia da quello ideologico, è radicalmente avverso ai metodi proposti da Luzzatto Fegiz; clima che nel 1942 poteva essere attribuito all'influenza della cultura del regime, alla quale Gini contribuiva con spirito autonomo e indipendente (Baffigi e Puggioni, 2015); la scarsa adesione

⁶ Gini (1942). Sulla polemica Gini-Luzzatto Fegiz si veda Baffigi e Puggioni (2015).

⁷ Si è parlato in questo senso di neo-organicismo di Corrado Gini. Si veda Cassata (2006), Baffigi (2007) e Baffigi e Puggioni (2015).

metodologica e ideologica alle rilevazioni campionarie, tuttavia, si protrasse a lungo anche dopo la guerra. La scuola italiana di statistica, ampiamente influenzata dalla posizione giniana approfondisce il tema della contrapposizione fra rilevazione campionaria e censuaria. in un interessantissimo volumetto, scritto insieme ad Alighiero Naddeo, Marcello Boldrini (1957) mette in guardia dalla possibile «grave rinuncia alla rilevazione totalitaria dei fenomeni statistici empirici, la sola adatta a caratterizzare – a fine memorativo e comparativo – una situazione storica per se stessa conclusa»⁸.

Anche dal punto di vista politico il clima di ostilità proseguì nel dopoguerra, seppur inserito nel nuovo contesto di competizione democratica fra partiti. Come ricorda Rinauro (2002), nel suo bel libro sulla storia del sondaggio d'opinione in Italia, i due partiti del fronte popolare, PCI e PSI, attaccarono con grande vigore lo statistico triestino, che nel gennaio e nel marzo 1948 aveva svolto due sondaggi elettorali che prevedevano la netta vittoria democristiana alle storiche elezioni del 18 aprile. Rinauro (2002, pp. 419 e ss.) ricorda articoli molto duri de *L'Unità* e dell'*Avanti!*. Quest'ultimo, ad esempio, attaccò la Doxa in due articoli nei quali si sottolineavano legami societari della Doxa con alcune aziende americane, con la Fiat, con la Montecatini, con la Pirelli, la Motta, ecc.; si gettava inoltre discredito sulla società di sondaggi sostenendo che “quei compiacenti pronostici elettorali” erano il frutto di uno scambio con il governo democristiano che aveva concesso alla Doxa 16 milioni di lire per una ricerca sulla distribuzione del reddito nazionale, la quale tra l'altro, secondo le insinuazioni dell'*Avanti*, avrebbe consentito alla Doxa di «usurpare le competenze ufficiali dell'Istat»⁹.

Questo era il difficile contesto, scientifico e politico al tempo stesso, in cui trovò a muoversi chi nel dopoguerra cercò di promuovere il metodo campionario per le rilevazioni statistiche. Vi furono eccezioni in questo clima di avversità; una delle quali, che si rileverà particolarmente importante in questa storia, fu quella di Luigi Einaudi che, già nel 1945, aveva scritto un articolo in favore del “metodo

⁸ Boldrini e Naddeo (1957, p. 51). Vale la pena leggere un brano più ampio delle considerazioni di Boldrini: «Il ricorso a campioni, nel campo delle statistiche empiriche, dovrebbe aver luogo solo in caso di necessità, quando cioè una conoscenza urgente non è realizzabile in alcun altro modo: ma guai a trasformarlo in faciloneria e in malvezzo, come non di rado è avvenuto e avviene. Questa appunto sembra essere la colpa di taluni trattatisti, quando essi si abbandonano a indagini, sia pure rispettabili ed utili, sulla tecnica per il prelevamento dei campioni; tecnica che, dopotutto, si riduce a un enunciato semplicissimo, vale a dire a ricercare le vie e i mezzi più economici e appropriati per realizzare indirettamente il criterio bernoulliano della estrazione a sorte con metodo elementare. Tutti i suggerimenti tendenti a sostituire siffatta procedura maestra, perché talora impraticabile, in tanto diventano giustificabili in quanto rappresentano espedienti che se ne allontanano il meno possibile. Ma non è questo ovviamente il tema centrale della cosiddetta teoria dei campioni. Insistendovi troppo, si perde di vista il *punctum saliens* di natura eminentemente concettuale, e si trascurano i non pochi inconvenienti che derivano dalla grave rinuncia alla rilevazione totalitaria dei fenomeni statistici empirici, la sola adatta a caratterizzare – a fine memorativo e comparativo – una situazione storica per se stessa conclusa» (Boldrini e Naddeo, 1957, p. 51).

⁹ Rinauro osserva che «la causa della commissione governativa del dicembre 1947 del sondaggio sulla distribuzione del reddito nazionale risiedeva, oltre che nel rapporto di amicizia di Luzzatto Fegiz con Einaudi e Del Vecchio, nel fatto che a quel tempo l'Istat non era ancora in grado di effettuare una simile rilevazione e il governo non intendeva pagare il prezzo di un censimento completo[...]; inoltre i 16 milioni di finanziamento pubblico non furono neppure sufficienti a coprire le spese della Doxa che, lungi da ricavarne un profitto, dovette accollarsi l'onere di un milione di lire, come si rileva dai verbali di bilancio dell'Istituto» (Rinauro, 2002, p. 419).

Gallup”, pubblicato dalla rivista *Risorgimento liberale* (Rinauro, 2002, p. 345). Come vedremo, Luigi Einaudi, con la sua concretezza empiristica posta al servizio delle istituzioni, fu uno dei personaggi chiave nel favorire l’affermazione del metodo campionario in Italia o, più specificamente, nel comprendere il valore sociale di un’attività come quella svolta dalla Doxa, istituto che nel 1951 svolgerà la prima indagine campionaria sulle famiglie italiane per conto della Banca d’Italia, i cui principali risultati furono riportati nella pubblicazione “Effetti di ipotetiche variazioni del reddito sulle uscite di un “campione” di famiglie italiane”.

La possibilità di ricostruire in dettaglio le vicende che portarono all’indagine del 1951 è legata all’esistenza di documenti conservati in vari fondi dell’Archivio storico della Banca d’Italia; i documenti utilizzati in questo lavoro sono descritti in una cronologia ragionata riportata in appendice; inoltre, il lettore più interessato può prenderne visione consultandone le copie digitali che si mettono a disposizione presso l’Archivio storico.

I primi passi delle indagini sulle famiglie in Italia

Einaudi non fu uno statistico ma il suo liberalismo e il suo riformismo furono strettamente legati al suo essere empirista, alieno dalle derive irrazionaliste e idealiste che hanno sempre influenzato la cultura italiana. Norberto Bobbio (1986), nel suo “Profilo ideologico del Novecento italiano” colloca Einaudi, insieme al solo Salvemini, in un capitolo intitolato “La lezione dei fatti”, dedicato ai «due empiristi di questa storia». «Quel poco (non pochissimo) di pensiero liberale e democratico, di liberalismo civile e di democratismo non demagogico, che sopravvisse nell’età del prefascismo [...], fu opera non certo dei neofiti dell’idealismo e tanto meno dell’irrazionalismo, ma dei "superstiti" o, più correttamente dei superati, del positivismo, che non avevano forse letto né Spencer né Ardigò, ma non avevano neppure sviscerato la filosofia di Hegel» (Bobbio, 1986, p. 86). Gli empiristi Luigi Einaudi (1874-1961) e Gaetano Salvemini (1873-1957), «accolti come maestri (si pensi a Gobetti), ebbero il coraggio e la prudenza di considerare la filosofia come una zona pericolosa e si tennero per tutta la vita ben fermi sul solido terreno dei problemi concreti» (ivi, p. 86).

Tra i problemi concreti che Einaudi si trovò ad affrontare nella sua attività istituzionale vi fu certamente la necessità di realizzare la sua massima “conoscere per deliberare”: l’incontro con Luzzatto Fegiz trovò quindi un terreno molto fertile. La conoscenza e la stima reciproca fra il giovane statistico e il più maturo economista risaliva all’inizio degli anni Trenta, quando Luzzatto Fegiz aveva pubblicato alcune ricerche statistiche su “La riforma sociale”, diretta dall’economista piemontese. Il rapporto tra i due doveva essersi consolidato se, nel 1938, al rientro dal suo viaggio negli Stati Uniti, Luzzatto Fegiz gli riferiva in una lettera, con speranza e ambizione, l’attività lì svolta, i suoi progetti di ricerca e le strade esplorate per finanziarli. Durante il soggiorno, Luzzatto Fegiz aveva approfondito le sue conoscenze sulla tecnica del sondaggio e di questo rendeva

conto a Luigi Einaudi, che veniva anche informato della richiesta di finanziamenti che aveva rivolto alla Rockefeller Foundation¹⁰.

Dovette essere naturale per Luzzatto Fegiz, nei mesi successivi alla fondazione della Doxa, avvenuta nel 1946, rivolgersi al suo illustre amico, divenuto Governatore della Banca d'Italia, per promuovere l'attività del neonato, ma quanto mai attivo, istituto. Il 18 marzo del 1947, lo statistico scrive all'economista «ricordando il suo incoraggiamento al tempo in cui il nostro Istituto era appena in gestazione»¹¹. Luzzatto Fegiz elenca le «importanti analisi di mercato» già svolte «per conto di grandi imprese italiane» e fa riferimento a trattative in corso con enti stranieri, per svolgere analoghi lavori; si augura infine che «l'Ufficio studi della Banca d'Italia vorrà seguire l'esempio degli analoghi uffici delle nostre maggiori imprese, e abbonarsi al bollettino, che nei prossimi numeri conterrà anche i risultati di sondaggi di carattere economico»¹². Poche settimane più tardi, Einaudi diviene Ministro del Bilancio nel secondo governo De Gasperi. Luzzatto Fegiz non attende molto: già il 19 luglio scrive al neo Ministro con consapevole informalità¹³, proponendo un progetto di sondaggio sul reddito nazionale¹⁴.

Il 20 luglio Luzzatto invia al Ministro del Tesoro, Gustavo Del Vecchio, un progetto di massima che definisce “approvato dal prof. Einaudi” (ASBI, Direttorio Einaudi, cart. 17, fasc. 11, sfasc. 9, p. 163).

La corrispondenza tra Einaudi e Luzzatto continua anche dopo il 1948, durante il settennato presidenziale: sempre cordiale e rispettoso, lo statistico sollecita commenti e valutazioni del suo lavoro al suo interlocutore. Einaudi non si sottrae, commenta e propone. Il 21 maggio 1949, il suo segretario Antonio D'Aroma spedisce a Luzzatto il volume “I redditi delle famiglie italiane nel 1948” «con alcune annotazioni del Presidente Einaudi»¹⁵.

Nel frattempo, Luzzatto intrattiene stretti rapporti anche con la Banca d'Italia. Incontra il Governatore Menichella, del quale, in una lettera del 23 luglio 1949, Luzzatto riferisce a Einaudi le «parole lusinghiere per l'indagine Doxa sul

¹⁰ «Per suggerimento del prof. Mills, il valoroso statistico della Columbia e membro del National Council of Economic Research, ho preso contatto con la Rockefeller Foundation a cui ho esposto i miei progetti per uno studio sui bilanci familiari, da iniziarsi intanto a Trieste. Io ho il pieno appoggio dell'Istituto Centrale di Statistica, ma mi mancano i mezzi. In complesso ho avuto un'ottima accoglienza, ma credo che la decisione finale spetti al sig. Kittredge, direttore per l'Europa» (Lettera di Luzzatto Fegiz a Einaudi, Trieste, 9 giugno 1938, in ALE, Fondazione Luigi Einaudi, Torino, archivio Luigi Einaudi, sez. 2, “Corrispondenza”, b. 2, f. “Luzzatto Fegiz Pierpaolo”).

¹¹ La lettera di Luzzatto Fegiz a Einaudi sta in ASBI, Banca d'Italia, Direttorio-Einaudi, cart. 17, fasc. 11, sfasc. 9. La citazione nel testo è a p. 104.

¹² Ivi.

¹³ «So bene che non è questa la forma in cui si fanno affari con lo Stato, ma penso che se Lei approva in linea di massima il progetto, esso potrebbe essere poi eseguito per conto di uno o più enti parastatali anziché per conto dello Stato» (ivi, p. 155). Lo statistico conclude la lettera ringraziando Einaudi per avergli dato «la possibilità di passare una serata con Lei e con i Suoi».

¹⁴ Sull'attività di pubbliche relazioni di Luzzatto Fegiz nell'immediato dopoguerra e, in particolare, sui rapporti fra Luzzatto Fegiz e Einaudi, si veda Rinauro (2002), pp. 343 e ss.

¹⁵ ASBI, Banca d'Italia, Direttorio Einaudi, cart. 17, fasc. 11, sfasc. 9, p. 102.

reddito nazionale» (ivi, p. 115), nonché il suggerimento «di studiare il problema delle scorte di biglietti presso i privati, con particolare riguardo alle variazioni stagionali nella consistenza di tali scorte nelle varie regioni e classi sociali» (p. 115)¹⁶. L'instancabile attività di *public relation* svolta dal direttore della Doxa l'aveva inevitabilmente portato anche a stabilire rapporti diretti con il Servizio studi della Banca¹⁷. È probabile che in occasione dell'incontro con Menichella, di cui aveva scritto a Einaudi, Luzzatto avesse incontrato il capo del Servizio studi, Paolo Baffi. Sta di fatto che il 7 gennaio 1950 Luzzatto scrive a Menichella riferendosi a una conversazione «con il Capo dell'Ufficio Studi della Banca d'Italia», avvenuta qualche mese prima, nel corso della quale Luzzatto si era «impegnato a inviare, a completamento della proposta già fatta, i risultati dell'elaborazione di un "sondaggio-pilota" sulle abitudini e preferenze dei consumatori»¹⁸. Il capo della Doxa, elenca alcune proposte di ricerca. Egli sembra insistere in particolare su un progetto di studio sulla distribuzione del reddito nazionale. Scrive Luzzatto:

Mi permetto di ricordarLe che la mia richiesta di un contributo per gli studi sul reddito nazionale (che, riallacciandosi a quelli eseguiti nel 1948 per il Ministero delle Finanze, dovrebbero fornire dati sulla distribuzione del reddito nazionale nel 1950) era stata da Lei presa in benevola considerazione. Ora, poiché in relazione a tale progetto, abbiamo già eseguito dei lavori preliminari, affrontando spese considerevoli, e poiché i vari enti che ci hanno promesso contributi attendono di vedere il piano finanziario completo prima di prendere una decisione finale sopra l'entità e la modalità della loro partecipazione, Le sarei estremamente grato se Ella volesse farci in proposito una comunicazione formale, che ci consentirà di predisporre il piano delle rilevazioni per il 1950¹⁹.

L'intesa già c'era, quindi, si trattava di formalizzarla. Esattamente un mese dopo, il 7 febbraio, Menichella risponde a Luzzatto Fegiz dicendo di essere «lieto di comunicarLe che proporrò al Consiglio superiore della Banca, di concorrere al finanziamento dell'indagine sulla distribuzione dei redditi in Italia nel 1950 con un complessivo di lire due milioni» (ivi, p. 78).

L'episodio è importante in quanto, in prospettiva, si tratterà della prima indagine sui bilanci delle famiglie finanziata dalla Banca. In effetti Menichella, nella stessa lettera del 7 febbraio, porrà alcune condizioni sull'esecuzione della ricerca:

gradirei che nel condurre questa indagine la Doxa cercasse, ove possibile, di venire incontro alle richieste formulate verbalmente dal Servizio Studi di questa Banca, Al

¹⁶ Luzzatto acclude alla lettera «due copie del questionario di prova del sondaggio-pilota». Il questionario è costituito da due pagine che comprendono domande di confronto fra prima e dopo la guerra sul potere d'acquisto della lira, sul contante tenuto a portata di mano, sulla collocazione delle eccedenze temporanee di contante, sul taglio di banconote preferito, ecc. La lettera sta in ASBI, Banca d'Italia, Direttorio-Einaudi, cart. 17, fasc. 11, sfasc. 9, pp. 115-121).

¹⁷ Del resto la Banca seguiva da vicino le ricerche in corso promosse dalla Doxa e finanziate dal Ministero. Non è un caso che fra le carte di Antonino Occhiuto, statistico che in quegli anni aveva un ruolo importante nel Servizio studi, sia conservata una copia della "Relazione preliminare sulle indagini in corso presso l'istituto Doxa". Il documento, riservato, era stato inviato da Luzzatto Fegiz al Ministero delle Finanze in data 16 giugno 1948 (ASBI, Banca d'Italia, Studi, Pratiche 379, fasc. 14).

¹⁸ ASBI, Banca d'Italia, Studi, Pratt. 284, 1, p.79.

¹⁹ Ivi, p. 80.

riguardo Ella potrà prendere contatti diretti con detto Servizio per le ulteriori precisazioni²⁰.

Come vedremo, il desiderio di Menichella verrà esaudito anche se non senza ritardi e difficoltà che di fatto spostarono l'anno di riferimento dell'indagine dal 1950 al 1951. Ma c'è di più: le circostanze vorranno che quella che si era presentata come una indagine Doxa sostenuta da una molteplicità di finanziatori, divenne di fatto la prima indagine campionaria condotta sulle famiglie per conto della Banca d'Italia.

L'indagine del 1951

La storia dell'indagine sui bilanci delle famiglie curata dalla Banca d'Italia, comincia quindi nel 1950, anche se in maniera non ufficiale: non esiste una pubblicazione da cui risulti il coinvolgimento della Banca. La metà del contributo della Banca alla Doxa venne versato tempestivamente, già nel marzo del 1950, un mese dopo la lettera con cui Menichella aveva accettato la proposta. Lo si capisce dalla comunicazione che Paolo Baffi, il 29 dicembre 1950, invia a Luzzatto; suona come un rimprovero:

Chiar.mo Professore,

dopo l'erogazione, avvenuta nello scorso marzo, della prima metà del contributo stanziato a favore del suo istituto per le note indagini, nessuna notizia è qui pervenuta sulla esecuzione delle medesime.

Poiché la richiesta dell'anticipo era stata giustificata con l'imminenza dei lavori, e considerato anche che era previsto che Ella mantenesse contatti con questo Servizio, Le sarei molto riconoscente se Ella volesse fornire qualche indicazione sul seguito avuto dalla iniziativa.

Le porgo i più distinti saluti.²¹

Cosa era accaduto? Lo spiega Luzzatto Fegiz in due lettere: una indirizzata a Menichella, di poco posteriore alla richiesta di chiarimenti di Baffi, e una inviata a fine marzo allo stesso Baffi.

Nella lettera a Menichella, il 10 gennaio 1951, Luzzatto sottolinea che l'obiettivo originario di ripetere «su scala più ampia e con metodi più raffinati, lo studio del 1948», non si era potuto realizzare «non essendosi raggiunto nel finanziamento l'ammontare previsto»; ciò aveva comportato la necessità di ridurre l'indagine «rispetto a quella progettata» e a rinunciare «a quelle elaborazioni dettagliate (per es. distribuzione dei redditi nelle singole regioni e classi di reddito) che presupponevano un "campione" di ampiezza considerevole»²².

Nella imbarazzata lettera a Paolo Baffi del 24 marzo, i chiarimenti forniti sui problemi di finanziamento sono più analitici e forniscono un quadro dei principali finanziatori della ricerca e di quelli che non avevano onorato le promesse iniziali. Luzzatto conclude la lettera sottolineando che il suo...

²⁰ Ivi, p. 78.

²¹ ASBI, Banca d'Italia, Studi, Pratt., 284, 1, p. 78.

²² Ivi, p. 76.

...principale desiderio è quello che l'Istituto di emissione, il Governatore e Lei stesso conservino della Doxa la buona opinione che ne avevano in passato. Quantunque non sia nostra colpa se siamo stati lasciati in asso da quello che doveva essere il principale finanziatore [l'Associazione industriale lombarda], siamo pronti, anzi desiderosi, di spendere integralmente il milione che ancora dovrebbe esserci versato per quelle nuove indagini che nel momento attuale dovessero apparire importanti e urgenti al Suo ufficio e al Governatore²³

La ricerca della Doxa dovette quindi ridurre le ambizioni iniziali, ma in compenso divenne più focalizzata, in virtù del fatto che gli eventi appena raccontati, concentrando la committenza alla sola Banca d'Italia, definirono con più precisione gli obiettivi perseguiti. Le istruzioni di Paolo Baffi giungono a Luzzatto con una lettera del 12 aprile, nella quale il Capo del Servizio Studi prende atto delle difficoltà e dichiara di concordare sul fatto che «il proseguimento dell'indagine riuscirà più efficace su base meno ampia e con scopi ben limitati».²⁴

Le indicazioni formulate da Baffi sono precise: suggerisce di limitare l'inchiesta aggiuntiva al comportamento finanziario dei consumatori prendendo in considerazione tre categorie di redditeri capi-famiglia, due di consumatori-produttori (agricoltori-conduttori e piccoli imprenditori) ed una di consumatori puri (operai e impiegati). Baffi fissa poi lo scopo principale dell'inchiesta:

l'accertamento delle opinioni prevalenti circa la distribuzione di un eventuale incremento di reddito (preferibilmente nelle due ipotesi di aumento, poniamo, del 10 e del 50 per cento) tra consumi e risparmi, distinguendo i primi in consumi correnti e di beni durevoli ed i secondi in investimenti diretti (in immobili o nell'azienda), tesaurizzazione ed altre forme di risparmio. Di un qualche interesse sarebbe altresì l'accertamento di eventuali mutamenti intervenuti nelle abitudini dei consumatori non solo rispetto al 1938, ma anche rispetto alla precedente inchiesta del marzo 1950.

Potrebbe anche essere utile di considerare come verrebbe distribuita tra minori consumi e minori risparmi una eventuale riduzione di reddito»²⁵.

L'inchiesta si svolgerà, sostanzialmente, lungo le linee tracciate da Baffi²⁶. Il 19 gennaio 1952 Luzzatto scrive al Governatore Menichella²⁷ facendo esplicito riferimento alle linee di ricerca che gli erano stati indicati nella citata lettera del 24 marzo 1951. Al fine di fornire i dati richiesti dalla Banca d'Italia, la Doxa aveva

²³ Ivi, p. 73. La lettera a Baffi, tre pagine dattiloscritte, mostra le difficoltà in cui la Doxa si era venuta a trovare nella gestione di una ricerca finanziata da più di una istituzione. Il Governatore e Baffi avevano manifestato esplicitamente la loro delusione per il fatto di «aver avuto la relazione insieme a una pubblicazione dell'Associazione delle Casse di Risparmio». Luzzatto assicura che in ogni caso non considera esauriti gli obblighi che lo legano alla Banca d'Italia in seguito alla concessione del contributo.

²⁴ Ivi, p. 69.

²⁵ Lettera di Paolo Baffi a Luzzatto Fegiz, Roma, 12 aprile 1951, in ASBI, Banca d'Italia, Studi, 284, 1, pp. 69-70.

²⁶ Come vedremo più avanti la relazione all'inchiesta, svolta fra il 15 ottobre e il 3 novembre 1951, recherà il titolo "Effetti di ipotetiche variazioni del reddito sulle uscite di un "campione" di famiglie italiane".

²⁷ Lettera di Luzzatto Fegiz a Donato Menichella del 19 gennaio 1952, in ASBI, Banca d'Italia, Studi, Pratt., 284, 1, p. 30.

condotto una nuova inchiesta. Alla lettera Luzzatto allega tavole statistiche riassuntive:

Non ho creduto di aggiungere alcun commento, né una nota introduttiva, perché mi riservo di redigere la relazione definitiva dopo aver avuto da Lei e dal prof. Baffi indicazioni sul numero delle copie che si desiderano e sull'uso cui esse saranno destinate²⁸.

Menichella comunica immediatamente a Baffi la ricezione della lettera: è interessante leggere la sintesi del loro colloquio che ne fa lo stesso Baffi, inizialmente annotata a margine della missiva, poi in un appunto un po' più esteso, scritto a mano, indirizzato a Salvatore Guidotti, dirigente del Servizio studi e statistico impegnato su più fronti, in quel periodo fondativo per la statistica italiana²⁹. La nota di Baffi si articola in tre punti. In primo luogo, bisognava sopprimere il primo quesito, una domanda in senso lato politica, che richiedeva agli intervistati un confronto fra la situazione economica del paese al momento della rilevazione con quella di sei mesi prima: circa la metà degli intervistati ravvisava un peggioramento. Il secondo punto era connesso alla strategia di diffusione del rapporto: occorre «pregare Luzzatto Fegiz di distribuire lui le copie agli indirizzi che a noi interessano.... senza indicare che l'indagine è stata svolta su nostro incarico». La terza questione riguardava l'interpretazione dei risultati, sulla quale Menichella e Baffi si riservavano di intervenire: «prima della diffusione, chiedere a Luzzatto Fegiz la relazione introduttiva riservando eventuali osservazioni»³⁰. È quindi evidente che l'indagine era ormai sotto la salda guida di Baffi e del suo principale collaboratore in campo statistico: non ci sono dubbi che l'indagine del 1951 possa considerarsi come la prima indagine Banca d'Italia sulle famiglie. Vi è però una sorta di consegna del silenzio: nella comunicazione che Baffi fa a Guidotti vengono indicati i principali enti e i personaggi cui la pubblicazione dovrà essere inviata; le spese di stampa e di spedizione furono a carico della Banca; ma il committente è tenuto nascosto: probabilmente i tempi non erano maturi, le polemiche politiche sulla Doxa non erano ancora sufficientemente distanti; la carenza di "coscienza statistica" in Italia era un dato troppo consolidato per non creare intralci e difficoltà in chi volesse invece innovare l'apparato statistico del Paese. Lo stesso Istituto centrale di statistica in

²⁸ Ivi.

²⁹ Per un breve profilo biografico di Guidotti, si veda Fazio (1997).

³⁰ «Caro Guidotti, avevo cercato di raggiungerti sabato per chiederti consiglio circa il seguito da dare alla lettera di LF; e ciò specialmente in previsione di una visita che il governatore mi avrebbe fatto, come ha fatto, ieri. L'intesa provvisoria, subordinata al tuo parere conforme e modificabile per accordo diretto tra lui e te, è stata di:

pregare LF di distribuire lui le copie agli indirizzi che a noi interessano (Ass Banc, Ass CCR, Ass BP, ECE, BRI, studiosi, parlamentari) senza indicare che l'indagine è stata svolta su nostro incarico;

far sopprimere il quesito 1;

autorizzare la pubblicazione a 1-2 mesi dalla diffusione anzidetta;

ma questa doveva essere la lettera a) – prima della diffusione, chiedere a LF la relazione introduttiva riservando eventuali osservazioni.

La mia impressione dell'indagine è stata buona. Ma non si può sapere qualcosa dell'ampiezza del campione, della sua distribuzione per classi sociali, ampiezza dei centri abitati ecc.?» (Nota manoscritta di Baffi a Guidotti, in ASBI, Banca d'Italia, pratt., n. 284, f. 1, pp. 28-29)

quegli anni introduceva la rilevazione campionaria fra le tecniche di rilevazione utilizzate, che però continuava a essere vittima di una strutturale disattenzione da parte dei politici e della burocrazia; la carenza di finanziamenti e la generale trascuratezza, tra l'altro, fecero sì che «il ricorso generalizzato dell'Istat alla tecnica del campione rappresentativo per le indagini socioeconomiche decollerà solo tra gli anni 80 e gli anni 90»³¹.

“Effetti di ipotetiche variazioni del reddito sulle uscite di un “campione” di famiglie italiane”

Il 30 maggio 1952 Luzzatto Fegiz fa sapere a Guidotti che «[f]ra pochi giorni saranno pronte le 150 copie della relazione sul risparmio e sulle spese»³². Il rapporto era presentato da una breve avvertenza, scritta da Luzzatto Fegiz, che inquadrava l'indagine all'interno di filoni di indagine sui cicli economici riconducibili agli «schemi di Kahn, Keynes, Hansen e di numerosi altri autori» (Doxa, 1952, p. I). I concetti elaborati da queste teorie, sottolinea il direttore della Doxa, hanno un preciso contenuto empirico che si presta, «entro certi limiti» ad essere rilevato statisticamente. Dal punto vista metodologico, enfatizza l'autore, «i dati analitici non sono raccolti per l'intera popolazione considerata, ma solo per una parte rappresentativa o campione» (ivi).

Conformemente alle indicazioni fornite da Paolo Baffi nella citata lettera del 12 aprile 1951, l'oggetto del “sondaggio” era costituito dalle «intenzioni di un “campione” di capifamiglia, e cioè la propensione verso certe forme di consumo e di investimento nelle tre ipotesi di un aumento durevole del reddito, di una diminuzione pure durevole di tale reddito, e nel caso di un'entrata straordinaria “una tantum”.» (ivi). Maggiori dettagli sui risultati erano sintetizzati nell'introduzione, oltre che in specifiche tabelle. L'universo di riferimento era costituito dai «capifamiglia italiani, d'ambo i sessi, e di tutte le categorie economico-sociali» (Doxa, 1952, p. 1), mentre il metodo seguito per l'indagine era quello del campione stratificato per regioni, per capoluoghi di provincia e altri comuni, per professione e “classe sociale”, costituito da 1089 persone.

Il nucleo essenziale del questionario era costituito da cinque delle undici domande complessive e riguardava per grandi linee quattro argomenti: le forme preferite di risparmio e gli effetti di un incremento di reddito, moderato o rilevante, sulle spese e sui risparmi e gli analoghi effetti di una diminuzione dei redditi.

Tra le forme preferite di risparmio emergevano chiaramente i libretti di risparmio, in particolare al Nord, e i buoni postali fruttiferi, particolarmente intensa al Sud; ma non era trascurabile l'importanza attribuita al contante tenuto in casa (Doxa, 1952, p. 2).

Le informazioni richieste con questa domanda non erano tra quelle indicate esplicitamente da Baffi nella lettera del 24 marzo, ma rientravano perfettamente tra gli argomenti di interesse per la Banca, sia per la conoscenza dell'economia sia

³¹ Rinauro (2002, p. 530). Si veda anche Zuliani (1996).

³² ASBI, Banca d'Italia, Studi, pratiche, 284, 1, p. 53.

per la conduzione della politica monetaria; del resto, esse dovevano interessare in modo particolare Paolo Baffi che da anni aveva ben chiara l'esigenza di approfondire lo studio del legame fra i fenomeni reali e quelli finanziari. Si trattava di un filone di ricerca che di fatto avrebbe condotto alla elaborazione teorica e alla costruzione statistica dei conti finanziari, la cui realizzazione costituirà un ulteriore contributo legato all'impegno statistico della Banca d'Italia³³.

L'argomento interessava inevitabilmente anche il Governatore Donato Menichella, come testimonia un episodio rivelatore: il 17 aprile 1954 l'agenzia di stampa "L'informazione", riferiva dati sulle forme di risparmio preferite in Italia. Menichella annota a margine del foglio giornalistico una richiesta rivolta al Servizio Studi: «dove sono tratti questi dati? Vorrei un calcolo sulla base dei dati del bilancio monetario come segue: su 100 lire di risparmio, se ne investono x in libretti e buoni; y presso le banche» e continua a esemplificare: il Servizio Studi predispone una tabella che riporta i dati sulla formazione delle disponibilità monetarie dal 1948 al 1953. L'appunto che accompagna la tabella informa il Governatore che «le cifre pubblicate sono state tratte da una indagine Doxa del 1951 (Effetti di ipotetiche variazioni del reddito sulle uscite di un "campione" di famiglie italiane – pag. 2 – Le forme di risparmio preferite)»³⁴. Non vi sono riferimenti alle circostanze descritte nel paragrafo precedente, le quali avevano portato alla realizzazione di quella indagine sotto la guida e il contributo finanziario della Banca, anche grazie all'intervento dello stesso Menichella.

Dunque, Menichella non ricordava dell'indagine campionaria svolta dalla Doxa due anni prima: potremmo dire che il suo obiettivo di non rilevare che «l'indagine è stata svolta su nostro incarico» era stato perfettamente raggiunto. Ma la evidente dimenticanza ha un significato più profondo e non può essere liquidata con semplice ironia: l'incarico conferito alla Doxa evidenziava una importante esigenza avvertita dai vertici della Banca negli anni della Ricostruzione, quella di conoscere i comportamenti delle famiglie in un contesto economico sempre più complesso, con redditi crescenti, caratterizzato da una economia di mercato che andava integrandosi negli scambi internazionali. Paolo Baffi, in particolare, era particolarmente consapevole di tale esigenza conoscitiva e, come abbiamo visto, considerava lo studio del legame fra aspetti finanziari e reali come un suo filone di ricerca, al quale aveva dedicato già il suo impegno di studioso e di intellettuale funzionario al servizio della Banca. Ma all'esigenza e alla competenza, di cui la Banca già disponeva all'inizio del decennio Cinquanta, mancava il supporto di un'organizzazione tecnica e istituzionale più strutturata, in grado di conferire risorse sistematicamente cospicue, e al tempo stesso qualificate, alla attività conoscitiva. Non è un caso che, come abbiamo visto, l'incarico alla Doxa per l'indagine del 1951 non si configurava come un'attività programmata,

³³ Si veda De Bonis e Gigliobianco (2012). L'importanza di Paolo Baffi nell'impegno statistico della Banca d'Italia era già emersa in un campo come quello della supervisione bancaria, nel quale, all'indomani della promulgazione della legge bancaria del 1936, venticinquenne aveva dato l'impronta alle nuove statistiche sul credito per rami di attività (Baffigi, 2010).

³⁴ La tabella richiesta e l'appunto che l'accompagna si trovano nello stesso fascicolo che conserva la pagina de "L'informazione" vergata da Menichella (ASBI, Carte Baffi, Servizio Studi, cart. 10, fasc. 4, pp. 2-6).

ma era stato il frutto di una proposta dell'instancabile Luzzatto Fegiz e, in parte, il portato di circostanze fortuite che avevano messo la Banca nelle condizioni di sfruttare una occasione utile ma sostanzialmente impreveduta, colta sicuramente con competenza ed impegno, ma che era andata a collocarsi in un contesto organizzativo non pronto a darle basi solide per il futuro. In questo quadro non stupisce la dimenticanza di Menichella.

Ed è proprio per questo motivo che l'episodio della prima indagine sulle famiglie appare storicamente interessante: esso ci mostra alcuni fotogrammi di un percorso complesso che condurrà negli anni Sessanta, Governatore Guido Carli, all'affermazione della Banca come tecnostuttura; un'evoluzione istituzionale e organizzativa che va inquadrata «nell'orizzonte politico dell'epoca, caratterizzato da ben noti problemi di governabilità del paese: grazie alla propria capacità di elaborazione e di previsione, unita a un più diretto rapporto con l'opinione pubblica, la Banca assunse un ruolo di primo piano nel tracciare la rotta della politica economica» (Gigliobianco, 2006, p. 262).

Le prime indagini della Banca d'Italia

Tra il marzo 1961 e il settembre 1962 la Banca d'Italia effettua sei indagini sul reddito, il consumo e il risparmio delle famiglie italiane. I risultati non vengono pubblicati; sono in parte ripresi e commentati nel Bollettino di luglio-agosto 1966, che illustra i risultati della prima indagine della nuova serie, effettuata nel marzo del 1966 e relativa, prevalentemente, ai dati di flusso del 1965 (Tavola 2.1)³⁵. La *Premessa* dell'Ufficio Ricerche del Servizio Studi Economia Italiana descrive le motivazioni della ricerca.

“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. L'insieme delle famiglie, però, ha un comportamento economico meno dinamico di altri operatori, quali ad esempio lo Stato, le imprese, il settore estero, cosicché stimoli espansivi o depressivi hanno difficilmente origine da esso; tuttavia, gli impulsi che partono dagli altri settori influiscono sulle decisioni delle famiglie. Data l'importanza quantitativa di questo operatore, il risultato finale di tali perturbazioni dipende in gran parte dal suo modo di reagire. La condotta economica delle famiglie, in generale, e la loro reazione agli stimoli provenienti dagli altri settori, in particolare, derivano in misura notevole da caratteristiche strutturali concernenti il reddito percepito e la sua distribuzione, la propensione al consumo e quella al risparmio finanziario, la natura e la concentrazione della ricchezza, ecc.. Laddove esistono serie di tali aggregati per il settore delle famiglie, è possibile con metodi econometrici rilevare gli aspetti principali del comportamento delle famiglie; tuttavia, anche in questa felice circostanza una conoscenza della struttura e del comportamento a livello microeconomico può servire a dare maggiore validità ai risultati ottenuti a livello aggregato, sui quali si suole basare in definitiva

³⁵ Alcuni documenti relativi a quelle indagini dei primi anni sessanta sono ora disponibili sul sito internet della Banca d'Italia, nella sezione dedicata ai documenti storici dell'indagine sui bilanci delle famiglie italiane.

molte decisioni di politica economica. 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.” (Banca d'Italia, 1966).

La pubblicazione dei risultati delle prime indagini non era stata accolta con entusiasmo dall'Istat e dal Ministero del Bilancio e della Programmazione Economica. I risultati dell'indagine divergevano da quelli della Contabilità Nazionale, per vari motivi (variabilità delle stime dovuta alla limitata ampiezza del campione di famiglie intervistate, modalità di stima del reddito familiare e differenti definizioni di reddito, reticenza delle famiglie a dichiarare l'ammontare dei redditi percepiti, ecc.) e vi era il timore di ingenerare confusione nell'opinione pubblica, in quanto non tutti i lettori potevano essere in grado di comprendere appieno le cause delle discrepanze e utilizzare pertanto i risultati in modo appropriato e inequivocabile³⁶.

La serie delle indagini comunque proseguì. Nel Bollettino del 1970 (Banca d'Italia, 1970) Adalberto Ulizzi, illustrando i risultati sul 1968, venne peraltro incontro all'esigenza di chiarire meglio ai lettori le motivazioni dell'indagine, gli errori cui le indagini campionarie sono soggette, le differenze di definizione rispetto alla CN e lo scarto tra il reddito monetario rilevato dall'indagine e quello desumibile dalla CN.

“Tali ricerche (ovvero la serie di indagini, ndr) costituiscono un'utile integrazione delle informazioni a livello aggregato desumibili dalle statistiche ufficiali. Da queste ultime è possibile rilevare, ad esempio, il reddito complessivo prodotto dalla collettività in un certo anno, ma non come questo si distribuisce tra gruppi di famiglie o di individui diversi per residenza, livello d'istruzione, genere di attività svolta, età e così via. Inoltre solo ricorrendo a rilevazioni campionarie è possibile analizzare le relazioni intercorrenti tra grandezze finanziarie nei diversi gruppi di popolazione (ad es. il risparmio in funzione del reddito e del patrimonio netto).”

Sulle stime e sulla loro affidabilità e precisione si osservava che:

“Le stime campionarie sono affette da errori di vario tipo, che possono ripartirsi in due categorie: quelli di natura probabilistica o accidentale, la cui ampiezza è determinabile a priori, fissati certi livelli di confidenza, e quelli di natura non probabilistica o sistematica. Quest'ultima categoria comprende errori di diversa origine. Una parte delle famiglie estratte a sorte non viene intervistata per assenza, rifiuto o altro, determinando nel campione una distorsione, che si riflette sulle stime (...). Un'altra fonte di errori è rappresentata dalle risposte degli intervistati; le inesattezze fornite in sede d'intervista non sempre riflettono il desiderio di nascondere a estranei dati di carattere più o meno riservato, ma talvolta sono dovute al fatto che l'intervistato non ricorda esattamente il dato richiesto (...) o non intende correttamente la domanda postagli. Altri errori, infine, possono verificarsi in sede di trascrizione dei dati sui questionari e in sede di codifica. La somma algebrica di tutti questi errori, insieme a quelli di natura probabilistica, rappresenta lo scarto della stima campionaria dal corrispondente valore riferito all'universo. Tra i vari errori menzionati, 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

³⁶ Il carteggio è conservato presso l'Archivio storico della Banca d'Italia: Asbi, Banca d'Italia, Studi, pratt., n.1039, fasc.2.

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. Nell'esposizione dei risultati si è evitato, di conseguenza, di dare stime dimensionali delle grandezze finanziarie sopra menzionate, limitandosi a mettere in evidenza la distribuzione di tali variabili tra gruppi diversi di popolazione.”

Quanto al reddito, l'indagine non si contrapponeva alle informazioni esistenti, ma le integrava con gli aspetti distributivi:

“La disponibilità per il reddito di valutazioni macroeconomiche ufficiali permette di effettuare dei confronti con le analoghe stime campionarie, misurandone il grado di attendibilità, fatte opportune correzioni per rendere omogenei i termini di confronto. (...) Lo scarto tra le stime campionarie del reddito monetario e l'altra valutazione (ovvero la contabilità nazionale, ndr) non appare molto forte e tale comunque da togliere significato anche alle successive analisi sulla distribuzione del reddito tra gruppi diversi di popolazione. Giova comunque sottolineare che scopo di questa indagine non è quello di fornire stime del reddito familiare in aggiunta o in opposizione a quelle già esistenti, ma di integrare le conoscenze già acquisite a livello aggregato con altre informazioni sulla distribuzione del reddito tra gruppi di famiglie e di individui, ripartiti in base a criteri geografici, demografici, economico-sociali.”

Tre aspetti dunque emergevano chiaramente: a) il fatto che l'indagine, per la contenuta dimensione del campione, potesse condurre a stime caratterizzate da un'ampia variabilità; b) l'esigenza di un più stretto raccordo tra le definizioni utilizzate nell'indagine e quella della CN; c) la presenza, soprattutto per le attività finanziarie, di errori non campionari tutt'altro che trascurabili. Su quest'ultimo punto la Banca si basava sulle ricerche di Ulizzi (1967), del tutto innovative per l'epoca.

Il rischio di una lettura impropria dei risultati da parte di un lettore non tecnico, paventato dall'Istat e dal Ministero del Bilancio e della Programmazione Economica, era però presente e si sarebbe concretizzato qualche anno più tardi. I risultati dell'indagine sul 1982 (Banca d'Italia, 1983), che in termini di variazione percentuale del reddito familiare rispetto all'anno precedente erano molto discosti dalla CN, furono probabilmente la principale causa dell'avvio di una riflessione metodologica sull'indagine. Nei giorni 8 e 9 febbraio 1985 il Servizio Studi organizzava un convegno presso la S.a.di.ba (Perugia), avente come oggetto le indagini campionarie sui bilanci delle famiglie italiane. Nella sua introduzione al convegno Carlo Santini, allora Capo del Servizio Studi, scriveva:

“La Banca d'Italia conduce da quasi 20 anni un'indagine sui bilanci familiari con l'obiettivo di raccogliere dati sulla distribuzione del reddito e sulle caratteristiche strutturali del risparmio; trovandosi nella peculiare posizione di produttore e al tempo stesso utilizzatore di statistiche, la Banca d'Italia ha sempre avvertito la necessità di integrare i dati disponibili a livello aggregato con informazioni microeconomiche.

Dopo un'esperienza che si è protratta così a lungo senza modifiche sostanziali nell'impianto della rilevazione, ci si è trovati in una fase di ripensamento sia sull'oggetto sia sul metodo di questa indagine; non solo a causa dell'insoddisfazione per alcuni risultati, e in particolare quelli sul risparmio, ma anche a causa della ripresa dell'interesse per l'argomento da parte di studiosi di altri Istituti. (...) Abbiamo organizzato un convegno, i cui atti sono pubblicati in questo volume, con la speranza che potesse costituire un momento di discussione e confronto delle esperienze compiute, un'occasione per esplorare le prospettive di queste indagini, le

possibilità di cooperazione per un più ampio utilizzo dei loro risultati e per l'affinamento dei loro fondamenti metodologici. (...) La potenziale ricchezza di tali fonti statistiche risulta evidente dallo spettro di utilizzazioni di queste fonti statistiche di cui si è discusso (...) Insieme a tale potenziale ricchezza, gli interventi menzionati hanno anche messo in luce la debolezza dei dati di fonte campionaria attualmente disponibili in Italia, soprattutto per quanto riguarda la numerosità dei campioni rilevati, il livello di disaggregazione delle informazioni raccolte e l'affidabilità dei dati ottenuti, che frequentemente appaiono viziati dalla reticenza delle famiglie intervistate a rivelare le proprie entrate e la propria situazione patrimoniale. (...) ha ricevuto particolare attenzione da parte dei partecipanti al convegno la necessità di armonizzare metodi, definizioni e presentazione dei risultati ottenuti, allo scopo di consentire il confronto tra i dati di indagini diverse così come tra dati campionari e dati aggregati, e quindi la reciproca validazione e la possibilità di integrazione delle informazioni provenienti da fonti statistiche indipendenti” (Banca d'Italia, 1986).

La stagione di cambiamento, che si era già avviata con le innovazioni introdotte nella prima metà degli anni ottanta alle metodologie di estrazione del campione (grazie alla collaborazione dell'Istat, a partire dal 1984 il campione veniva estratto dalle liste anagrafiche invece che da quelle elettorali, metodo, quest'ultimo, che – in assenza di metodi di correzione – determinava una sovrastima delle famiglie più numerose) traeva particolare momento dalla costituzione di un gruppo di lavoro che si avvaleva della consulenza del prof. Luigi Biggeri.

Il gruppo concludeva i lavori nel marzo del 1986, proponendo numerose innovazioni: rilevazioni con periodicità e caratteristiche differenti (un'indagine annuale con la quale rilevare il reddito e il patrimonio immobiliare delle famiglie e alla quale affiancare sezioni monografiche di particolare interesse nell'anno; un'indagine benchmark periodica da realizzare ogni 3-5 anni, sul risparmio e la ricchezza finanziaria delle famiglie); aumento della numerosità campionaria (10.000-12.000 famiglie per l'indagine annuale, 25.000-30.000 per quella periodica); revisione del disegno campionario, con un'estrazione casuale delle unità di primo stadio da quelle selezionate per l'indagine sulle forze di lavoro e l'estrazione delle famiglie dalle liste anagrafiche; modifiche delle definizioni delle variabili, per facilitare il confronto con la CN.

Tenuto conto della fattibilità sotto il profilo della gestione e dei costi, la soluzione adottata fu un raddoppio della dimensione del campione (da 4.000 a 8.000 famiglie) e, a partire dall'indagine sul 1987, una riduzione della frequenza di rilevazione (da annuale a biennale). Le restanti proposte (revisione delle definizioni di reddito, delle modalità di selezione delle unità di primo stadio e anche ulteriori innovazioni) venivano via via attuate.

Tavola 2.1

Pubblicazioni ufficiali della Banca d'Italia sull'IBF

Pubblicazione, anno, anno di edizione, numero e pagine	Curatori	Anno di riferimento
BS, 1966, 21, n. 4, 441-54	A. Ulizzi	1965/66
BS, 1967, 22, n. 4, 441-72	non indicato	1966
BS, 1968, 23, n. 5-6, 643-75	A. Ulizzi	1967
BS, 1970, 25, n. 1, 103-67	A. Ulizzi	1968
BS, 1971, 26, n. 1, 125-78	A. Ulizzi	1969
BS, 1973, 28, n. 3-4, 501-60	F. M. Frasca e D. Qualeatti	1970, 1971
BS, 1974, 29, n. 3-4, 449-500	F. M. Frasca e D. Qualeatti	1972
BS, 1976, 31, n. 4, 709-59	F. M. Frasca e R. A. Pirrotta	1973, 1974
BS, 1977, 32, n. 1, 169-219	R. A. Pirrotta	1975
BS, 1977, 32, n. 4, 827-913	R. A. Pirrotta	1976
BS, 1978, 33, n. 2-3, 287-363	R. A. Pirrotta	1977
BS, 1979, 34, n. 3, 189-261	R. A. Pirrotta	1978
BS, 1980, 35, n. 3-4, 311-74	R. A. Pirrotta e G. Zen	1979
BS, 1981, 36, n. 1-4, 539-607	G. Zen	1980
SBS, 1983, 36, n. 14	G. Zen	1981
SBS, 1983, 36, n. 57	G. Zen	1982
BS, 1984, 39, n. 3-4, 273-315	D. Gressani	1983
BS, 1985, 40, n. 3-4, 433-64	L. Cannari e D. Gressani	1984
BS, 1987, 42, n. 1-2, 327-60	L. Cannari	1986
SBS, 1989, 42, n. 5	C. A. Bollino, L. Cannari e G. D'Alessio	1987
SBS, 1991, 1 (n.s.), n. 26	G. D'Alessio e A. I. Rinaldi	1989
SBS, 1993, 3 (n.s.), n. 44	G. D'Alessio	1991
SBS, 1995, 5 (n.s.), n. 9	G. D'Alessio	1993
SBS, 1997, 7 (n.s.), n. 14	G. D'Alessio	1995
SBS, 2000, 10 (n.s.), n. 22	G. D'Alessio e I. Faiella	1998
SBS, 2002, 12 (n.s.), n. 6	G. D'Alessio e I. Faiella	2000
SBS, 2004, 14 (n.s.), n. 12	C. Biancotti, G. D'Alessio, I. Faiella e A. Neri	2002
SBS, 2006, 16 (n.s.), n. 7	I. Faiella, R. Gambacorta, S. Iezzi e A. Neri	2004
SBS, 2008, 18 (n.s.), n. 7	I. Faiella, R. Gambacorta, S. Iezzi e A. Neri	2006
SBS, 2010, 20 (n.s.), n. 8	L. Bartiloro, G. Cappelletti, F. D'Amuri, R. Gambacorta, S. Iezzi, S. Magri, A. Neri e C. Rondinelli	2008
SBS, 2012, 22 (n.s.), n. 6	C. Biancotti, F. D'Amuri, R. Gambacorta, G. Ilardi, A. Neri e C. Rondinelli	2010
SBS, 2014, 24 (n.s.), n. 5	F. Carta, R. Gambacorta, G. Ilardi, A. Neri e C. Rondinelli	2012
SBS, 2015, 25 (n.s.), n. 64	R. Gambacorta, S. Iezzi, G. Ilardi, A. Neri e A. Rosolia	2014

(*) BS: *Bollettino Statistico*; SBS: *Supplemento al Bollettino Statistico (Supplementi al Bollettino Statistico dopo il 1991)*.

Le indagini più recenti

Con l'indagine sul 1987 viene sperimentato un sovracampionamento di famiglie ad alto reddito³⁷, selezionando un campione integrativo di imprenditori, dirigenti e liberi professionisti da elenchi professionali. Ex post i soggetti sovracampionati mostrano tuttavia caratteristiche differenti dalle corrispondenti categorie rilevate nel campione ordinario; l'unione dei due campioni si presenta

³⁷ Esperimenti di sovracampionamento erano stati condotti nel 1971 e 1972 con riferimento alle famiglie meridionali e nel 1973, 1974 e 1975 con riferimento alle famiglie ad alto reddito (si veda Brandolini, 1999; Banca d'Italia, 1976).

come problematica e richiede aggiustamenti ad hoc che nella successiva rilevazione scoraggiano la ripetizione del sovracampionamento.

Nell'intento di limitare la reticenza degli intervistati a dichiarare le attività finanziarie nell'indagine sul 1987 si sperimenta la rilevazione di queste quantità sotto forma di percentuali rispetto al reddito complessivo della famiglia anziché in valore assoluto. I risultati non sono particolarmente soddisfacenti; le stime risentono infatti dei problemi di misura sia nelle attività finanziarie sia nel reddito.

Nell'indagine sul 1987 viene inoltre introdotta la sezione sui non rispondenti, che consente di misurare i tassi di risposta e valutare i motivi della mancata partecipazione all'indagine.

L'indagine sul 1987 è anche quella che registra una sostanziale modifica nella struttura della pubblicazione che ne riporta i principali risultati. La pubblicazione si amplia, passando da circa 30 a 80 pagine, inglobando una più ampia sezione metodologica, un insieme di tavole standardizzate e riportando in appendice il questionario utilizzato nella rilevazione.

A partire dall'indagine sul 1987 la rilevazione si susseguirà a un ritmo biennale, salvo l'eccezione dell'ulteriore slittamento di un anno nel 1998. Questa revisione è il risultato di una riflessione complessiva sugli obiettivi dell'indagine, che si andranno a concentrare da allora in avanti principalmente su temi di natura strutturale.

Nell'indagine sul 1989 viene introdotta una componente panel, al fine di migliorare la stima delle variazioni intercorrenti tra un'indagine e la successiva. La componente panel era inizialmente di dimensioni limitate: al termine della rilevazione, solo 1.200 delle circa 8.000 famiglie intervistate per l'indagine sul 1989 avevano partecipato anche all'indagine precedente. Numerosi erano stati i problemi sorti dal tentativo di reintervistare, senza alcun preavviso, le stesse famiglie a due anni di distanza. La disponibilità di quei dati costituiva però un fecondo terreno per la ricerca economica e offriva l'opportunità di svolgere riflessioni sulla qualità dei dati raccolti, che potevano essere controllati a due anni di distanza (anche mediante la rilevazione degli stock di fine periodo e delle variazioni intervenute tra un'indagine e la successiva).

Per le attività finanziarie, si sperimentano la rilevazione delle quote percentuali in valore delle diverse attività finanziarie, seguite dall'ammontare in valore dei soli depositi, in modo da poter ricostruire successivamente gli ammontare di tutti gli altri strumenti. Nella stessa indagine si sperimenta la somministrazione di una busta chiusa che l'intervistato doveva compilare con l'effettivo valore delle attività finanziarie possedute e spedire alla società di rilevazione in forma anonima (ma con l'indicazione di un codice di aggancio al questionario). Purtroppo inviarono la busta solo 3.500 delle 8.000 famiglie intervistate e la qualità delle risposte contenute nelle buste si rivelò bassa.

Nell'indagine sul 1989 venne inoltre introdotta la rilevazione delle caratteristiche degli intervistatori, in modo da poterne valutare eventuali effetti sulle stime.

All'inizio degli anni novanta viene avviato il progetto di costituzione dell'archivio storico dell'indagine. I microdati delle indagini a partire dal 1977 in

avanti, dispersi in banche dati ormai obsolete, vengono recuperati e documentati³⁸. Inizia un'opera di ricodifica delle variabili rilevate con una certa continuità fino agli anni più recenti, in modo da consentirne il confronto temporale. I criteri di imputazione dei dati mancanti e di calcolo degli aggregati più rilevanti, come il reddito e la ricchezza, vengono resi, per quanto possibile, omogenei nel tempo.

L'operazione richiese la collaborazione di personale assegnato ad hoc al progetto; l'archivio fu completato e reso pubblico solo a metà degli anni novanta, dopo una lunga fase di verifica tra gli utenti interni delle varie procedure utilizzate³⁹.

L'archivio storico rispondeva allo stimolo dei ricercatori che richiedevano in misura via via crescente dati micro per le loro analisi. Sulla scia di questa tendenza e dell'innovazione tecnologica, matura inoltre la decisione di ridurre al minimo la procedura amministrativa di rilascio dei dati; nel giro di pochi anni si passa da un rilascio fortemente selettivo dei nastri magnetici contenenti la base dati, a una distribuzione sostanzialmente libera di dischetti, poi CD-ROM e, più tardi, file su Internet⁴⁰.

A partire dall'indagine sul 1991 la Banca d'Italia ha introdotto meccanismi di controllo dell'operato della società incaricata di svolgere l'indagine e degli intervistatori. In particolare nelle indagini sul 1991, sul 1993 e sul 1998 la Banca si è avvalsa dell'operato di una società di certificazione, il cui controllo mirava a verificare sia che gli intervistatori, nella fase di contatto delle famiglie e durante lo svolgimento delle interviste, seguissero tutte le indicazioni loro fornite, sia che la società, nelle fasi di codifica e di *cleaning*, adottasse strumenti adeguati per l'intercettazione e la rimozione di errori e incoerenze nei dati. Nelle altre occasioni di indagine le verifiche sono state effettuate direttamente dalla Banca. I risultati di tali attività hanno fornito utili indicazioni nel ridisegnare le procedure e gli incentivi per la società e gli intervistatori.

Dall'indagine sul 1991 la rilevazione si stabilizza, sia in termini di questionario sia di modalità di rilevazione. Le attività finanziarie vengono rilevate richiedendo la specifica della classe di importo. Attingendo ai suggerimenti presenti nella letteratura che si occupa della rilevazione di fenomeni sensibili, a partire da quella rilevazione, per limitare la reticenza, il cartellino che illustra le classi di importo prevede numerose classi, in particolare per gli importi più elevati. Le classi con gli importi maggiori sono inoltre mostrate prima delle altre. Sebbene gli effetti di tali accorgimenti appaiano modesti, questa modalità rimane in vigore negli anni successivi.

³⁸ I microdati delle indagini precedenti al 1977 sono purtroppo andati perduti; i Supplementi che ne riportano i principali risultati sono stati di recente messi a disposizione in forma elettronica nelle pagine dedicate all'indagine sui bilanci delle famiglie del sito Internet della Banca, insieme a quelli più recenti.

³⁹ Il lavoro di raccordo storico fu curato da Luigi Cannari, Giovanni D'Alessio e Massimo Gallo, inviato dal Nucleo di Ricerca di Venezia presso il Servizio Studi.

⁴⁰ I microdati relativi alle indagini successive al 1977 sono da tempo disponibili gratuitamente sul sito Internet della Banca, per lo svolgimento di ricerche di carattere scientifico.

Dal 1995 la rilevazione delle attività finanziarie si concentra in una grossa tavola che rileva conoscenza, possesso, classe di ammontare e ammontare preciso. Seguendo una strategia di rilevazione denominata *unfolding brackets* (Juster e Smith, 1997), nel caso di indisponibilità a fornire quest'ultimo dato, si richiede una sottoclasse dell'importo.

Nel corso degli anni novanta si stabilizza sia la struttura del questionario di base sia il ricorso all'impiego di sezioni monografiche per la rilevazione di particolari argomenti nonché la modalità di rilevazione. A partire dalla rilevazione sul 1995 si adotta inoltre lo schema delle "rotazioni", ovvero parti di questionario che vengono sottoposte a sottoinsiemi casuali del campione. L'accorgimento si dimostrerà di notevole utilità nel dare risposta alle crescenti richieste degli utenti che premevano per ampliare il questionario in varie direzioni, limitando al tempo stesso l'onere di risposta dei rispondenti. Il prezzo da pagare, oltre a una certa complicazione organizzativa, è quello (spesso accettabile) di disporre di queste informazioni su campioni meno numerosi.

Al fine di disporre di informazioni per le analisi intergenerazionali a partire dall'indagine sul 1995 vengono inoltre sottoposte all'intervista per la prima volta famiglie formate dalle originarie unità panel incluse nel campione (tipicamente, quelle formate dai figli che costituivano una nuova famiglia).

Nell'indagine sul 1995 si svolge inoltre un test su un campione limitato di 200 famiglie per la rilevazione delle informazioni tramite l'impiego del computer (*Computer-Assisted Personal Interviewing - CAPI*). La sperimentazione viene giudicata positivamente e la modalità di rilevazione viene estesa alla maggior parte delle famiglie rilevate a partire dall'indagine sul 1998.

Nelle indagini successive la metodologia è rimasta in larga misura invariata; le principali modifiche hanno riguardato il questionario, sia negli aspetti di natura monografica sia nella rilevazione di informazioni utili per migliorare la qualità delle stime sulla ricchezza e la comparabilità con aggregati macroeconomici⁴¹ e in particolare con quelli della contabilità finanziaria⁴².

Negli ultimi anni l'indagine del nostro istituto si è inoltre inserita nel più ampio progetto dell'HFCS in ambito eurosistema, che mira a collegare le indagini sulle famiglie condotte dai paesi dell'area dell'euro. Tale iniziativa, se da un lato amplia le opportunità di analisi, fornendo elementi di comparazione internazionale dei fenomeni, dall'altro tende a ridurre gli spazi di autonomia nella conduzione della rilevazione. Il questionario, infatti, prevedendo un'ampia sezione obbligatoria comune, lascia meno spazio alle iniziative nazionali. Per questo motivo si sperimentano due indagini intermedie sul 2013 e sul 2015, su campioni ridotti di 2.000 famiglie panel, con lo scopo di approfondire particolari aspetti che, per motivi di spazio, non è possibile rilevare nelle usuali indagini IBF.

⁴¹ Sulle iniziative di ricostruzione della ricchezza e di confronto con altre fonti si veda Banca d'Italia (2008); per una panoramica sulla ricchezza delle famiglie italiane, si veda Cannari e D'Alessio (2006).

⁴² Sui CF si veda in particolare Banca d'Italia (2006) e Banca d'Italia (2003).

L'utilizzo dell'indagine per la politica economica e la ricerca

Le indagini campionarie sono essenziali per svolgere analisi di policy. Come osservava in un intervento del 2009 l'allora Governatore Mario Draghi (Draghi, 2009), le famiglie, le imprese, gli operatori economici in generale, hanno comportamenti diversi, che riflettono numerose determinanti. La risposta a eventi esterni o a misure di politica economica non è uguale per tutti; per comprendere l'impatto aggregato occorrono informazioni sull'intera distribuzione dei fenomeni. Questo è ancora più necessario quando si studiano interventi rivolti a specifiche fasce della popolazione.

L'attenzione è sempre stata rivolta ad aspetti che stanno al centro dell'azione di una Banca centrale, come ad esempio l'evoluzione del sistema dei pagamenti e il risparmio delle famiglie.

Man mano che la società è divenuta più complessa, si è venuta rafforzando la necessità di rispondere a quesiti specifici. Per questo, dagli anni novanta si è consolidata una struttura modulare del questionario che affianca parti monografiche alle tradizionali sezioni sul reddito e sul patrimonio. Queste sono state utilizzate sia per migliorare la comprensione su aspetti come l'avversione al rischio, le aspettative di reddito, le eredità sia per assicurare un set informativo adeguato su questioni più strettamente legate alla policy, come le aspettative di pensione, i comportamenti finanziari e la relativa conoscenza in materia, o i comportamenti e giudizi sui temi del civismo, dell'evasione fiscale e della corruzione.

Lo sviluppo delle tecnologie informatiche ha inoltre reso possibile usare grandi banche dati microeconomiche per simulare, con un certo grado di complessità, i comportamenti degli agenti economici sotto diverse ipotesi di evoluzione delle grandezze economiche e demografiche. Le informazioni dell'indagine sono quindi divenute la base dati dei modelli di micro-simulazione utili alla valutazione di scenari alternativi di policy così come gli effetti di modifiche ai trasferimenti e all'imposizione fiscale. Non è dunque un caso che l'indagine costituisca una fonte spesso richiamata nelle audizioni parlamentari o nei documenti in cui la Banca (ma anche altre istituzioni) svolge i suoi compiti di consulenza nei confronti del Parlamento e del Governo in materia di politica economica e finanziaria.

All'interno del nostro Istituto i dati dell'indagine hanno più volte costituito la base di articolati progetti di ricerca. Solo per accennare ai principali, alla fine degli anni ottanta ricordiamo il progetto *Saving and the Accumulation of Wealth*, coordinato da Albert Ando, Luigi Guiso e Ignazio Visco (Ando, Guiso e Visco, 1994), che mirava allo studio delle determinanti del risparmio, dei riflessi dell'invecchiamento sui sistemi pensionistici, della distribuzione della ricchezza, dei vincoli di liquidità. Negli anni novanta l'indagine sulle famiglie è stata protagonista del progetto di ricerca coordinato da Amartya Sen sulla misurazione del benessere umano dalla prospettiva alternativa dell'approccio delle *capabilities* e sulle sue implicazioni per il disegno delle politiche economiche⁴³. Negli anni

⁴³ Per quel progetto di ricerca non fu pubblicato un rapporto unitario ma diversi contributi di ricerca, tra i quali - ad esempio - quello di Brandolini e D'Alessio (1998).

duemila gli studi sui dati dell'indagine hanno costituito la parte centrale del progetto di ricerca *Household Wealth in Italy* (Banca d'Italia, 2008b).

Le stime dell'indagine sono spesso utilizzate nelle pubblicazioni periodiche della Banca (la Relazione Annuale, il Bollettino Economico e il Rapporto sulla Stabilità Finanziaria) unitamente ai dati macroeconomici, per valutare l'evoluzione dell'indebitamento delle famiglie e della loro vulnerabilità.

I dati raccolti nell'IBF sono usati, sia all'interno sia all'esterno della Banca d'Italia, per lo svolgimento di ricerche di varia natura. Secondo le segnalazioni effettuate dagli utenti, ai quali vengono chiesti i riferimenti bibliografici delle loro pubblicazioni, i lavori basati sull'indagine a fine 2015 ammontano a circa 800 unità. La quota di utenti della Banca d'Italia è di circa il 15 per cento, circa il 30 per cento in termini di lavori di ricerca.

L'importanza relativa dei vari campi di ricerca (Figura 2.1) è cambiata nel corso del tempo. Nei primi anni gli studi erano concentrati sui temi del reddito, del risparmio e della politica fiscale. Successivamente la ricerca si è rivolta anche a nuovi argomenti, come il lavoro, la famiglia, l'incertezza, la povertà, le pensioni. Nel nuovo secolo emerge una maggiore attenzione per i temi della ricchezza e della disuguaglianza, dell'analisi territoriale e dell'istruzione. Negli anni più recenti, con il protrarsi della crisi economico e finanziaria, la ricerca si è concentrata sui temi dei rischi delle famiglie, della vulnerabilità e della sostenibilità dei debiti. Lungo l'intero periodo esaminato, rilevante è stato l'impegno di studio per gli aspetti metodologici, con circa 120 lavori (circa 2-3 lavori all'anno).

L'estrema varietà degli utilizzi dell'indagine può essere illustrata esaminando alcuni lavori scientifici presentati nel corso della recente Conferenza sul cinquantenario dell'indagine. Neri, Rondinelli e Scoccianti (2015) illustrano gli effetti sui consumi del cosiddetto bonus Renzi; D'Alessio e Iezzi (2015b) esaminano la diffusione del fenomeno del sovraindebitamento delle famiglie italiane dopo la crisi mentre Bartiloro, Michelangeli e Rampazzi (2015) comparano la vulnerabilità delle famiglie indebitate nei vari paesi europei; Cappariello, Marino e Zizza (2015) esaminano le dimensioni dell'economia sommersa.

Per migliorare l'accesso della collettività ai risultati dell'indagine, è stato di recente messo a disposizione un set di tavole e grafici che ne sintetizzano i principali risultati. Sul sito sono inoltre presenti anche numerosi documenti sull'indagine di carattere storico, come ad esempio le tavole statistiche delle indagini che precedono il 1977, i questionari utilizzati, i rapporti dell'epoca. Si tratta di materiale che viene messo a disposizione dei ricercatori e dei cittadini e che va a integrare il vasto patrimonio informativo disponibile sulle famiglie italiane.

Figura 2.1

Cloud degli argomenti trattati nei lavori basati sull'IBF, 1966-2014



3. Le principali caratteristiche dell'indagine⁴⁴

La composizione del campione

L'indagine si è tradizionalmente posta l'obiettivo di rappresentare la popolazione residente in Italia. Da un punto di vista operativo, la selezione delle unità da intervistare è stata effettuata, per gli anni fino al 1983, tramite le liste elettorali. L'utilizzo di tali liste generava tuttavia una sovrarappresentazione delle famiglie con più soggetti maggiorenni; oggi negli archivi che utilizzano quei dati tale distorsione è in parte corretta tenendo sotto controllo la distribuzione per età dei componenti⁴⁵.

A partire dal 1984, l'estrazione delle famiglie è stata effettuata sulle liste anagrafiche. Anche le liste anagrafiche tuttavia sono soggette ad approssimazioni ed errori, con conseguenze sulla copertura dei campioni e sulle proprietà degli stimatori. Alcune verifiche empiriche, condotte principalmente sulla base della comparazione dei dati sulla popolazione desumibili dai censimenti con quelli della

⁴⁴ Una descrizione dei metodi impiegati nella rilevazione è fornita nei Supplementi che riportano i principali risultati di ciascun anno. Per l'indagine sul 2014 si veda, ad esempio, Banca d'Italia (2015).

⁴⁵ Questa soluzione è parsa preferibile rispetto ai metodi basati sull'inverso della probabilità di estrazione, che davano luogo a una riduzione della dimensione media della famiglia poco plausibile.

anagrafi, hanno messo in evidenza alcuni criticità di quest'ultima fonte (Commissione per la Garanzia dell'Informazione Statistica, 2002):

- a) le liste anagrafiche non rappresentano adeguatamente gli stranieri in Italia. Ciò accade sia perché l'anagrafe non tiene conto dell'immigrazione clandestina sia perché sovente questo gruppo di cittadini tende a non rispettare l'obbligo di comunicare cambi di residenza in Italia o il ritorno in patria;
- b) alcuni comuni, in particolare quelli aventi natura turistica, tendono ad avere residenti fittizi (di comodo), cioè coloro che fissano la residenza nel luogo della loro casa di vacanza per ottenere vantaggi fiscali o di altra natura;
- c) in altri casi, le anagrafi possono risultare scarsamente accurate sui legami familiari, come, ad esempio, nel caso di soggetti che dichiarano – fittiziamente – di non costituire una famiglia, pur convivendo.

Queste caratteristiche producono effetti diversi e difficilmente quantificabili sui risultati dell'indagine. Per quanto riguarda il punto a), è plausibile che la sottorappresentazione degli stranieri di recente immigrazione conduca a una sottostima della parte più povera della popolazione. I problemi richiamati nel punto b) tendono invece a rendere più probabile l'inclusione nel campione teorico di famiglie più agiate (con più abitazioni), anche se tale effetto dovrebbe essere compensato, almeno in parte, dalla maggiore difficoltà di ottenere l'intervista da parte di queste stesse famiglie (perché di fatto assenti nell'abitazione secondaria). Le dichiarazioni non veritiere relative ai legami familiari conducono invece a un'alterazione delle probabilità di selezione delle famiglie (quelle che operano suddivisioni fittizie hanno maggiore probabilità di essere estratte e possono essere sovrarappresentate). Nel complesso il primo effetto (quello relativo agli stranieri) è forse quello più rilevante per gli anni più recenti, ma anche gli altri aspetti richiamati potrebbero essere non trascurabili.

Va inoltre tenuto presente che l'accuratezza delle liste anagrafiche si riverbera sulle stime non solo tramite la composizione del campione, ma anche tramite i valori della popolazione cui le stime vengono ancorate (la cosiddetta distorsione da conteggio). Se, ad esempio, i comuni turistici hanno maggiore popolazione residente (a causa dei residenti fittizi), i valori medi rilevati in quei comuni saranno ponderati per una popolazione maggiore di quella effettiva. Tale distorsione è nulla solo quando il numero delle famiglie in anagrafe coincide con quello della famiglie effettivamente residenti (cioè quando i residenti fittizi eguagliano gli esclusi dalle liste).

Oltre agli aspetti collegati all'accuratezza delle liste anagrafiche, va tenuto presente che esse escludono i cittadini italiani che vivono nelle istituzioni (carceri, caserme, case di riposo, conventi). Di ciò si tiene conto all'atto della riponderazione, sebbene solo a livello nazionale. In alcuni casi peraltro i soggetti esclusi dalle liste possono comunque essere oggetto di indagine (per esempio, la famiglia intervistata può fornire informazioni su un componente residente in carcere, se la reclusione non è di lungo termine o assume forme ibride come ad esempio gli arresti domiciliari).

Il campo di osservazione dell'indagine differisce quindi da quello utilizzato nella CN. Il settore delle Famiglie e istituzioni senza scopo di lucro al servizio delle famiglie comprende individui o gruppi di individui nella loro funzione di consumatori e nella eventuale funzione di produttori di beni e servizi non finanziari e servizi finanziari destinabili alla vendita purché, in quest'ultimo caso, il loro comportamento economico e finanziario non sia tale da configurare una quasi-società.

Le risorse principali di queste unità provengono da redditi da lavoro dipendente, da redditi da capitale, da trasferimenti effettuati da altri settori, da entrate derivanti dalla vendita della produzione o da entrate imputate per i prodotti destinati all'autoconsumo. Fanno parte delle Famiglie (in particolare del sottosectore Famiglie produttrici) le società semplici, le società di fatto, le imprese individuali, la cui funzione principale sia produrre beni e servizi non finanziari destinabili alla vendita, con un numero di dipendenti fino a cinque unità. Si noti che la fissazione della soglia dei cinque dipendenti per discriminare le famiglie produttrici dalle quasi-società non è imposta dal SEC2010; si tratta altresì di un criterio operativo adottato dall'Istat, ovviamente nel rispetto dei principi generali del SEC2010.

Volendo riclassificare i dati dell'indagine utilizzando il discrimine tra famiglie e quasi-società basato sul numero dei dipendenti, come accade nella CN, non si incontrerebbero particolari difficoltà: l'informazione viene infatti rilevata per le famiglie che svolgono in proprio un'attività produttiva e per i gestori di società. Si tratta semplicemente di rielaborare in modo diverso informazioni già rilevate. Permangono però alcune differenze di cui bisogna tenere conto ai fini di eventuali comparazioni e in particolare l'inclusione nel settore famiglie della CN, delle Istituzioni senza scopo di lucro al servizio delle famiglie (ISP) e della popolazione residente in istituti; queste componenti sono escluse dall'IBF.

Il panel

L'indagine è stata effettuata, fino al 1987, sulla base di rilevazioni indipendenti nel tempo. Dall'indagine sul 1989, per favorire l'analisi dell'evoluzione dei fenomeni oggetto di indagine, è stato introdotto uno schema (detto *split sample*) che prevede a fianco di un campione nuovo estratto dalle liste anagrafiche la presenza nel campione di unità già intervistate in occasione di precedenti indagini (famiglie *panel*).

La tavola 3.1 mostra la numerosità campionaria delle indagini dal 1987 al 2014, evidenziando il numero di famiglie intervistate più di una volta nel corso del tempo. Si rileva, ad esempio, che delle 8.156 famiglie del campione intervistate nell'indagine sul 2014, 13 partecipano sin dal 1987, 64 dal 1989, 168 dal 1991 e così via; le famiglie intervistate per la prima volta nella rilevazione sul 2014 sono 3.697.

Tavola 3.1

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	7.328	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	1.565
2014														3.697
Numerosità campionaria	7.328	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

(*) Nel 1987 non sono considerate le famiglie del sovra-campionamento

A partire dall'indagine sul 1993, la quota di famiglie panel sul totale risulta intorno al 45-50 per cento, quota ottimale per la stima dei valori medi, in presenza di correlazione degli aggregati compresa tra 0,5 e 0,7. Nelle ultime rilevazioni la quota di famiglie panel sul totale è andata anche sopra il 50 per cento, riflettendo la maggiore difficoltà relativa che si incontra nell'intervistare famiglie nuove.

La disponibilità di dati panel è importante sotto diversi punti di vista. Siccome i principali fenomeni sono correlati nel corso del tempo (la ricchezza, il reddito e gli altri fenomeni analizzati hanno elevati livelli di inerzia), è possibile sfruttare tale caratteristica per rendere meno variabili le stime dell'indagine.

I dati raccolti in un panel inoltre, contrariamente a quelli raccolti con indagini cross-section ripetute nel tempo su unità indipendenti, consentono di studiare la dinamica dei fenomeni su singoli individui o famiglie⁴⁶. I dati panel sono inoltre largamente impiegati nell'analisi econometrica per tenere sotto controllo eventuali effetti individuali.

A fronte di queste opportunità offerte dai dati panel, è però opportuno rilevare anche alcuni aspetti problematici.

⁴⁶ L'analisi delle evoluzioni nel tempo è ovviamente valutabile anche utilizzando domande retrospettive, che peraltro sono presenti non di rado nell'indagine. Va però rilevato che, in generale, la qualità dei dati raccolti con domande retrospettive tende ad essere inferiore a quella delle informazioni raccolte nel corso del tempo con il panel.

In primo luogo la popolazione di un panel può non essere rappresentativa del complesso della popolazione. Ad esempio, la popolazione formata dalle unità co-presenti nel 2012 e nel 2014 esclude – per definizione – le unità che nel periodo in esame sono entrate (nate o immigrate) o uscite (morte o emigrate) dalla popolazione di riferimento. Quando si analizzano panel di una certa lunghezza questo fenomeno può avere notevole rilievo.

Al di là degli aspetti che riguardano la popolazione, il processo di non risposta che si osserva per le singole indagini tende ad essere amplificato nel panel. La diversa propensione che i segmenti della popolazione manifestano per la partecipazione alle indagini conduce a un progressivo allontanamento del campione effettivo da quello teorico, e questo può generare una distorsione delle stime.

Un ulteriore aspetto problematico che riguarda i dati panel, in particolare quando vengono utilizzati per stimare le variazioni nel tempo, è rappresentato dagli errori di misura. Come è noto, ogni variabile rilevata è soggetta a un errore di misura. Biancotti, D'Alessio e Neri (2008) hanno mostrato che persino variabili apparentemente semplici da rilevare, come il sesso, l'anno e il luogo di nascita presentano alcuni errori di rilevazione. In questi casi, di frequente si tratta di imprecisioni legate ad equivoci, come quando i nomi dei figli non individuano direttamente il sesso, o errori di codifica dei rilevatori, o errori legati a una incerta memoria del rispondente; l'ambiguità di certe domande contribuisce ad acuire il problema. Mentre però nelle usuali stime dei livelli cross-section questi errori tendono sovente ad annullarsi, poiché ad esempio gli errori di codifica che modificano una variabile in un senso sono spesso compensati da altrettanti errori nella direzione opposta (ad esempio i maschi che sono codificati come femmine e viceversa)⁴⁷, nelle stime delle variazioni (cioè di quanti dipendenti sono diventati autonomi o viceversa) gli errori di misura tendono a sovrastimare i flussi lordi, cioè i cambiamenti di stato. Ai reali cambiamenti tendono infatti a sommarsi anche gli errori che si manifestano in una delle due rilevazioni⁴⁸.

Sebbene nel corso degli anni si sia prestata sempre più attenzione alla qualità dei microdati, l'errore di misura è incompressibile oltre una certa soglia. Le matrici di transizione stimate su dati panel tendono quindi a sovrastimare i cambiamenti di stato⁴⁹.

Va inoltre tenuto presente che la composizione della famiglia varia nel tempo. In molte circostanze le variazioni in aggregati rilevanti possono dipendere da questi cambiamenti, di cui è necessario tenere conto nelle analisi.

⁴⁷ Gli errori di misura che nelle cross-section non comportano distorsione hanno comunque spesso effetti negativi sull'analisi dei legami tra le variabili (c.d. attenuation bias).

⁴⁸ Ovviamente vi possono anche essere casi in cui, per errore, si manifesta una persistenza in luogo di una variazione; questi casi sono però in generale meno frequenti di quelli di segno opposto, quando il fenomeno in esame non è troppo frequente.

⁴⁹ Esempi di stima della mobilità, effettuata tenendo conto del problema degli errori di misura, si trova in Neri (2009) e in D'Alessio e Iezzi (2015b).

Il questionario e la fase di rilevazione

Il questionario è predisposto seguendo una struttura modulare. Si compone 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.

Nelle ultime indagini le sezioni stabilmente oggetto di rilevazione sono le seguenti:

- Struttura della famiglia: caratteristiche socio-demografiche dei componenti della famiglia intervistata e delle famiglie di origine del capofamiglia e del coniuge;
- Occupazione e redditi: condizione occupazionale, aspetti riguardanti il lavoro (o la pensione), i redditi e la storia lavorativa dei componenti;
- Strumenti di pagamento e forme di risparmio: rapporti con banche o posta, uso dei vari strumenti (assegni, bonifici, ecc.), possesso di carte di credito ecc., attività finanziarie possedute;
- Abitazioni di residenza e altri beni immobili: caratteristiche dell'abitazione, mutui, caratteristiche degli altri immobili posseduti;
- Indebitamento;
- Consumi e altre spese familiari: rilevazione sommaria delle spese per beni durevoli e non durevoli, assegni alimenti;
- Previdenza complementare e forme assicurative;
- Notizie a cura dell'intervistatore: informazioni sull'andamento dell'intervista (tempo impiegato, clima dell'intervista, comprensione delle domande, presumibile veridicità delle risposte).

Con riferimento ai contenuti dell'indagine, si è assistito negli ultimi anni a un ampliamento del focus nella direzione della ricchezza e dell'indebitamento. La maggiore rilevanza delle componenti di reddito dovute al capitale e il cambiamento strutturale dei sistemi economici tendono, infatti, ad accrescere l'importanza relativa della ricchezza quale indicatore delle condizioni economiche delle famiglie. In una situazione in cui aumentano le forme di precarietà e i vincoli di finanza pubblica richiedono un contenimento delle prestazioni sociali, il reddito tende a perdere di importanza quale indicatore della capacità di far fronte a eventi imprevisti⁵⁰.

Per ridurre l'onere di risposta per gli intervistati, negli ultimi anni alcune sezioni del questionario sono somministrate solo a un sottoinsieme casuale del

⁵⁰ Cfr. Draghi (2007) e Visco (2008). Sulle forme di ricchezza rilevate è oggetto di valutazione la possibilità di ampliare il campo di osservazione a componenti finora relativamente trascurate, come la ricchezza pensionistica.

campione. In generale, a ogni famiglia si sottopone solo una delle due sezioni monografiche, sulla base dell'anno di nascita (dispari o pari) del capofamiglia⁵¹.

Nell'ultimo ventennio le sezioni monografiche hanno riguardato un'ampia varietà di fenomeni: capital gains, eredità, avversione al rischio, lavoro domestico, mobilità intergenerazionale, uso di servizi pubblici, capitale sociale, evasione fiscale, aspettative di reddito e occupazione, aspettative di pensionamento, scelte e informazioni finanziarie, uso delle nuove tecnologie.

Il questionario viene sottoposto a verifica prima della rilevazione estensiva, allo scopo di limitare difficoltà nella comprensione delle domande e nell'elaborazione delle risposte da parte degli intervistati⁵².

Dal 1998 la rilevazione dei dati viene effettuata in prevalenza con l'aiuto del computer (CAPI)⁵³. 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, alla presenza della famiglia, le eventuali incoerenze⁵⁴. Le interviste che vengono condotte utilizzando il questionario cartaceo, che nel corso degli anni sono via via sempre di meno, vengono trasferite su supporto elettronico, utilizzando il programma CAPI come procedura di immissione dei dati, di modo che i due blocchi di dati siano sottoposti allo stesso insieme di controlli.

Nelle indagini più recenti, la durata media dell'intervista è di circa un'ora. Essa è peraltro abbastanza variabile all'interno del campione, risultando positivamente correlata al reddito e al numero dei componenti della famiglia.

Nel corso del tempo la crescente domanda di informazioni sui comportamenti economici delle famiglie e la pluralità di utilizzi dell'indagine hanno condotto a un progressivo appesantimento del questionario, passato dalle 23 pagine dell'indagine sul 1987 alle oltre 60 in quella sul 2014; l'incremento è avvenuto principalmente tra il 1987 e il 1991 e tra il 2006 e il 2010 (Figura 3.1).

La durata media dell'intervista, che non veniva misurata prima del 1993, registra invece un decremento tra il 1993 e il 1998 (passando da 64 a 54 minuti) e

⁵¹ Tale soluzione, oltre a fornire stime basate su una numerosità campionaria inferiore, non consente di disporre congiuntamente delle risposte alle due sezioni.

⁵² Per molti anni, le interviste di prova sono state condotte seguendo la tecnica suggerita da Oksenberg, Cannel e Kalton (1991). Su un numero esiguo di famiglie (talvolta meno di 100) distribuite sull'intero territorio nazionale, venivano condotte interviste di prova da coppie di intervistatori; il primo eseguiva regolarmente l'intervista, mentre il secondo annotava punti critici su un apposito questionario. Questa fase ha spesso fornito indicazioni utili per riformulare in modo più chiaro alcuni quesiti. Data una certa stabilità del questionario questa tecnica è stata abbandonata negli ultimi anni, a favore di test che forniscano anche una piena assicurazione del funzionamento del programma di rilevazione CAPI.

⁵³ Nell'indagine sul 2014 le interviste effettuate tramite la metodologia CAPI sono risultate il 92,9 per cento.

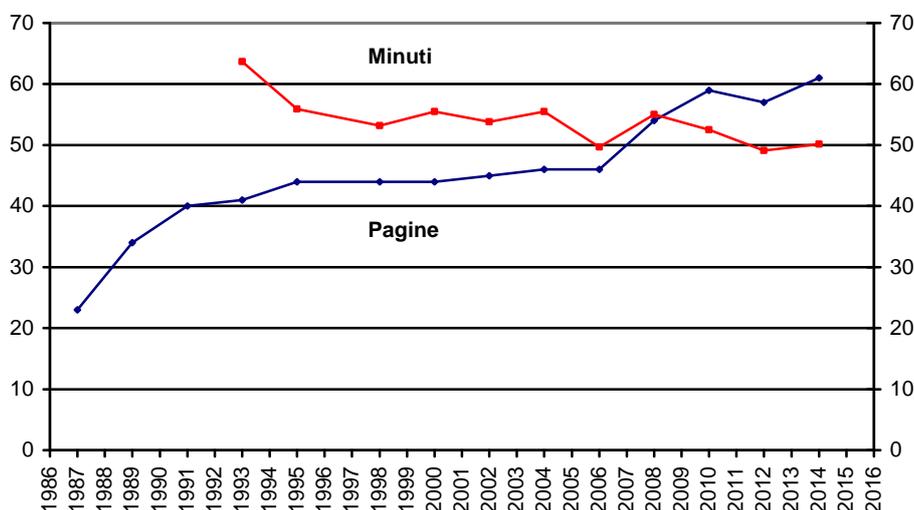
⁵⁴ Numerose sono le cause alla base delle incoerenze che caratterizzano i dati rilevati: da parte dell'intervistato possono esserci problemi di comprensione della domanda, di memoria nel fornire una risposta adeguata o anche reticenza nel fornire informazioni ritenute riservate; da parte dell'intervistatore 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.

poi tra il 2004 e il 2006 (da 55 a 50 minuti), risalendo a 55 minuti negli ultimi anni.

L'ampliamento del questionario deriva dall'introduzione, avvenuta nell'indagine sul 1995, di sezioni di questionario sottoposte a sottoinsiemi del campione totale. A parità di questionario, la durata media delle interviste tende a riflettere anche l'evoluzione socio-demografica intervenuta nel periodo; ad esempio la riduzione del numero medio di componenti per famiglia tende a determinare una diminuzione della durata dell'intervista mentre nella direzione opposta agisce l'incremento del tassi di occupazione. Nelle ultime rilevazioni, il questionario è divenuto più minuzioso su alcuni aspetti rivolti a segmenti specifici della popolazione, ad esempio l'indebitamento, con un incremento sensibile nel numero di pagine, a cui non ha fatto però riscontro un corrispondente aumento della durata dell'intervista.

Figura 3.1

Lunghezza del questionario e durata dell'intervista, 1987-2014



La rilevazione dei dati è affidata a una società specializzata, che si avvale di intervistatori professionali. 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.

Le famiglie contattate per l'intervista, cui viene garantito il completo anonimato, ricevono un opuscolo che illustra le finalità dell'indagine e fornisce alcuni esempi di utilizzo dei dati raccolti. Viene inoltre messo a disposizione un numero verde a cui gli intervistati possono rivolgersi per delucidazioni. Le famiglie non ricevono alcun compenso per l'intervista rilasciata; quelle interessate possono richiedere una copia della pubblicazione contenente i risultati di una precedente indagine. In occasione della pubblicazione dei risultati dell'indagine ad esse viene inviata una lettera di ringraziamento alla quale sono allegati alcuni articoli di giornale che riportano i commenti della stampa.

Campionamento e pesi

In un'indagine come l'IBF, per la natura degli argomenti trattati è preferibile che le interviste siano svolte di persona (*face to face*)⁵⁵. Questo aspetto costituisce un vincolo operativo importante, in quanto impone di concentrare le interviste in un numero limitato di punti di campionamento (comuni), cui assegnare altrettanti intervistatori.

Per tale motivo lo schema di campionamento dell'IBF prevede un procedura di selezione a due stadi. Vengono cioè dapprima selezionati alcuni comuni (unità di primo stadio) e, all'interno di quei comuni vengono poi campionate alcune famiglie (unità di secondo stadio). Sulla base di questa procedura, le circa 8.000 famiglie del campione vengono selezionate all'interno dei circa 300 comuni italiani che costituiscono le unità di primo stadio dell'indagine.

Lo schema di campionamento a due stadi comporta, in generale, una perdita di efficienza, nel confronto con un campione casuale semplice; le stime prodotte con campioni di questo tipo hanno cioè errori standard più elevati di quelli che si avrebbero con campioni casuali semplici di uguale numerosità. Questo fenomeno deriva dal fatto che le unità di una stessa città presentano caratteri che tendono ad essere tra loro positivamente correlati; più elevata è la correlazione tra queste unità, maggiore è la perdita di efficienza che si associa a questo schema campionario.

Questa osservazione vale a maggior ragione se si considera il campione di individui, appartenenti alle famiglie intervistate. Gli individui costituiscono le unità di terzo stadio, tutte selezionate all'interno delle unità di secondo stadio (le famiglie). Le statistiche calcolate sugli individui, relative a caratteri che tendono ad essere correlati all'interno della famiglia, come ad esempio il titolo di studio dei suoi componenti, dovrebbero tenere conto di questo aspetto, calcolando nuovi errori standard, in generale più ampi di quelli che si otterrebbero con campioni di individui indipendenti della stessa numerosità complessiva.

La perdita di efficienza che in generale si associa alla procedura a due stadi può essere talvolta limitata adottando particolari forme di campionamento delle unità di primo stadio (i comuni). Nell'IBF prima di procedere all'estrazione dei comuni, questi vengono suddivisi in base alla regione e alla classe di ampiezza demografica (stratificazione delle unità di primo stadio).

I comuni con oltre 40.000 abitanti sono sempre inclusi nel campione (sono detti comuni auto-rappresentativi); pertanto il campione finale di famiglie includerà comunque alcune unità delle città più grandi del Paese. La numerosità campionaria prefissata è pari a 20 famiglie per i comuni che hanno fino a 350.000 abitanti; per gli 8 comuni con oltre 350.000 abitanti il campione teorico è di 120

⁵⁵ Nel corso del 2016 è stata condotta, in collaborazione con l'Istat, un'indagine sperimentale sulle famiglie con una rilevazione tramite il web, allo scopo di valutarne le problematiche in termini di partecipazione ed errori di risposta. I risultati dell'esperimento non sono al momento disponibili; è presumibile tuttavia attendersi che le distorsioni del campione e le altre criticità possano nel corso degli anni ridursi con il sempre maggiore uso del computer da parte della popolazione nel suo complesso, aprendo nuove opportunità per lo svolgimento di indagini sulle famiglie, meno vincolate al contatto diretto.

famiglie. Va rilevato che, per motivi operativi, la numerosità effettiva può poi discostarsi anche di parecchio dall'obiettivo; di questo si tiene ovviamente conto nella fase di calcolo dei pesi.

I comuni con meno di 40.000 abitanti sono estratti secondo una procedura che assegna a quelli di maggiore dimensione una probabilità più elevata di essere inclusi nel campione (PPS - *Probability Proportional to Size*). Se nei comuni selezionati con il metodo PPS si fissa pari a una costante la numerosità campionaria delle famiglie da intervistare, per esempio ipotizzando 30 interviste per comune, si ottiene – oltre che una semplice regola operativa e un carico ragionevole per un intervistatore – che tutte le famiglie hanno la stessa probabilità di essere incluse nel campione; 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.

Il diverso criterio adottato per i comuni con oltre 40.000 abitanti, nonché altri motivi che verranno trattati più avanti, rendono necessario l'uso di pesi di riproporzionamento; l'adozione del criterio PPS e del numero di interviste costante a livello comunale limita però la variabilità degli stessi, cosa che a sua volta tende a ridurre l'errore standard delle stime.

In accordo con il piano di campionamento, a ciascun individuo della famiglia viene attribuito un peso iniziale, definito come l'inverso della probabilità di inclusione nel campione. Il coefficiente di ponderazione risulta costante a livello di comune e pari a:

$$w_{h\alpha} = \begin{cases} \frac{P_h}{\tilde{P}_h} \frac{P_{h\alpha}}{n_{h\alpha}} & \text{per i comuni con più di 40.000 abitanti} \\ \frac{1}{m_h} \frac{P_h}{n_{h\alpha}} & \text{per i comuni fino a 40.000 abitanti} \end{cases}$$

dove P_h, \tilde{P}_h e m_h sono rispettivamente la popolazione residente, quella dei comuni oggetto di rilevazione e il numero di comuni campione nell'*h-esimo* strato, e $P_{h\alpha}$ e $n_{h\alpha}$ rispettivamente la popolazione e il numero di componenti intervistati nell'*α-esimo* comune dell'*h-esimo* strato.

Si noti che la probabilità di estrazione di una famiglia all'interno di un comune selezionato è approssimativamente pari a $n_{h\alpha}/P_{h\alpha}$. Per i comuni con oltre 40.000 abitanti, che sono sempre inclusi nel campione teorico, è necessario considerare che – per problemi organizzativi – non è sempre possibile condurre interviste in tutti i comuni dello strato. Il primo termine dell'equazione precedente consente pertanto di tenere conto di tale evenienza. I comuni fino a 40.000 abitanti sono selezionati con probabilità proporzionale alla dimensione; la probabilità di selezione dell'*α-esimo* comune nell'*h-esimo* strato è pertanto pari a $m_h P_{h\alpha}/P_h$; la probabilità di inclusione di una famiglia nel campione può quindi essere scritta come: $m_h n_{h\alpha}/P_h$.

È inoltre opportuno rilevare che, essendo n_{ha} il numero di componenti effettivamente intervistato, i pesi tengono conto della mancata risposta. In ogni cella di campionamento, il peso del disegno è pari al rapporto tra la numerosità delle unità nella popolazione e quella nel campione teorico (il numero di unità che ci attende di selezionare *ex-ante*). Poiché la numerosità del campione effettivo (il numero di unità che si è di fatto rilevato) può differire da quello teorico, un tipo di correzione della mancata risposta consiste nell'inflazionare i pesi per il rapporto tra dimensione del campione teorico e dimensione del campione effettivo.

Fino al 1987, il peso campionario era determinato dall'inverso della probabilità di inclusione (stimatore di Horwitz-Thompson). La procedura di ponderazione è stata modificata in seguito all'introduzione del panel, avvenuta nell'edizione del 1989. A partire da questa indagine il coefficiente pari all'inverso della probabilità di estrazione è soltanto il valore iniziale del coefficiente di ponderazione in una procedura che prevede diversi passi.

Le famiglie *panel* possono presentare caratteristiche socio-demografiche in parte diverse rispetto a quelle dell'intero campione, sostanzialmente a causa delle mancate interviste (*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 (classi di reddito e condizione professionale del capofamiglia), che modifica il peso iniziale di tale sottoinsieme di famiglie⁵⁶. A seguito della stratificazione a posteriori, per le principali variabili oggetto di rilevazione, si verifica che il valor medio delle famiglie panel per il periodo precedente coincide con quello rilevato per l'intero campione.

In presenza di correlazione tra i valori assunti da una variabile y rilevata in due occasioni consecutive, lo stimatore ottimo della media assegna alla parte *panel* del campione un peso relativo in generale superiore rispetto alla quota di interviste, riducendo corrispondentemente il peso relativo delle famiglie non *panel*.

Giacché tale riponderazione potrebbe alterare la struttura del campione, i pesi finali vengono allineati alle caratteristiche della popolazione per quanto riguarda il sesso, la classe di età, l'area geografica e la dimensione del comune di residenza.

Pesi individuali e pesi familiari

Nell'IBF tutti i componenti delle famiglie del campione sono oggetto di intervista, anche se non necessariamente in via diretta; un componente può infatti rispondere anche per conto di altri, se a conoscenza delle relative informazioni. All'interno della famiglia, quindi, tutti i componenti hanno la stessa probabilità di estrazione e dunque lo stesso coefficiente di ponderazione di partenza.

⁵⁶ Per tenere conto dell'*attrition* si sarebbe potuta modellare la mancata partecipazione, come proposto da Giraldo, Rettore e Trivellato (2001). 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 vedano Kalton e Flores Cervantes (2003).

A causa di errori campionari e non campionari, tuttavia, la struttura della popolazione rilevata dall'indagine differisce da quella osservata da altre fonti (per esempio censuarie o anagrafiche).

Adottando i pesi che tengono conto solo della probabilità di estrazione e della post-stratificazione delle famiglie panel, l'IBF, per esempio, mostra una maggiore incidenza di soggetti anziani, non occupati e con licenza elementare rispetto alla rilevazione sulle forze di lavoro condotta dall'Istat; sono al contrario sottorappresentati i giovani, i laureati e soprattutto i lavoratori indipendenti. La fase finale di ponderazione, che allinea la composizione del campione a quella della popolazione, opera pertanto consistenti ridimensionamenti di tali segmenti (Tavola 3.2).

In tale situazione l'allineamento delle distribuzioni del campione alle corrispondenti distribuzioni note della popolazione consente di tenere sotto controllo una possibile fonte di distorsione (e di varianza) delle stime. Inoltre, dal punto di vista della presentazione, i risultati dell'indagine si inscrivono coerentemente all'interno di un set informativo noto.

L'allineamento, nel caso in esame, presenta però anche alcuni problemi specifici. In primo luogo le informazioni note sulla popolazione riguardano alcune distribuzioni marginali ma non le distribuzioni congiunte; non è dunque possibile applicare una semplice post-stratificazione. Inoltre le informazioni note sono a livello di individuo, mentre la ponderazione viene condotta a livello di famiglia, attribuendo cioè all'interno della famiglia lo stesso peso a tutti i componenti⁵⁷.

La procedura utilizzata per la correzione (applicata sia ai dati dell'archivio storico sia a quelli degli archivi annuali) è il cosiddetto *raking*; siccome la correzione riguarda più fenomeni, la procedura è iterativa⁵⁸.

⁵⁷ Questa soluzione è coerente con il fatto che all'interno di una famiglia tutti i componenti sono intervistati ed equivale ad assumere che la probabilità di risposta di una famiglia sia funzione delle caratteristiche dei propri membri.

⁵⁸ L'*Iterative Proportional Fitting* (o *raking*) è una tecnica che consente di allineare in modo simultaneo i pesi campionari alla distribuzione di alcune caratteristiche note da fonti esterne. Si veda per esempio Verma (2000) o Kalton e Flores Cervantes (2003).

Tavola 3.2

Effetto delle fasi supplementari di ponderazione, 2000-2010
(punti percentuali di variazione rispetto al peso campionario)

Individui									
Sesso	2000	2002	2004	2006	2008	2010	2012	2014	Media
Maschio	0,8	0,8	0,2	0,1	0,2	0,0	-0,7	0,1	-0,3
Femmina	-0,8	-0,8	-0,2	-0,1	-0,2	0,0	0,7	-0,1	0,3
Età									
Fino a 30 anni	2,1	2,1	1,3	2,0	1,3	2,0	3,4	3,8	3,6
Da 31 a 40 anni	2,5	2,5	2,9	3,1	3,1	1,9	3,8	2,9	2,9
Da 41 a 50 anno	0,6	0,6	0,5	0,4	0,4	1,2	1,6	2,7	1,8
Da 51 a 65 anni	-3,2	-3,2	-3,0	-3,0	-2,5	-3,0	-4,3	-3,7	-4,0
Oltre 65 anni	-1,9	-1,9	-1,7	-2,6	-2,4	-2,0	-4,7	-5,7	-5,2
Titolo di studio									
Senza titolo	0,4	0,4	0,5	0,8	0,8	1,3	2,1	2,2	1,9
Licenza elementare	-2,2	-2,2	-1,5	-2,0	-1,7	-1,1	-1,5	-2,3	-1,6
Licenza media inferiore	0,8	0,8	0,2	0,4	0,1	0,0	-0,1	0,5	0,2
Licenza media superiore	0,8	0,8	0,3	0,4	0,4	-0,6	-1,0	-0,1	-0,6
Laurea	0,2	0,2	0,4	0,4	0,5	0,4	0,5	-0,4	0,2
Condizione professionale									
Lavoratore dipendente	0,2	0,2	-0,3	0,0	0,3	-0,4	3,3	1,6	1,5
Lavoratore indipendente	2,8	2,8	3,1	2,9	2,7	2,2	-0,2	-0,3	0,6
Non occupati	-3,0	-3,0	-2,8	-2,9	-3,0	-1,8	-3,1	-1,5	-2,1

Capifamiglia (maggiori percettori di reddito)									
Sesso	2000	2002	2004	2006	2008	2010	2012	2014	Media
Maschio	-1,0	2,0	0,6	0,8	0,6	0,2	-2,9	0,5	-0,7
Femmina	1,0	-2,0	-0,6	-0,8	-0,6	-0,2	2,9	-0,5	0,7
Età									
Fino a 30 anni	0,4	0,6	0,5	1,1	0,9	0,8	0,8	0,6	0,7
Da 31 a 40 anni	2,1	4,2	4,1	4,7	4,5	3	0,2	6,0	3,1
Da 41 a 50 anno	-0,9	1,8	1,7	1,5	1,5	2,7	-1,6	5,0	2,0
Da 51 a 65 anni	-3,1	-3,6	-3,9	-3,7	-2,9	-3,6	0,5	-5,2	-2,8
Oltre 65 anni	1,5	-2,9	-2,3	-3,7	-4,1	-3	0,1	-6,3	-3,1
Titolo di studio									
Senza titolo	0,3	-1,0	-0,7	-0,8	-0,6	-0,4	-0,9	0,8	-0,2
Licenza elementare	-0,1	-3,3	-2,3	-2,7	-2,7	-1,6	0,5	-2,5	-1,2
Licenza media inferiore	0,0	1,9	1,0	1,2	1,2	0,7	1,0	1,5	1,1
Licenza media superiore	0,1	1,8	1,1	1,5	1,6	0,4	-3,0	1,5	-0,4
Laurea	-0,2	0,6	0,9	0,8	0,5	0,8	2,4	-1,3	0,6
Condizione professionale									
Lavoratore dipendente	-2,9	1,5	0,1	1,1	1,4	0,6	1,7	5,8	2,7
Lavoratore indipendente	3,6	4,9	5,5	5,2	4,7	4,5	-0,4	-0,1	1,3
Non occupati	-0,7	-6,3	-5,6	-6,4	-6,2	-5,2	-1,4	-5,7	-4,1

Un esempio può aiutare a chiarire il metodo. Supponiamo di voler allineare i dati dell'indagine alla distribuzione per classi di età e per titolo di studio della popolazione utilizzando come fonte esterna i dati delle due distribuzioni marginali del censimento. In primo luogo (a) si calcola la distribuzione della popolazione per classi di età nel censimento e nell'indagine; poi, (b), per ciascuna classe di età, si calcolano i coefficienti di aggiustamento che consentono di allineare la

distribuzione dell'indagine a quella censuaria. Quindi si ripetono le operazioni (a) e (b) con riferimento ai livelli di istruzione. A questo punto si calcola la media dei pesi individuali (aggiustati per età e istruzione) all'interno di ciascuna famiglia e si attribuisce questo valore a tutti i componenti della famiglia. L'insieme di queste operazioni viene ripetuto fino a convergenza.

L'operazione è iterativa sia perché l'allineamento per la seconda variabile (l'istruzione) modifica i risultati relativi alla prima delle due variabili (l'età), sia perché l'attribuzione dello stesso peso a tutti i componenti appartenenti alla stessa famiglia modifica i risultati sia per età sia per livello di istruzione, per cui è necessario iterare la procedura.

Nell'applicazione pratica del *raking* è utile tenere presente che l'allineamento della distribuzione di uno o più fenomeni a distribuzioni provenienti da altre fonti tende a modificare la struttura del campione, per cui è consigliabile introdurre, tra i vincoli, anche la composizione del campione per le principali variabili socio demografiche. Per fare un esempio, se volessimo allineare la distribuzione delle attività finanziarie a dati di fonte esterna (per esempio le segnalazioni di vigilanza per classi di importo), dovremmo tenere presente che questo allineamento tenderebbe ad accrescere il peso della popolazione residente al Centro Nord, per cui sarebbe utile introdurre come vincolo anche il rispetto della composizione del campione per area di residenza⁵⁹.

Nell'applicazione di questa procedura bisogna inoltre tenere presente che non è consigliabile ricorrervi per allineare fenomeni "rari" a dati di fonte esterna, perché ciò tende spesso a generare pesi assai elevati per un numero limitato di soggetti. Inoltre la convergenza non è assicurata in linea teorica; un esame della velocità con cui la procedura consente di soddisfare i vincoli posti fornisce utili indicazioni nel rapporto tra le caratteristiche del campione e i vincoli stessi; tanto più i vincoli sono compatibili con i dati osservati, tanto più velocemente la procedura converge restituendo distribuzioni dei componenti delle famiglie allineate a quelle desiderate.

L'editing dei dati

L'indagine prevede una serie di controlli di qualità sulle informazioni rilevate; questi controlli sono eseguiti dall'intervistatore, dalla società di rilevazione e dalla Banca d'Italia.

L'intervistatore, qualora rilevi un dato anomalo o un'incoerenza tra risposte fornite a diverse domande richiede - se possibile - spiegazioni all'intervistato. Ormai, poiché le interviste sono svolte quasi sempre tramite computer, ci si preoccupa di progettare il questionario CAPI in modo che è, a fronte di risposte anomale, sia lo stesso programma a proporre domande di conferma. Naturalmente è bene anche non esagerare nella richiesta di conferme, poiché l'eccessiva pedanteria potrebbe infastidire il rispondente e produrre un effetto contrario a quello desiderato. Nel passato, nel caso di questionari cartacei ci si fondava sulla

⁵⁹ Questo approccio è ad esempio adottato in D'Alessio e Neri (2015), che sperimentano allineamenti dei risultati campionari a quelli di reddito e ricchezza noti da fonte esterna.

capacità dell'intervistatore di cogliere incoerenze o problemi nelle risposte, tramite annotazioni sul questionario. In ogni caso, la società di rilevazione controlla singolarmente ogni questionario. Con i questionari cartacei, ciò avveniva con un processo manuale prima di trasferire i dati su supporto elettronico. Con i questionari elettronici, vengono egualmente effettuate verifiche ulteriori rispetto a quelle già previste all'atto dell'intervista. I questionari privi dei fondamentali requisiti di qualità vengono eliminati. In una fase successiva, mediante programmi di controllo, la stessa società individua i dati anomali e le incoerenze; per questi casi, si verifica dapprima se l'errore ha origine nel processo interno (ad esempio nella fase di codifica o di digitazione dei dati), nel qual caso viene rimosso. Se, viceversa, l'errore risiede nel questionario, viene contattata direttamente la famiglia (in generale telefonicamente) per chiedere conto della reale situazione. La Banca d'Italia effettua ulteriori verifiche di qualità; i dati anomali riscontrati sono sottoposti all'attenzione della società di rilevazione che talvolta provvede alla verifica diretta presso le famiglie.

Ciò nonostante, talvolta nella fase di analisi ci si imbatte in dati statistici che appaiono non realistici. È ovvio che queste informazioni, nonostante tutti gli accorgimenti descritti, possono essere effettivamente errate. Assicurare l'assenza totale di errori in un'indagine così complessa non è ovviamente possibile. In numerose circostanze, tuttavia, i dati anomali o incoerenti derivano da situazioni effettivamente inusuali o comunque non previste adeguatamente dal questionario, che costituisce una griglia necessariamente semplificata della realtà.

Biancotti, D'Alessio e Neri (2004) riportano numerosi esempi delle annotazioni riportate dagli intervistatori sui loro questionari (o in alcuni casi dei controllori della società di rilevazione). Per esempio, a prima vista può sembrare un errore una spesa per alimentari della famiglia pari a zero; dalle annotazioni sul questionario si può però apprendere che si tratta di una persona sola che va regolarmente a mangiare nell'albergo di una figlia non convivente. In un altro caso, l'affitto risulta pari a zero perché la famiglia è morosa. Questi esempi mostrano quanto possa essere rischioso procedere ad aggiustamenti sommari di dati elementari ritenuti anomali.

Nell'IBF per lungo tempo gli interventi di correzione sono stati limitati ai casi in cui era del tutto evidente l'origine dell'incoerenza, lasciando al ricercatore, in relazione al caso specifico da analizzare, la valutazione in merito all'eventuale trattamento da riservare ai dati anomali.

Negli ultimi anni, sono state dedicate maggiori risorse alla verifica dei dati, e in particolare alle variazioni dei principali aggregati osservati sulle famiglie panel, facendo anche uso di informazioni che su base qualitativa forniscono un supporto alla soluzione di eventuali incoerenze. Nel questionario sono state introdotte una serie di domande qualitative sullo stato e sull'andamento di alcuni aggregati che vengono utilizzate per individuare l'incoerenza e per attivare un ricontatto della famiglia intervistata o, in alcuni casi più semplici, una diretta correzione del dato errato.

La stima degli aggregati

La definizione di reddito - I principali aggregati economici vengono costruiti a partire da singole voci elementari del questionario.

Il reddito familiare è la somma dei redditi degli individui che ne fanno parte; i redditi individuali, a loro volta, sono la somma di elementi il cui numero e tipo è variabile da caso a caso. A livello individuale vengono infatti rilevati i redditi da lavoro, dipendente o autonomo, i redditi da pensione o da altri trasferimenti. I redditi da capitale, invece, vengono rilevati a livello familiare, soprattutto per la difficoltà nell'imputare la proprietà del capitale finanziario all'interno della famiglia.

Dal punto di vista della stima, le poste sono talvolta chieste direttamente agli intervistati, come ad esempio il salario annuo o i proventi da lavoro autonomo; talvolta sono invece stimate sulla base di altre informazioni richieste, come i redditi da attività finanziaria, per i quali si applica un tasso di interesse allo stock di fine anno.

La definizione di reddito dell'indagine, il cui schema di aggregazione è mostrato nella tavola 3.3 (con riferimento all'indagine sul 2014), presenta alcune differenze rispetto a quella della CN.

In primo luogo l'indagine, ritenendo eccessivamente oneroso chiedere agli intervistati sia i redditi lordi sia gli oneri fiscali e contributivi, rileva direttamente i redditi al netto delle imposte e dei contributi sociali, mentre invece in CN sono disponibili i redditi da varie fonti e le imposte complessivamente pagate. Conseguentemente, la comparazione tra le due fonti può essere effettuata solo per l'insieme dei redditi; a livello di tipo di reddito, infatti, i valori netti dell'indagine non possono essere agevolmente comparati con quelli lordi della CN.

La rilevazione dei singoli redditi al netto delle imposte e dei contributi rappresenta una semplificazione non priva di ambiguità in alcune circostanze. Ad esempio, nel caso di più redditi da lavoro in capo alla stessa persona, uno da lavoro dipendente e uno secondario autonomo, gli importi netti per tipo non sono univocamente definiti. È plausibile che il rispondente, che per i redditi da lavoro dipendente è soggetto a ritenuta alla fonte e per gli altri provvede a pagamenti di imposte successivamente, indichi quale reddito netto da lavoro dipendente quello ricevuto in busta paga e quale reddito netto del lavoro secondario quello dato dall'importo lordo meno le imposte e i contributi calcolati in sede di dichiarazione dei redditi sulla base dell'aliquota marginale. Ovviamente questa ripartizione delle imposte tra le varie fonti è arbitraria e riflette circostanze del tutto incidentali che riguardano le modalità di pagamento delle imposte, la considerazione di quale reddito sia prevalente per l'individuo, ecc..

I redditi da lavoro indipendente presentano nell'indagine una definizione in parte diversa rispetto alla CN. Mentre la CN include i redditi prelevati dai membri delle quasi-società ovvero gli importi che le famiglie prelevano effettivamente per i propri bisogni sui profitti conseguiti dalle quasi-società di loro proprietà, nell'indagine si attribuisce alla famiglia l'intero profitto dell'attività imprenditoriale (per la quota spettante alla famiglia), anche quello eventualmente non distribuito e reinvestito.

Tavola 3.3

Piano di aggregazione delle variabili: conto del reddito, 2014

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.

Per quanto riguarda gli ammortamenti, la differenza tra CN e IBF riguarda il metodo di calcolo; mentre nella CN si fanno ipotesi sulla durata dei beni capitali, nell'indagine la valutazione del deprezzamento nel corso dell'anno è lasciata all'intervistato.

Complessità di rilevazione e di stima caratterizzano anche i redditi da capitale. La parte più cospicua è rappresentata dagli affitti imputati corrispondenti alle abitazioni occupate dal proprietario. Nell'indagine la stima di questa componente è basata sulla valutazione soggettiva che l'intervistato fornisce riguardo a quanto potrebbe essere l'affitto da pagare per quella stessa abitazione. Tale metodo di valutazione ha il vantaggio di riflettere con una certa rapidità i mutamenti dei valori di mercato; d'altra parte, basandosi su valutazioni soggettive, il metodo presuppone la conoscenza da parte del rispondente delle quotazioni correnti. Inoltre, il metodo comporta un certo disallineamento rispetto agli affitti effettivi, i quali da un lato si riferiscono anche a contratti stipulati a condizioni non di mercato⁶⁰ e dall'altro tendono ad adeguarsi più gradualmente alle condizioni di mercato, a causa di una certa rigidità dei contratti in essere. Per quanto riguarda gli affitti effettivamente percepiti, essi sono rilevati senza alcuna

⁶⁰ Ad esempio le case degli enti pubblici sono sovente affittati a prezzi inferiori a quelli di mercato.

menzione alle imposte sul reddito; tenuto conto delle modalità di pagamento delle relative imposte sul reddito e sul patrimonio, e potrebbero anche essere dichiarati dai rispondenti al lordo di imposta.

Come si è già anticipato, i redditi da capitale finanziario vengono stimati applicando un tasso di interesse allo stock di fine anno⁶¹. Per i depositi, i titoli di stato e i mutui passivi (che costituiscono una componente negativa del reddito) si applica il relativo tasso medio osservato nell'anno di riferimento; per le azioni e le quote di fondi comuni, il relativo reddito viene calcolato escludendo le variazioni dei valori degli asset⁶².

Rispetto alla CN vi sono inoltre ulteriori differenze di minore rilievo. In primo luogo l'indagine non considera alcune forme di autoconsumo e in particolare quelli relativi a beni e servizi prodotti da imprese non costituite nella forma di società e appartenenti alle famiglie e destinati ad essere consumati dai membri della famiglia. Una quota rilevante di questa posta è rappresentata dai consumi di prodotti agricoli, posta che ha però progressivamente perso rilevanza in connessione con la riduzione del peso dell'agricoltura nell'economia.

L'indagine non considera inoltre i trasferimenti sociali in natura, ovvero i beni e servizi individuali forniti alle famiglie a titolo di trasferimento in natura dall'amministrazione pubblica.

L'indagine considera invece nel conto del reddito i trasferimenti tra famiglie dovute per esempio ai divorzi e alle separazioni: tali poste sono invece assenti a livello aggregato, dove le poste attive e passive si annullano.

La definizione di ricchezza - Secondo le indicazioni di CN, le attività di natura economica sono distinte in due grandi aggregati: le attività non finanziarie e le attività finanziarie⁶³. Le prime sono a loro volta suddivise in attività non finanziarie prodotte, ovvero quelle che sono state ottenute quale prodotto dei processi di produzione, e le attività non finanziarie non prodotte, ovvero non ottenute tramite processi di produzione. Tra le prime rientrano: il capitale fisso, le scorte, gli oggetti di valore. Tra le seconde rientrano i beni materiali non prodotti, come ad esempio i terreni e i giacimenti, e i beni immateriali non prodotti, come

⁶¹ Nelle indagini dal 1973 al 1975 e in altre rilevazioni fino al 1986 agli intervistati veniva esplicitamente richiesto l'ammontare dei dividendi e degli interessi percepiti.

⁶² Per la remunerazione delle azioni, fondi comuni e altri titoli di questo tipo, l'IBF ha adottato criteri non univoci nel tempo e non sempre coincidenti con quelli della CN. In alcuni anni, ad esempio, sono stati inclusi nel reddito anche i capital gains, per tenere conto del fatto che alcuni asset (ad esempio, i fondi comuni) remunerano il capitale, almeno in parte, tramite la rivalutazione delle quote. Altre volte il rendimento di questi asset è stato stimato con un criterio cautelativo, applicando il tasso di interesse dei titoli di Stato.

⁶³ I criteri di definizione dei settori istituzionali interagiscono con la rappresentazione contabile delle attività e delle passività. Il patrimonio netto delle società e quasi società, per esempio, è un'attività finanziaria degli altri settori; ad esso corrispondono attività e passività di vario tipo nel bilancio delle società e quasi-società. Se il settore delle società e quasi società venisse per ipotesi eliminato (rimuovendo così il *corporate veil*), il patrimonio netto delle società e quasi società e le attività finanziarie degli altri settori ad esse corrispondenti si eliderebbero e i settori diversi dalle società e quasi società si ritroverebbero nel portafoglio, in luogo delle attività finanziarie emesse dalle società e quasi società, le attività reali e finanziarie e le altre passività che queste avevano in bilancio. Cfr., per approfondimenti, IFC, 2007).

ad esempio i brevetti e l'avviamento di attività commerciali. Le attività finanziarie sono suddivise in: oro monetario e diritti speciali di prelievo; biglietti, monete e depositi; titoli diversi dalle azioni; prestiti; azioni e altre partecipazioni; riserve tecniche di assicurazione; altri conti attivi e passivi.

Nell'indagine la ricchezza familiare netta è calcolata come la somma delle attività reali (immobili – ovvero abitazioni, fabbricati non residenziali e terreni - aziende, oggetti di valore) e delle attività finanziarie (depositi, titoli di Stato, obbligazioni, azioni, ecc.), al netto delle passività finanziarie (mutui e altri debiti) (Tavola 3.4). Non sono inclusi i contanti (biglietti e monete), la quota maturata del trattamento di fine rapporto (TFR) e il valore attuale della futura pensione pubblica o privata, in quanto non rilevate nell'indagine. Le componenti della ricchezza sono in generale, ma non sempre, valutate al prezzo di realizzo, stimato soggettivamente dall'intervistato⁶⁴.

Tavola 3.4

Piano di aggregazione delle variabili: conto del patrimonio, 2014

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 famiglie
PF	Passività finanziarie (-)
PF1	Debiti verso banche e società finanziarie
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.	

Da lato delle attività reali, la più rilevante differenza rispetto alla classificazione adottata dal SEC, che non ne prevede una diretta controparte, è

⁶⁴ Per le attività finanziarie il questionario non specifica il criterio di valutazione: mentre per le azioni, i fondi comuni e gli altri asset quotati è presumibile che gli intervistati indichino il prezzo di mercato alla fine dell'anno di riferimento, si può ritenere che per alcuni asset (ad esempio i titoli di Stato) venga indicato il valore nominale. Nel caso delle passività, la domanda fa invece esplicito riferimento al valore nominale del capitale residuo.

costituita dalla componente "ricchezza in aziende"⁶⁵. Dal lato finanziario le principali differenze riguardano: (i) le riserve tecniche di assicurazione, che nei Conti Finanziari (d'ora in avanti CF) includono l'accantonamento al fondo trattamento di fine rapporto, non rilevato nell'IBF, e le modalità di misurazione delle componenti relative alle assicurazioni sulla vita e ai fondi pensione; (ii) le azioni ed altre partecipazioni, che nei CF escludono il valore delle partecipazioni in quasi-società, rilevate invece, anche se con particolari modalità, nell'IBF; (iii) i prestiti alle cooperative, inclusi nell'IBF ed esclusi dai CF.

Nell'IBF la ricchezza in aziende è data dalla somma di tre componenti: il valore della ditta/azienda di liberi professionisti, imprenditori individuali e lavoratori autonomi, rilevato sulla base della domanda "*Quanto crede che potrebbe valere la sua ditta/azienda se lei volesse venderla, cessando di svolgervi l'attività, considerando gli eventuali strumenti utilizzati per l'esercizio dell'attività, le scorte di magazzino e l'avviamento ed escludendo il valore degli immobili?*"; il valore delle imprese familiari, rilevato con una domanda analoga alla precedente; il valore delle partecipazioni in società di persone o di capitale, dove il membro della famiglia svolge una attività di controllo o di gestione. In questo caso il valore della partecipazione è rilevato sulla base della domanda "*Qual è il valore di mercato della società, con riferimento alla sola quota a lei spettante?*".

Le definizioni dell'indagine e quelle della CN non sono dunque sovrapponibili. Per la seconda sono infatti da classificare tra le Società e quasi-società e non tra le Famiglie non soltanto le imprese che abbiano la forma giuridica di società di persone o di capitale ma anche le imprese individuali e le società semplici e di fatto con un numero di dipendenti superiore alle 5 unità, le quali, per convenzione, si suppongono dotate di autonomia decisionale e che l'indagine considera invece indistinte dalle famiglie che le possiedono. Un'ulteriore peculiarità della definizione di "ricchezza in aziende" usata nell'indagine, rispetto alla CN, sta nel fatto che tale definizione include le partecipazioni in quelle imprese nelle quali la famiglia esercita un ruolo di controllo o di gestione (che nei CF sono contabilizzate nelle voce Azioni e partecipazioni, assieme a tutte le altre attività della specie, indipendentemente dal ruolo di controllo o di gestione esercitato dalla famiglia).

Per accrescere la comparabilità tra i dati dell'indagine e le stime macro, i dati dell'IBF relativi alla ricchezza in aziende devono dunque essere riclassificati tenendo conto della forma giuridica e della dimensione delle imprese di proprietà delle famiglie (cfr. Banca d'Italia, 2008b).

La definizione di consumo – L'indagine rileva in modo sommario i consumi delle famiglie (tavola 3.5). A parte i beni durevoli, di cui viene rilevata la spesa netta nell'anno, le spese per l'affitto (effettivo o imputato), le spese alimentari e i fringe benefit dei lavoratori dipendenti, il resto viene stimato richiedendo

⁶⁵ Analogamente alla distinzione tra investimenti diretti e di portafoglio presente nella bilancia dei pagamenti, il valore della proprietà di imprese è in parte tra le attività reali, quando questa è collegata all'attività lavorativa del proprietario, e in parte tra le attività finanziarie, nel caso di possesso di azioni o quote di società, quando questa attività costituisce un mero impiego di risparmio.

l'importo totale mensile alla stessa famiglia. Si tratta di una stima di qualità relativamente modesta che ha però lo scopo di costituire un ulteriore punto di riferimento delle condizioni economiche delle famiglie.

Tavola 3.5

Piano di aggregazione delle variabili: conto dell'utilizzazione reddito, 2014

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.	

Gli archivi annuali e l'archivio storico

L'archivio storico raccoglie le informazioni relative alle principali caratteristiche delle famiglie rilevate nel corso degli anni:

- i caratteri anagrafici e lo status occupazionale dei singoli componenti;
- i redditi (da lavoro dipendente e indipendente, da pensione, da trasferimenti, da capitale) percepiti dai componenti del nucleo familiare;
- i consumi (durevoli e non durevoli);
- i beni immobili abitati o posseduti dai membri della famiglia;
- le attività e passività finanziarie detenute dalla famiglia;

Poiché il questionario ha subito nel corso del tempo numerose modifiche, l'archivio raccoglie informazioni talvolta ricodificate rispetto a quelle presenti negli archivi annuali, al fine di ottenere la massima omogeneità nell'intero periodo considerato. Informazioni più disaggregate vengono invece fornite, quando disponibili, all'interno di sottoperiodi omogenei di rilevazione.

Nell'archivio storico non sono incluse le variabili rilevate nelle sezioni monografiche e quelle rilevate in una sola indagine o anche in un numero limitato di indagini. Non sono inoltre incluse quelle variabili per le quali si sono registrati cambiamenti nell'impostazione della domanda tali da rendere impossibili i confronti temporali.

Le stime ottenute sulla base dei dati dell'archivio storico possono essere diverse da quelle ottenute mediante i dati rilasciati negli archivi annuali. Ciò può accadere perché nell'archivio storico gli aggregati sono calcolati privilegiando la comparabilità nel tempo. Per le indagini meno recenti i coefficienti di ponderazione presenti nell'archivio storico sono stati rivisti rispetto a quelli utilizzati al tempo della pubblicazione originaria, per assicurare l'allineamento

della struttura del campione a quella della popolazione, nella ricostruzione resa nota di recente dall'Istat.

Errori standard delle stime

L'analisi dei dati campionari ha lo scopo di ottenere stime di uno o più parametri della popolazione (come ad es. medie, totali e coefficienti di regressione). Una valutazione della variabilità campionaria è necessaria per poter costruire intervalli di confidenza delle stime puntuali.

Per determinare quale sia la variabilità campionaria degli stimatori è necessario utilizzare metodi che consentano di tenere conto sia della procedura con cui gli elementi del campione sono stati selezionati sia della natura di tali stimatori⁶⁶.

La complessità insita nel disegno campionario e la non linearità della maggior parte degli stimatori utilizzati non consente l'utilizzo di formule analitiche, rendendo necessario il ricorso a tecniche di linearizzazione della varianza. Questi metodi non tengono conto dell'influenza sulla variabilità delle stime dovuta agli aggiustamenti dei coefficienti di ponderazione e al successivo processo di post-stratificazione.

Per ovviare a questo problema si può ricorrere a *metodi di replicazione*, che consistono nell'ottenere informazioni sulla variabilità riproducendo le fasi di stima su "repliche" del campione originario. Negli ultimi anni la varianza è computata utilizzando il metodo *jackknife*, metodo che viene preferito in presenza di un disegno di campionamento complesso (stratificazione e selezione a due stadi)⁶⁷.

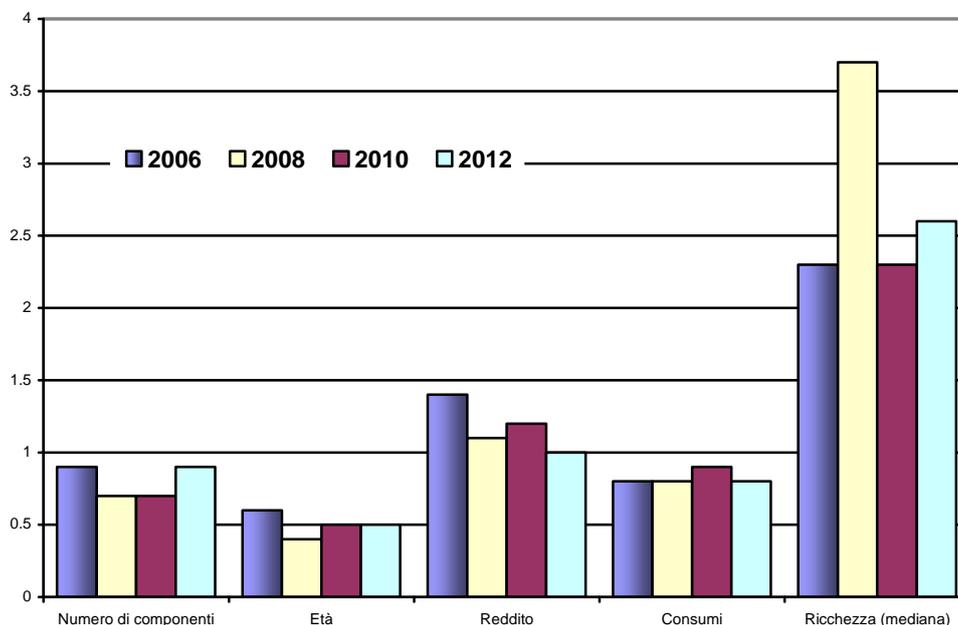
I coefficienti di variazione dei valori medi delle principali variabili sono riportati nella Figura 3.2 (Banca d'Italia 2008a, 2010, 2012, 2014a).

Le medie delle variabili demografiche sono caratterizzate da una ridotta variabilità, anche a causa della stratificazione a posteriori effettuata. Quanto alle variabili economiche di maggiore interesse, si osserva come gli errori standard delle medie del consumo e del reddito siano notevolmente inferiori rispetto a quelli relativi alla ricchezza netta.

Gli errori standard delle stime a livello di dominio risultano in generale superiori a quelli dell'intero campione, in funzione della numerosità campionaria e della variabilità del fenomeno nel dominio.

⁶⁶ Per ulteriori dettagli si veda Faiella (2008).

⁶⁷ Si vedano Wolter (1985) e Shao e Tu (1995).

Figura 3.2**Errori standard delle principali grandezze rilevate, 2006-2012**
(in percentuale della media o mediana)***Sovracampionamento***

L'indagine, nella sua configurazione attuale, non prevede il sovracampionamento di segmenti della popolazione. Tale scelta non è determinata dalla mancanza di interesse per questo approccio; anzi, è generalmente riconosciuto che una maggiore frazione sondata delle famiglie più ricche consentirebbe di ottenere stime più stabili di certi aggregati. Inoltre, il controllo che la quota di famiglie ricche nel campione corrisponda a quella della popolazione potrebbe consentire di migliorare la qualità delle stime non solo riducendo la varianza ma anche riducendo la distorsione che deriva dalla minore probabilità a collaborare delle famiglie più ricche. Il motivo per cui tale tecnica non ha trovato applicazione è la mancanza di liste affidabili di famiglie da sovracampionare.

Nelle indagini della Banca d'Italia un esperimento di sovracampionamento è stato realizzato nel 1987 con l'obiettivo di pervenire a stime maggiormente affidabili del reddito. Il sovracampionamento veniva effettuato su un campione di circa 700 famiglie di dirigenti, imprenditori e liberi professionisti, tratte da elenchi professionali e commerciali; la quota di tali categorie veniva stimata, ex ante, sulla base delle informazioni rilevate nel 1986.

L'integrazione del campione aggiuntivo con quello ordinario presentava due tipi di problemi: a) quale peso assegnare alle famiglie sovracampionate, dato che la procedura utilizzata non consentiva di determinare le probabilità di inclusione delle unità della popolazione; b) come gestire le discrepanze tra le caratteristiche del campione aggiuntivo e quelle delle corrispondenti tipologie nel campione ordinario. Qualche notazione può aiutare a comprendere i due tipi di difficoltà.

Supponiamo che nella collettività di riferimento di dimensione N il numero di soggetti di interesse per il sovracampionamento sia pari a N_r . Adottando una frazione di campionamento costante pari a f , nel complesso ci aspetteremmo di avere nel campione $f N_r$ soggetti della categoria di interesse. Supponiamo che per migliorare la precisione delle stime si voglia accrescere la numerosità di questo sotto campione di interesse.

Se la caratteristica da sovracampionare può essere isolata in uno strato a sé, di cui è nota la numerosità della popolazione N_r e per il quale è disponibile una lista da cui estrarre il campione di dimensione n_r , allora il sovracampionamento equivale semplicemente ad assegnare una frazione sondata maggiore per lo strato in questione rispetto alla media del campione, cioè $f_r > f$. Per ottenere stime corrette, si utilizzeranno coefficienti di ponderazione inversamente proporzionali alla maggiore probabilità di inclusione delle unità dello strato.

Sovente, tuttavia, non è facile isolare, all'interno del campione ordinario, gli elementi di cui si desidera una sovra-rappresentazione nel campione. Ad esempio, se si desidera avere nel campione una maggiore presenza di persone ricche e si dispone di un elenco esaustivo di persone ricche da cui estrarre un campione supplementare di numerosità n_r^s , ma al tempo stesso non si è in grado di evitare la presenza di persone ricche nel campione ordinario (n_r^o), allora bisogna gestire la contemporanea presenza nel campione di famiglie ricche provenienti dal campione ordinario e da quello supplementare. Poiché si è ipotizzato di avere una lista completa delle persone con la caratteristica richiesta, e dunque di conoscerne la numerosità totale N_r , sarà sufficiente assegnare alle unità campionate della specie il coefficiente di riproporzionamento $p_r = N_r / (n_r^o + n_r^s)$, e alle altre un peso del tipo $p = (N - N_r) / (n^o - n_r^o)$, che escluda cioè dal riproporzionamento le famiglie ricche già considerate nell'altro coefficiente.

Se l'elenco di nominativi da cui estrarre il campione supplementare non è esaustivo ma è un campione casuale rispetto alla (sotto) popolazione, e dunque non si conosce l'informazione circa la dimensione N_r della popolazione, allora è possibile stimare la quota dei soggetti in questione sul campione ordinario e poi ripartirne il peso con le unità del campione supplementare: $p_r = N n_r^o / (n_r^o + n_r^s) n^o$.

Questo metodo fornisce risultati in media corretti, ma presenta il problema che se N_r è piccolo, la sua stima $N n_r^o / n^o$ può essere assai imprecisa; siccome quella stima viene poi utilizzata per attribuire un peso alle osservazioni sia del campione ordinario sia di quello supplementare, la qualità delle stime ne risente negativamente rispetto alla situazione in cui si conosce il numero complessivo N_r di soggetti di interesse.

La situazione si complica se l'insieme dei soggetti di interesse presenti nelle liste utilizzate per il sovracampionamento non è né esaustivo né casualmente estratto dalla popolazione di riferimento. In tal caso il campione aggiuntivo non è rappresentativo della corrispondente tipologia di soggetti presenti nel campione ordinario e non è in generale possibile correggere la distorsione tramite l'uso di pesi.

Nell'esperienza del sovracampionamento effettuato nel 1987 l'insieme dei dirigenti, imprenditori e liberi professionisti da cui venne estratto il campione aggiuntivo non era né esaustivo, né casualmente estratto dalla popolazione di

riferimento. Dirigenti, imprenditori e liberi professionisti sovra campionati avevano nel complesso un reddito medio superiore a quello delle corrispondenti categorie del campione ordinario. La differenza si concentrava tra gli imprenditori, che risultavano utilizzare un numero di dipendenti significativamente superiore a quello degli imprenditori presenti nel campione ordinario. Per tenere conto di questo aspetto venne effettuata una stratificazione a posteriori delle osservazioni, allineando il campione aggiuntivo degli imprenditori a quello ordinario, per classi di dipendenti. Ma la distribuzione degli imprenditori per classi di dipendenti nel campione ordinario era soggetta a un'ampia variabilità campionaria, determinata dalla modesta dimensione di quella categoria; l'uso dei pesi calcolati con una procedura di stratificazione ex post come sopra descritto tendeva dunque a vanificare il beneficio del sovra campionamento⁶⁸. In considerazione di questi aspetti esso non venne ripetuto nella successiva indagine sul 1989.

In prospettiva, comunque, sarebbe utile sperimentare di nuovo forme di sovra campionamento per i fenomeni caratterizzati da una minore frequenza (in particolare redditi e ricchezza molto elevati). Varie sono, in particolare, le fonti che potrebbero rivelarsi utili: le dichiarazioni dei redditi delle persone fisiche; le informazioni catastali sulla proprietà immobiliare; le segnalazioni di vigilanza sull'indebitamento delle famiglie (essendo i debiti positivamente correlati con la dimensione della ricchezza e con il reddito). Nella pratica, tuttavia, ci si scontra con la difficoltà di individuare una procedura che consenta di utilizzare nel campionamento liste di particolari tipologie di cittadini nel pieno rispetto della privacy.

Un procedura che non avrebbe questo genere di inconveniente è quella che fa riferimento a campioni areali; un sovra campionamento delle famiglie residenti nelle zone a più elevato reddito o ricchezza, per esempio nelle zone urbane di particolare pregio, potrebbe essere effettuato disponendo solo di informazioni legate alle caratteristiche degli immobili (ad esempio i relativi prezzi al metro quadrato) o statistiche aggregate sulle caratteristiche della popolazione (ad esempio la quota di laureati o di dirigenti). Questo approccio, tuttavia, essendo basato su indicatori meno direttamente legati agli individui che si cerca di sovra rappresentare, presenta in generale una minore efficacia nel procedimento di selezione di quei gruppi.

L'IBF e l'HFCS

L'Household Finance and Consumption Survey (HFCS) è una indagine campionaria armonizzata su ricchezza, reddito e consumi delle famiglie dell'area dell'euro condotta su base volontaria dalle Banche Centrali Nazionali (BCN)⁶⁹. Essa è stata predisposta con l'obiettivo di fornire informazioni sul comportamento delle famiglie dell'area dell'euro, utili per una migliore comprensione dei meccanismi di trasmissione della politica monetaria e per la valutazione di profili

⁶⁸ Problemi analoghi hanno caratterizzato le esperienze di sovracampionamento dei primi anni settanta.

⁶⁹ Per una presentazione completa delle caratteristiche specifiche di ogni indagine, si veda BCE (2013a).

di stabilità finanziaria. L'IBF costituisce il contributo italiano a questa rilevazione⁷⁰.

Le BCN rilevano, con una metodologia per quanto possibile omogenea e secondo definizioni armonizzate, le variabili necessarie a ricostruire i bilanci delle famiglie sia in termini di consistenze sia di flussi, con particolare attenzione alle componenti della ricchezza; tre quarti delle informazioni sono rilevate a livello familiare, le restanti a livello personale.

Tra i 15 paesi partecipanti alla prima edizione, 8 hanno adattato indagini già esistenti e 7 ne hanno avviate di nuove. Il campione complessivo della prima rilevazione, che include indagini riferite al 2010/2011 è pari a oltre 60.000 famiglie (tavola 3.6). Nel complesso si tratta di un campione piuttosto numeroso; a livello di singolo paese, invece, le numerosità non sono sempre adeguate a fornire stime dotate della precisione necessaria.

Tavola 3.6

Caratteristiche del campione e della popolazione, I edizione
(numeri)

Paese	Famiglie campione	Famiglie popolazione	Individui popolazione
Belgio.....	2.327	4.692.601	10.839.900
Germania	3.565	39.673.000	81.086.000
Grecia.....	2.971	4.114.150	10.860.000
Spagna	6.197	17.017.706	45.632.200
Francia	15.006	27.860.408	62.464.200
Italia	7.951	23.817.962	60.309.900
Cipro.....	1.237	303.242	836.600
Lussemburgo.....	950	186.44	462.600
Malta.....	843	143.677	410.000
Paesi bassi.....	1.301	7.386.144	16.366.300
Austria	2.38	3.773.956	8.021.900
Portogallo.....	4.404	3.932.010	10.637.400
Slovenia	343	777.777	2.000.400
Slovacchia.....	2.057	1.911.664	5.412.300
Finlandia	10.989	2.531.500	5.271.500
Area dell'euro.....	62.521	138.122.237	320.611.200

Poiché l'HFCS prevede che i dati relativi al reddito siano al lordo delle imposte e dei contributi sociali, i dati IBF (che rilevano i valori netti) hanno dovuto subire un trattamento preliminare che consiste nel calcolare e aggiungere ai dati osservati una stima dei contributi e delle imposte (c.d. "lordizzazione"). La metodologia utilizzata per calcolare l'imposta sul reddito delle persone fisiche e le relative addizionali comunale e regionale è costituita da un calcolo ricorsivo che

⁷⁰ Come per l'IBF, i microdati anonimi sono a disposizione degli studiosi per finalità di ricerca.

consente di trovare il reddito lordo tale per cui, date le caratteristiche familiari e le diverse tipologie di reddito guadagnate e di patrimonio possedute da ciascun individuo, applicando la struttura dell'imposta si ottiene un reddito netto pari a quello dichiarato nell'indagine. Per i contributi sociali si applica semplicemente l'aliquota proporzionale specifica di ciascuna tipologia di lavoratore⁷¹.

Le prime analisi condotte⁷² mostrano l'interesse ma anche la complessità nel comparare i risultati di indagini di questo tipo tra paesi con strutture demografiche ed economiche profondamente diverse e con indagini che presentano problematiche riguardanti la qualità dei dati non sempre omogenee.

Tavola 3.7

Caratteristiche del campione e della popolazione, I edizione
(numeri)

	Reddito medio pro capite HFCS (A)	Reddito medio pro capite di CN (B)	Rapporto tra stime HFCS e CN (A/B)	Indice di Gini – HFCS (D)
Belgio	21.444	20.239	106,0	0,46
Germania	21.298	20.910	101,9	0,39
Grecia	10.478	15.932	65,8	0,35
Spagna	11.683	15.712	74,4	0,38
Francia	16.466	21.316	77,2	0,35
Italia	13.563	17.838	76,0	0,36
Cipro	15.679	15.820	99,1	0,41
Lussemburgo	33.715	32.233	104,6	0,40
Malta	9.267	12.142	76,3	0,33
Paesi bassi	20.666	17.122	120,7	0,31
Austria	20.667	22.655	91,2	0,37
Portogallo	7.507	11.916	63,0	0,43
Slovenia	8.684	11.668	74,4	0,41
Slovacchia	4.757	7.784	61,1	0,29
Finlandia	21.678	20.217	107,2	0,32

Nel complesso i valori medi nazionali stimati dall'HFCS appaiono mediamente più bassi di quelli di CN, sebbene in misura variabile da paese a paese (tavola 3.7). L'Italia presenta valori simili alla Francia e alla Spagna e con questi occupa una posizione all'incirca mediana; paesi come Belgio, Germania, Finlandia e Lussemburgo presentano valori medi campionari superiori a quelli di CN mentre altri paesi (Grecia, Portogallo e Slovacchia) presentano sottostime più marcate di quelle relative all'Italia. I livelli di correlazione delle stime HFCS con quelli di CN sono comunque elevati (0,943).

Nel corso degli ultimi mesi del 2016 dovrebbero rendersi disponibili i risultati relativi alla seconda wave dell'HFCS, riferita a un numero più ampio di

⁷¹ Per una descrizione della procedura di lordizzazione si veda Marino e Rapallini (2003).

⁷² Si veda BCE (2013b); con particolare riferimento all'Italia, si veda Gambacorta et al. (2013).

paesi rispetto alla prima edizione⁷³; le analisi consentiranno di mettere a fuoco ulteriormente i pregi e gli eventuali limiti di queste analisi comparative tra paesi.

4. Mancate risposte, under-reporting e confronto con altre fonti

Mancate risposte parziali

In linea generale, le mancate risposte alle domande poste in un'indagine possono dipendere da vari fattori: gli intervistati possono infatti non essere a conoscenza di determinati particolari che vengono loro chiesti, ad esempio per motivi di memoria o perché l'informazione è nota a un membro della famiglia non presente al momento dell'intervista. In altri casi, gli intervistati possono non essere disposti a fornire l'informazione richiesta, per motivi di riservatezza o altri motivi.

Nell'IBF gli intervistati sono informati in anticipo sull'oggetto dell'indagine; la partecipazione è pertanto subordinata alla disponibilità a fornire informazioni sul reddito, la ricchezza e i comportamenti in campo finanziario. Questa strategia ha vantaggi di natura pratica; si riesce infatti ad evitare di dover gestire un largo numero di rifiuti a proseguire l'intervista, quando questa è già stata ampiamente svolta. È inoltre possibile definire con precisione con la società incaricata di svolgere la rilevazione quante interviste complete sono necessarie per concludere l'incarico. Dal punto di vista operativo le procedure di imputazione dei dati mancanti finiscono con il riguardare un numero limitato di famiglie, e raramente per le variabili chiave. Ovviamente questa scelta comporta l'inconveniente di ampliare la quota di famiglie non disposte a partecipare all'indagine (includendovi anche coloro che sarebbero stati disposti a fornire qualche informazione secondaria), aspetto che viene trattato più avanti.

Le mancate risposte parziali generano vari problemi che possono essere di differente gravità a seconda delle caratteristiche probabilistiche del meccanismo che genera i dati mancanti. Di norma, se i dati mancanti sono il risultato di un processo casuale che si realizza con la stessa probabilità in tutte le osservazioni, i problemi sono meno gravi che se la mancata risposta è funzione del fenomeno oggetto di indagine.

Il problema più ovvio che sorge in presenza di missing data è l'impossibilità di calcolare correttamente gli aggregati che includono tra le loro componenti l'informazione mancante. Per esempio, se uno dei componenti la famiglia non dichiara il reddito da lavoro indipendente non è possibile calcolare il reddito complessivo di quel componente e della sua famiglia.

Si possono poi generare varie forme di incoerenza tra i risultati di un'indagine. Supponiamo che la dimensione del fenomeno x non possa concettualmente superare quella del fenomeno y (per fare un esempio, l'età dei figli non può superare quella delle madri), ovvero $x < y$, per ogni famiglia. Se fossero missing buona parte delle età sia dei figli giovani sia delle madri anziane

⁷³ La II wave dell'HFCS comprenderà i dati di 20 paesi; oltre ai 15 partecipanti alla prima edizione (si veda tav. 3.6), saranno inclusi anche i dati di Estonia, Irlanda, Lettonia, Polonia e Ungheria.

potrebbe ipoteticamente accadere che il valore medio di x , calcolato solo su coloro che forniscono la risposta, superi quello di y , generando valori medi tra loro incoerenti.

Se la variabile mancante è una variabile di classificazione possono generarsi incoerenze nella presentazione o nell'interpretazione dei dati; per esempio, la media ponderata dei redditi delle famiglie del Centro Nord e di quelle del Mezzogiorno potrebbe differire dalla media italiana laddove non fosse nota l'area di residenza di un certo insieme di famiglie, incluso nelle statistiche relative all'intero Paese, ma escluso dalle statistiche relative alle due ripartizioni territoriali.

Infine vi è il problema della possibile distorsione delle stime effettuate con dati mancanti. Ovviamente, se i dati mancanti sono il risultato di un processo casuale, o comunque la loro distribuzione è indipendente dal fenomeno oggetto di indagine, le stime ricavate sui soli dati disponibili sono corrette. Le stime continuano ad essere corrette anche se la distribuzione dei dati mancanti dipende da alcune variabili correlate con il fenomeno esaminato, purché si disponga di queste variabili e le si utilizzi nella imputazione dei dati mancanti. Se, ad esempio, il reddito degli autonomi è più frequentemente mancante di quello dei dipendenti, la stima del reddito complessiva sarà corretta purché si imputino i dati mancanti sulla base della condizione professionale. Chiaramente, la stima cessa di essere corretta se la probabilità di non risposta dipende, oltre che dalla condizione professionale (che abbiamo ipotizzato di conoscere e di poter controllare), anche da altre variabili correlate con il reddito (o anche dallo stesso ammontare del reddito) che non è possibile controllare e inserire nel modello di imputazione.

L'imputazione dei dati mancanti è dunque una fase importante che può essere condotta in vari modi. In letteratura si fa riferimento o a metodi che individuano osservazioni dalle caratteristiche simili a quelle che posseggono dati mancanti, da utilizzare come donatori nella fase di imputazione (in alcuni casi il record donatore è relativo alla stessa unità in tempi precedenti), o metodi basati su modelli statistici, che ad esempio assegnano alle unità con dati mancanti i valori previsti da una regressione. Sono casi particolari di quest'ultimo metodo la stima con medie semplici o condizionate⁷⁴.

Nell'IBF il numero dei dati imputati è in generale modesto, dell'ordine di qualche decina di casi per la maggior parte delle variabili. Come si è detto, ciò è anche il risultato di una scelta sulle modalità di rilevazione: si preferisce acquisire

⁷⁴ Una strategia alternativa all'imputazione è quella della ponderazione; in presenza di dati mancanti si riponderano i record i cui dati sono disponibili in modo da che questi possano rappresentare anche quelli con dati mancanti (che invece si escludono). È possibile trovare equivalenze formali tra le due procedure; la riponderazione dell'intero campione con un unico coefficiente moltiplicativo equivale a forme di imputazione dei dati mancanti con modelli incondizionati. Ugualmente, una riponderazione condizionata equivale ad adottare un metodo di imputazione condizionato con le stesse variabili di classificazione. I due metodi possono però divergere notevolmente sul piano pratico. La riponderazione ha il vantaggio di operare sempre con dati osservati, senza la necessità cioè di inventare un numero "verosimile" per i record con dati mancanti, in cui le relazioni di coerenza tra le diverse variabili sono mantenute; il campione, però, cambia continuamente, per numerosità e caratteristiche, in relazione ai dati mancanti di ciascuna variabile. L'imputazione ha il vantaggio di mantenere la struttura del campione inalterata; ha però il difetto di richiedere un'imputazione per ogni dato mancante, operazione che non sempre consente di preservare la coerenza dei dati.

subito la disponibilità dell'intervistato a fornire informazioni su temi delicati quali quelli del reddito e della ricchezza, in modo da evitare questionari incompleti, che si otterrebbero probabilmente intervistando i soggetti senza una preventiva ed esplicita illustrazione dei contenuti dell'intervista. Inoltre, nel contratto con la società di rilevazione si specifica che la validità dell'intervista è subordinata alla completezza dei dati su reddito e ricchezza. Questa strategia ovviamente diretta verso la mancata partecipazione coloro che sarebbero disposti a fornire solo alcune delle informazioni richieste, spingendo i rispondenti che hanno informazioni anche non del tutto precise a fornire delle stime dei fenomeni richiesti.

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. Per quanto detto, si tratta spesso di valori mancanti dovuti alla non conoscenza, o alla incapacità di fornire il dato, e non a una volontaria indisponibilità a fornire l'informazione. Per queste variabili, in media, i dati mancanti rappresentano meno del 10 per cento dei casi rilevati.

La metodologia utilizzata per l'imputazione fa in generale 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 il più delle volte 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.

Questa procedura consente di preservare la variabilità dei dati effettivamente rilevati; dal punto di vista inferenziale, tuttavia, il dato imputato non può essere trattato alla stregua di un dato effettivamente osservato. Procedure che tengono conto di questo aspetto sono quelle basate sulla cosiddetta imputazione multipla, che però non sono utilizzate nell'indagine⁷⁵.

Mancate risposte totali e selection bias

Le mancate interviste rappresentano un problema nelle indagini statistiche in quanto possono condurre a campioni nei quali i segmenti di popolazione meno disposti a collaborare 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.

Le informazioni sul comportamento di risposta delle famiglie contattate sono raccolte stabilmente nell'IBF. Ad esempio, per il 2014 la tavola 4.1 mostra

⁷⁵ L'imputazione multipla viene adottata da quasi tutte le indagini che partecipano al progetto dell'HFCS. La procedura prevede una replicazione dei record relativi alle unità rilevate e l'imputazione di valori diversi ai dati mancanti nelle varie repliche, in modo da rappresentare l'incertezza associata al dato imputato. Poiché il metodo comporta notevoli svantaggi in termini di maneggevolezza dei dati e di complessità dei criteri di stime esso non viene adottato dall'IBF che presenta un limitato numero di dati mancanti.

che gli intervistatori hanno contattato 15.302 famiglie, delle quali solo il 53,3 per cento ha fornito l'intervista. Il tasso di partecipazione è più elevato per le famiglie panel (84,8 per cento) che per le altre (36,9 per cento).

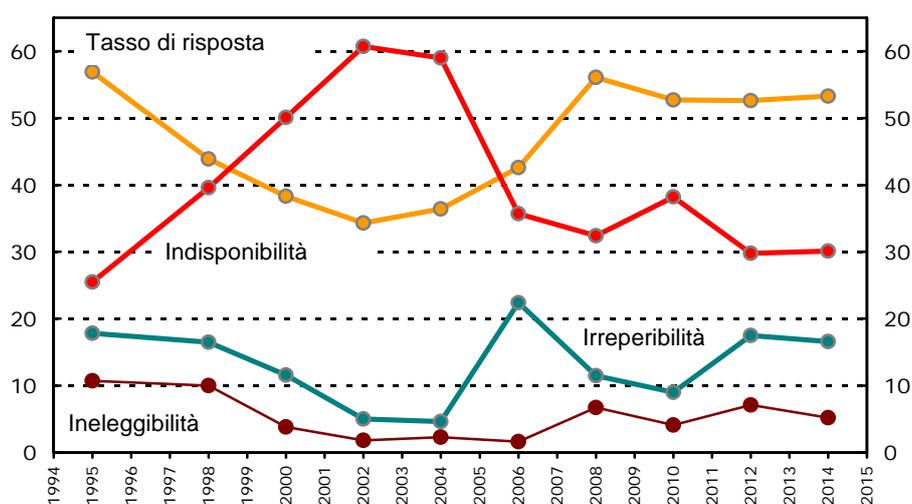
Tavola 4.1
Famiglie contattate e motivi della mancata intervista, 2014

Famiglie:	Panel		Non panel		Totale	
	(unità)	(percentuali)	(unità)	(percentuali)	(unità)	(percentuali)
Intervistate	4.447	84,8	3.709	36,9	8.156	53,3
Indisponibili	626	11,9	3.975	39,5	4.601	30,1
Irreperibili	169	3,2	2.376	23,6	2.545	16,6
Totale	5.242	100	10.060	100	15.302	100
Inleggibili (*)	180	3,4	618	6,1	798	5,2

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

La figura 4.1 mostra come il tasso di risposta sia calato pressoché stabilmente tra il 1995 e il 2002, passando da quasi il 60 al 35 per cento, con un recupero negli anni successivi. Il miglioramento dei tassi di risposta degli ultimi anni riflette soprattutto una riduzione dell'indisponibilità delle persone a concedere l'intervista (a sua volta attribuibile a una maggiore attenzione al fenomeno dei rifiuti) mentre l'irreperibilità, ovvero l'impossibilità di contattare telefonicamente la famiglia o di trovare qualcuno in casa in tre visite effettuate dagli intervistatori in giorni e orari diversi, è leggermente calata fino al 2004 ma negli ultimi anni si è di nuovo manifestata con intensità.

Figura 4.1
Andamento dei tassi di risposta, 1995-2014
(in percentuale dei contatti)



Nel tentativo di accrescere il tasso di risposta, a partire dall'indagine sul 2006 sono stati forniti meno nominativi di riserva alla società di rilevazione, che aveva dunque un maggiore incentivo a ottenere l'intervista. Tale strategia ha

prodotto, come si vede, risultati positivi riportando il tasso di risposta sopra il 50 per cento.⁷⁶

Sebbene un non elevato tasso di risposta costituisca motivo di preoccupazione, va precisato che esso non implica necessariamente una distorsione delle stime; affinché questa si realizzi è infatti necessario che la mancata partecipazione si accompagni a un divario dei valori medi tra rispondenti e non rispondenti.

Per limitare gli effetti negativi della mancata partecipazione sono adottate varie misure. 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 fonte di distorsione importante, 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 una componente di distorsione permanga, 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. Diversi metodi sono stati sperimentati nel corso del tempo per approfondire questo aspetto:

- nella conduzione dell'IBF si raccolgono correntemente informazioni sullo 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). La tabulazione di queste informazioni, sebbene sia riferita a famiglie intervistate, segnala una maggiore difficoltà nell'ottenere l'intervista dalle famiglie più ricche, residenti nei grandi centri urbani e il cui capofamiglia è più istruito; minori difficoltà si incontrano con le famiglie con un ridotto numero di componenti, con capofamiglia pensionato o residenti in comuni di piccole dimensioni. Nel 2014, ad esempio, per intervistare 100 famiglie del quinto con maggiore reddito sono stati necessari 181 azioni di contatto (visite e telefonate) laddove per intervistare 100 famiglie del quinto più povero sono state sufficienti 151 azioni di contatto (tavola 4.2). Sebbene tali informazioni non si concretizzino in una precisa stima della distorsione delle stime IBF, esse sono sufficienti a rendere evidente il problema. Elaborando informazioni di questo tipo, D'Alessio e Faiella (2002) mostrano che i valori medi di reddito e ricchezza delle famiglie differiscono in funzione del livello di facilità con cui sono intervistate. Ad esempio, posto pari a 100 il valore medio del totale campione, coloro che collaborano dopo aver ricevuto la telefonata della società che ha convertito

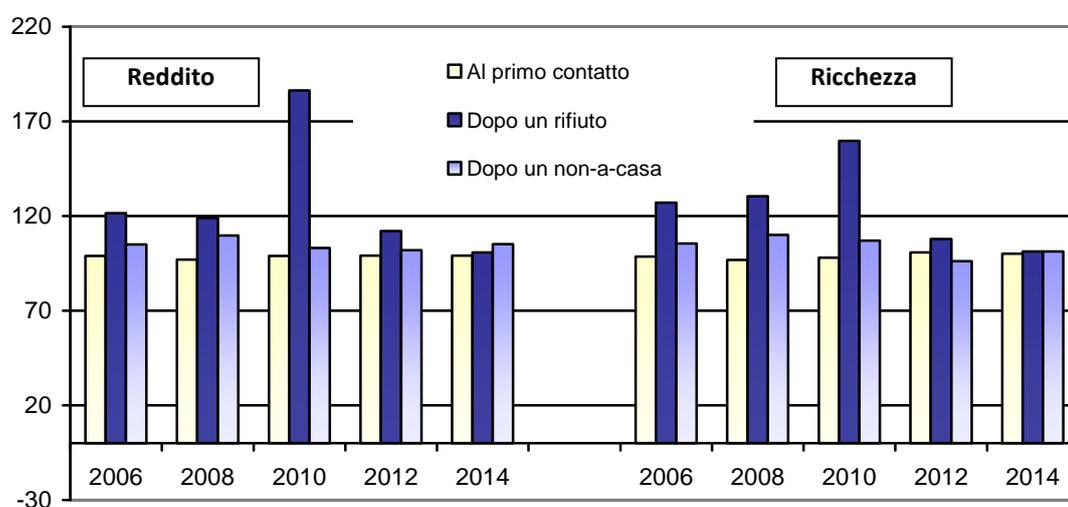
⁷⁶ La crescente difficoltà nell'ottenere informazioni tramite intervista è comune ad altri paesi. Si veda de Leeuw e de Heer (2002) e Groves, Fowler, Couper, Lepkowsky, Singer e Tourangeau (2004).

un rifiuto in partecipazione presentano un valore medio di circa 125 in termini di reddito e di ricchezza netta; quelli che partecipano dopo che alla prima visita non sono stati trovati a casa hanno, in media, un indice dei redditi e della ricchezza pari a circa 105 (Figura 4.2). Sebbene i risultati non siano sempre confermati nei vari anni e mostrino una certa instabilità (anche a causa della ridotta numerosità soprattutto delle famiglie che sono intervistate dopo un rifiuto), si ricava un'indicazione di un possibile bias da selezione.

Figura 4.2

Reddito e ricchezza familiari secondo le modalità di intervista, 2006-2014

(media sul totale interviste dell'anno = 100)

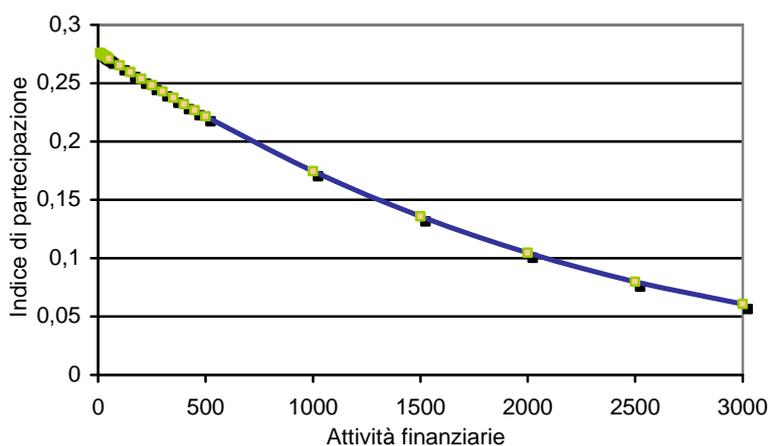


- in un esperimento condotto degli anni novanta da Cannari e D'Alessio (1992), per l'insieme di famiglie panel, che aveva cioè già partecipato a una o più indagini, sono state valutate le caratteristiche di coloro che interrompevano la loro collaborazione rispetto al gruppo di quelle che la proseguivano (*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;
- in un esperimento condotto da D'Alessio e Faiella (2002), circa 2000 famiglie sono state selezionate per l'intervista dell'IBF conoscendone ex-ante, in modo anonimo e nel rispetto della riservatezza, alcune informazioni di natura bancaria. È stato quindi possibile valutare sia il comportamento di partecipazione in funzione della ricchezza finanziaria posseduta, sia quello di risposta, riferito al rapporto tra gli importi dichiarati nell'intervista e quelli effettivamente posseduti, di cui si parlerà oltre. Sulla base di questo esperimento, gli autori concludono che la non risposta non è casuale ma caratterizza più frequentemente le famiglie più ricche (Figura 4.3). 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.

Figura 4.3

Partecipazione stimata all'indagine secondo le attività finanziarie possedute, 1998



- nelle indagini sul 2006 e sul 2008 i non rispondenti sono stati oggetto di una rilevazione telefonica supplementare da parte della società di rilevazione. Circa un quarto delle famiglie non rilevate dagli intervistatori ha risposto a un breve questionario contenente alcune domande in grado di identificare, sia pure grossolanamente, il loro tenore di vita. Le elaborazioni segnalano una maggiore difficoltà di ottenere le interviste dalla famiglie più agiate (Neri e Ranalli, 2011);
- D'Alessio e Iezzi (2015a) confermano una maggiore difficoltà nell'intervista delle famiglie più agiate con riferimento alla loro minore reperibilità nei giorni festivi e prefestivi e nei periodi dedicati alle vacanze. Correggendo per le diverse probabilità di partecipazione dei segmenti della popolazione a partecipare all'indagine in funzione del periodo e del momento scelto per lo svolgimento dell'intervista, gli autori quantificano in circa il 2 e il 4 per cento la sottostima dovuta alla selezione rispettivamente del reddito e della ricchezza.

Tavola 4.2

Contatti, durata media dell'intervista e attendibilità delle risposte, 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.

In conclusione, la distorsione delle stime dovuta alle mancate interviste spesso è presente ma si ritiene di entità contenuta, anche in seguito agli accorgimenti adottati. I divari, talvolta anche marcati, che è possibile osservare rispetto a stime di fonte macro economica possono dunque trovare più ampia giustificazione considerando i comportamenti di under-reporting.

Il confronto tra stime micro e stime macro e l'under-reporting

I redditi: confronto con i dati della Contabilità Nazionale

Le analisi effettuate da Brandolini (1999), riferite al periodo 1970-1995, confrontano le diverse componenti di reddito con quelle della CN. Siccome nel corso del tempo le fonti di reddito rilevate con l'indagine sono state sottoposte a cambiamenti, Brandolini effettua i corrispondenti aggiustamenti sulle serie di CN.

Sotto il profilo dell'eshaustività delle componenti rilevate dall'indagine, l'autore segnala che il rapporto tra le serie armonizzate e le serie originali risulta pari a circa il 90 per cento nel 1970-72 e al 94 nel 1976-1981; è attorno al 100 per cento negli anni successivi. In altri termini per la maggior parte degli anni settanta l'indagine escludeva dal proprio campo di osservazione componenti del reddito disponibile il cui peso era compreso tra il 6 e il 10 per cento dell'aggregato complessivo.

Per quanto attiene invece alla comparazione delle stime degli ammontari nelle due fonti, tra il 1970 e il 1995 il reddito dell'indagine, riportato al complesso delle famiglie italiane, era di circa il 30 per cento inferiore a quello di CN, con notevoli variazioni da un anno all'altro. La discrepanza tra le stime basate sui valori pubblicati e i dati di CN tende a ridursi negli anni settanta. Il divario si amplia di nuovo nei primi anni ottanta e supera il 32 per cento da allora, con l'eccezione del 1987 e del 1989, quando cala al 27 e al 24 per cento. Utilizzando le stime basate sull'Archivio storico il profilo è più stabile: la discrepanza tra le serie che escludono interessi e dividendi è in media pari al 28 per cento e varia dal 24 al 32 per cento. Includendo interessi e dividendi la discrepanza tende ad aumentare.

La scomposizione per fonti di reddito e il relativo confronto con la CN è resa difficoltosa dal fatto che l'indagine rileva i dati al netto delle imposte sul reddito e sul patrimonio, mentre nella CN non si dispone della suddivisione delle imposte per fonti di reddito, per cui è necessario adottare metodi di imputazione in qualche misura arbitrari. Brandolini preferisce calcolare aggregati al netto delle tasse attribuendo ad ogni tipologia di reddito di CN una quota proporzionale di imposte. I valori così ottenuti sono stati poi confrontati con i dati dell'indagine. Con l'eccezione di pochi anni il reddito da lavoro dipendente risultava misurato in modo soddisfacente fino al 1989; nei primi anni novanta la discrepanza cresceva fino a superare il 10 per cento. Al contrario il reddito da lavoro indipendente e gli interessi netti e i dividendi presentano discrepanze rilevanti (dell'ordine del 50-60 per cento per il reddito e più elevate per gli interessi e dividendi) con i dati di CN. Il divario tra IBF e CN si riduce per i trasferimenti, inferiori di un quarto a quelli

di CN. Gli affitti (comprensivi di quelli imputati ai proprietari dell'abitazione di residenza) risultano invece più elevati nell'indagine rispetto ai CN⁷⁷.

Un più recente confronto tra le stime del reddito dell'indagine e quelle aggregate è riportato nella Tavola 4.3 (Banca d'Italia, 2012). Secondo questi dati, l'indagine coglie meglio i redditi da pensione e quelli da lavoro dipendente; maggiore è la sottostima per i redditi da lavoro autonomo, da società e da capitale finanziario. La componente degli affitti imputati appare invece sovrastimata rispetto alla CN.

Tavola 4.3

**Confronto tra le stime macroeconomiche
e risultati dell'indagine: conto del reddito**
(valori percentuali)

Voci	Rapporto fra stima campionaria e aggregata	
	2008	2010
Reddito da lavoro dipendente	89,0	87,8
Affitti imputati ⁽¹⁾	211,9	224,7
Reddito da lavoro indipendente in imprese individuali con meno di 6 dipendenti e affitti effettivi ⁽²⁾	53,3	53,1
Reddito da lavoro indipendente in imprese individuali con più di 5 dipendenti ⁽³⁾	11,2	8,4
Reddito da società e da capitale finanziario ⁽⁴⁾	12,3	6,2
Reddito da pensione e trasferimenti ⁽⁵⁾	75,1	72,4
Totale reddito disponibile	76,2	79,4

I valori di contabilità nazionale sono al netto di contributi sociali e imposte (attribuite in proporzione alle varie voci di reddito esclusi affitti imputati).

Le corrispondenti voci di contabilità nazionale sono: (1) risultato lordo di gestione delle famiglie consumatrici; (2) quota di reddito misto trasferito dalle famiglie produttrici alle famiglie consumatrici; (3) redditi prelevati dai membri delle quasi-società; (4) redditi da capitale (dividendi, altri utili distribuiti dalle società, interessi); (5) prestazioni sociali e altri trasferimenti correnti netti.

Numerosi studi hanno esaminato il fenomeno dell'under-reporting nei dati IBF. Per quanto riguarda l'under-reporting del reddito, alcuni lavori (Cannari e Violi, 1995) si basano su metodi «indiretti» di ricostruzione del reddito «vero», traendo spunto da lavori condotti su dati inglesi (Pissarides e Weber, 1989). Tali metodi sono fondati 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 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 consumi. Gli autori sperimentano due modelli: il primo è basato sulla stima del valore attuale dei redditi futuri attesi e utilizza tale variabile come determinante

⁷⁷ Ciò avviene nonostante il numero delle abitazioni di proprietà delle famiglie sia sottostimato dall'indagine rispetto ai dati di Censimento (si veda più avanti). I diversi metodi di calcolo relativi agli affitti imputati sembrano essere responsabili di questo divario.

della spesa per consumi alimentari; il secondo si basa invece sulla relazione tra spesa per consumi (alimentari e complessiva) e reddito permanente, trattando quest'ultimo come variabile non osservabile.

Per quanto riguarda il sottoinsieme del campione che si assume non caratterizzato da fenomeni di *under-reporting* del reddito, gli autori fanno riferimento al complesso delle famiglie con capofamiglia lavoratore dipendente. Poiché l'analisi riguarda essenzialmente i redditi da lavoro, le famiglie con capofamiglia pensionato o percettore di altri redditi non da lavoro vengono escluse dal campione.

I parametri che evidenziano il comportamento di *under-reporting* delle famiglie di lavoratori indipendenti risultano significativamente diversi da 0. In termini aggregati la sottostima del reddito da lavoro indipendente rilevato con l'indagine equivale a poco più del 20 per cento dell'aggregato dichiarato. Dalla ricerca emerge anche la presenza di comportamenti di *over-reporting*, per i percettori di redditi da lavoro indipendente (approssimativamente) inferiori alla media⁷⁸.

I risultati ottenuti vanno comunque interpretati con una certa cautela. In primo luogo i comportamenti di *under-reporting* potrebbero caratterizzare anche i lavoratori dipendenti (come il confronto con la CN suggerisce). Inoltre, la stima della funzione di *reporting* potrebbe risentire notevolmente del comportamento di reticenza delle famiglie ad alto reddito. In tal caso, il comportamento di *over-reporting* stimato per le restanti famiglie potrebbe risultare più accentuato di quello effettivo. Infine, i modelli sperimentati sono basati sull'ipotesi che le preferenze dei lavoratori autonomi non siano differenti da quelle dei lavoratori dipendenti; tale ipotesi, più realistica per i consumi alimentari, appare più discutibile per la spesa complessiva.

Per quanto riguarda il modello basato sulla stima del valore attuale dei redditi futuri, esso potrebbe risentire degli errori di misurazione non solo dei redditi futuri, ma anche della ricchezza reale e finanziaria. Inoltre, è da tenere presente che i parametri della relazione stimata vengono a dipendere dal tasso di interesse utilizzato per determinare il valore attuale dei redditi futuri attesi, dalle aspettative sull'età di pensionamento, sul rapporto tra pensione e reddito e sulla vita attesa, tutti fenomeni difficili da prevedere.

Più di recente Neri e Zizza (2010) hanno proposto un articolato metodo per la correzione dei redditi familiari.

Per correggere i redditi da lavoro autonomo è stata stimata la relazione fra redditi e valore dell'abitazione di residenza per i soli dipendenti pubblici, che si ritiene dichiarino correttamente il loro reddito; tale relazione è stata poi applicata ai lavoratori autonomi, ipotizzando che il valore dell'abitazione sia fedelmente riportato da entrambe le tipologie di lavoratori. Ne è conseguita una revisione al rialzo del reddito degli autonomi in media del 36 per cento.

Per l'aggiustamento dei redditi da capitale finanziario è stata adottata la metodologia proposta in D'Aurizio, Faiella, Iezzi e Neri (2006), basata sulle informazioni ottenute da un'indagine su un campione di clienti di un grande gruppo bancario italiano. I relativi redditi sono risultati in media triplicati per

⁷⁸ Tali risultati sembrano suggerire l'ipotesi che le famiglie tendano a dichiarare agli intervistatori i fenomeni che sembrano loro appropriati per il gruppo sociale nel quale esse si identificano.

effetto della correzione, che è stata invece di entità più modesta per gli interessi sulle passività (+9 per cento). Per i redditi da capitale reale è stata usata la procedura di Cannari e D'Alessio (1990), che si concentra sulla omessa dichiarazione delle seconde case; ne è seguita una revisione al rialzo dei redditi per affitti effettivi e imputati in media del 23 per cento.

L'integrazione dei redditi derivanti da attività secondarie è stata effettuata a partire dall'Indagine Istat sulle Condizioni di vita (Eusilc), che, come ricordato, incorpora informazioni di natura amministrativa e fiscale. Per effetto della correzione, la quota dei percettori di un reddito da lavoro dipendente o autonomo come fonte addizionale è aumentata rispettivamente di circa 3 e 4 punti percentuali.

Nel complesso, le procedure di aggiustamento hanno permesso di stimare un errore complessivo sui redditi familiari pari a circa il 12 per cento del valore dichiarato; le procedure impiegate risultano tuttavia ancora non del tutto stabilizzate e più vicine a quelle che un ricercatore può legittimamente impiegare, sotto la propria responsabilità, che a procedure tipiche di una produzione statistica standard (come la pubblicazione dei Supplementi al bollettino statistico). Del resto, non esiste al momento nessun istituto di statistica che utilizzi in modo ufficiale indagini campionarie sottoposte a procedure di correzione dei dati di simile complessità.

I redditi: confronto con i dati fiscali

Una prima comparazione dei IBF con quelli di fonte fiscale è stata effettuata da Cannari, Ceriani e D'Alessio (1997), confrontando i dati fiscali e quelli dell'indagine per i redditi netti relativi alle varie categorie di contribuenti. Il lavoro non è recente, ma le questioni affrontate sono ancora attuali.

Il lavoro considera in primo luogo la composizione delle due popolazioni di riferimento. I dati del campione IBF vengono dapprima corretti per tenere conto di problemi di mancata partecipazione all'indagine⁷⁹. Le famiglie del campione sono poi selezionate depurando le osservazioni che sulla base dei redditi riportati nell'indagine, non sono obbligati a presentare la dichiarazione⁸⁰.

⁷⁹ Rispetto all'indagine dell'Istat sulle forze di lavoro, per esempio, il numero di lavoratori dipendenti risulta sovrastimato mentre quello degli indipendenti sottostimato. Le famiglie con redditi elevati, inoltre, mostrano una minore propensione a collaborare. Per tenere conto di questi fattori gli autori effettuano una stratificazione a posteriori, sulla base di informazioni di fonte Istat, utilizzando la composizione della popolazione per sesso, età e area geografica, la composizione degli occupati (dipendenti e indipendenti), il settore di attività (agricoltura, industria, altri settori) e il livello di istruzione.

⁸⁰ Questa procedura non elimina del tutto i problemi di comparabilità in quanto è possibile che alcuni contribuenti, pur essendo nella condizione di esonero, presentino comunque la dichiarazione, ad esempio per usufruire di detrazioni/oneri deducibili; di questa circostanza gli autori tengono conto con opportune elaborazioni effettuate su un campione di contribuenti predisposto dal Ministero delle Finanze.

Le definizioni delle componenti di reddito dell'indagine sono inoltre rese il più possibile comparabili con i dati fiscali⁸¹.

Il confronto tra le due fonti viene effettuato sia sul numero di contribuenti sia sull'ammontare di reddito. Riguardo a quest'ultimo aspetto, viene utilizzato l'ammontare di reddito spendibile totale e le stime delle componenti di reddito al netto delle imposte, che per i dati fiscali sono state ottenute detraendo da ciascuna componente di reddito lordo una quota proporzionale dell'imposta netta pagata⁸².

Permangono peraltro alcuni problemi nel confronto; il più rilevante è connesso con l'ambiguità circa il metodo utilizzato dai soggetti intervistati con più fonti di reddito per depurare dall'imposta le singole componenti.

Il confronto effettuato dagli autori con le statistiche fiscali è relativo alle dichiarazioni presentate nel 1990, suddivise per categorie omogenee di contribuenti (Ministero delle Finanze, 1992). In totale l'indagine stima un numero di contribuenti con redditi da lavoro dipendente o indipendente, tenuti all'invio della dichiarazione, leggermente superiore (circa il 3 per cento) rispetto a quello fornito dal Ministero delle Finanze⁸³.

Per quanto riguarda i lavoratori dipendenti, alla notevole somiglianza delle stime relative al numero complessivo di contribuenti, si contrappongono alcune differenze per quanto riguarda i redditi diversi dal lavoro dipendente. Nell'IBF i lavoratori dipendenti che percepiscono anche redditi di lavoro autonomo e d'impresa sono largamente sottostimati⁸⁴. Anche per gli imprenditori l'indagine appare sottorappresentare i soggetti che percepiscono redditi da più di una attività lavorativa⁸⁵. Per contro, analizzando i soggetti che hanno un solo tipo di reddito

⁸¹ Per i redditi dei fabbricati, per esempio, gli autori tengono conto del fatto che l'indagine valuta gli affitti imputati, che costituiscono una quota rilevante dell'intero aggregato, al prezzo di mercato (ovvero quanto il proprietario riterrebbe di ottenere affittando l'immobile), che è in generale superiore alla rendita catastale utilizzata a fini fiscali. Come si è già osservato, inoltre, la formulazione della domanda sugli affitti lascia ritenere che le imposte non siano detratte.

⁸² La scelta di operare il confronto sui redditi netti consente di evitare la ricostruzione del reddito complessivo nei dati dell'indagine della Banca d'Italia, inevitabilmente soggetta ad imprecisioni sia per la mancanza di alcune variabili, che dovrebbero essere ricostruite indirettamente (oneri deducibili/detrazioni d'imposta), sia per i limiti cui sono soggette le procedure di ricostruzione del reddito al lordo dell'imposta in presenza di fenomeni di evasione.

⁸³ Tale accostamento deriva più dal fatto che le divergenze tra le due fonti tendono a compensarsi che da una piena convergenza. Infatti, mentre secondo l'indagine i percettori di reddito da lavoro dipendente (esclusi i pensionati) ammontano a 14,612 milioni, una cifra leggermente superiore a quella desunta dalle statistiche fiscali (13,996 milioni), per gli altri tipi di reddito si osservano differenze piuttosto marcate a sfavore dell'indagine, che sottostima il numero di percettori di redditi da impresa e da libera professione rispettivamente del 12 e del 24 per cento.

⁸⁴ Nell'indagine solo 91.000 lavoratori percepiscono anche redditi di lavoro autonomo (contro i 422.000 delle statistiche fiscali) e solo 152.000 percepiscono reddito d'impresa (contro 387.000); per contro i percettori di redditi da fabbricati superano del 6 per cento circa il corrispondente dato dell'Anagrafe Tributaria. I percettori di redditi di terreni, infine, risultano sottorappresentati dall'indagine (767.000 contro 1.771.000).

⁸⁵ Secondo la rilevazione della Banca d'Italia, gli imprenditori con redditi di lavoro autonomo sono solo 6.000 (contro i 226.000 del fisco); quelli con redditi di lavoro dipendente sono solo 399.000 (contro 826.000). Come nel caso precedente, si ottiene un maggior numero di percettori di redditi da fabbricati e un minor numero di contribuenti con redditi da terreni. Analogamente nel caso dei percettori di redditi professionali, la compresenza di redditi professionali, redditi di lavoro dipendente e redditi d'impresa è assai inferiore nell'indagine rispetto alle statistiche fiscali.

da lavoro, e considerando anche coloro che hanno congiuntamente redditi da fabbricati e terreni, nell'IBF si registra una costante sovrastima del numero di contribuenti.

Nel complesso l'indagine sottostima i contribuenti con due o più attività lavorative e i percettori di redditi da terreni; il numero di percettori di redditi da fabbricati risulta invece superiore alle statistiche dell'anagrafe tributaria.

Queste divergenze potrebbero essere spiegate da fattori sia di natura campionaria sia fiscale. Per esempio, i soggetti con più di una attività potrebbero mostrare una minore propensione a collaborare all'indagine a causa di una minore disponibilità di tempo libero o potrebbero dimenticare, in sede d'intervista, le attività secondarie marginali o quelle svolte per un periodo di tempo limitato. Queste motivazioni non sembrano, tuttavia rendere pienamente conto di tutti i divari osservati. Per i lavoratori dipendenti alcune divergenze che si osservano per le caratteristiche dei contribuenti tra i dati dell'indagine e i dati fiscali sono difficilmente imputabili solo ai secondi lavori; sebbene alcune di queste siano probabilmente da attribuirsi ai problemi di qualità che caratterizzano tanto l'indagine quanto i dati fiscali, il più ridotto numero di contribuenti rilevato dall'anagrafe tributaria nelle regioni meridionali appare coerente con la presenza in queste aree di un'economia sommersa di dimensioni maggiori che nel resto del Paese.

Per i lavoratori indipendenti il minor numero di contribuenti stimato dall'indagine potrebbe derivare, soprattutto per gli imprenditori, dai comportamenti di *splitting*, vale a dire di divisione dei redditi tra il reale percettore e il coniuge (ed eventualmente altri componenti della famiglia) al fine di minimizzare l'imposta. Questa ipotesi sembra compatibile con alcuni divari che si osservano nella composizione socio-demografica dei sotto insiemi di contribuenti, in particolare con la maggiore presenza di contribuenti di sesso femminile riscontrata nei dati fiscali.

Le divergenze riguardanti la composizione delle due fonti rendono più problematico il confronto relativo ai dati di reddito in quanto i valori medi sono riferiti a insiemi che hanno talvolta diversa numerosità e differenti caratteristiche. Ad esempio, ipotizzando che la differenza riscontrata nel numero di lavoratori indipendenti sia effettivamente dovuta a comportamenti di *splitting*, ne deriva che la semplice comparazione tra i dati medi delle due fonti tende a sovrastimare l'evasione in quanto, per coloro che adottano questa strategia, compara i dati relativi all'impresa nel suo complesso con quelli artificiosamente suddivisi tra i componenti della famiglia. Analogamente, ipotizzando che il minor numero di contribuenti riscontrato nel Mezzogiorno corrisponda effettivamente a comportamenti di evasione totale, la comparazione dei redditi medi perde parzialmente di significato in quanto raffronta insiemi non totalmente omogenei. Per questi motivi, dopo aver effettuato la comparazione dei dati medi di reddito, CCD conducono alcuni tentativi per tenere conto, sia pure in modo approssimativo, della diversa composizione dei due insiemi posti a confronto.

Nell'IBF, il reddito spendibile dei lavoratori dipendenti risulta leggermente inferiore al corrispondente valore riportato nelle statistiche fiscali. Per i redditi da lavoro dipendente la discrepanza tra i dati d'indagine e quelli fiscali è assai

modesta (0,2 per cento a favore dell'indagine). I redditi da lavoro autonomo e d'impresa percepiti dai dipendenti risultano invece notevolmente inferiori a quelli fiscali; ciò deriva dal fatto che l'indagine sottorappresenta i soggetti che percepiscono redditi da più attività lavorative.

Nel caso degli imprenditori il reddito spendibile dell'indagine risulta invece assai più elevato di quello fiscale, con una differenza percentuale del 41,7 per cento. Se si tiene conto dei possibili fenomeni di *splitting* e si effettua una stratificazione a posteriori per tenere conto della differente composizione dei due insiemi, l'ammontare di reddito spendibile risultante nell'indagine supera del 31,5 per cento il corrispondente dato fiscale. Per i professionisti il reddito spendibile dell'indagine è inferiore a quello fiscale (-6,8 per cento).

In sintesi l'evasione fiscale risulta concentrata tra le categorie di lavoro indipendente; questo risultato si rafforza ulteriormente se si considera che i redditi da lavoro indipendente sono sottostimati dall'indagine rispetto alla CN.

Stime dell'evasione sono ottenute dal confronto dei dati IBF e fiscali anche da Fiorio e d'Amuri (2005). Questi autori rilevano che la percentuale di reddito nascosto al fisco è costantemente più elevata nel gruppo dei lavoratori autonomi rispetto ai lavoratori dipendenti: a parità di livello di reddito considerato, i lavoratori autonomi evadono tra il 7 il 27 per cento in più rispetto ai lavoratori dipendenti.

La percentuale di reddito non dichiarato è, per entrambe le tipologie lavorative, più elevata per i livelli di reddito più bassi e decresce al crescere del reddito.

Per i lavoratori autonomi, la percentuale di reddito nascosto è dell'ordine del 70 per cento nel primo decile, 54 per cento nel secondo, oltre il 30 per cento nel terzo e quarto. Tra i lavoratori dipendenti l'evasione si manifesta soprattutto tra i redditi più bassi, probabilmente a causa della presenza di lavoro nero tra i lavoratori meno qualificati, che accettano di lavorare senza un regolare contratto.

Più di recente Marino e Zizza (2011) hanno stimato la propensione all'evasione dell'Irpef confrontando i dati IBF con quelli forniti dal Ministero dell'Economia e delle finanze riferiti all'universo delle dichiarazioni dei redditi relativi al 2004.

Il confronto, basato sui redditi netti pro capite, è effettuato a un elevato livello di dettaglio, utilizzando le variabili di stratificazione comuni alle due fonti (sesso, età, area geografica e tipologia di reddito).

Per minimizzare i rischi associati all'uso di dati campionari, tipicamente affetti da under-reporting, mancate risposte distribuite non casualmente, reticenza e scarsa rappresentatività di alcune tipologie di percettori, il lavoro effettua una serie di trattamenti preliminari sui dati IBF (descritti in dettaglio in Neri e Zizza, 2008), volti ad aumentarne la rappresentatività.

La metodologia di stima è simile a quella usata da Cannari, Ceriani e D'Alessio (1997) e si fonda sulla stessa ipotesi che i redditi dichiarati in IBF siano più veritieri di quelli forniti all'Amministrazione finanziaria, poiché la

partecipazione all'indagine è anonima e volontaria e non è indirizzata a controlli con finalità specifiche.

La stima dell'evasione dell'Irpef è effettuata confrontando i redditi netti pro capite ottenuti dall'IBF e dai dati fiscali per diverse categorie di contribuenti, dopo avere opportunamente lavorato sui dati IBF per selezionare le famiglie di contribuenti e per definire aggregati quanto più possibile simili nelle due fonti⁸⁶. La comparazione viene condotta sulla base delle caratteristiche comuni ai due data-set: genere, età, area geografica di residenza, tipologia di reddito (o di redditi, nel caso il contribuente abbia più fonti di reddito).

Per l'intera popolazione si stima un tasso medio di evasione pari al 13,5 per cento dei redditi. I lavoratori dipendenti e i pensionati avrebbero una tendenza pressoché inesistente a evadere il fisco. I lavoratori autonomi, gli imprenditori e coloro che hanno solo redditi da fabbricati evadono invece di più: il tasso di evasione è pari al 56,3 per cento per i primi e a oltre l'80 per cento per i secondi. La propensione a evadere risulta maggiore per i contribuenti più giovani (19,9 per cento), per quelli che vivono nel Centro (17,4 per cento, contro 14,8 nel Nord e 7,9 nel Sud) e per gli uomini (17,3 per cento contro 9,9 delle donne).

Di recente è stato condotto un esteso confronto tra i dati IBF e i dati fiscali, con la collaborazione dell'Istat che ha reso possibile l'aggancio dei dati in forma anonima (Iezzi e Neri, 2014). I risultati possono così riassumersi:

- non si registrano significative differenze fra le famiglie contattate e quelle non contattate in termini di redditi e di numero di percettori, cioè non sembra vi siano problemi di selezione nell'applicazione delle procedure di contatto da parte della società di rilevazione e degli intervistatori.
- Le famiglie che hanno rifiutato di partecipare all'indagine hanno un reddito netto familiare fiscale che in media è lievemente superiore (di circa il 3 per cento) a quello delle famiglie intervistate; la differenza non risulta tuttavia significativa.
- Le famiglie che hanno rifiutato di partecipare hanno un valore medio dei redditi da pensione superiore (circa il 7 per cento) rispetto alle famiglie intervistate. La differenza dipende principalmente dal maggior numero di percettori, solo in parte compensato da pensioni di importo unitario inferiore. Anche per i redditi da impresa, le famiglie che hanno rifiutato di partecipare presentano valori medi superiori a quelli degli intervistati (14 per cento circa), principalmente a causa di valori pro capite più elevati. I lavoratori dipendenti intervistati hanno sostanzialmente redditi in linea con quelli non intervistati.
- Le famiglie irreperibili presentano un reddito fiscale netto complessivo sostanzialmente equivalente con quello degli intervistati. Questi risultati sono frutto però di una compensazione di tendenze opposte: gli irreperibili hanno infatti redditi da lavoro dipendente e da impresa significativamente superiori a quelli dei rispondenti (di circa il 7 e il 58 per cento rispettivamente), mentre i redditi da pensione risultano decisamente minori

⁸⁶ Ad esempio, gli affitti imputati sono stati stimati nuovamente facendo ricorso ai dati rilevati sul pagamento dell'IMU, che consente di ricavare informazioni vicine a quelle di fonte fiscale.

di quelli delle famiglie intervistate (di circa il 30 per cento), soprattutto a causa del minor numero di percettori di pensione. Le famiglie irreperibili sembrano presentare caratteristiche demografiche diverse da quelle intervistate: sono spesso formate da individui con redditi da lavoro dipendente e da impresa medio alti, che vivono in famiglie con pochi componenti; probabilmente questi individui trascorrono poco tempo a casa e dunque sono più difficili da contattare.

- Le famiglie che sono state dichiarate ineleggibili dagli intervistatori risultano nel complesso più povere rispetto a quelle intervistate, con un reddito complessivo inferiore alla media generale di circa il 20 per cento. Questa differenza è principalmente dovuta al fatto che queste famiglie hanno un numero di percettori decisamente inferiore a quelle intervistate. Inoltre, la percentuale di ineleggibili fra gli individui di cittadinanza straniera è oltre due volte quella rilevata per gli italiani (8,4 contro 3,5 per cento). Questi risultati potrebbero indicare la difficoltà ad intervistare alcuni segmenti della popolazione (quelli più mobili e più poveri, ad esempio) selezionandoli attraverso le liste anagrafiche. Al momento comunque, il fenomeno interessa un numero ridotto di famiglie.
- Il reddito medio familiare di fonte IBF risulta inferiore a quello di fonte fiscale di circa il 5,8 per cento. I redditi da lavoro dipendente e da pensione stimati da IBF sono inferiori a quelli fiscali rispettivamente di quasi il 12 e il 5,1 per cento. Per quanto riguarda invece i redditi da lavoro autonomo e quelli da impresa i dati IBF sono superiori a quelli fiscali rispettivamente dell'8 e del 23 per cento.
- Nel complesso i dati di fonte MEF presentano una maggiore concentrazione del reddito netto familiare rispetto a quelli dell'indagine (l'indice di Gini risulta rispettivamente pari a 0,432 e 0,372).

Scomponendo la differenza tra il valore medio dei redditi MEF per tutte le famiglie contattate e quello dei redditi IBF si verifica che l'errore di risposta (principalmente l'under-reporting) è la principale causa delle differenze osservate. Per quanto riguarda il reddito da lavoro dipendente e da lavoro autonomo l'analisi segnala la quasi totale responsabilità di tale errore; per i redditi di impresa tale fonte di errore determina oltre il 60 per cento della differenza complessiva. Al contrario, per il reddito da pensione è la mancata partecipazione ad avere un peso maggiore e pari a due volte quello dell'errore di risposta. In sintesi, l'analisi ha messo in luce come i fenomeni di non risposta abbiano effetti relativamente modesti sulle stime del reddito familiare complessivo. Una significativa fonte di distorsione nelle stime di reddito dell'indagine deriva invece dal comportamento di reporting degli intervistati.

I redditi: confronto con i dati dell'indagine Eu-Silc

L'indagine EU-SILC (*Statistics on Income and Living Conditions*) è un'indagine sulle condizioni delle famiglie condotta dagli istituti di statistica dei paesi membri della comunità europea. L'indagine rileva informazioni sui redditi netti (a partire dal 2007 anche sui redditi lordi) e su vari indicatori di esclusione sociale.

In Italia l'indagine viene condotta dall'Istat con un'indagine annuale a partire dal 2004 che rileva le informazioni su un campione formato da 26.000 famiglie (per un totale di quasi 70mila individui), distribuite in circa 800 comuni italiani.

Nonostante alcune differenze nelle modalità di calcolo degli aggregati, come ad esempio quella che riguarda i redditi da lavoro autonomo⁸⁷, le stime ottenute nell'indagine EU-SILC possono essere utilizzate per confronto con quelle HFCS (e IBF in particolare), con risultati abbastanza soddisfacenti.

Se, ad esempio, consideriamo i valori medi di reddito familiare per l'insieme dei paesi rilevato sia in HFCS sia in EU-SILC, si osserva una correlazione tra le due stime pari a 0,97 (Figura 4.4). Nel complesso, i livelli di sottostima rispetto ai dati di CN sono mediamente comparabili con quelli osservati nelle indagini HFCS; per l'Italia il valor medio di EU-SILC è più prossimo a quello di CN di circa 5 punti percentuali, su una sottostima complessiva valutata nel 20-25 per cento.

Gli indici di concentrazione di Gini calcolati nelle due indagini presentano una correlazione decisamente inferiore, pari a 0.147 (Figura 4.5); se però si eliminano dal calcolo la Slovenia e il Belgio⁸⁸ che hanno valori eccessivamente diversi nelle due indagini, la correlazione sale a 0.61.

I profili che si ricavano dalle due indagini sono nel complesso piuttosto simili (Figura 4.6). A livello territoriale, l'IBF presenta valori relativamente più alti per il Nord-est, il Centro e per le Isole e più bassi per il Nord-ovest e il Sud. Per categoria professionale i lavoratori indipendenti sono relativamente più agiati in IBF rispetto a EU-SILC (con un indice che supera 140 contro poco più di 120) (Tavola. 4.4).

⁸⁷ Nell'indagine EU-SILC i redditi da lavoro autonomo rilevati presso gli intervistati sono posti a confronto con quelli di fonte amministrativa. Nella stima del reddito familiare viene imputato il massimo tra i due valori. Alcune altre difformità nelle modalità di calcolo dei redditi si osservano per altre componenti, come per esempio nel calcolo dei redditi da capitale finanziario.

⁸⁸ La Slovenia presenta una numerosità campionaria di sole 350 unità; il Belgio appena sopra 2.000.

Figura 4.4
Valori medi di reddito pro capite nelle indagini HFCS e EU-SILC, 2010
(migliaia di euro)

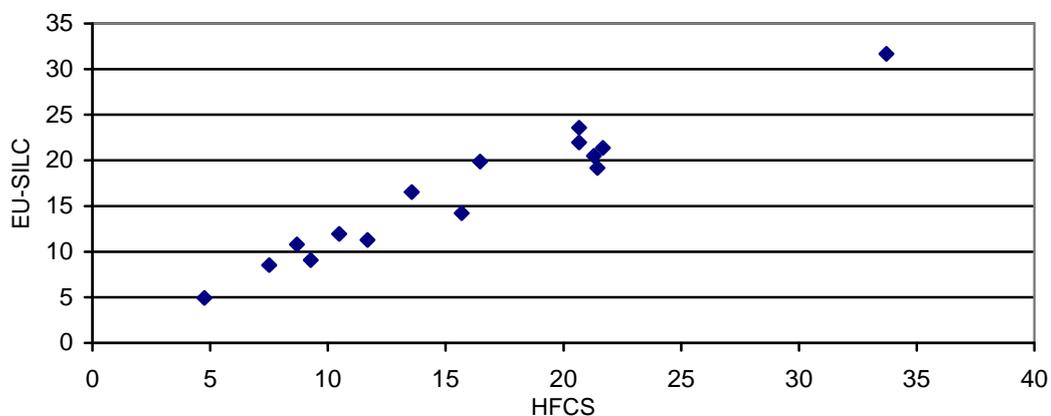


Figura 4.5
Concentrazione dei redditi pro capite nelle indagini HFCS e EU-SILC, 2010

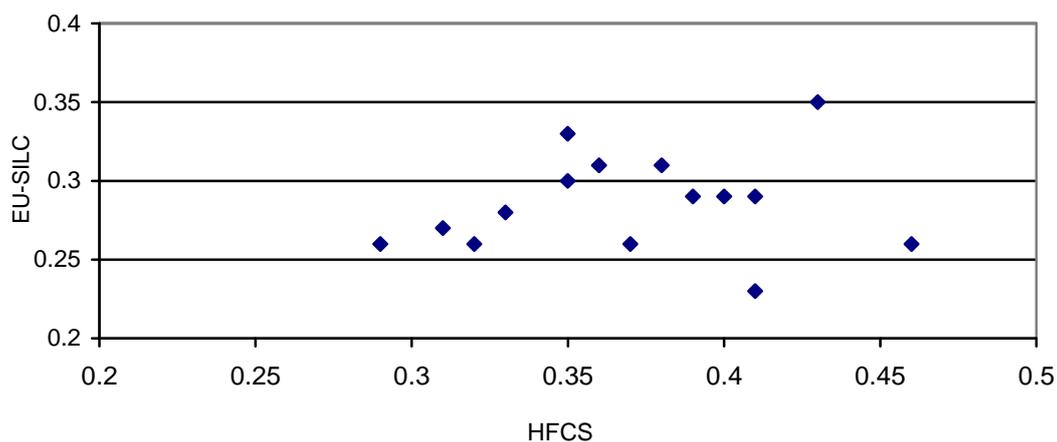


Figura 4.6
Valori medi di reddito familiare nelle indagini HFCS e EU-SILC, 2012
(numeri indice 100=Italia)

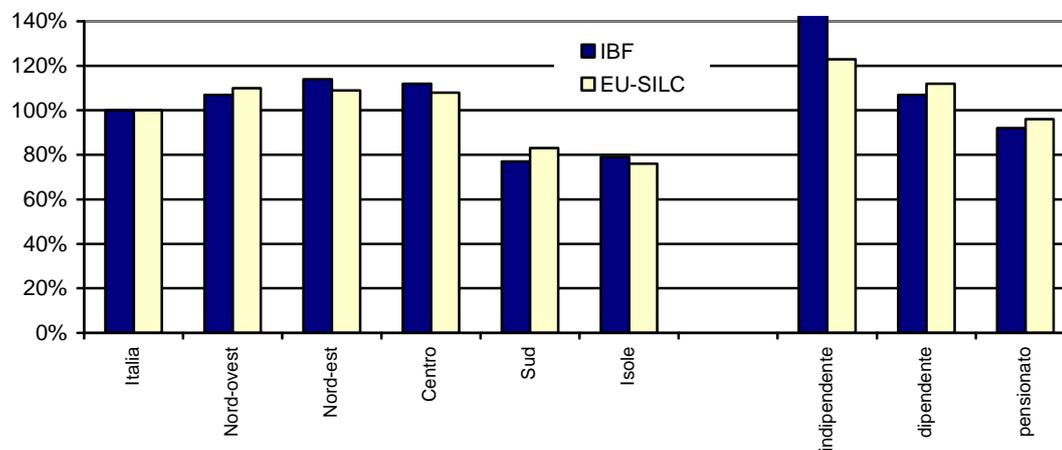


Tavola 4.4

**Valori medi di reddito familiare nelle indagini
campionarie HFCS e EU-SILC, 2012**

(euro, valori percentuali)

	IBF			
	Totale famiglie	di cui con p.r. indipendente	di cui con p.r. dipendente	di cui con p.r. pensionato
Nord-ovest	32.627	49.683	33.363	30.129
Nord-est	34.664	51.506	34.447	31.491
Centro	34.000	43.793	35.639	32.263
Sud	23.336	35.417	27.485	20.744
Isole	24.079	35.229	28.338	22.536
Italia	30.380	44.903	32.560	28.039
	Eu-SILC			
	Totale famiglie	di cui con p.r. indipendente	di cui con p.r. dipendente	di cui con p.r. pensionato
Nord-ovest	38.253	49.597	41.744	34.638
Nord-est	37.679	48.721	40.887	33.152
Centro	37.378	42.338	41.194	35.725
Sud	28.717	33.090	33.103	30.317
Isole	26.280	28.619	31.090	29.545
Italia	34.681	42.546	38.887	33.188
	IBF/EU-SILC (in percentuale)			
	Totale famiglie	di cui con p.r. indipendente	di cui con p.r. dipendente	di cui con p.r. pensionato
Nord-ovest	85,3	100,2	79,9	87,0
Nord-est	92,0	105,7	84,2	95,0
Centro	91,0	103,4	86,5	90,3
Sud	81,3	107,0	83,0	68,4
Isole	91,6	123,1	91,1	76,3
Italia	87,6	105,5	83,7	84,5

I consumi: confronto con dati dell'indagine Istat

Il lavoro di Battistin, Miniaci e Weber (2003) esamina la qualità delle risposte riguardanti i consumi durevoli e non durevoli rilevati nell'indagine. L'analisi viene condotta con riferimento al 1995, comparando le risposte dell'IBF, basate su domande retrospettive, con quelle dell'indagine sui consumi dell'Istat, basate in misura rilevante sull'uso del diario.

Il lavoro mostra che i dati sui consumi dell'IBF sono affetti da alcuni tipici difetti delle domande retrospettive, quali l'arrotondamento delle risposte, che tendono a concentrarsi su valori tipici (1000, 1500, 2000, ecc.). In particolare, i dati sui consumi alimentari risultano di qualità paragonabile a quelli dell'Istat, una volta che si tenga in considerazione il fenomeno di arrotondamento delle risposte; i dati riguardanti i consumi non durevoli e totali, che nell'IBF sono rilevati in modo piuttosto sommario, sono invece più seriamente affetti da errori di misura.

Lo studio di Cifaldi e Neri (2013) esamina il grado di accuratezza delle informazioni sul reddito e sui consumi. Il principale risultato che emerge è che i consumi sembrano essere mediamente più sottostimati del reddito, il quale, essendo il tema centrale dell'indagine, è misurato con maggiore precisione.

Sui consumi alimentari lo studio conferma che essi, a parte gli arrotondamenti, appaiono in linea con quelli riportati dall'indagine dell'Istat. La voce che risente maggiormente degli errori di misura sembra invece essere quella degli altri consumi non durevoli, quali ad esempio le spese per la casa, l'istruzione e il tempo libero. L'indagine contiene, in effetti, una sola domanda che chiede alle famiglie di quantificare il totale di queste spese, producendo dunque risposte poco accurate.

Il grado più elevato di sottostima dei consumi rispetto al reddito implica che il risparmio misurato dall'indagine risulti sovrastimato rispetto a quello effettivo, una distorsione nota, che emerge anche dal confronto con i dati provenienti dai conti nazionali.

L'errore di misura sulla stima dei consumi, inoltre, non aumenterebbe al crescere del livello del consumo dichiarato nell'indagine, al contrario di quello che si verifica per il reddito. Di conseguenza, il risparmio risulterebbe sovrastimato soprattutto per le famiglie con redditi modesti.

Il lavoro di Pisano e Tedeschi (2014), dopo aver evidenziato la maggiore accuratezza dei dati sui consumi non durevoli dell'indagine Istat rispetto a quelli di IBF, segnala che invece per quanto riguarda i beni durevoli sarebbero i dati IBF ad essere più accurati. Ciò sarebbe dovuto al fatto che nella domanda IBF si utilizza un periodo di riferimento (l'anno) più adeguato al tipo di spesa meno frequente rispetto all'Istat, che utilizza un riferimento temporale principalmente mensile.

Le attività reali: confronto con dati di CN

Il confronto tra le stime macroeconomiche e quelle dell'IBF richiede la riclassificazione dei dati relativi alla ricchezza in aziende rilevati nell'indagine, adottando definizioni più coerenti con quelle di CN. Operando tale riclassificazione sui dati del 2000 e del 2002 la ricchezza in aziende viene riallocata per il 40 per cento circa tra le attività reali e per la parte restante nelle attività finanziarie, che complessivamente si rivalutano di circa un terzo (Cannari, Faiella, Marchese e Neri, in Banca d'Italia, 2008).

Ponendo a confronto i dati dell'indagine con quelli di fonte macroeconomica si osserva che per il complesso delle attività reali il rapporto tra le due stime sfiora il 90 per cento: l'indagine sembra dunque in grado di cogliere larga parte di questa componente della ricchezza delle famiglie. Per le abitazioni, che costituiscono la parte preponderante delle attività reali, il grado di copertura supera il 90 per cento.

La discrepanza tra le due fonti deriva prevalentemente dalla sottostima del numero di abitazioni non occupate dal proprietario. La dimensione media delle case in termini di superficie è sovrastimata dall'indagine così come anche il prezzo medio delle abitazioni.

Per lo stock di beni durevoli, il tasso di copertura dell'indagine è inferiore all'80 per cento; tassi di copertura ancor più bassi si riscontrano per i terreni e per i fabbricati non residenziali.

Le abitazioni: coerenza delle risposte dell'indagine e confronto con dati di Censimento

In uno studio condotto sulle abitazioni, Cannari e D'Alessio (1990) segnalano la possibilità di svolgere una verifica di coerenza delle risposte degli intervistati a partire dai soli dati dell'indagine. 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 stimato in circa 3 milioni. Secondo questa stima, le abitazioni date in affitto sarebbero dichiarate nell'indagine nella misura di circa il 30-40 per cento (tavola 4.5).

Tavola 4.5**Abitazioni affittate dichiarate dai proprietari e dagli affittuari, 1991-2014***(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	Quota percentuale	Quota percentuale	Quota percentuale
			(b) / (a)	(b) / (a) Nord	(b) / (a) Centro	(b) / (a) Sud
1991.....	3.291.258	983.777	29,9	15,2	17,5	13,0
1993.....	3.220.253	1.391.772	43,2	33,9	32,1	22,0
1995.....	3.360.512	1.533.344	45,6	46,0	44,4	38,0
1998.....	3.255.218	1.112.374	34,2	47,5	60,5	36,4
2000.....	3.182.180	1304.149	41,0	39,3	39,7	22,6
2002.....	2.970.913	978.709	32,9	61,8	23,0	22,3
2004.....	3.304.629	967.758	29,3	42,4	28,3	21,0
2006.....	3.360.706	861.826	25,6	30,6	47,8	18,6
2008.....	3.320.834	1.529.607	46,1	41,5	82,2	38,0
2010.....	3.646.078	1.205.595	33,1	38,4	40,9	20,4
2012.....	3.850.413	1.225.610	31,8	35,2	39,7	20,7
2014.....	3.391.587	1.083.131	31,9	36,1	44,0	16,1
Media	-	-	36,1	39,0	41,7	24,1

La ripartizione di questo indicatore per area geografica mostra valori più elevati per il Nord e il Centro rispetto al Sud e Isole. Sebbene non sia da escludere la possibilità che alcune case date in affitto siano possedute da famiglie residenti in aree geografiche diverse, è presumibile che il divario osservato sia da attribuire al più elevato livello di under-reporting che caratterizza le famiglie meridionali⁸⁹.

Il confronto dei dati relativi alle abitazioni dichiarate dalle famiglie con i dati di censimento segnala simili misure di under-reporting (tavola 4.6). Secondo i dati dell'indagine sul 1991, ci sarebbero 15,3 milioni di abitazioni di proprietà delle famiglie; il censimento ne stima 22,9 milioni⁹⁰. 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

⁸⁹ 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 da famiglie.

⁹⁰ È possibile che una parte del divario sia da attribuire alla presenza di abitazioni in usufrutto o in uso gratuito.

cento. Simili livelli di under-reporting richiedono trattamenti adeguati, soprattutto quando i dati vengono utilizzati per la produzione di stime aggregate⁹¹.

Tavola 4.6

Abitazioni dichiarate nell'indagine e dati di censimento, 1991-2014

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
1993.....	12.340.184	4.676.460	17.016.644	2.315.954	23.437.958	72,6
1995.....	13.040.473	5.302.615	18.343.089	2.166.325	23.915.457	76,7
1998.....	13.819.808	4.640.864	18.460.672	2.246.121	24.499.098	75,4
2000.....	14.579.414	4.464.775	19.044.189	2.124.591	24.840.866	76,7
2001.....	-	-	-	-	25.035.166	-
2002.....	14.825.485	3.823.484	18.648.969	2.151.803	25.390.315	73,4
2004.....	15.491.800	3.958.271	19.450.071	2.347.636	26.167.661	74,3
2006.....	15.961.712	3.859.934	19.821.646	2.261.599	27.103.231	73,1
2008.....	16.964.765	4.759.211	21.723.976	2.037.326	28.077.781	77,4
2010.....	16.461.997	4.572.917	21.034.915	2.310.729	28.840.727	72,9
2012.....	16.052.196	5.494.757	21.546.953	2.570.592	29.372.670	73,4
2014.....	16.619.504	4.674.897	21.294.400	2.654.056	29.790.201	71,5

(*) 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. Gli altri anni sono stimati interpolando (e estrapolando) i dati di Censimento con l'utilizzo dei dati delle nuove costruzioni di fonte Cresme (cfr. La ricchezza delle famiglie italiane, Banca d'Italia, 2014b).

Le abitazioni: confronto con i dati fiscali

Neri e Monteduro (2013) hanno condotto una comparazione tra i dati IBF e quelli di fonte fiscale riguardanti il patrimonio immobiliare⁹².

⁹¹ Si veda ad esempio Cannari e D'Alessio (1990) e Brandolini, Cannari, D'Alessio e Faiella (2004).

⁹² A partire dal 2008 il Dipartimento delle Finanze (MEF) e l'Agenzia del Territorio hanno creato una banca dati immobiliare integrata che abbina gli archivi dei fabbricati censiti in catasto, quella delle dichiarazioni dei redditi (nelle diverse forme: UNICO, 730, CUD) e quella dell'Osservatorio del Mercato Immobiliare. L'integrazione di queste banche dati consente oggi, grazie anche alle nuove tecnologie informatiche, di elaborare un'enorme quantità di informazioni e di effettuare analisi su un patrimonio informativo che coinvolge ben oltre 20 milioni di soggetti persone fisiche e quasi 60

Naturalmente le fonti amministrative del Catasto e quelle fiscali si riferiscono agli individui percettori di reddito o ai proprietari di quote di immobili. Nell'indagine invece, le informazioni sulla ricchezza sono raccolte a livello familiare, richiedendo solo la quota di proprietà della famiglia nel suo complesso e chi fra i componenti risulta proprietario dell'immobile. Per effettuare la comparazione è dunque necessario effettuare alcune ipotesi, come ad esempio assumere una distribuzione egalitaria delle quote fra i possessori che consenta di stimare le quote individuali anche nell'indagine. Ciò introduce un margine di incertezza nel confronto. Va inoltre tenuto presente che l'attribuzione delle quote di proprietà nei dati amministrativi può essere legata ad aspetti fiscali o convenzionali, che non riflettono l'effettiva disponibilità dei beni posseduti; tale tendenza potrebbe risultare attenuata nell'indagine campionaria, determinando alcune divergenze tra le due fonti.

Per quanto attiene ai valori, invece, le stime effettuate dagli intervistati nell'IBF si confrontano con quelle presenti negli archivi fiscali (basati sulle valutazioni dell'OMI).

I risultati mostrano come l'indagine tenda a sottostimare il numero dei contribuenti che possiedono quote di immobili. In particolare, risulta sottostimato il numero dei contribuenti che possiedono quote molto piccole (inferiori a un sesto del valore dell'immobile) o che invece possiedono più di una abitazione (e soprattutto quelle con un numero elevato di abitazioni, oltre 5).

In termini di ricchezza abitativa la sottostima nei dati dell'indagine è maggiore sotto i 50.000 euro e sopra i 500.000 euro. La differenza più consistente fra le due fonti deriva proprio da quest'ultima classe. Dei circa 1.200 miliardi di euro di sottostima dell'IBF rispetto ai dati MEF, il 70 per cento (circa 805 miliardi) è attribuibile alle classi di ricchezza più elevata⁹³.

La distribuzione della ricchezza abitativa fra i contribuenti appare più concentrata nei dati MEF rispetto ai dati IBF (indice di Gini pari rispettivamente a 0,493 e 0,416).

L'associazione fra ricchezza in abitazioni e reddito risulta più forte nei dati dell'indagine rispetto a quelli del MEF (coefficiente di associazione Φ , pari rispettivamente a 0,53 e 0,31). La maggiore associazione fra reddito e ricchezza in abitazioni osservata nell'indagine potrebbe essere dovuta sia a comportamenti elusivi che si riflettono nelle dichiarazioni rese al fisco, per effetto di intestazioni di quote di proprietà a figli o coniugi con redditi modesti, sia a diversi meccanismi di under-reporting nelle due fonti.

Correggendo i dati campionari con le informazioni amministrative disponibili, la ricchezza immobiliare delle famiglie risulterebbe in media superiore di almeno un quarto a quella dichiarata nell'indagine; non con tutte le

milioni di unità immobiliari, collegando puntualmente ogni proprietario a ciascuna quota immobiliare di cui è titolare.

⁹³ Le differenze nelle altre caratteristiche socio-demografiche dei contribuenti (almeno quelle disponibili) non sembrano invece poter determinare un impatto rilevante nello spiegare le differenze fra le fonti. I due campioni risultano abbastanza allineati per quanto riguarda l'area geografica dei contribuenti (che in entrambi i casi si trovano per il 50 per cento al Nord) e il genere.

metodologie impiegate si otterrebbe invece un aumento della concentrazione della ricchezza.

Le attività finanziarie e la ricchezza nel suo complesso

Per la ricchezza finanziaria un confronto puntuale tra CF e IBF è stato effettuato da Bonci, Marchese e Neri (2005; in seguito BMN), che quantificano le discrepanze tra le due fonti attribuendole ai vari possibili fattori: differenze di definizione, errori di misura, errori campionari e non campionari, in particolare mancata partecipazione e under reporting.

Per comparare le due fonti BMN operano nel modo seguente:

- riconciliano la definizione settoriale, escludendo dai CF i dati relativi alle ISP;
- rielaborano i dati dei CF in accordo con i criteri di valutazione adottati nell'indagine;
- escludono le attività finanziarie rilevate in una sola delle due fonti. Riguardo a quest'ultimo aspetto, sono stati esclusi da un lato le partecipazioni in quasi-società e i debiti commerciali (rilevati solo dall'IBF) e dall'altro il TFR (rilevato solo dai CF). Per le gestioni patrimoniali, che non hanno una controparte nella classificazione dei CF, i dati dell'IBF sono stati riclassificati sulla base della composizione del portafoglio delle gestioni⁹⁴.

I risultati di BMN indicano che l'importanza delle ISP in termini di attività e passività finanziarie è marginale: il loro peso sul totale (Famiglie e ISP) è inferiore al 2 per cento quando si prende in esame il totale delle attività finanziarie; la quota è leggermente superiore per i depositi (circa 3 per cento) e le passività finanziarie (circa 5 per cento).

Quanto al secondo aspetto, il criterio generale seguito nei CF è la registrazione al valore di mercato (con l'eccezione dei depositi e prestiti) e in base al principio della competenza economica (in contrapposizione alla registrazione per cassa). Entrambi i fattori contribuiscono a una differente valutazione nei CF rispetto all'IBF⁹⁵. Per uniformare la valutazione delle attività e delle passività

⁹⁴ Gli strumenti esclusi, perché non rilevati in entrambe le fonti, hanno un peso relativamente contenuto sul totale della ricchezza finanziaria, con l'eccezione delle partecipazioni in quasi-società. Nel 2002, sul totale delle attività finanziarie dell'indagine, i debiti commerciali rappresentavano lo 0,7 per cento. Fornire una stima del peso delle partecipazioni in quasi-società risulta invece più complicato, per la mancanza nell'indagine di alcune informazioni cruciali quali il valore del TFR e la forma giuridica delle imprese familiari (si veda Bonci, Marchese, Neri, 2005). In modo approssimato, stimando un'incidenza media del TFR a partire dagli archivi della Centrale dei Bilanci e includendo fra le quasi-società solo le imprese familiari con oltre 5 dipendenti, si giunge comunque a un peso di circa il 20 per cento. Sul totale delle attività finanziarie dei CF il TFR pesa per meno del 5 per cento.

⁹⁵ La valutazione al mercato fa sì che il valore degli strumenti finanziari tenga conto delle variazioni di prezzo dovute, tra gli altri, ai movimenti dei tassi d'interesse (in particolare per i titoli di Stato e le obbligazioni). La registrazione per competenza comporta la contabilizzazione degli interessi nel momento della loro maturazione, sotto forma di reinvestimento nel relativo strumento finanziario.

finanziarie disponibili in entrambe le due fonti, BMN agiscono dal lato dei CF (i prestiti, ad esempio, sono ricondotti al valore nominale detraendo i ratei di interesse). Gli importi degli strumenti finanziari vengono pertanto confrontati adottando i criteri di valutazione (in certi casi implicitamente) utilizzati nell'IBF⁹⁶. Il peso delle correzioni per le voci analizzate, stimato come rapporto fra valore della correzione e valore dello strumento prima della correzione, è dell'ordine del 9 per cento.

I risultati del confronto per i singoli strumenti è illustrato nel seguito.

I depositi e prestiti – Comprendono i depositi bancari, i certificati di deposito, le operazioni di pronti contro termine, i libretti postali e i buoni fruttiferi postali. La stima dei CF è basata su informazioni settorizzate provenienti dalle segnalazioni statistiche di vigilanza: è pertanto ragionevole pensare che in questo caso gli errori di misura siano poco significativi.

A partire dal 1998, la stima campionaria dei depositi bancari rappresenta circa il 60 per cento della corrispondente stima aggregata. I depositi postali stimati dall'indagine sono circa un terzo di quelli dei CF; la discrepanza aumenta per i buoni fruttiferi postali.

I titoli di Stato – Tale voce comprende i titoli a breve e a medio/lungo termine emessi dalle Amministrazioni centrali (BOT, CCT, BTP, CTZ, ecc.) o dalle Amministrazioni locali. La stima delle consistenze di titoli di Stato detenuti dal settore delle famiglie risente, nei CF, di una componente stimata in modo residuale. Nel confronto con l'indagine, dal 1995 al 1998 il rapporto fra stime campionarie e CF oscilla in media attorno a circa il 40 per cento. I BOT presentano la minore distanza fra i due criteri di valutazione e nel 2000 sono stimati dall'indagine in valore quasi doppio a quello dei CF. Questo risultato deriva probabilmente dalla tendenza delle famiglie a classificare sotto la voce BOT anche altri titoli di Stato.

La distanza fra le due stime aumenta per gli strumenti con un minor grado di diffusione fra le famiglie e in particolare per i BTP e CCT.

Le obbligazioni e i fondi comuni – La stima delle obbligazioni possedute dalle famiglie risente, nei CF, di una componente calcolata in maniera residuale (che si somma all'ammontare detenuto a custodia presso le banche). Le differenze con l'IBF sono particolarmente rilevanti: le stime campionarie sono pari a circa il 16 per cento delle stime macro.

Il valore aggregato delle quote di fondi comuni di pertinenza del settore Famiglie è invece basato su informazioni di vigilanza prodotte dalle società di gestione dei fondi. Anche in questo caso le distanze fra le due fonti sono significative. In media i valori di natura campionaria rappresentano circa il 28 per cento delle stime macro.

⁹⁶ Per il valore delle azioni e delle quote di fondi comuni, ad esempio, si è ipotizzato che nell'indagine campionaria la valutazione sia ai prezzi di mercato (come nei CF): in questo caso non è stato dunque necessario porre in atto correttivi alle due fonti.

Le azioni e altre partecipazioni – Nei CF la stima delle azioni e altre partecipazioni è ottenuta a partire dai dati disponibili sulle imprese di capitali contenuti nell'archivio Cerved. Al momento, dunque, esistono stime macro solo per il valore di imprese in forma giuridica di società di capitali; mancano invece stime per il valore delle quasi-società⁹⁷. Il valore attribuito al settore delle Famiglie è ottenuto in modo residuale, sottraendo al totale in circolazione le quantità attribuite ai rimanenti settori istituzionali, per i quali sono invece disponibili informazioni dirette.

Dal lato dell'IBF gli investimenti in azioni e altre partecipazioni sono distinti (seguendo la terminologia utilizzata nella compilazione della Bilancia dei Pagamenti) in investimenti diretti costituiti dalle partecipazioni in società nelle quali la famiglia detiene una quota di controllo o che comunque permetta di svolgere un ruolo attivo di gestione e investimenti indiretti che comprendono le restanti partecipazioni che costituiscono il portafoglio finanziario delle famiglie.

Per ragioni di comparabilità con i CF, la stima campionaria è costituita dal valore degli investimenti in azioni (di società quotate e non) e in partecipazioni in società a responsabilità limitata che le famiglie detengono come investimenti indiretti (di cui alla sezione C del questionario) e dal valore delle partecipazioni in società di capitali per la quota di competenza (di cui all'allegato B4 del questionario). Sono quindi escluse le partecipazioni in quasi-società.

In media, nel periodo considerato, le stime campionarie costituiscono il 23 per cento delle corrispondenti stime aggregate. Le differenze risultano minori per le partecipazioni in società quotate, che in media nel periodo sono circa il 30 per cento dei CF.

I titoli esteri – La voce include i titoli di stato e le obbligazioni emesse da Amministrazioni estere, le azioni e partecipazioni in società estere e gli altri titoli emessi dal Resto del mondo. Le statistiche macro sono principalmente basate su informazioni settorizzate fornite dall'Ufficio Italiano dei Cambi (oggi incorporato nella Banca d'Italia) e le stime campionarie coprono circa il 3,3 per cento dei valori aggregati, la maggiore distanza fra gli strumenti finanziari.

Le assicurazioni vita e i fondi pensione – La voce comprende il montante accumulato dalle famiglie per investimenti in polizze assicurative e fondi pensione (come detto, si è escluso il TFR dall'analisi per mancanza del dato nell'indagine). Il dato dei CF è basato sulle statistiche ottenute dai bilanci delle compagnie di assicurazione e dei fondi pensione (forniti dall'ISVAP, oggi IVASS, e dalla COVIP) e integrate da informazioni Istat.

In media, nel periodo considerato, la stima campionaria è circa pari al 60 per cento della stima macro. Il divario fra le due stime mostra una tendenza a crescere nel tempo. Per tali strumenti le stime campionarie sembrano negli anni allontanarsi sempre di più da quelle aggregate. Come notano BMN, quindi, potrebbe dunque essere opportuna una più chiara individuazione nel questionario del tipo di polizze e/o di piani integrativi disponibili sul mercato.

⁹⁷ Esperimenti di stima di questa componente sono stati condotti da Rodano e Signorini (2008).

I crediti commerciali – Le stime dei CF, ottenute aggregando i dati dell'archivio Cerved sui bilanci delle società di capitali, risentono di diversi problemi: (i) non sono incluse le quasi-società non finanziarie (organismi senza personalità giuridica che redigono una contabilità completa e il cui comportamento economico e finanziario si differenzia da quello dei proprietari); (ii) l'archivio Cerved non riporta informazioni sui crediti e debiti commerciali per circa il 60 per cento delle imprese che redigono il bilancio in forma abbreviata; (iii) mancano inoltre i rapporti con gli altri settori istituzionali, primo tra tutti il settore delle Amministrazioni pubbliche. Ai problemi elencati si somma la natura residuale dell'ammontare dei crediti commerciali detenuti dalle famiglie.

La corrispondente stima campionaria, d'altra parte, è ottenuta dai crediti commerciali dichiarati dalle famiglie produttrici che, sulla base della definizione di CN, rientrano nel settore delle famiglie (liberi professionisti, imprenditori individuali, imprese familiari fino a cinque dipendenti).

Nel periodo in esame, la stima campionaria risulta pari ad oltre 3 volte quella dei CF riferita alle famiglie produttrici. Di conseguenza il peso sul totale attività finanziarie è stimato in maniera superiore dall'indagine (nei CF è invece prossimo a zero).

I prestiti alle cooperative – La stima aggregata dei prestiti effettuati dalle famiglie alle società cooperative è basata sugli archivi della Cerved e della Centrale dei Bilanci. A partire da quest'ultimo archivio è stata stimata, per classi di indebitamento, l'incidenza media dei prestiti da soci/azionisti sul totale dei prestiti consolidati (per le cooperative). Le stime ottenute sono state poi applicate alle società cooperative di Cerved (che contiene l'universo delle imprese in esame). Il confronto con i dati dell'IBF mostra un sostanziale allineamento per il 1995 e 1998. Per il 2002 invece la stima campionaria costituisce circa la metà della corrispondente stima aggregata.

Le passività finanziarie – Sono considerati i prestiti sia a breve sia a medio e lungo termine concessi alle famiglie dalle istituzioni finanziarie (banche, assicurazioni e altre società finanziarie), escludendo le voci relative ai prestiti da altri soggetti appartenenti al settore delle famiglie, quali parenti e amici, poiché il fenomeno non è rilevato nei CF.

Dal lato dei CF, la stima è ottenuta a partire dalle segnalazioni di vigilanza, che indicano in modo esplicito il settore di destinazione dei finanziamenti. La stima campionaria è pari a circa il 45 per cento di quella aggregata.

Un più recente confronto tra le stime dell'indagine della ricchezza e quelle aggregate è riportato nella Tavola 4.7 (Banca d'Italia, 2012). Secondo questi dati, la componente reale appare sottostimata di circa il 6 per cento nel 2010; la sottostima è maggiore per le componenti attive e passive della ricchezza finanziaria.

Tavola 4.7

**Confronto tra le stime macroeconomiche
e risultati dell'indagine: conto del patrimonio**
(valori percentuali)

Voci	Rapporto fra stima campionaria e aggregata	
	2008	2010
Abitazioni.....	97,8	101,2
Terreni.....	53,1	61,3
Stock di capitale ⁽¹⁾ , scorte e avviamento	63,0	43,4
Totale attività reali.....	92,3	93,6
Depositi bancari e postali	22,3	21,5
Titoli di stato	27,7	44,4
Obbligazioni,azioni, fondi comuni.....	24,4	46,0
Totale attività finanziarie	23,8	33,8
Passività finanziarie ⁽²⁾	37,6	34,6
Totale ricchezza netta	72,2	78,3

(1) Lo stock di capitale comprende i fabbricati non residenziali.

(2) Include tutti i debiti verso banche e altre istituzioni finanziarie.

Lo studio dei comportamenti di reticenza sulle attività finanziarie

Le discrepanze tra i dati dell'IBF e quelli dei CF possono dipendere da vari fattori: non risposte, non reporting, under-reporting. Alcuni lavori hanno cercato nel corso del tempo di approfondire il peso di queste componenti sul divario.

Il primo studio sul tema dell'under-reporting nell'IBF viene svolto a metà anni sessanta da Ulizzi (1967), che condusse una ricerca basata su 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 non risposta totale 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 realmente posseduti). 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. Tra i restanti intervistati, buona parte dichiarava un valore inferiore a quello vero o si rifiutava di rispondere. I fenomeni di under-reporting e non-reporting risultavano più marcati tra le famiglie più ricche. Questi comportamenti sono ancora oggi diffusi e costituiscono il principale problema di qualità delle indagini sul reddito e sulla ricchezza.

Nel lavoro di Cannari, D'Alessio, Raimondi e Rinaldi (1990) si svolgeva un esercizio di matching statistico tra i dati rilevati nell'IBF e quelli ricavati da una specifica rilevazione svolta dalla BNL sui propri clienti. L'assunzione era che l'indagine della BNL poteva considerarsi esente da under-reporting a causa del rapporto di fiducia con la propria clientela; l'ampia diffusione della banca sul territorio assicurava inoltre una buona rappresentatività della popolazione dei clienti bancari.

L'esperimento mise in luce che, a parità di caratteristiche socio-demografiche, i rispondenti dell'IBF dichiaravano assai meno attività finanziarie di quante ne dichiarassero i clienti della BNL. In particolare, gli autori illustrano come il comportamento di reticenza sia articolato in diversi stadi.

Vi è una quota di famiglie che è totalmente reticente, non dichiarando alcun rapporto con le banche o la posta e quindi nessuna attività finanziaria. La sottostima delle attività finanziarie attribuibile a tale comportamento non è elevata nel complesso, risultando in media intorno al 5 per cento; essa tuttavia è relativamente più consistente per le famiglie più povere e con basso titolo di studio.

La sottostima delle attività finanziarie attribuibile a comportamenti di non reporting, ovvero di mancata dichiarazione dei titoli posseduti, è pari a un ulteriore 10 per cento⁹⁸.

La parte più consistente della sottostima è tuttavia attribuibile alla dichiarazione degli importi posseduti; in media, per ogni 100 di attività finanziarie dichiarate le famiglie ne possiederebbero effettivamente circa 170.

Complessivamente, le tre fasi di correzione conducono a una rivalutazione media delle attività finanziarie pari al 94,4 per cento. La rivalutazione risulta crescente al crescere dell'età e decrescente al crescere del livello di istruzione e del reddito familiare; è particolarmente rilevante per le famiglie con capofamiglia in condizione non professionale (173,9 per cento) o lavoratore autonomo (116 per cento).

Questi risultati confermano quelli di Ulizzi (1967) secondo cui passando dalle classi sociali inferiori alle superiori le informazioni sui portafogli risultano errate in misura percentualmente decrescente.

Va rimarcato che, anche a seguito delle correzioni adottate dagli autori, i valori medi rimangono sensibilmente inferiori a quelli che si desumono dividendo i dati dei CF per il numero delle famiglie. Come è emerso da studi successivi, pesano le diverse definizioni degli insiemi di riferimento, gli errori di misura che caratterizzano anche i dati macroeconomici (per esempio per le poste calcolate a residuo), la presenza di un ridotto strato di famiglie molto ricche, che non sono rappresentate adeguatamente da indagini campionarie ma che posseggono grandi ammontari.

⁹⁸ Tale comportamento veniva valutato nel lavoro con riferimento ai soli titoli di Stato, che allora rappresentavano la parte preponderante delle attività finanziarie possedute dalle famiglie.

Il lavoro di Cannari e D'Alessio (1993) svolgeva un esperimento di matching a partire dagli stessi dati della BNL ma con una tecnica diversa. In particolare la correzione veniva effettuata mediante imputazione, e non più mediante riponderazione delle unità come nel lavoro precedente; ciò aveva il vantaggio di condurre a nuove stime delle attività finanziarie senza che questa procedura avesse impatto sulle distribuzioni delle restanti variabili. I risultati sulle entità delle rivalutazioni, essendo basati sugli stessi dati di base, confermavano sostanzialmente quelli del lavoro precedente. Il lavoro mostrava inoltre che l'indice di concentrazione di Gini risultava solo moderatamente modificato dalla correzione (per il 1991, da 0,644 a 0,635); se alla correzione tramite modello veniva poi fatto seguire un aggiustamento proporzionale dei valori tale da ricondursi agli importi totali della CF l'indice di Gini risultava pressoché identico a quello dei dati non aggiustati (0,649). Questo risultato indicava che gli errori presenti nei dati, da questo punto di vista, tendevano complessivamente a compensarsi.

In occasione dell'indagine sul 1998 era stato progettato uno studio fondato su un matching delle informazioni di fonte indagine e di fonte bancaria. Era stato appositamente sviluppato un protocollo che consentiva di utilizzare le informazioni presenti negli archivi di un'importante banca commerciale e di porle a confronto con quelle dichiarate nell'indagine, senza violare la privacy dei rispondenti.

I dati bancari estratti tuttavia risultarono non di buona qualità in quanto la procedura che li aveva generati conteneva alcuni errori; d'altra parte anche l'indagine aveva avuto alcuni problemi con la rilevazione quell'anno; ciò suggerì di limitare l'analisi (riportata in D'Alessio e Faiella, 2002) alla non-response, per la quale i dati di cui si disponeva erano sufficientemente accurati.

Il più recente lavoro sull'under-reporting delle attività finanziarie, che propone anche stime corrette che tengono conto di questo fenomeno, è quello di D'Aurizio, Faiella, Iezzi e Neri (2006). Rispetto ai lavori precedenti questo si caratterizza per una maggiore finezza negli strumenti finanziari esaminati e per la diversa metodologia impiegata. I dati dell'IBF sono confrontati con quelli, anonimi, messi a disposizione dal gruppo Unicredito.

La media delle attività finanziarie corrette per l'under-reporting è pari a più del doppio rispetto ai dati non aggiustati, raggiungendo l'85 per cento della stima dei CF. La maggiore correzione si riscontra per le obbligazioni e per i fondi comuni. L'intensità della correzione risulta relativamente superiore per le famiglie con un solo componente e cresce con l'età del capofamiglia. Aumenta inoltre per i capofamiglia con basso titolo di studio o in condizione non professionale.

In sintesi, i lavori effettuati sulle componenti della discrepanza tra l'indagine e i CF individuano la compresenza di fenomeni di non risposta, di non reporting e di under reporting. La sottostima delle attività e passività finanziarie determinata da comportamenti di non risposta totale sembra quantitativamente meno importante di quella che deriva dalle restanti componenti. Insomma, la ricerca più recente pare confermare quanto ottenuto da Ulizzi nelle prime indagini.

Le attività finanziarie: confronto con le segnalazioni di Vigilanza

In questo paragrafo si riportano alcuni confronti per classi di importo tra i valori delle attività finanziarie e dei debiti rilevati nell'IBF sul 2006 e i corrispondenti valori tratti dalle Segnalazioni di vigilanza. Poiché queste ultime sono una delle più importanti fonti di informazione per la costruzione delle stime macroeconomiche, il confronto effettuato è sostanzialmente tra dati micro e dati macro; a nostra conoscenza un confronto del genere è del tutto innovativo nel panorama delle statistiche italiane sulla ricchezza finanziaria delle famiglie.

Per i depositi bancari l'indagine rileva il numero dei conti e l'importo in essi detenuto complessivamente dalla famiglia. Per effettuare la comparazione con le segnalazioni di vigilanza, relative ai conti e non alle famiglie, è necessario fare alcune ipotesi. In particolare si considerano i due casi estremi: nel primo, l'importo detenuto dalla famiglia è suddiviso in parti uguali tra tutti i conti della famiglia; nella seconda, tutto l'importo viene attribuito a un conto principale, mentre agli altri viene attribuito un saldo pari a zero.

Il confronto tra le due fonti sul numero dei conti di deposito segnala un chiaro comportamento di reticenza delle famiglie. Nell'IBF il numero di conti stimato è pari a 23 milioni, mentre nelle segnalazioni di vigilanza supera i 36 milioni. È peraltro possibile che un certo numero di questi conti posseduti dalle famiglie e non dichiarati siano di importo modesto o nullo.

Il confronto tra le distribuzioni (figura 4.7) mostra che le stime campionarie sono leggermente sbilanciate verso gli importi di minore dimensione a scapito di quelli medio alti. I depositi di dimensione più elevata non mostrano invece differenze di particolare rilievo tra le due fonti nell'ipotesi di concentrazione degli importi in un unico conto.

Nel complesso, larga parte della differenza tra i dati dell'indagine e quelli di vigilanza sembra derivare dal minor numero dei conti di deposito nei dati IBF.

Per i titoli, l'indagine non rileva se e in che misura questi siano ripartiti in più conti di custodia; la ricostruzione della distribuzione per classi di importo è quindi effettuata in base a due ipotesi. Nella prima, l'importo totale detenuto in titoli dalla famiglia è attribuito a un unico conto; nel secondo, in presenza di un rapporto di clientela con più banche, l'importo totale della famiglia è suddiviso in tanti conti di custodia, di uguale importo, quanti sono questi rapporti. Nel primo caso si ottengono 5 milioni di conti di custodia titoli, nel secondo circa 7 milioni. Entrambi questi valori sono assai inferiori ai numeri di conti bancari di custodia titoli delle famiglie che risultano dalle segnalazioni di vigilanza (circa 18,6 milioni nel 2006).

La distribuzione dell'ammontare detenuto per classi di importo complessivo del conto ottenuta dall'indagine nella prima delle due ipotesi tende a sovrastimare l'importanza dei conti di importo modesto nel confronto con le segnalazioni di vigilanza. I conti di importo elevato sono invece significativamente sottostimati dall'indagine. La sottostima delle classi di importo elevato tende ovviamente ad aumentare adottando l'ipotesi di ripartizione dei valori detenuti tra le diverse banche (figura 4.8).

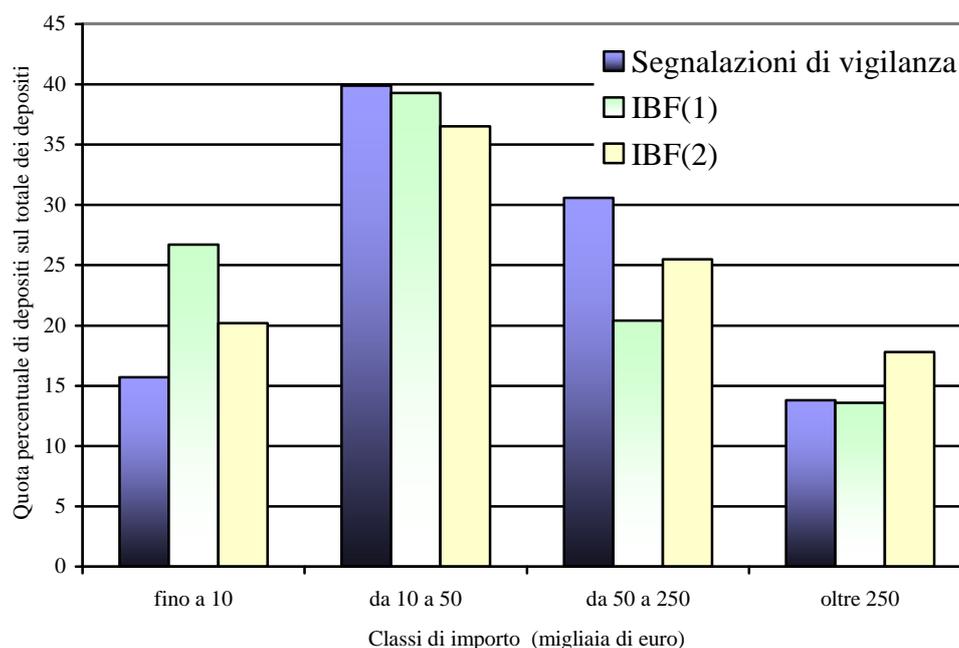
Indicazioni simili si traggono dai dati sull'indebitamento, largamente sottostimato dall'indagine nelle fasce di importo elevato (figura 4.9). Anche in questo caso, l'adozione di ipotesi di suddivisione dell'ammontare totale del debito in più rapporti con istituzioni finanziarie tenderebbe ad accrescere la distanza tra le due distribuzioni.

Nel complesso l'indicazione che si trae da questi dati è che una parte considerevole della discrepanza tra l'indagine e le stime macroeconomiche deriva dalla sottostima campionaria del numero di soggetti con attività e/o passività finanziarie; questa sottostima può essere dovuta sia a comportamenti di non reporting (ovvero famiglie che, pur possedendo strumenti, dichiarano di non possederli) sia a una minore probabilità di collaborazione delle famiglie con attività finanziarie e/o debiti. La parte restante della discrepanza può derivare sia da fenomeni di under-reporting più accentuati per le famiglie più ricche o più indebitate, sia da una probabilità di collaborare che tende a diminuire al crescere della ricchezza o dell'indebitamento.

Figura 4.7

Distribuzione dei depositi per classi di importo, 2006

(valori percentuali)

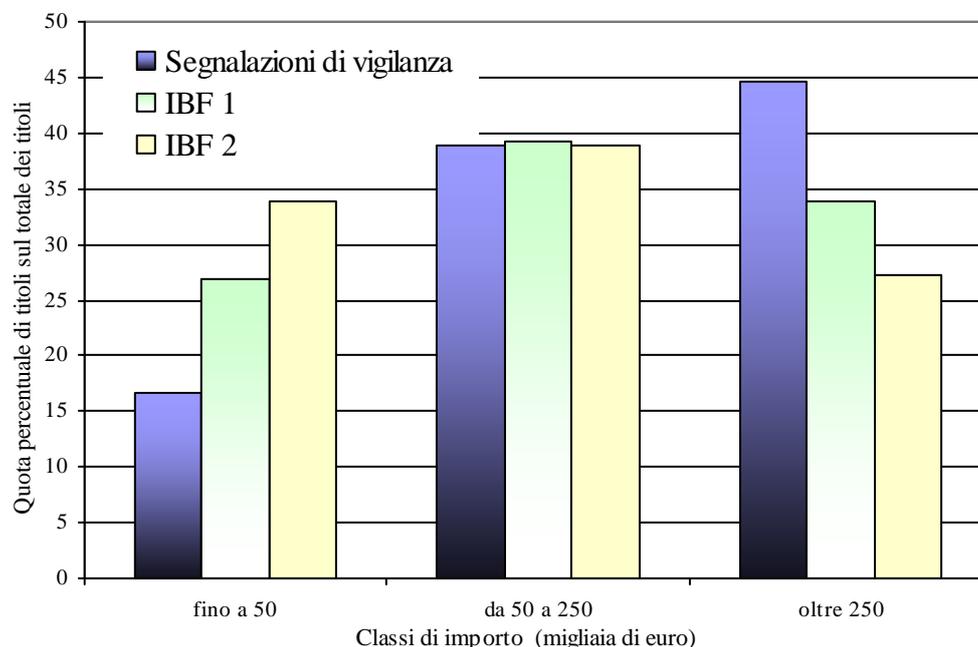


Fonte: elaborazioni su dati Banca d'Italia, Segnalazioni di vigilanza e IBF, 2006.

(1) Nell'ipotesi di equi-ripartizione dei depositi nei conti della famiglia. - (2) Nell'ipotesi che tutto sia detenuto in un unico conto per famiglia.

Figura 4.8

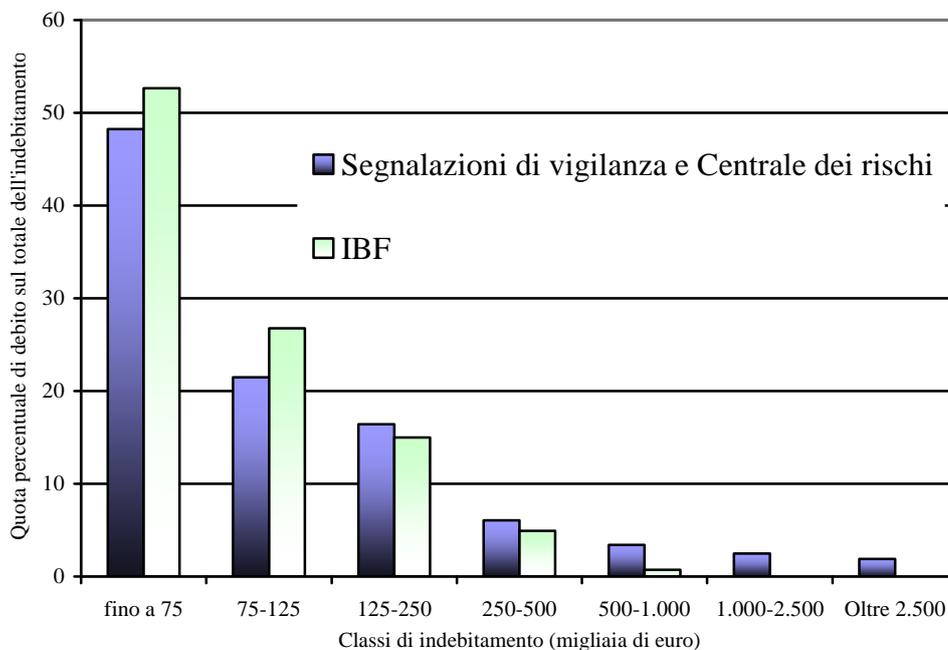
Distribuzione dei titoli in custodia per classi di importo
(valori percentuali)



Fonte: elaborazioni su dati Banca d'Italia, Segnalazioni di vigilanza e IBF, 2006.
Nell'ipotesi che nell'IBF tutti i titoli siano detenuti in un unico conto per famiglia.

Figura 4.9

**Indebitamento delle famiglie (consumatrici e produttrici):
confronto tra segnalazioni di vigilanza e IBF**
(quote percentuali di debito per classi di indebitamento)



Fonte: Elaborazioni su dati IBF, Segnalazioni di vigilanza e Centrale dei rischi. Indebitamento delle famiglie nei confronti di banche e intermediari finanziari di cui all'art. 107 del T.U.. I debiti comprendono le sofferenze e gli effetti insoluti e al protesto. Dati relativi al dicembre 2004. – (1) Segnalazioni di vigilanza e Centrale dei rischi. – (2) Indagine sui bilanci delle famiglie.

Il contante e le monete

Nell'IBF per molti anni è stata presente una domanda che richiedeva alle famiglie di indicare la scorta di contante mediamente detenuta. Ovviamente si trattava di un dato indicativo, utile a rendere conto principalmente del progredire dell'uso di strumenti di pagamento più evoluti, come le carte di credito e il bancomat.

Secondo l'indagine, tra il 1995 e il 2004, la cifra detenuta in media dalle famiglie oscillava tra i 350 e i 400 euro; la comparazione con la CF (possibile per il 1996) mostra un netto divario tra le due stime ed è probabile che il dato dei conti finanziari relativo alle famiglie (2 mila euro nel 1996) sia ampiamente sopravvalutato, come accade anche in altri paesi⁹⁹.

Nel contante sono incluse le monete, sebbene queste ultime ne rappresentino una quota – in valore – modesta. Le monete furono oggetto di specifica analisi in occasione del passaggio all'euro, quando fu necessario procedere a una valutazione dei pezzi in circolazione per orientare la produzione delle nuove monete. Poiché le monete, al contrario delle banconote, non sono oggetto di ritiro da parte delle banche, tutto ciò che viene emesso dovrebbe essere in circolazione. Nello schema dei conti finanziari, per convenzione, il valore delle monete circolanti viene attribuito interamente alle famiglie. Sulla base del criterio enunciato, ogni famiglia avrebbe dovuto avere in media circa 100 mila lire in monete, un valore scarsamente plausibile. Studi condotti all'epoca mostrarono che in effetti non si teneva conto del fatto che le banche e altri operatori possedevano un certo ammontare di pezzi "dormienti", nonché dei fattori di decadimento come i pezzi che i turisti portano nel loro paese, gli smarrimenti e le distruzioni, i possibili utilizzi industriali illegittimi. Si tratta di un caso particolare, oltretutto su una voce quantitativamente poco rilevante del bilancio familiare. Esso tuttavia è esemplificativo delle difficoltà di stima che possono caratterizzare anche i dati macroeconomici.

Il giudizio degli intervistatori sulla veridicità delle risposte

La reticenza delle famiglie a dichiarare le proprie fonti di reddito o le forme di attività finanziarie o reali possedute influisce in modo significativo sulla qualità delle stime. Sebbene la partecipazione all'indagine sia volontaria e il contenuto della rilevazione noto all'intervistato sin dall'inizio, alle domande più "delicate", relative al reddito o alla ricchezza, le famiglie spesso non rispondono in modo veritiero.

Per disporre di elementi di giudizio su tale fenomeno, che per sua natura risulta difficilmente investigabile, viene richiesto agli intervistatori di formulare, immediatamente dopo l'intervista, un giudizio sintetico 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

⁹⁹ Cfr. Sprenkle (1993).

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 4.8): 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.

Tavola 4.8

Veridicità delle risposte sul reddito e la ricchezza, 1993-2014
(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

Anno	Punteggio in decimi sulla veridicità delle risposte (secondo l'intervistatore)											Totale	Media
	1	2	3	4	5	6	7	8	9	10			
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	
2014...	0,2	0,4	0,5	1,4	4,7	10,2	15,2	20,1	24,9	22,5	100,0	8,1	

Ciò nonostante, il livello di attendibilità non è omogeneo all'interno del campione. Giudizi più elevati caratterizzano stabilmente le famiglie con capofamiglia con elevato titolo di studio, lavoratore dipendente, residente al Centro e al Nord. Un'attendibilità inferiore si riscontra per le famiglie con capofamiglia con modesto titolo di studio, lavoratore autonomo, residente al Sud e nelle Isole. L'attendibilità risulta inoltre crescente al crescere del reddito dichiarato nell'indagine (tavola 4.9)¹⁰⁰.

È interessante osservare che la correlazione tra indicatore di attendibilità e valori dichiarati di reddito, attività o passività finanziarie è in generale positiva, sebbene di entità molto modesta. I tentativi di utilizzare questa informazione per ricavare delle stime aggiustate per l'under-reporting danno luogo in generale a correzioni molto limitate.

¹⁰⁰ La relazione tra livello di attendibilità e reddito "vero" è ovviamente ignota.

Considerando solo il sottoinsieme di famiglie il cui livello di attendibilità nell'IBF 2014 è superiore alla sufficienza (circa il 90 per cento dell'intero campione) si ottiene una stima del reddito familiare superiore a quello relativa all'intero campione solo dell'1,5 per cento. La correzione è appena superiore (2,9 per cento) se si effettua la stima solo con le famiglie che hanno un punteggio di attendibilità superiore o uguale a 7 decimi. Per la ricchezza le variazioni nei due casi considerati sono egualmente limitate (1,3 e 3,6 per cento); per le attività finanziarie l'entità della correzione è maggiore (tra il 3,1 e il 7,6 per cento).

Un'interessante eccezione agli andamenti mostrati riguarda il reddito da attività autonoma. Per questa variabile si rileva una correlazione debolmente negativa con il grado di attendibilità: in altri termini coloro che hanno dichiarato redditi più bassi sono ritenuti più affidabili di quelli che hanno dichiarato redditi più elevati. Di conseguenza le stime sui sottocampioni delle famiglie giudicate meno reticenti conducono addirittura a stime inferiori a quelle originarie per una misura che varia tra il 3,8 e l'8,1 per cento.

Tavola 4.9

Giudizio dell'intervistatore e stima del reddito e della ricchezza, 2014

	Correlazione con il giudizio dell'intervistatore	Solo giudizio maggiore o uguale a 6 ⁽¹⁾	Solo giudizio maggiore o uguale a 7 ⁽¹⁾
Reddito da lavoro dipendente.....	0,1531	3,6	7,0
Reddito da lavoro autonomo.....	-0,0395	-3,8	-8,1
Reddito da trasferimenti.....	0,0247	0,7	1,0
Reddito da capitale.....	0,0954	1,3	3,4
Reddito familiare.....	0,1244	1,5	2,9
Attività reali.....	0,0645	1,0	3,1
Attività finanziarie.....	0,0745	3,1	7,6
Passività finanziarie.....	0,0620	1,0	3,5
Ricchezza familiare.....	0,0693	1,3	3,6
Numerosità campionaria (unità).....	8.156	7.574	6.742
Media del punteggio espresso dall'intervistatore.....	8,1	8,4	8,7

⁽¹⁾Variazione percentuale rispetto alla stima ottenuta con il totale del campione.

D'Alessio e Neri (2015) stimano il reddito familiare in funzione di una serie di caratteristiche e del giudizio dell'intervistatore sulla veridicità delle risposte fornite sul reddito e la ricchezza. Una volta individuato il contributo attribuibile a quest'ultima componente, è possibile stimare il reddito che la famiglia avrebbe dovuto dichiarare in presenza di un indice di veridicità pari al suo massimo (cioè un punteggio di 10).

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 (segnalando che la relazione tra giudizio degli intervistatori e reddito della specie prima evidenziata viene invertita in presenza di altre covariate) 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.

Stime micro coerenti con stime macro

Nei paragrafi precedenti si sono effettuati vari confronti tra stime micro e stime macro, in qualche caso cercando di interpretare i motivi che potessero spiegarne la distanza. Come si è detto, non sempre i motivi dei divari sono da attribuirsi a problemi nelle stime campionarie; anche le stime macro hanno infatti i loro problemi e non possono essere assunte come punti di riferimento privi di errore. Non si può negare, però, che se il dato micro fornisce il suo contributo di migliore qualità nell'analisi distributiva, tra le categorie di famiglie, il dato macro è in generale preferibile per quanto riguarda l'ordine di grandezza dei fenomeni e la loro dinamica nel tempo.

Questa considerazione suggerisce di approfondire il tema della riconciliazione micro-macro, in particolare sviluppando modelli che consentano di ottenere stime micro coerenti con le stime macro disponibili, in modo da inscrivere le informazioni distributive all'interno del set informativo disponibile da fonte macro.

In un recente lavoro D'Alessio e Neri (2015) mostrano come la combinazione di modelli di correzione costruiti sulla base di specifiche basi informative e tecniche di calibrazione possono essere utilizzate per ottenere stime campionarie coerenti con quelle di fonte macro.

Gli autori rilevano che, nella pratica applicazione di queste tecniche, è problematico identificare 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.

Nel lavoro citato, le correzioni danno nel complesso risultati che, in termini di profili, sono piuttosto simili tra di loro e rispetto ai dati di partenza; in altri termini, spesso i rapporti relativi tra le diverse categorie di famiglie non sono alterati in misura importante nei dati corretti. Diversi indizi lasciano però immaginare che le stime IBF sottostimino la concentrazione del reddito e della ricchezza.

Si tratta ancora di esperimenti che meritano di essere approfonditi in futuro. Ulteriori studi sembrano necessari, in particolare con riferimento all'effetto congiunto di più procedure di correzione, che attualmente sono impiegate secondo modalità sostanzialmente indipendenti. I risultati ottenuti fin qui sembrano comunque costituire un passo importante per la riconciliazione di tre fonti: l'IBF, la CN e i dati fiscali.

Altri errori di misura

Sebbene per un'indagine sul reddito e la ricchezza la mancata risposta e l'*under-reporting*, di cui si è già detto, costituiscano senz'altro i fenomeni che più di altri possono determinare problemi di qualità delle stime, numerosi altri fattori sono indicati in letteratura come potenziali elementi di disturbo.

Il questionario, in primo luogo, non è uno strumento neutro: il modo in cui le domande sono formulate, come sono ordinate e poste in relazione tra di loro, le modalità di risposta previste, sono tutti elementi in grado di influire sulle risposte degli intervistati, in particolare – ma non solo – nella rilevazione di opinioni, aspettative e altri elementi di natura soggettiva.

Un altro fattore importante è costituito dagli intervistatori, il cui atteggiamento o modo di illustrare le domande può influire sulla risposta degli intervistati.

Altri problemi possono derivare dai processi cognitivi del rispondente; al di là degli elementi intenzionali riconducibili ai fenomeni di *under-reporting*, le risposte a determinate domande sono influenzate dalla capacità di elaborazione (ad esempio per le domande ipotetiche) o di memoria (per le domande retrospettive) del rispondente. Vanno inoltre considerati gli elementi di conoscenza vera e propria, soprattutto quando, per questioni pratiche, è talvolta consentito a un soggetto di rispondere per conto di un altro. Aspetti più generali, come la motivazione del rispondente, il tempo e l'impegno a lui richiesto sono ulteriori elementi che sono in grado di influire sulla qualità delle risposte fornite¹⁰¹.

Sulla base di tali considerazioni va pertanto segnalato che le stime fornite presentano una fonte di variabilità aggiuntiva rispetto a quella campionaria.

L'analisi degli errori di misura viene svolta con una diversa metodologia a seconda che si abbia a che fare con variabili che variano nel tempo oppure no e a seconda della natura della variabile (Biancotti, D'Alessio e Neri, 2004).

Per le variabili *time-invariant*, come ad esempio il sesso, l'anno di nascita o l'età alla quale si è iniziato a lavorare, una valutazione sulla qualità del dato può essere svolta confrontando le risposte fornite dai soggetti panel in due indagini consecutive. Le misurazioni ottenute possono essere infatti considerate indipendenti, giacché si può ragionevolmente ipotizzare che le famiglie ricordino di aver partecipato all'indagine ma non ricordino le loro risposte.

In simili condizioni l'indice di correlazione tra le due misure è il quadrato dell'indice di affidabilità di Heise (1969)¹⁰², il quale esprime la quota di variabilità della variabile rilevata che origina dal fenomeno che si intende misurare (Lord e Novick, 1968).

Per le variabili qualitative, l'indice di affidabilità è la frazione di unità classificate coerentemente nelle due rilevazioni; talvolta in luogo di questa misura se ne usa una versione modificata, che tiene conto del fatto che due misure possono dare luogo alla stessa classificazione casualmente.

Nell'IBF, tutte le domande, anche quelle prive di ambiguità di definizione o di risposta, come ad esempio il sesso, il luogo e l'anno di nascita, presentano alcune incoerenze, con percentuali nei primi nell'ordine del 2-3 per cento e ormai

¹⁰¹ Si veda, ad esempio, Groves e Couper (1998).

¹⁰² L'indice di affidabilità di Heise può essere interpretato come il coefficiente di correlazione tra una misurazione con errore di una variabile e un'ipotetica misurazione corretta della stessa.

sotto l'1 per cento, come riflesso della maggiore attenzione prestata al fenomeno delle discrepanze. L'analisi dei casi errati evidenzia che le discrepanze si manifestano 3 volte su 4 per i figli in giovane età, soggetti che – a parte le caratteristiche demografiche – non sono più oggetto di analisi nel corso dell'intervista (tavola 4.10). L'indagine sul 2010 mostra livelli di coerenza delle risposte con quelle fornite nella rilevazione precedente inferiori rispetto al passato.

Le incoerenze aumentano quando la risposta fa riferimento a elementi che possono essere percepiti come delicati dagli intervistati o non perfettamente conosciuti da tutti i membri della famiglia, come ad esempio il tipo di diploma o il titolo di studio (tavola 4.11).

Tavola 4.10

Affidabilità delle risposte relative al sesso, luogo e data di nascita, 1989-2014
(percentuali)

Indagini	Sesso ⁽¹⁾	Anno di nascita ⁽¹⁾	Luogo di nascita ⁽¹⁾
1989-1991.....	98,2	95,5	98,4
1991-1993.....	98,1	96,9	98,4
1993-1995.....	99,7	98,8	98,2
1995-1998.....	99,7	98,8	97,5
1998-2000.....	99,9	98,3	97,3
2000-2002.....	99,4	99,1	98,4
2002-2004.....	99,7	99,2	98,7
2004-2006.....	99,7	98,2	97,7
2006-2008.....	99,9	99,1	98,1
2008-2010.....	96,7	93,8	96,0
2010-2012.....	99,8	98,4	96,7
2012-2014.....	99,8	98,9	98,6

(1) Quota delle risposte coerenti sul totale.

L'incoerenza è invece il segnale di una difficoltà oggettiva di individuare la risposta corretta (ad esempio, nel caso dell'età alla quale i soggetti hanno iniziato a lavorare) quando questa richiede uno sforzo di memoria o quando la domanda non specifica chiaramente come trattare alcune situazioni ambigue (ad esempio i periodi di apprendistato e il lavoro saltuario).

Gli errori di risposta possono manifestarsi più intensamente se la stessa domanda fa riferimento a concetti vaghi, come ad esempio quando si chiede se l'abitazione è ubicata nel "centro" o in "periferia". In questo caso l'indice di affidabilità (risposte coerenti) presenta un valore (nel 2006/2008) di 0,56, che si riduce a 0,4 tenendo conto delle casuali risposte coincidenti.

Tavola 4.11

Affidabilità della variabile “diploma”, 2006-2008

(percentuali)

2006	2008						Totale
	A	B	C	D	E	F	
A. Istituto professionale	4,9	4,2	0,5	0,1	0,4	0,4	10,5
B. Istituto tecnico	4,1	44,1	2,4	0,3	0,6	1,1	52,7
C. Liceo (classico, scientifico e linguistico)	0,7	1,8	15,6	0,3	0,4	0,1	19,0
D. Liceo artistico e istituti d'arte	0,1	0,1	0,2	2,0	0,2	0,0	2,6
E. Magistrali	0,5	0,5	0,4	0,0	11,8	0,1	13,3
F. Altro	0,3	0,5	0,4	0,0	0,2	0,4	1,9
Totale	10,6	51,3	19,5	2,7	13,7	2,2	100,0
Indice di affidabilità λ^2 (r. coerenti)	88,7	84,3	92,7	98,6	96,6	96,8	78,9

La valutazione dell'affidabilità di grandezze che variano nel tempo può essere effettuata facilmente utilizzando il metodo di Heise che, a condizione di disporre di almeno tre rilevazioni sulle stesse unità panel e sotto alcune ipotesi sulla dinamica temporale delle variabili, permette di separare l'effettiva variazione di ciascuna grandezza esaminata dal relativo errore di misura e ricavare una stima dell'affidabilità¹⁰³.

Con riferimento ai principali aggregati, l'indice calcolato per le terne di indagini consecutive, a partire da quella 1989-1991-1993 fino all'ultima disponibile 2010-2012-2014 (tavola 4.12), segnala una maggiore affidabilità dei valori rilevati per il reddito, con un valore medio dell'indice pari a 0,94, rispetto alla ricchezza netta e ai consumi (con un indice medio intorno a 0,83 e 0,87).

Tavola 4.12

Indice di affidabilità di Heise per le principali variabili, 1989-2014

Grandezze	1989	1991	1993	1995	1998	2000	2002	2004	2006	2008	2010	Media
	1991	1993	1995	1998	2000	2002	2004	2006	2008	2010	2012	
	1993	1995	1998	2000	2002	2004	2006	2008	2010	2012	2014	
Reddito disponibile netto	0,89	0,94	0,89	0,84	0,89	0,85	0,80	0,82	0,91	0,95	0,96	0,94
Reddito netto lavoro dipendente	0,97	0,98	0,96	0,94	0,94	0,94	0,91	0,95	0,96	0,97	0,96	0,96
Pensioni e trasferimenti netti ..	0,97	0,99	0,96	0,93	0,93	0,94	0,97	0,90	0,83	0,99	0,97	0,93
Reddito netto da lavoro autonomo	0,89	0,97	0,84	0,71	0,75	0,81	0,73	0,82	0,93	0,95	0,94	0,94
Reddito netto da capitale	0,82	0,79	0,81	0,77	0,79	0,78	0,74	0,81	0,92	0,89	0,91	0,91
Ricchezza netta	0,80	0,74	0,76	0,87	0,86	0,82	0,86	0,85	0,55	0,99	0,96	0,83
Attività reali	0,80	0,72	0,73	0,88	0,89	0,82	0,89	0,85	0,53	0,99	0,95	0,82
Attività finanziarie	0,66	0,81	0,93	0,68	0,46	0,62	0,46	0,61	0,86	0,9	0,83	0,86
Passività finanziarie	0,67	0,84	0,88	0,67	0,73	0,81	0,79	0,77	0,72	0,9	0,72	0,78
Consumi	0,85	0,81	0,79	0,74	0,82	0,77	0,76	0,86	0,85	0,88	0,88	0,87

103

Un indice di affidabilità elevato non implica che la stima non possa risentire di altri problemi, come ad esempio la distorsione derivante da *under-reporting*. È infatti sufficiente che esso sia sistematico nel corso del tempo perché non venga registrato dall'indice di affidabilità di Heise.

Negli ultimi anni, tuttavia, la tavola mostra indicatori ampiamente in crescita, probabilmente a causa del più incisivo trattamento di editing dei dati che, per la parte panel, tiene in conto delle informazioni ottenute nelle precedenti indagini. Ciò da un lato migliora l'aderenza tra le dichiarazioni delle famiglie in anni contigui; d'altra parte rende le misurazioni nel tempo non indipendenti. Pertanto l'indice di Heise non può più essere utilizzato per misurare l'affidabilità dei dati (almeno quelli per i quali il panel non è disponibile).

Può essere utile osservare che negli anni precedenti emergeva, ad esempio, che i redditi da pensioni e da lavoro dipendente erano caratterizzati da una maggiore affidabilità rispetto ai redditi da lavoro autonomo e da capitale. Sulle componenti di ricchezza, maggiore era l'affidabilità delle componenti reali rispetto a quelle finanziarie. È presumibile che, anche negli anni recenti, i dati raccolti (sulle unità non panel) siano caratterizzati da affidabilità di questo tipo.

Il ruolo dell'intervistatore

Nell'indagine si tiene traccia di molti aspetti operativi connessi alla rilevazione; si raccolgono ad esempio informazioni sulle caratteristiche degli intervistatori, sull'uso del CAPI, sulla presenza o meno di altri componenti della famiglia oltre il capofamiglia al momento dell'intervista, sulla data e ora dell'intervista e sulla sua durata.

Un fattore che influisce in modo rilevante sulla qualità dei dati raccolti in un'indagine è costituito dagli intervistatori. L'impegno, le caratteristiche e l'atteggiamento degli intervistatori possono influire sia sul tasso di partecipazione all'indagine sia sulle risposte fornite degli intervistati (tavola 4.13).

È quindi possibile effettuare una serie di considerazioni nel rapporto tra queste caratteristiche e i dati raccolti anche se non sono rispettati i criteri di randomizzazione che sarebbero necessari per condurre esperimenti controllati. Le conclusioni vanno dunque interpretate con cautela, tenendo presente che i modelli esplicativi potrebbero non tenere conto di tutti i fattori rilevanti.

In media, ormai oltre il 90 per cento delle rilevazioni avviene con metodologia CAPI (tavola 4.14). Il questionario elettronico è largamente prevalente nelle aree ad alta intensità di intervistatori di professione; ciò riflette sia la propensione della società di rilevazione nel fornire una dotazione costosa principalmente a collaboratori stabili, sia il minore interesse da parte dell'intervistatore occasionale ad apprendere il funzionamento dello strumento informatico. Nell'indagine sul 2014 il numero di intervistatori è inferiore rispetto al 2006; negli ultimi anni, infatti, si è scelto di selezionare un gruppo più ristretto di intervistatori, che sono anche più esperti, come si evince dal più elevato numero di anni di esperienza che si riscontra in media nell'ultima rilevazione, e più maturi.

Tavola 4.13

Caratteristiche degli intervistatori, 2006 e 2014

(valori percentuali, unità)

	2006				2014			
	Nord	Centro	Sud e Isole	Totale	Nord	Centro	Sud e Isole	Totale
Sesso								
Maschi.....	14,3	13,5	26,3	18,3	10,2	14,3	24,3	17,3
Femmine.....	85,7	86,5	73,7	81,7	89,8	85,7	75,7	82,7
Titolo di studio								
Media inferiore.....	18,1	10,8	3,9	11,9	10,2	5,7	4,1	6,5
Media superiore.....	64,8	62,2	75,0	67,9	76,3	68,6	79,7	76,2
Laurea e oltre.....	17,1	27,0	21,1	20,2	13,6	25,7	16,2	17,3
Classe di età								
fino a 30 anni.....	7,6	8,1	10,5	8,7	0,0	2,9	0,0	0,6
da 31 a 40 anni.....	12,4	27,0	28,9	20,6	8,5	17,1	16,2	13,7
da 41 a 50 anni.....	34,3	21,6	39,5	33,9	33,9	25,7	40,5	35,1
da 51 a 65 anni.....	40,0	43,2	19,7	33,5	40,7	37,1	41,9	40,5
oltre 65 anni.....	5,7	0,0	1,3	3,2	16,9	17,1	1,4	10,1
Attività intervistatore								
Non professionista.....	41,0	27,0	27,6	33,9	25,4	25,7	28,4	26,8
Professionista.....	59,0	73,0	72,4	66,1	74,6	74,3	71,6	73,2
Totale.....	100,0							
Numero Intervistatori	108	42	77	227	59	35	74	168
Anni di esperienza.....	10,8	12,0	11,9	11,4	15,3	16	18,2	16,7
Precedenti indagini Banca d'Italia.....	2,3	2,9	3,3	2,7	4,5	5,4	6,4	5,5

Tavola 4.14

Interviste CAPI per area geografica, 1998-2014

(valori percentuali, unità)

	Nord	Centro	Sud	Totale interviste CAPI
1998.....	60,8	64,7	75,7	67,3
2000.....	52,2	79,8	78,3	67,0
2002.....	54,4	63,8	80,0	64,9
2004.....	64,6	78,9	81,0	73,1
2006.....	74,7	77,8	87,5	79,4
2008.....	79,3	62,0	90,8	79,5
2010.....	85,0	68,7	93,2	84,4
2012.....	88,1	86,7	96,1	90,7
2014.....	91,6	91,1	95,6	92,9

Secondo Biancotti, D'Alessio e Neri (2008) le modalità operative e le caratteristiche degli intervistatori hanno un impatto limitato sulla probabilità di osservare variazioni anomale. Solo l'esperienza dell'intervistatore e la buona qualità del clima in cui si svolge l'intervista riducono la probabilità di osservare differenze elevate fra i redditi misurati nei due periodi.

Gli intervistatori professionisti hanno in media risultati migliori in termini di qualità delle risposte; anche l'aver collaborato a precedenti rilevazioni dell'indagine della Banca d'Italia ha effetti migliorativi.

L'esperienza maturata nell'ambito della stessa indagine dall'intervistatore ha un effetto positivo sull'accuratezza. Le ultime interviste di ogni rilevatore sono in media significativamente migliori delle precedenti, per maggiore dimestichezza con i contenuti e con le procedure.

Il rischio di commettere errori aumenta se l'intervistatore opera in una provincia diversa da quella di residenza. Tale risultato può dipendere da numerosi fattori. Le interviste realizzate in trasferta sono in media più brevi delle altre, controllando per il numero di componenti e di percettori di reddito; chi si sposta ha probabilmente maggiore fretta e questo può riflettersi sulla qualità dei dati raccolti.

Nel lavoro di Ilardi e Neri (2014) si mostra che le capacità dell'intervistatore di ottenere l'intervista e una partecipazione ampia in termini di disponibilità e accuratezza delle risposte, sono tra loro correlate. L'impiego di bravi intervistatori agisce dunque su entrambi questi fattori. Secondo le stime, però, il ruolo dell'intervistatore è maggiore nell'ottenere l'intervista (vale circa il 20 per cento della varianza) rispetto alla influenza sulla qualità dei dati (circa il 5 per cento).

5. Problemi di qualità legati a particolari utilizzi dei dati

Dati pooled, stime regionali o per piccoli domini

Le recenti evoluzioni del contesto amministrativo, con il conseguente affermarsi della dimensione regionale nell'analisi dei problemi, nella programmazione e nella valutazione delle politiche, accresce l'esigenza di disporre di dati a livello regionale. D'altra parte le informazioni tratte dall'IBF non possono essere direttamente utilizzate per stime regionali a causa della eccessiva variabilità delle stime campionarie per quel livello di disaggregazione.

Le argomentazioni che andiamo a svolgere in questo paragrafo con riferimento alle stime regionali possono essere replicate, mutatis mutandis, nel caso di stime campionarie relative a domini con numerosità limitata o per fenomeni relativamente rari nel campione complessivo.

La tavola 5.1 mostra in che modo i campioni di famiglie rilevate nel periodo 1993-2014 si ripartiscono nelle diverse regioni. La numerosità campionaria risulta più elevata (oltre le 500 unità) in Lombardia, Emilia-Romagna, Piemonte, Valle d'Aosta, Campania e Sicilia. Su livelli intermedi si colloca nelle altre regioni con l'eccezione del Molise e della Basilicata, dove il numero di famiglie campione è decisamente esiguo (quasi sempre inferiore a 100 unità).

Come mostrano Cannari e D'Alessio (2003), se si tiene conto dello schema di campionamento adottato, le stime campionarie del reddito e della ricchezza netta elaborate per i singoli anni presentano errori standard ragguardevoli. Oltre all'intrinseca variabilità del reddito e della ricchezza, va considerato che la stratificazione delle unità di primo stadio (comuni) non riesce a compensare la

perdita di precisione dovuta al disegno a due stadi, che riflette la circostanza che le famiglie residenti all'interno dello stesso comune presentano una relativa omogeneità. Il livello di precisione delle stime, inoltre, è inferiore per la ricchezza rispetto al reddito.

Tavola 5.1

Numerosità del campione IBF, 1993-2014

(numero di famiglie intervistate)

Regioni	Numerosità campionaria										
	1993	1995	1998	2000	2002	2004	2006	2008	2010	2012	2014
Piemonte – Val d'Aosta.....	680	662	577	757	802	769	776	834	757	726	768
Lombardia	820	824	820	860	905	843	873	844	803	895	944
Trentino Alto Adige	205	220	146	161	129	151	173	174	172	178	238
Veneto	435	476	351	439	521	578	596	601	512	536	499
Friuli Venezia Giulia	257	313	270	255	236	250	255	253	214	196	214
Liguria	407	386	306	316	372	371	340	315	311	300	347
Emilia Romagna	741	725	526	751	716	678	699	720	708	681	677
Toscana	592	589	477	598	559	641	620	607	615	636	605
Umbria	307	288	263	271	270	283	256	267	277	269	277
Marche	366	373	324	328	379	389	356	354	355	357	345
Lazio	431	411	460	425	457	425	343	413	452	458	452
Abruzzo	273	311	275	228	231	220	204	201	202	214	204
Molise	78	85	85	83	85	99	86	138	116	119	111
Campania	682	709	687	815	703	626	597	627	752	751	716
Puglia	591	520	463	471	413	450	415	451	454	516	453
Basilicata	95	127	89	95	78	126	128	128	126	128	128
Calabria	302	262	218	210	203	193	190	190	196	230	217
Sicilia	539	559	598	630	655	590	527	516	587	619	618
Sardegna	288	295	212	308	297	330	334	344	342	342	343
Totale Italia.....	8.089	8.135	7.147	8.001	8.011	8.012	7.768	7.977	7.951	8.151	8.156

Volendo ottenere informazioni a livello regionale sui dati IBF è possibile fare ricorso a un accorpamento delle osservazioni provenienti da indagini contigue in modo da effettuare stime con un maggior numero di osservazioni¹⁰⁴. Eventualmente si possono utilizzare anche metodi di stima robusti alla presenza di outliers oltre che informazioni di fonte esterna per aspetti inerenti la struttura regionale del campione.

L'accorpamento, ad esempio, di tre rilevazioni consente di ottenere campioni regionali di dimensione superiore alle 1.000 unità in gran parte delle regioni. Se, nonostante l'accorpamento, in alcune regioni la numerosità

¹⁰⁴ Si veda ad esempio Cannari, D'Alessio e Mori (2011).

campionaria continua a essere insufficiente, è possibile considerare più rilevazioni oppure accorpare alcune regioni, sulla base di un criterio di distanza geografica.

Tavola 5.2

Numerosità del campione IBF, 1993-2014

(numero di famiglie intervistate)

Regioni	Numerosità campionaria								
	1993-1995-1998	1995-1998-2000	1998-2000-2002	2000-2002-2004	2002-2004-2006	2004-2006-2008	2006-2008-2010	2008-2010-2012	2010-2012-2014
Piemonte – Val d'Aosta.....	1.919	1.996	2.136	2.328	2.347	2.379	2.367	2.317	2.251
Lombardia	2.464	2.504	2.585	2.608	2.621	2.560	2.520	2.542	2.642
Trentino Alto Adige	571	527	436	441	453	498	519	524	588
Veneto	1.262	1.266	1.311	1.538	1.695	1.775	1.709	1.649	1.547
Friuli Venezia Giulia	840	838	761	741	741	758	722	663	624
Liguria	1.099	1.008	994	1.059	1.083	1.026	966	926	958
Emilia Romagna	1.992	2.002	1.993	2.145	2.093	2.097	2.127	2.109	2.066
Toscana	1.658	1.664	1.634	1.798	1.820	1.868	1.842	1.858	1.856
Umbria	858	822	804	824	809	806	800	813	823
Marche	1.063	1.025	1.031	1.096	1.124	1.099	1.065	1.066	1.057
Lazio	1.302	1.296	1.342	1.307	1.225	1.181	1.208	1.323	1.362
Abruzzo	859	814	734	679	655	625	607	617	620
Molise	248	253	253	267	270	323	340	373	346
Campania	2.078	2.211	2.205	2.144	1.926	1.850	1.976	2.130	2.219
Puglia	1.574	1.454	1.347	1.334	1.278	1.316	1.320	1.421	1.423
Basilicata	311	311	262	299	332	382	382	382	382
Calabria	782	690	631	606	586	573	576	616	643
Sicilia	1.696	1.787	1.883	1.875	1.772	1.633	1.630	1.722	1.824
Sardegna.....	795	815	817	935	961	1.008	1.020	1.028	1.027
Totale Italia.....	23.371	23.283	23.159	24.024	23.791	23.757	23.696	24.079	24.258

Limitando l'attenzione a una generica regione, e omettendo il corrispondente indice per semplicità, le stime relative alla media della variabile y_{it} (dove i rappresenta l' i -esima osservazione della t -esima indagine) nel periodo che include T indagini di numerosità n_t ($t=1, \dots, T$) possono essere ricavate come media semplice delle stime annuali nell'intervallo considerato¹⁰⁵:

$$Y = \sum_{it} (y_{it} / n_t) / T \quad \text{per } i = 1, \dots, n_t \text{ e } t=1, \dots, T$$

Lo stimatore, combinazione lineare di stimatori corretti, è esso stesso corretto e ottimale, in termini di errore standard, quando le stime annuali sono tra loro indipendenti. Questa condizione, considerando le indagini successive al 1987, non è soddisfatta. Una parte delle famiglie del campione viene infatti reintervistata da un'indagine all'altra, per cui la varianza di Y , $V(Y)$, risente della correlazione tra le risposte che le stesse famiglie hanno fornito in più rilevazioni. Per il reddito e la ricchezza questa correlazione è positiva e dell'ordine di 0,5-0,6;

¹⁰⁵ I valori nei vari anni possono essere rapportati alla media nazionale di ciascun anno, in modo da ottenere indici che non risentono, tra l'altro, della variazione dei prezzi.

V(Y) è dunque maggiore rispetto a una situazione in cui i campioni annuali sono indipendenti.

In presenza di una correlazione non nulla tra i fenomeni osservati in due indagini contigue, lo stimatore di cui sopra presenta un'efficienza inferiore a quella di stimatori alternativi che di quella correlazione tengono conto.

Nel caso di due soli periodi (T=2), con campioni di uguale numerosità (n), caratterizzati da una quota di unità reintervistate pari a p, Cannari e D'Alessio (2003) mostrano che lo stimatore ottimo si ottiene assegnando a ciascuna unità del panel un peso proporzionale a $1/(1+\rho)$. Si tratta quindi di un'operazione semplice dal punto di vista computazionale¹⁰⁶.

Quando le indagini considerate sono più di due, la quota di unità panel è variabile nel tempo e alcune unità sono intervistate in più di due periodi, diviene più complesso individuare, per via analitica, i pesi che minimizzano la varianza della media riferita all'intero periodo. La soluzione può essere ottenuta per via numerica, con algoritmi iterativi. Cannari e D'Alessio (2003) suggeriscono un'approssimazione applicabile al caso di tre indagini contigue: i pesi delle osservazioni presenti in due indagini (necessariamente contigue, non essendo previsti rientri delle famiglie che escono dal panel) vanno divisi per $(1+\rho)$; quelli delle osservazioni presenti in tutte e tre le indagini vanno invece divisi per $[1+(4/3)\rho + (2/3)\rho^2]$, assumendo che la correlazione tra t_1 e t_3 sia uguale al prodotto delle correlazioni tra t_1 e t_2 e tra t_2 e t_3 , uguali tra di loro.

Tavola 5.3

Coefficienti di aggiustamento dei pesi delle famiglie rilevate in 3 waves
(in funzione del coefficiente di autocorrelazione)

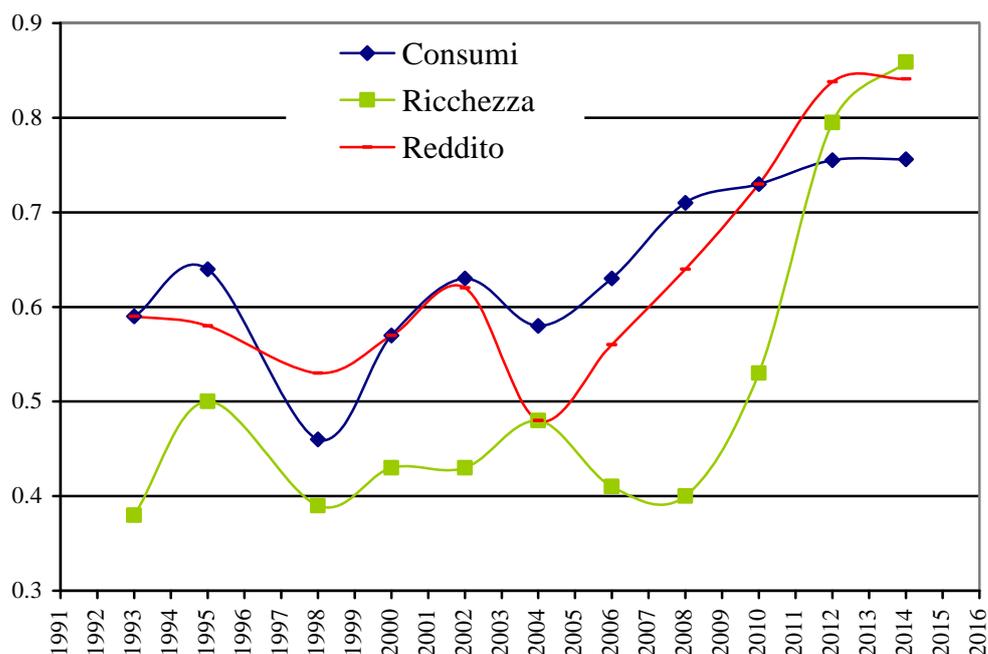
Coefficiente di autocorrelazione	Famiglie intervistate nel triennio		
	1 volta	2 volte	3 volte
	-	$1 / (1 + \rho)$	$1 / [1 + (4/3)\rho + (2/3)\rho^2]$
0,0	1,00000	1,00000	1,00000
0,1	1,00000	0,90909	0,87719
0,2	1,00000	0,83333	0,77320
0,3	1,00000	0,76923	0,68493
0,4	1,00000	0,71429	0,60976
0,5	1,00000	0,66667	0,54545
0,6	1,00000	0,62500	0,49020
0,7	1,00000	0,58824	0,44248
0,8	1,00000	0,55556	0,40107
0,9	1,00000	0,52632	0,36496
1,0	1,00000	0,50000	0,33333

¹⁰⁶ In un disegno campionario a due stadi, le considerazioni sopra riportate valgono anche, *mutatis mutandis*, per le unità di primo stadio; i comuni oggetto di rilevazione più volte nel corso del tempo, in presenza di autocorrelazione positiva tra le rispettive medie, dovrebbero contribuire meno alla stima della media di strato. Nel caso in esame tuttavia, essendo i comuni in larga parte tenuti fissi da una indagine all'altra per consentire la rilevazione delle famiglie panel, il guadagno che è possibile ottenere tramite una ponderazione alternativa delle unità è modesto.

Il livello di autocorrelazione ovviamente dipende dalla variabile considerata e dall'anno di rilevazione. Nel 2010 esso risulta pari a circa 0,59 per la ricchezza, 0,79 per il reddito e 0,73 per i consumi; negli anni più recenti i valori sono un po' più alti. Si possono quindi adottare i coefficienti di aggiustamento dei pesi che corrispondono al valore di 0,6-0,7, intermedi tra quelli stimati.

Figura 5.1

Correlazione del reddito, del consumo e della ricchezza familiare con lo stesso aggregato nella wave precedente



Se ad esempio si considera il valore di 0,6 il record relativo alla famiglia rilevata una sola volta nelle tre waves avrà un peso unitario; i due record relativi alle famiglie intervistate in due delle tre waves avranno un peso di 0,625 in ciascuna occasione; i tre record delle famiglie sempre presenti nelle tre rilevazioni avranno invece un peso di 0,49 in ciascuna occasione. Complessivamente i pesi assunti da queste tre tipologie di famiglie sono dunque rispettivamente pari a 1, 1,25 e 1,47, testimoniando il fatto che, a causa dell'autocorrelazione positiva, ogni ulteriore presenza nel campione di una famiglia già intervistata comporta un minore contributo informativo.

Questo sistema di ponderazione sarà adottato tenendo presente il vincolo che le somma dei pesi dei record rilevati in ogni anno devono essere uguali tra di loro; appare infatti ragionevole richiedere che il valore medio che si ricava dai dati ponderi in egual modo le 3 differenti waves.

Per migliorare la qualità delle stime regionali possono inoltre essere utilizzate tecniche di stratificazione a posteriori. Sebbene in linea teorica le

operazioni di post stratificazione non assicurino necessariamente un aumento dell'efficienza degli stimatori, di norma questo si realizza¹⁰⁷.

È possibile, ad esempio, utilizzare come fonti esterne le statistiche censuarie della popolazione o le statistiche sulle forze di lavoro, che pervenendo da un'indagine campionaria assai più numerosa, dovrebbero meglio rappresentare la composizione della popolazione per alcune sue caratteristiche importanti ai fini della stima del reddito e della ricchezza.

Data la limitata numerosità di alcuni campioni regionali, piuttosto che limitare notevolmente il numero dei caratteri da tenere sotto controllo, è preferibile fare ricorso a una post-stratificazione sulle sole distribuzioni marginali, tramite tecniche di raking¹⁰⁸.

In tutte le indagini sul reddito e la ricchezza è necessario confrontarsi con il problema degli *outliers*, vale a dire i valori estremi presenti nel campione in grado di influenzare pesantemente la stima. Questi in alcuni casi derivano da errori di misura (di risposta, di codifica, ecc.)¹⁰⁹; in altri casi si tratta di unità effettivamente presenti nella popolazione, sia pure in una percentuale molto ridotta.

L'utilizzo ai fini della microsimulazione

Un utilizzo particolarmente rilevante dell'indagine è quello inerente la valutazione e la simulazione di misure di policy, resa possibile dalla disponibilità di una base dati ricca e articolata. Esempi di questo genere sono l'analisi dei comportamenti di risparmio delle famiglie al variare della legislazione sui trattamenti pensionistici o l'impatto redistributivo di modifiche dell'imposizione fiscale di differenti cespiti.

La banca dati dell'IBF è stata spesso utilizzata a tali fini, in alcuni casi mediante elaborazioni relativamente semplici sugli archivi di base, in altre circostanze sviluppando veri e propri modelli di microsimulazione.

Al di là di esperimenti condotti da singoli ricercatori già alla fine degli anni ottanta, come ad esempio in Marenzi (1989), altri modelli di microsimulazione in Italia hanno utilizzato la banca dati dell'IBF come, ad esempio, ITAXMOD (Lugaresi, 1990; Di Biase et al., 1995), Euromod (Atella, Berliri e Parisi, 1999), Econlav (Coromaldi e Guerrera, 2009).

¹⁰⁷ Tali stimatori, che preservano la proprietà di non distorsione, realizzano in generale un modesto guadagno di efficienza, essendo gran parte della variabilità dovuta all'imprecisione nella stima dei valori medi per i singoli strati del disegno di campionamento, per i quali – a livello regionale - sono spesso presenti poche unità. Il guadagno di efficienza è tanto più elevato quanto più alta è la correlazione tra la variabile di studio e quella di post-stratificazione (si veda per esempio Cicchitelli et al., 1994).

¹⁰⁸ Si tenga presente che i vincoli sono definiti a livello individuale mentre il calcolo dei pesi è a livello familiare; ciò comporta una maggiore difficoltà di convergenza negli algoritmi iterativi di *raking*.

¹⁰⁹ Nelle fasi di predisposizione del questionario, di acquisizione e di controllo dei dati si predispongono una serie di accorgimenti al fine di intercettare questi tipi di errore, in particolare quelli corrispondenti a valori anomali che comportano maggiori problemi in fase di stima. Naturalmente nessun processo di controllo di qualità, per quanto accurato, può assicurare la totale assenza di errori.

I lavori di Ando e Nicoletti Altimari (2004) e quelli di Ando et al. (1999) e Cannari, Nicoletti Altimari (1997), che ne definiscono una variante denominata DYNAMITE, illustrano, inoltre, la struttura e le proprietà di un modello di microsimulazione dinamico¹¹⁰ che descrive in maniera coerente e integrata i principali aspetti socio-demografici ed economici delle diverse tipologie familiari presenti nella società italiana e che permette di analizzarne la possibile evoluzione futura. Tali modelli consentono di studiare l'impatto dei cambiamenti demografici sulle principali grandezze macroeconomiche, di prevederne l'evoluzione e di analizzare le conseguenze di particolari politiche economiche, come riforme pensionistiche e fiscali.

Nel modello di Ando e Nicoletti Altimari (2004) le osservazioni hanno peso unitario (ogni individuo è rappresentativo solo di se stesso) e si evolvono nel tempo sulla base di eventi casuali con probabilità preassegnate. Ad esempio, il modulo demografico del modello prevede che ogni individuo non sposato abbia una probabilità di sposarsi, uguale a quella degli individui simili nel complesso della popolazione; lo stato successivo dipende dalla realizzazione dell'evento che viene simulato dal sistema. Da un punto di vista campionario, ogni individuo mantiene nel tempo il suo peso. Per rappresentare la varietà delle situazioni che si viene a realizzare con il passare del tempo, è necessario disporre di un vasto insieme di osservazioni elementari, che nel caso specifico venivano ottenute tramite ricampionamento.

Altri autori, tra cui lo stesso Ando (1996), hanno utilizzato, su dati diversi da quelli dell'IBF, meccanismi diversi per tenere conto di tutte le possibili realizzazioni di un evento. Si possono infatti elencare i possibili stati in cui un individuo può trovarsi al tempo successivo e assegnare a ciascuno di essi un peso proporzionale alla probabilità di realizzazione; tale procedura risulta però operativamente più complessa in quanto accresce nel tempo in maniera esponenziale il numero dei record necessari per descrivere i possibili casi.

È inoltre evidente che i modelli del primo tipo incorporano un elemento di variabilità in più rispetto ai secondi, che deriva dal fatto che non si rappresentano tutte le possibili situazioni, ma solo quelle che casualmente si realizzano. Simulazioni diverse danno origine a risultati diversi, il che presenta pro (replicando le simulazioni si hanno informazioni sulla variabilità dei risultati) e contro (da un lato la replicazione è costosa, dall'altro con una sola simulazione si potrebbe pervenire a risultati anche molto diversi da quelli medi). Quale dei due metodi sia preferibile dipende ovviamente dall'oggetto dell'analisi e dalla numerosità e complessità dei fenomeni da simulare. Tanto maggiore è la complessità tanto più preferibile tende ad essere il modello del primo tipo.

¹¹⁰ I modelli di microsimulazione si suddividono in due grandi categorie: i modelli statici e quelli dinamici. I primi consentono di valutare l'impatto diretto determinato da una o più variabili di policy sui fenomeni oggetto di interesse, senza tenere conto dell'evoluzione dei comportamenti degli agenti economici nel corso del tempo. Modelli di questo tipo possono essere utilizzati, per esempio, per valutare la variazione di gettito attribuibile a un cambiamento nella curva delle aliquote dell'imposta sui redditi o nelle detrazioni. A differenza dei modelli statici, quelli dinamici tengono conto dell'evoluzione del comportamento degli agenti nel corso del tempo; modelli di questo tipo possono per esempio tenere conto del fatto che una variazione delle imposte può riflettersi, nei periodi successivi, sull'offerta di lavoro o sui comportamenti di evasione/elusione.

L'IBF è stata largamente utilizzata per la valutazione degli effetti determinati da variabili fiscali. Per tale valutazione, essendo le misure generalmente basate sul concetto di reddito lordo, occorre effettuare alcuni interventi sui dati che, come si è detto, rilevano i redditi netti: in particolare è necessario ricostruire i redditi al lordo delle imposte.

In assenza di evasione (trascurando per semplicità gli oneri deducibili e le detrazioni d'imposta), il reddito netto (RN) è pari al reddito lordo (RL) meno le imposte (I). Poiché le imposte sono una funzione nota del reddito lordo ($I=I(RL)$) si può scrivere la relazione: $RN= RL - I(RL)= f(RL)$. Per ottenere una stima del reddito lordo a partire dal netto è quindi sufficiente invertire la funzione $f(\cdot)$: $RL=f^{-1}(RN)$.

In presenza di evasione, le imposte sono calcolate sul reddito dichiarato, per cui la relazione $RL=f^{-1}(RN)$ tende a sovrastimare sia l'imposta pagata che il reddito lordo. Per ovviare a questo problema alcuni autori assumono che il reddito dichiarato sia una funzione monotona crescente del reddito vero e associano ai redditi netti dell'indagine una stima delle imposte pagate utilizzando i dati del Ministero delle Finanze¹¹¹. Anche questa procedura presenta alcuni limiti. L'ipotesi sul legame funzionale tra reddito vero e reddito dichiarato non è verificabile e la robustezza dei risultati è influenzata sia dagli errori di misurazione del reddito d'indagine sia dalla imperfetta comparabilità dei gruppi di soggetti posti a confronto.

Nei dati sul reddito al lordo delle imposte distribuiti dalla Banca all'interno della rilevazione armonizzata europea (HFCS) la ricostruzione al lordo delle imposte viene effettuata sulla base della metodologia illustrata in Marino e Rapalini (2003).

Più di recente i dati IBF sono stati utilizzati per sviluppare un modello di microsimulazione riguardante la vulnerabilità delle famiglie (Michelangeli e Pietrunti, 2014). Gli autori, utilizzando le previsioni a livello macro relative ai debiti e ai redditi delle famiglie, stimano gli andamenti futuri dei livelli di indebitamento e di servizio del debito delle famiglie indebitate. In particolare, sulla base di tali informazioni forniscono valutazioni sulla vulnerabilità attesa delle famiglie.

6. Conclusioni

Come ha anche sottolineato L.F. Signorini (2016) nella sua relazione di apertura della Conferenza per il cinquantenario dell'indagine, la funzione statistica di una banca centrale, e le indagini campionarie che ne sono uno strumento essenziale, sono in continua evoluzione, sia per incorporare gli avanzamenti tecnologici nel processo di produzione dei dati sia per rispondere alle sollecitazioni di una domanda di informazione sempre più esigente. Sono molte le sfide che vanno affrontate: la comparabilità internazionale; il raffronto dei dati elementari raccolti nell'indagine con quelli desumibili da altre indagini campionarie o dagli archivi amministrativi; la coerenza tra le informazioni

¹¹¹ Cfr., per esempio, Cannari e Violi (1995).

microeconomiche e le stime aggregate; la tempestività delle stime; la diffusione dei dati a un pubblico ampio e dalle necessità diverse.

Per quanto riguarda la comparabilità a livello internazionale, il nostro Istituto ha sostenuto fin dalla metà degli anni novanta il “Luxembourg Income Study” e ha promosso nel decennio successivo il “Luxembourg Wealth Study”, due progetti per la costruzione di un archivio di dati microeconomici sul reddito e sulla ricchezza familiare basato sulla raccolta e armonizzazione ex post delle rilevazioni esistenti in vari paesi del mondo. Negli ultimi anni l’indagine è divenuta parte della “Household Finance and Consumption Survey” (HFCS) dell’eurosistema, un progetto che mira a sviluppare le indagini sulle finanze familiari condotte dai paesi dell’area dell’euro. Tale iniziativa amplia le opportunità di analisi, fornendo elementi di comparazione internazionale dei fenomeni cruciali per la condotta della politica monetaria e l’analisi della stabilità finanziaria.

Il miglioramento della qualità delle rilevazioni, l’integrazione con le altre fonti campionarie e amministrative e la coerenza con le statistiche aggregate sono l’oggetto di un grande sforzo delle agenzie statistiche nazionali e internazionali. Come accade in molti paesi, le indagini sul reddito e sulla ricchezza forniscono risultati differenti dagli aggregati macroeconomici. Considerando le indagini condotte dal sistema europeo delle banche centrali, per esempio, in ciascun paese la ricchezza netta media stimata sulla base dei dati campionari è minore di quella della contabilità nazionale, in una misura che varia tra il 10 e il 30 per cento. Le discrepanze sono più sensibili per le attività finanziarie. In parte le differenze rilevate tra stime micro e dati aggregati possono risentire di differenti definizioni. Più spesso però le cause sono da ricercare nella reticenza delle famiglie circa gli aspetti economici e nella tendenza delle indagini a sotto-rappresentare alcuni segmenti della popolazione, soprattutto le famiglie più ricche, poche di numero ma titolari di un ammontare significativo di certe attività.

A questo fine vanno considerate tre possibili strategie: a) sviluppare l’integrazione dei dati campionari con quelli di fonte amministrativa, dove ritenuti sufficientemente affidabili; b) adottare forme di campionamento in grado di rappresentare adeguatamente la coda destra della distribuzione del reddito e della ricchezza (sovracampionamento dei ricchi); c) rafforzare l’analisi dei dati micro mediante procedure che consentano di tenere conto degli aspetti problematici tipici delle stime campionarie e di costruire dati microeconomici sempre meglio integrati con quelli aggregati (contabilità nazionale o dati di fonte amministrativa).

Il raccordo tra dati micro e macro consente inoltre la costruzione di modelli di simulazione grazie ai quali si possono anticipare gli effetti delle misure di policy sui diversi segmenti della popolazione.

Per quanto riguarda la rilevazione dei dati, negli ultimi anni si stanno rapidamente diffondendo le indagini sulle famiglie basate sull’uso di Internet (web/mail). Queste rilevazioni presentano notevoli vantaggi in termini di costo e tempestività, anche se l’ancora limitata diffusione di Internet, soprattutto in alcuni strati della popolazione, rimane un ostacolo a una loro maggiore adozione. È tuttavia prevedibile che la crescente alfabetizzazione digitale della popolazione

italiana renda la rilevazione via web uno strumento appropriato, anche se non necessariamente esclusivo, per lo svolgimento di indagini presso le famiglie da parte del nostro Istituto. L'indagine sulle famiglie andrà probabilmente ridisegnata per cogliere al meglio queste opportunità.

Concludiamo questa rassegna con le parole del Governatore Visco (2015) che, in occasione dell'intervento conclusivo alla Conferenza sul cinquantenario dell'indagine, ha ringraziato le oltre 154.000 famiglie che hanno partecipato all'indagine nel corso di questi decenni, alcune di loro molte volte, sugli aspetti difficili e riservati che l'indagine ha approfondito nel corso del tempo.

Appendice: Cronologia dei documenti utilizzati nel testo

I documenti storici utilizzati nel testo sono stati conservati nell'Archivio storico della Banca d'Italia. In questa appendice li descriviamo sommariamente, indicandone la collocazione archivistica ed elencandoli in ordine cronologico. Per brevità ci riferiamo a Pierpaolo Luzzatto Fegiz con l'acronimo "LF".

18 marzo 1947 LF scrive al Governatore della Banca d'Italia, Luigi Einaudi, al fine di promuovere l'attività della DOXA; nella lettera egli ricorda l'incoraggiamento ricevuto da Einaudi nel periodo in cui la DOXA «era appena in gestazione», nel 1946 (Direttorio-Einaudi, cart. 17, fasc. 11, sfasc. 9, pp. 103 e ss.).

19 luglio 1947 LF prosegue con la sua opera di promozione delle indagini Doxa con una nuova lettera a Luigi Einaudi, nel frattempo divenuto Ministro del Bilancio (ivi, p.155).

20 luglio 1947 lettera al Ministro del Tesoro con allegato un progetto di massima «approvato dal prof. Einaudi» (ivi. p. 163).

16 giugno 1948 "Relazione preliminare sulle indagini in corso presso l'Istituto DOXA", inviata al Ministero delle Finanze, relativa alla distribuzione delle famiglie italiane con due o più membri secondo l'entità del reddito familiare (Banca d'Italia, Studi, Pratiche 379, fasc, 14).

29 dicembre 1950 Baffi a LF; sollecita l'esecuzione dei lavori concordati che la Banca ha già finanziato con un contributo nel marzo 1949 (Banca d'Italia, Studi, pratiche, 284, 1, p. 77).

24 marzo 1951 LF a Baffi: si dice dispiaciuto della delusione sulla conduzione dell'inchiesta espressa da Baffi al telefono, anche a nome del Governatore Menichella (ivi, p. 71).

12 aprile 1951 Baffi a LF: risponde alla lettera del 24 marzo.. Tono molto critico sull'inchiesta in corso (sulla distribuzione del RN. Suggerimenti sulla struttura del questionario volti a rendere l'inchiesta più focalizzata (ivi p. 69).

19 gennaio 1952 LF comunica a Menichella che è stata eseguita una nuova inchiesta sulle linee richieste da Baffi nella lettera del 12 aprile 1951 (ivi, p. 5 e p. 30, quest'ultima con commenti a mano di Baffi).

17 maggio 1952 LF a Salvatore Guidotti, dirigente del Servizio studi e statistico: trasmette copia di lettera inviata a Baffi nello stesso giorno nella quale chiede l'approvazione della bozza dell'Introduzione-Riassunto alle tavole sul risparmio (ivi, p. 58).

3 luglio 1952 LF a BI; fattura relativa alla stampa di 150 copie di "Effetti di ipotetiche variazioni del reddito sulle uscite di un campione di famiglie italiane (ivi, p. 3).

12 febbraio 1953 LF a Salvatore Guidotti: sollecita il pagamento per le spese di stampa delle 150 copie di "Effetti...." (ivi, p. 2).

17 aprile 1954 annotazione di Menichella su "L'informazione" in cui chiede chiarimenti sui dati lì riportati (Carte Baffi, Servizio Studi, 10, 4, ivi, p. 6).

27 aprile 1954 nota di risposta alla richiesta formulata da Menichella il 17 aprile (ivi, pp. 2-5).

1962 (data incerta) discussione metodologica, sul più opportuno metodo di campionamento da adottare per l'indagine sul fenomeno "possesso titoli azionari e del possesso di c/c bancari o postali" (Carte Rey, Pratiche, 17, 1 p. 43).

1966 Appunto al Governatore in cui si propone l'indagine per il 1966 (Carte Rey, Pratiche, 17, 2).

1967-1968 Appunti e facsimili questionario indagine CISER 1968 (Carte Rey, Pratiche, 17, 3).

2 gennaio 1969 Lettera del Presidente dell'Istat Giuseppe De Meo al Governatore Guido Carli (Studi, Pratiche, 1039, 1, 1)

10 gennaio 1969 lettera Del Ministro del Bilancio Luigi Preti al Governatore Guido Carli (Direttorio- Carli, 70, 1, 9, pp. 5-6).

15 gennaio 1969 lettera di Guido Carli a Luigi Preti (ivi, pp. 3-4).

22 gennaio 1969: lettera del Ministro del Bilancio Preti a Carli (ivi, p. 2).

Riferimenti bibliografici

- Ando A. (1996), “*Micro Simulation Analysis of Aggregate Savings Behaviour in Japan*”, NIRA Research Output, 9(1).
- Ando A., A. Brandolini, G. Bruno et al. (1999), *The Bank of Italy's Dynamic Microsimulation Model of the Italian Economy (DYNAMITE): Recent Developments*, S.a.di.ba, Perugia, Banca d'Italia.
- Ando A., L. Guiso, I. Visco (1994), *Saving and the Accumulation of Wealth, Essays on Italian Household and Government Saving Behavior*, Cambridge University Press.
- Ando A., S. Nicoletti Altamari, (2004), *A micro simulation model of demographic development and households' economic behavior in Italy*, Temi di discussione, Banca d'Italia, n. 533.
- Atella V., C. Berliri, V. Parisi (1999), *Il modello Italiano in Euromod: problemi metodologici e primi risultati, Euromod: un modello di microsimulazione su scala europea*, Ceis-CNEL, Roma, 16 settembre.
- Baffigi A. (2007), *Cultura statistica e cultura politica: l'Italia nei primi decenni unitari*, in “Quaderni dell'Ufficio ricerche storiche”, Banca d'Italia, 15.
- Baffigi A. (2010), *Giorgio Mortara e la statistica sul credito per rami di attività*, in “Quaderni storici”, 134, XLV, 2, agosto, pp. 419-443.
- Baffigi A., A. Puggioni (2015), “*Il polso di Gini*”: *the controversy with Luzzatto Fegiz about opinion-polls*, mimeo.
- Banca d'Italia (1966), *Reddito, risparmio e alcuni consumi delle famiglie italiane*, a cura di Ufficio Ricerche Econometriche, Bollettino, n. 4, luglio-agosto, pp. 442-454.
- Banca d'Italia (1970), *Risparmio e struttura della ricchezza delle famiglie italiane nel 1968*, a cura di A. Ulizzi, Bollettino, Banca d'Italia, n. 1, pp. 103-167, Gennaio-Febbraio.
- Banca d'Italia (1983), *I bilanci delle famiglie italiane nell'anno 1982*, a cura di G. Zen, Supplemento al Bollettino, Banca d'Italia, n. 57, Dicembre.
- Banca d'Italia (1986), *Le indagini campionarie sui bilanci delle famiglie italiane*, Atti del convegno S.a.di.ba., 8-9 Febbraio 1985. Numero speciale dei contributi all'analisi economica.
- Banca d'Italia (2006), *I conti finanziari: la storia, i metodi, l'Italia e i confronti internazionali*, Atti del convegno Perugia, S.a.di.ba., 1-2 Dicembre 2005.
- Banca d'Italia (2008a), *I bilanci delle famiglie italiane nell'anno 2006*, a cura di I. Faiella, R. Gambacorta, S. Iezzi, A. Neri, Supplementi al Bollettino Statistico, Indagini campionarie (nuova serie), Banca d'Italia, n. 7, gennaio.
- Banca d'Italia (2008b), *Household wealth in Italy*, Papers presented at the conference held in Perugia, 16-17 October 2007, Banca d'Italia.

- Banca d'Italia (2010), *I bilanci delle famiglie italiane nell'anno 2008*, a cura di L. Bartiloro, G. Cappelletti, F. D'Amuri, R. Gambacorta, S. Iezzi, S. Magri, A. Neri, C. Rondinelli, *Supplementi al Bollettino Statistico, Indagini campionarie (nuova serie)*, Banca d'Italia, n. 8, febbraio.
- Banca d'Italia (2012), *I bilanci delle famiglie italiane nell'anno 2010*, a cura di C. Biancotti, F. D'Amuri, R. Gambacorta, G. Ilardi, A. Neri e C. Rondinelli, *Supplementi al Bollettino Statistico (nuova serie)*, Banca d'Italia, n. 6, Gennaio.
- Banca d'Italia (2014a), *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.
- Banca d'Italia (2014b), *La ricchezza delle famiglie italiane - 2013*, *Supplementi al Bollettino statistico* n. 69, dicembre.
- Banca d'Italia (2015), *I bilanci delle famiglie italiane nell'anno 2014*, R. Gambacorta, S. Iezzi, G. Ilardi, A. Neri e A. Rosolia (a cura di), *Supplementi al Bollettino Statistico (nuova serie)*, Banca d'Italia, n. 64, Dicembre.
- BCE (2013a), *The Eurosystem Household Finance and Consumption Survey - Methodological report for the first wave*, ECB Statistics Paper Series, n.1, April (www.ecb.europa.eu/pub/pdf/other/ecbsp1en.pdf).
- BCE (2013b), *The Eurosystem Household Finance and Consumption Survey - Results from the first wave*, ECB Statistics Paper Series, n. 2, April 2013 (www.ecb.europa.eu/pub/pdf/other/ecbsp2en.pdf).
- Bartiloro L., V. Michelangeli, C. Rampazzi (2015), *The vulnerability of indebted households during the crisis: some evidence from the euro area*, Paper presented at the Conference "The Bank of Italy's Analysis of Household Finances" - Fifty Years of The Survey on Household Income and Wealth and the Financial Accounts, Bank of Italy, Rome, 3-4 December.
- Battistin E., R. Miniaci, G. Weber, (2003), *What do we learn from recall consumption data?*, *Temi di Discussione*, Banca d'Italia, n. 466.
- Bellhouse D. R. (1988), *A Brief History of Random Sampling Methods*, in Krishnaiah e Rao (1988), pp.1-14.
- Biancotti C., G. D'Alessio, A. Neri (2004), *Errori di misura nell'indagine sui bilanci delle famiglie italiane*, *Temi di discussione*, Banca d'Italia, n. 520.
- Biancotti C., G. D'Alessio, A. Neri (2008), *Measurement errors in the Bank of Italy's survey of household income and wealth*, *Review of Income and Wealth*, Series 54, Number 3, September, pp.466-493.
- Bobbio N. (1986), *Profilo ideologico del Novecento italiano*, Einaudi, Torino.
- Boldrini M., A. Naddeo (1957), *Le statistiche empiriche e la teoria dei campioni*, Giuffrè, Milano.

- Bonci R., G. Marchese, A. Neri (2005), *La ricchezza finanziaria nei conti finanziari e nell'indagine sui bilanci delle famiglie italiane*, Temi di discussione, Banca d'Italia, n. 565, novembre.
- Bowley A. L. (1906), *Address to the Economic Science and Statistics Section of the British Association for the Advancement of Science, York, 1906*, in "Journal of the Royal Statistical Society", vol. 69, n. 3, pp. 540-58.
- 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*, Temi di discussione, Banca d'Italia, n.530.
- Brandolini A., G. D'Alessio (1998), *Measuring Well-Being in the Functioning Space*, mimeo, Bank of Italy, pubblicato poi nel 2009 in *Debating Global Society: Reach and Limits of the Capability Approach*, a cura di E. Chiappero Martinetti, Giangiacomo Feltrinelli Foundation, Milano.
- Cannari L., V. Ceriani, G. D'Alessio (1997), *Il recupero degli imponibili sottratti a tassazione*, in "Ricerche quantitative per la politica economica - 1995", Banca d'Italia, Roma.
- 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*, Temi di discussione, Banca d'Italia, 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, Banca d'Italia, n.482.
- Cannari L., G. D'Alessio (2006), *La ricchezza degli italiani, il Mulino*.
- Cannari L., G. D'Alessio, A. Mori (2011), *La situazione economica e finanziaria delle famiglie lombarde*, Questioni di Economia e Finanza (Occasional papers), Banca d'Italia, n. 93, Giugno.
- Cannari L., G. D'Alessio, G. Raimondi, A.I. Rinaldi (1990), *Le attività finanziarie delle famiglie italiane*, Temi di discussione, Banca d'Italia, n. 136.
- Cannari L., S. Nicoletti Altamari (1997), *A Dynamic Microsimulation Model of the Italian Households' Sector*, Atti del Workshop "Le previsioni della spesa per pensioni", ISTAT, Roma, dicembre.

- 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.
- Cappariello R., M.R. Marino, R. Zizza (2015), The SHIW as a tool to measure unobserved economic activities, Paper presented at the Conference “The Bank of Italy’s Analysis of Household Finances” - Fifty Years of The Survey on Household Income and Wealth and the Financial Accounts, Bank of Italy, Rome, 3-4 December.
- Cassata, F. (2006), *Il fascismo razionale. Corrado Gini fra scienza e politica*, Roma, Carocci.
- Cifaldi G., A. Neri (2013), Ci sono rischi nel porre domande su reddito e consumo nella stessa indagine?, *Temi di discussione*, Banca d'Italia, n. 908.
- Commissione per la Garanzia dell’Informazione Statistica (2002), *Il campionamento da liste anagrafiche: analisi degli effetti della qualità della base di campionamento sui risultati delle indagini*, Rapporto di Ricerca, Presidenza del consiglio dei Ministri, dicembre.
- Coromaldi M, D. Guerrera (2009), Modello di Microsimulazione EconLav: la costruzione del data-set di input, Working papers del MEF – Dipartimento del Tesoro, n. 4, giugno. D’Alessio G., I. Faiella (2002), *Nonresponse behaviour in the Bank of Italy's Survey of Household Income and Wealth*, *Temi di discussione*, Banca d’Italia, n. 462.
- D’Alessio G., S. Iezzi (2015a), How the time of interviews affects estimates of income and wealth, *Questioni di Economia e Finanza (Occasional papers)*, Banca d'Italia, n. 273.
- D’Alessio G., S. Iezzi (2015b), Over-indebtedness in Italy: how widespread and persistent is it?, Paper presented at the Conference “The Bank of Italy’s Analysis of Household Finances” - Fifty Years of The Survey on Household Income and Wealth and the Financial Accounts, Bank of Italy, Rome, 3-4 December.
- D’Alessio G., A. Neri (2015), Income and wealth sample estimates consistent with macro aggregates: some experiments (2015), *Questioni di Economia e Finanza (Occasional papers)*, Banca d'Italia, n. 272.
- 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*, Banca d'Italia, n. 610.
- De Bonis R., A. Gigliobianco (2012), *The Origins of Financial Accounts in the United States and Italy: Copeland, Baffi and the Institutions*, in De Bonis R. e Pozzolo A. *The Financial Systems of Industrial Countries, Evidence from Financial Accounts*, Springer-Verlag Berlin Heidelberg, pp. 15 – 49.
- de Leeuw E., W. de Heer (2002), *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, pp. 41-54.
- Desrosières A. (1998), *The Politics of Large Numbers. A History of Statistical Reasoning*, Harvard University Press: Cambridge, Massachusetts, and London, England
- Di Biase R., M. Di Marco, F. Di Nicola, G. Proto (1995), *Itaxmod: A Microsimulation Model of the Italian Personal Income Tax and of Social Security Contributions*, Documenti di lavoro, ISPE, n. 16.
- Doxa (1952), *Effetti di ipotetiche variazioni del reddito sulle uscite di un "campione" di famiglie italiane*, Milano, Novembre.
- Draghi M. (2007), *Household wealth in central bank policy analysis*, in Banca d'Italia (2007), Atti della Conferenza "*The Luxembourg wealth study: enhancing comparative research on household finance*", Roma, 5-7 luglio.
- Draghi M. (2009), *Conoscere per deliberare*, Intervento del Governatore della Banca d'Italia, Padova, 18 dicembre.
- Faiella I. (2008), *Accounting for Sampling Design in the Survey on Household Income and Wealth*, Temi di discussione, Banca d'Italia, n. 662.
- Favero G., U. Trivellato (2000), *Il lavoro attraverso gli "Annali": dalle preoccupazioni sociali alla misura della partecipazione e dei comportamenti nel mercato del lavoro*, in Geretto (2000), pp. 225-304.
- Fazio A. (1997), *Salvatore Guidotti nel ricordo della Banca d'Italia*, in "Rivista economica del Mezzogiorno", XI, 2 (<http://www.rivisteweb.it/doi/10.1432/3717>).
- Fiorio C. V., F. d'Amuri (2005), *Workers' Tax Evasion in Italy*, CRPS Working Paper, n. 104.
- Gambacorta R., G. Ilardi, A. Locatelli, R. Pico, C. Rampazzi (2013), *Principali risultati dell'Household Finance and Consumption Survey: l'Italia nel confronto internazionale*, Questioni di Economia e Finanza (Occasional papers), Banca d'Italia, n. 161.
- Gemelli G. (2003), *Gli scienziati*, Melis G. (a cura di), "Le élites nella storia dell'Italia Unita", CUEN.
- Gigliobianco A. (2006), *Via Nazionale, Banca d'Italia e classe dirigente. Cento anni di storia*, Donzelli, Roma.
- Gini C., L. Galvani (1929), *Di una applicazione del metodo rappresentativo all'ultimo censimento italiano della popolazione (1° dicembre 1921)*, in "Annali di Statistica", serie VI, vol. IV.
- Gini C. (1942), *Polemiche. Il polso della nazione*, in "Archivio di studi corporativi", 13, IV, pp. 465-473.

- Giraldo A., E. Rettore, U. Trivellato (2001), *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.
- Groves R.M., F.J. Fowler, M.P. Couper, J.M. Lepkowski, E. Singer e R. Tourangeau (eds.) (2004), *Survey methodology*, New York, Wiley.
- Groves R.M. e M.P. Couper (1998), *How Survey Design Features Affect Participation, Nonresponse in Household Interview Surveys*, New York, Wiley, pp. 269-293.
- Heise D. (1969), *Separating Reliability and Stability in Test-Retest Correlation*, in *American Sociological Review*, vol. 34 (1), pp.93-101.
- Henock E.P. (1987), *The Measurement of Urban Poverty: from the metropolis to the nation, 1880-1920*, in "Economic History Review", vol. 40, issue 2, pages 208-227.
- Iezzi S., A. Neri (2014), I redditi delle famiglie Italiane: un confronto tra i dati dell'indagine sui bilanci delle famiglie e i dati del Ministero dell'economia e delle finanze, mimeo, Banca d'Italia.
- IFC (2007), *Measuring the financial position of the household sector*, IFC Bulletin, n.25 e n. 26.
- Ilardi G., A. Neri (2014), Interviewers, co-operation and data accuracy: is there a link?, mimeo, Banca d'Italia.
- Juster, T., J.P. Smith (1997), Improving the quality of Economic data: Lessons from the HRS and AHEAD, *Journal of the American Statistical Association*, 92, 1268-1278.
- Kalton G., I. Flores Cervantes (2003), *Weighting Methods*, in *Journal of Official Statistics*, Vol.19, No.2, 2003, pp. 81-97.
- Kruskal W., F. Mosteller (1979a), *Representative Sampling, I: Non-scientific Literature*, in "International Statistical Review", 47, pp. 13-24.
- Kruskal W., F. Mosteller (1979b), *Representative Sampling, II: Scientific Literature, Excluding Statistics*, in "International Statistical Review", 47, pp. 111-127, 245-265.
- Kruskal W., F. Mosteller (1979c), *Representative Sampling, III: The Current Statistical Literature*, in "International Statistical Review", 47, pp. 245-265.
- Kruskal W., F. Mosteller (1980), *Representative Sampling, IV: The History of the Concept in Statistics, 1895-1939*, in "International Statistical Review", 48, pp. 169-185.
- Lord F. M., M. R. Novick (1968), *Statistical Theories of Mental Test Scores*, Addison-Wesley.
- Lugaresi S. (1990), *L'impatto redistributivo dell'indicizzazione IRPEF: una microsimulazione con ITAXMOD*, Materiali di lavoro, ISPE, n. 4, aprile.

- Marenzi A. (1989), *La distribuzione del carico fiscale in Italia: un modello di microsimulazione*, Dipartimento di Economia Pubblica, Pavia.
- Marino M.R., C. Rapallini (2003). La composizione familiare e l'imposta sul reddito delle persone fisiche: un'analisi degli effetti redistributivi e alcune considerazioni sul benessere sociale, *Temi di discussione*, Banca d'Italia, n.477.
- Marino M.R., R. Zizza (2011), Evasion of the personal income tax in Italy: an estimate by taxpayer's type", in M. Pickhardt e A. Prinz (eds.).
- Michelangeli V., M. Pietrunti (2014), A microsimulation model to evaluate Italian households' financial vulnerability, *Questioni di Economia e Finanza (Occasional Papers)*, Banca d'Italia, n. 255, Settembre.
- Neri, A. (2009), *Measuring wealth mobility in Italy*, *Temi discussione*, Banca d'Italia, n. 703, gennaio.
- Neri A., T. Monteduro (2013), La ricchezza immobiliare delle famiglie italiane: un confronto fra dati campionari e censuari, *Questioni di Economia e Finanza (Occasional papers)*, Banca d'Italia, n. 146.
- Neri A., M.G. Ranalli (2011), To misreport or not to report? The case of the Italian Survey on Household Income and Wealth, mimeo Banca d'Italia.
- Neri A., C. Rondinelli, F. Scoccianti (2015), The marginal propensity to consume out of a tax rebate: the case of Italy, Paper presented at the Conference "The Bank of Italy's Analysis of Household Finances" - Fifty Years of The Survey on Household Income and Wealth and the Financial Accounts, Bank of Italy, Rome, 3-4 December.
- Neri A., R. Zizza (2010), *Income reporting behaviour in sample surveys*, *Temi di discussione*, Banca d'Italia, n. 777, November.
- Neyman J. (1934), *On the Two Different Aspects of the Representative Method: the Method of Stratified Sampling and the Method of Purposive Selection*, [Read before the Royal Statistical Society, June 19th, 1934], in "Journal of the Royal Statistical Society", Vol. 97, No. 4, pp. 558-625.
- Oksenberg L., C. Cannel, G. Kalton (1991), *New Strategies for Pretesting Survey Questions*, in *Journal of Official Statistics*, vol. 7, n.1, pp.349-365.
- Pisano E., S. Tedeschi (2014), Micro Data Fusion of Italian Expenditures and Incomes Surveys, Working Paper del Dipartimento di Economia Pubblica, n. 164, Università La Sapienza, Roma.
- Pissarides C.A., G. Weber (1989), *An Expenditure Based Estimate of Britain's Black Economy*, *Journal of Public Economics*, 39(1), Giugno.
- Rinauro S. (2002), *Storia del sondaggio d'opinione in Italia 1936-1994. Dal lungo rifiuto alla repubblica dei sondaggi*, Istituto Veneto di Scienze, Lettere ed Arti, Venezia.

- Rodano M. L., L.F. Signorini (2008), *Measuring the value of micro enterprises*, in Banca d'Italia (2008).
- Seng Y. P. (1951), *Historical Survey of the Development of Sampling Theories and Practice*, in "Journal of the Royal Statistical Society". Series A (General), Vol. 114, No. 2, pp. 214-231.
- Shao J., J. Tu (1995), *The Jackknife and Bootstrap*, Berlin, Springer Verlag.
- SIS (1996), *Cento anni di indagini campionarie*, Atti del convegno – Roma, 31 maggio – 1 giugno 1995, Centro di Informazione e Stampa Universitaria, Roma.
- Sprenkle C. M. (1993), *The Case of the Missing Currency*, Journal of Economic Perspectives, vol. 7, n. 4, pp. 175-184.
- Stephan F.F. (1948), *History of Modern Sampling Procedures*, in "Journal of the American Statistical Association", 43, 241, Mar., pp. 12-39.
- Ulizzi A. (1967), *La reticenza degli intervistati nella rilevazione campionaria dei titoli*, mimeo, Banca d'Italia.
- Verma V. (2000), *Advanced Sampling Methods*, Manual for Statistical Trainers, Statistical Institute for Asia and the Pacific, Tokyo, 2000, p. 6.13-6.21.
- Visco I. (2008), *Introductory Remarks*, in Banca d'Italia (2008), *Household wealth in Italy*, Roma.
- Visco I. (2015), *Fifty years of The Survey on Household Income and Wealth and the Financial Accounts*, Banca d'Italia, Roma.
- Wolter K.M. (1985), *Introduction to Variance Estimation*, Berlin, Springer Verlag.
- Zuliani A. (1996), *L'utilizzo della tecnica campionaria nelle indagini dell'Istituto Nazionale di Statistica*, in SIS (1996), pp. 35 – 40.