



BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

L'uso di internet nelle indagini sulle condizioni economiche delle famiglie: futuro prossimo o remoto?

di Romina Gambacorta, Martina Lo Conte, Manuela Murgia, Andrea Neri, Roberta Rizzi e Francesca Zanichelli

Giugno 2018

Numero

437



BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

L'uso di internet nelle indagini sulle condizioni economiche delle famiglie: futuro prossimo o remoto?

di Romina Gambacorta, Martina Lo Conte, Manuela Murgia,
Andrea Neri, Roberta Rizzi e Francesca Zanichelli

Numero 437 – Giugno 2018

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

L'USO DI INTERNET NELLE INDAGINI SULLE CONDIZIONI ECONOMICHE DELLE FAMIGLIE: FUTURO PROSSIMO O REMOTO?

di Romina Gambacorta*, Martina Lo Conte[♦], Manuela Murgia[♦], Andrea Neri*,
Roberta Rizzi[†] e Francesca Zanichelli*

Sommario

Le indagini sul reddito e la ricchezza delle famiglie vengono in genere realizzate attraverso interviste personali. Negli ultimi anni è aumentata la diffusione di indagini che utilizzano Internet come strumento di rilevazione, sia per motivi economici sia per la rapidità con cui i dati raccolti sono disponibili. Ad oggi non esistono però studi sugli effetti dell'uso di tale strumento nelle indagini sul reddito e sulla ricchezza delle famiglie. Il lavoro contribuisce a colmare questa lacuna illustrando i risultati di una indagine sperimentale condotta nel 2016 da Banca d'Italia in collaborazione con l'Istat. La qualità delle informazioni rilevate è valutata tramite il confronto con le statistiche fiscali e con le evidenze aggregate desumibili dall'indagine intermedia condotta tramite interviste personali. Il lavoro si concentra sui problemi di copertura, su quelli di mancata risposta e di misurazione. I risultati mostrano come lo strumento web possa rappresentare una valida alternativa per domande di natura qualitativa e per raccogliere informazioni sulle fonti di reddito meno sensibili. Per ridurre gli effetti distorsivi legati all'uso di questo strumento è fondamentale però disporre di informazioni ausiliarie su tutto il campione selezionato.

Classificazione JEL: C81, D31.

Parole chiave: indagine web, modalità di indagine, errori non campionari.

Indice

1. Introduzione	5
2. L'uso di Internet nelle indagini campionarie	5
3. L'indagine web sulle famiglie italiane (WEBIT).....	7
4. Il quadro teorico per la valutazione dei risultati dell'indagine.....	8
4.1. Gli errori di copertura	10
4.2. Gli errori di mancata risposta.....	11
4.3. Gli errori di misura.....	14
4.3.1. La misurazione dei redditi da lavoro e da pensione.....	15
4.3.2. Le valutazioni qualitative sulle condizioni economiche della famiglia.....	16
4.3.3. L'effetto delle modalità di risposta "Non so" / "Non intendo rispondere".....	17
5. Le misure adottate per aumentare il tasso di partecipazione.....	18
6. Conclusioni.....	19
Bibliografia.....	22
Appendice A - Tavole e figure.....	25
Appendice B - I pesi campionari.....	46

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

[♦] Istat, Direzione Centrale per la Raccolta Dati.

[†] Istat, Direzione Centrale per le Statistiche Sociali e il censimento della popolazione.

1. Introduzione¹

Le indagini condotte tramite un rilevatore (*face-to-face*) sono potenzialmente quelle in grado di assicurare la maggiore accuratezza delle risposte fornite dall'intervistato, soprattutto quando l'intervista avviene con ausilio di un questionario elettronico dotato di tutti gli opportuni controlli applicabili in tempo reale (*Computer Assisted Personal Interviewing*, CAPI). Questa tecnica di rilevazione permette inoltre di raggiungere una qualsiasi popolazione di interesse in quanto ogni unità selezionata può essere contattata, raggiunta e intervistata.

Una modalità di rilevazione alternativa, sempre più diffusa negli ultimi anni, è quella basata su questionari elettronici autocompilati in ambiente web (*Computer Assisted Web Interviewing*, CAWI). Questa modalità presenta indubbi vantaggi in termini di costi e di tempestività con cui le informazioni rilevate sono disponibili ed è destinata a diffondersi sempre di più vista la crescente penetrazione di Internet fra tutte le fasce di popolazione². Oltre ai vantaggi ricordati, ci sono almeno altri tre aspetti che la rendono interessante in chiave prospettica. In primo luogo, Internet consente la trasmissione anche di materiale audio e video, oltre che di testo. L'uso ad esempio di VoIP (*Voice over Internet Protocol*), quali Skype, permetterebbe di condurre "interviste di persona" anche tramite Internet. Un secondo aspetto è che Internet si sta rapidamente evolvendo da un mezzo informativo basato su testo ad uno strumento multimediale per il *social networking*. Questi *network* potrebbero ad esempio essere utili per realizzare disegni di campionamento del tipo *snowball* in cui un campione iniziale di individui viene utilizzato per individuare fra i loro conoscenti altri individui che sono d'interesse per le finalità dell'indagine (su tale metodologia si vedano ad esempio Salganik e Heckathorn, 2004, Heckathorn 2002, Poynter 2010). Un terzo elemento di interesse verso questo strumento emerge da alcuni studi che hanno evidenziato come la mancanza di un intervistatore possa generare risposte più veritiere nel caso di domande ritenute sensibili (Tourangeau e Yan, 2007; Kreuter et al., 2008).

Per tali ragioni, Banca d'Italia e Istat hanno deciso di collaborare nella realizzazione di una indagine sperimentale web finalizzata a raccogliere informazioni socio-economiche sulle famiglie italiane. La sperimentazione ha un duplice scopo. Il primo è quello di verificare gli effetti dell'impiego di una diversa modalità di indagine sulla qualità delle risposte a domande complesse come quelle sul reddito e sulle forme di risparmio e, dall'altro, di individuare le caratteristiche sociodemografiche e reddituali dei sottogruppi di popolazione con maggiore propensione a partecipare alle indagini web.

2. L'uso di Internet nelle indagini campionarie

Nonostante la rapida diffusione di Internet come strumento di rilevazione, il suo uso nelle indagini su larga scala basate su campioni probabilistici è ancora limitato (Tourangeau et al. 2013). Una delle principali difficoltà è legata alla mancanza di una lista esaustiva della popolazione di interesse che, oltre alle classiche informazioni demografiche, contenga anche informazioni sull'accesso o meno ad Internet. Le possibili strategie per aggirare tale ostacolo sono essenzialmente quattro (Couper, 2000). Una prima strada consiste nel selezionare in

¹ Gli autori desiderano ringraziare Gabriele Mambrini, Giuseppina Papadia e Gianfranco Stanziale per il supporto fornito durante la predisposizione dell'indagine.

² Uno degli obiettivi della Commissione Europea è quello della realizzazione entro il 2020 del *Digital single market*, ossia di un mercato digitale in cui gli individui e le imprese (indipendentemente dalla loro nazionalità o luogo di residenza) possano accedere senza ostacoli ed esercitare attività *online* in condizioni di concorrenza e con un elevato livello di sicurezza. Questo mercato è un punto cardine della *Digital Agenda for Europe* che ha lo scopo di sfruttare le potenzialità delle tecnologie dell'informazione e delle comunicazioni (TIC) al fine di promuovere l'innovazione, la crescita economica e il progresso dei paesi europei.

modo casuale un campione di visitatori di uno specifico sito web invitandoli a prendere parte all'indagine tramite dei *pop-up*. Un secondo approccio è quello di selezionare il campione da una lista disponibile (come ad esempio le liste anagrafiche) invitando alla partecipazione tramite l'invio di lettere o e-mail. Una terza possibilità è quella di offrire ad una parte di un campione, selezionato per partecipare ad una indagine condotta con altra metodologia tradizionale (ad esempio con quella CAPI), la possibilità di rispondere tramite web. Una quarta strategia è quella di selezionare un ampio campione probabilistico, condurre una breve indagine individuando coloro che hanno accesso ad Internet per poi intervistarli tramite web. L'unico esempio (di cui siamo a conoscenza) di indagine web su larga scala sul reddito e la ricchezza delle famiglie è rappresentata dalla componente olandese dell'indagine armonizzata europea HFCS (*Household Finance and Consumption Survey*). L'indagine è condotta su un panel di individui selezionati inizialmente a partire da un campione casuale di indirizzi postali³.

Per una discussione dettagliata sui vantaggi e gli svantaggi legati all'uso di Internet come modalità di intervista si rinvia al lavoro di Couper (2011) e agli atti del convegno organizzato da Eurostat nel 2014⁴. Gli aspetti più rilevanti ai fini del nostro lavoro possono riassumersi nei seguenti punti.

La mancanza dell'intervistatore trasferisce sul rispondente la responsabilità piena della comprensione delle domande, del completamento dell'indagine e della trasmissione finale dell'intervista. Ciò implica, in primo luogo, che si debba prestare particolare attenzione alla formulazione delle domande, alle istruzioni fornite al rispondente (evitando il più possibile ambiguità) e al flusso del questionario (evitando ad esempio che possa scorrerlo tutto, decidendo a quali domande e in che ordine rispondere). Inoltre, la mancanza di un intervistatore che incentivi e motivi gli individui, rende più complesso nelle indagini web convincere gli individui a partecipare. Risulta dunque essenziale il ricorso a incentivi o altre strategie, come ad esempio i solleciti, per ottenere tassi di partecipazioni accettabili.

Le indagini web sembrano invece particolarmente adatte alla raccolta di informazioni sensibili poiché, prescindendo dalla presenza dell'intervistatore, riducono i fenomeni di distorsione delle risposte fornite di fronte ad una terza persona per timore di giudizio sociale (*social desirability bias*) e in questo senso hanno mostrato un rendimento migliore rispetto ad altre forme di raccolta tramite questionari autogestiti (Tourangeau e Yan, 2007; Kreuter et al., 2008).

Le indagini web possono contribuire alla raccolta di dati accurati anche attraverso altri canali (Dillman et al. 2009, 2014; Fricker et al., 2005, Martin e Lynn, 2011). L'assenza dell'intervistatore e la possibilità di compilare il questionario nel momento della giornata desiderato danno agli intervistati tutto il tempo per consultare eventuali documenti (estratti conti bancari, dichiarazione dei redditi, ...) nonché di consultarsi con gli altri membri della famiglia senza alcuna pressione. Un altro aspetto importante è che normalmente le indagini web sono svolte in un arco temporale più ridotto rispetto alle indagini con intervistatore. La distanza temporale fra le interviste può causare differenze nelle risposte (D'Alessio e Iezzi, 2015), specialmente quando si tratta di informazioni di tipo economico (come ad esempio

³ Questo campione viene poi agganciato con un registro di numeri telefonici (con percentuale di successo intorno al 50 per cento). Gli individui sono quindi invitati tramite telefono o lettera a partecipare ad una breve rilevazione. A coloro che accettano di partecipare viene chiesta la disponibilità a partecipare ad un panel di lungo termine. Per incentivare la partecipazione, i panelisti ricevono in primo luogo il computer per poter partecipare alle indagini e un rimborso delle spese di connessione a Internet. Inoltre, ricevono dei punti per ogni questionario compilato che periodicamente possono monetizzare o comunque trasformare in qualche altro vantaggio economico (Teppa, 2012).

⁴ Eurostat ha fotografato lo stato dell'arte delle diverse tecniche d'indagine in un recente progetto ESSnet, conclusosi nel 2014 con il workshop "*Data collection for social surveys using multiple mode*" presso l'Istituto di Statistica Tedesco Destatis.

quello sullo stato occupazionale o sul valore della propria abitazione) o quando si tratta di domande su aspettative future (andamento del mercato azionario, prospettive di cambiamento del proprio status occupazionale). Infine, alcuni studi hanno mostrato come la presenza di un intervistatore possa incentivare comportamenti opportunistici (*satisficing behaviour*) tendenti a produrre risposte non accurate nel tentativo di ridurre il tempo e, in generale, l'impegno necessario alla compilazione del questionario (Simon 1957; Zhang 2013).

Anche se esiste già una vasta letteratura che si concentra sul confronto fra la metodologia web e altri metodi di indagine, a nostra conoscenza, non esistono invece studi specifici relativi a indagini su reddito e ricchezza delle famiglie. Una eccezione è costituita da un lavoro sull'indagine olandese che, utilizzando come termine di confronto i dati ufficiali dell'Istituto di statistica (Statistics Netherland), mostra come l'indagine rappresenti adeguatamente alcuni aspetti socio-demografici (età, area geografica e genere) della popolazione ma non altri (titolo di studio, composizione familiare, nazionalità, stato civile) (Teppa, 2012). Il presente lavoro ha come obiettivo quello di contribuire a colmare questa lacuna informativa.

3. L'indagine Web sulle Famiglie Italiane (WEBIT)

L'indagine web sulle famiglie italiane (WEBIT) è stata condotta nel primo semestre del 2016 su un campione probabilistico di circa 1.000 individui. Il campione inizialmente selezionato per partecipare all'indagine è stato di circa 10.000 famiglie che sono state casualmente estratte dall'archivio delle Liste Anagrafiche Comunali (LAC) di 250 comuni.

Le famiglie selezionate sono state invitate a partecipare all'indagine tramite una lettera a firma congiunta dei due Istituti. La lettera conteneva una breve presentazione dell'iniziativa, il collegamento al sito web dell'indagine ed un codice personale da utilizzare per accedere al questionario. L'indagine è stata realizzata tramite la piattaforma *LimeSurvey* predisposta dal Servizio Sviluppo informatico della Banca d'Italia ed accessibile anche tramite *smartphone* o *tablet*.

Una parte del campione (circa l'80%) è stata casualmente selezionata per partecipare ad un concorso a premi per la vincita di 5 *Ipad*. Le estrazioni si sono svolte in modo progressivo durante il periodo di rilevazione e dando la possibilità agli individui di partecipare a tutte le estrazioni successive alla data di completamento dell'intervista, in modo da incentivare a completare il prima possibile l'indagine.

Per vincere la diffidenza dei rispondenti sono state realizzate delle pagine web dedicate all'indagine sul sito dell'Istat e della Banca d'Italia, contenenti tutte le informazioni utili sulle finalità della rilevazione, sulle modalità di selezione del campione, sul trattamento dei dati, sull'integrazione con gli archivi fiscali detenuti dall'Istat nonché una copia della lettera inviata alle famiglie estratte a partecipare all'indagine. Inoltre, per offrire ai rispondenti un supporto più capillare, sono stati coinvolti gli Uffici Regionali dell'Istat ai quali è stata inviata un'email descrittiva di tutti gli aspetti contenutistici ed organizzativi dell'indagine. Infine, durante la rilevazione il personale dell'Istat ha condotto solleciti telefonici allo scopo di sensibilizzare e tranquillizzare ulteriormente le famiglie selezionate.

Un importante valore aggiunto dell'indagine WEBIT è che il campione selezionato è stato agganciato, tramite codice fiscale, alle informazioni amministrative di fonte fiscale⁵. L'aggancio è stato possibile per quasi il 95 per cento delle persone selezionate per partecipare all'indagine. La principale causa del mancato aggancio è la mancata presentazione della dichiarazione dei redditi da parte di alcune tipologie di individui.

⁵ L'aggancio è stato effettuato dai ricercatori Istat nel rispetto dell'attuale disciplina sulla privacy e secondo quanto stabilito nell'Accordo di collaborazione per la realizzazione di un progetto di ricerca finalizzato a valutare gli effetti dell'uso della tecnica di rilevazione CAWI, Repertorio Istat n.11 del 18 Febbraio 2016. L'archivio integrato è rimasto a disposizione dei soli ricercatori Istat.

Inoltre, l'indagine si è svolta in parallelo alla tradizionale indagine CAPI sulle condizioni economiche delle famiglie Italiane nel 2015 (IBF-I), condotta su un campione di circa 2.000 famiglie selezionate fra quelle che avevano partecipato all'indagine sui bilanci delle famiglie italiane nel 2014. Le due indagini sono state realizzate negli stessi comuni, utilizzando sostanzialmente lo stesso questionario e anche una identica formulazione delle domande. In particolare, il questionario utilizzato includeva domande sulle caratteristiche demografiche dei componenti della famiglia, sull'occupazione, sui redditi e le condizioni di vita; sui rapporti con le banche, sulle attività e passività finanziarie e sui consumi. Per rendere le due indagini più comparabili, nell'indagine IBF-I sono state incluse le seguenti domande: "Lei o un componente della Sua famiglia, a casa o altrove, "naviga" in Internet utilizzando un computer?" e "Lei o un componente della Sua famiglia, a casa o altrove, "naviga" in Internet utilizzando uno *smartphone* o un *tablet*?". Queste domande hanno permesso di definire la popolazione eleggibile per una indagine web e di studiarne le caratteristiche. Inoltre, nell'indagine web, per le domande più complesse come quelle sul reddito e sulla ricchezza sono stati inseriti dei glossari per sopperire all'assenza di un intervistatore. Infine, sempre nell'indagine web, in alcuni casi sono state testate diverse formulazioni della stessa domanda per verificare l'effetto sui risultati finali.

4. Il quadro teorico per la valutazione dei risultati dell'indagine

Il *Total Survey Error approach* fornisce un quadro teorico utilizzabile per valutare i risultati dell'indagine web (TSE, Bethlehem 2010, Biemer 2010).

L'obiettivo di ogni indagine campionaria è quello di minimizzare la discrepanza fra la stima campionaria di interesse e il valore vero della caratteristica della popolazione oggetto di analisi (*total survey error*, TSE). Questa discrepanza è il risultato finale dell'accumulo degli errori che possono verificarsi in fase di disegno, di intervista e di analisi dei dati campionari. Gli errori possono essere raggruppati in due tipologie: quelli di natura campionaria e quelli di natura non campionaria. I primi sono legati al fatto che l'analisi è condotta su un campione e sono facilmente quantificabili.

Secondo gli studi disponibili in letteratura (si vedano ad esempio Bethlehem 2010 e Tourangeau et al. 2013) le principali tipologie di errore non campionario che influenzano le indagini web sono gli errori di copertura, quelli di mancata risposta e quelli di misurazione. Ciascuna di queste fonti di errore può incidere sia sulla distorsione sia sulla efficienza degli stimatori. Il presente lavoro si concentra sul problema della distorsione.

Si supponga di voler stimare la media di una caratteristica \bar{Y} riferita all'intera popolazione italiana (come ad esempio il reddito familiare) e che la popolazione sia suddivisa in due gruppi mutualmente esclusivi (di dimensione ignota N_I e N_{NI}) a seconda che abbiano o meno accesso ad Internet. Siano \bar{Y}_I e \bar{Y}_{NI} i valori medi (ignoti) del reddito nei due sottogruppi. Inoltre, sia $\bar{Y}_{I,R}$ il valore medio del reddito nella sotto-popolazione dei rispondenti con accesso ad Internet ($N_{I,R} \subseteq N_I$). Estrahendo un campione probabilistico dalla popolazione con Internet e in presenza di mancata partecipazione di una parte del campione e di errori di misurazione, la distorsione dello stimatore della media campionaria \bar{y}_r è pari a:

$$[1] \quad B(\bar{y}_r) = E(\bar{y}_r) - \bar{Y} = B_{COP} + B_{NR} + B_{MIS}$$

dove B_{COP} , B_{NR} e B_{MIS} rappresentano le distorsioni legate rispettivamente a problemi di copertura, di mancata partecipazione e di misurazione.

La distorsione legata a problemi di copertura è data dalla differenza fra il valore atteso della media che si otterrebbe da un campione estratto dalla popolazione con Internet e quello dell'intera popolazione di interesse ($B_{COP} = E(\bar{y}_I) - \bar{Y}$, con $E(\bar{y}_I) = \bar{Y}_I$). Questa distorsione

dipende dal fatto che lo strumento non permette di raggiungere l'intera popolazione (a differenza di quanto avviene invece nelle indagini con intervistatori) e si verifica anche se non vi fossero problemi di non risposta o di misurazione.

Il secondo tipo di errore è legato al rifiuto di parte del campione selezionato a partecipare all'indagine (non risposta totale). Lo stimatore basato sul campione dei rispondenti $\bar{y}_{I,r}$ ha un valore atteso pari a quello della popolazione dei rispondenti $\bar{Y}_{I,R}$, che può essere diverso da quello del totale della popolazione che ha accesso a Internet ($B_{NR} = E(\bar{y}_{I,r}) - \bar{Y}_I$, con $E(\bar{y}_{I,r}) = \bar{Y}_{I,R}$). La presenza di un intervistatore in genere contribuisce a ridurre questo errore in quanto è fortemente motivato a convincere le famiglie a partecipare all'indagine. Utilizzando il web al contrario non esiste uno strumento di convincimento altrettanto efficace e dunque risulta più difficile ottenere tassi di partecipazione elevati.

Il terzo tipo di errore è quello di misura. Rientrano in questa categoria una molteplicità di casistiche quali gli errori compiuti dagli intervistati nel rispondere alle domande, gli errori di digitazione o la falsificazione dei dati. A seguito di questi fenomeni il valore atteso dello stimatore \bar{y}_r risulta diverso da quello vero ($B_{MIS} = E(\bar{y}_r) - \bar{Y}_{I,R}$).

Il principale errore di misura studiato in letteratura è legato al comportamento dei rispondenti. Tourangeau et al. (2000) distinguono tre fasi del processo cognitivo richiesto per rispondere alle domande: comprensione, recupero delle informazioni e formulazione di una strategia di risposta. Ognuna di queste fasi può potenzialmente generare un errore anche in funzione del metodo di intervista utilizzato. Una prima fonte potenziale di errore è legata alla mancata comprensione dei testi delle domande, per ambiguità grammaticali, eccessiva complessità, concetti vaghi o presenza di termini sconosciuti. La presenza di un intervistatore può in tali casi essere di aiuto, mentre attraverso lo strumento WEB aumenta il rischio che i rispondenti interpretino in modo diverso le domande.

Una seconda possibile fonte di errore ha a che fare con i problemi di recupero dell'informazione richiesta (Groves 2004). La presenza di un intervistatore costringe gli individui a fornire una risposta in tempi ridotti e può spingere ad adottare una strategia di ricostruzione approssimata, ricorrendo ad esempio a forme di arrotondamento (Pudney, 2008). Nel caso di intervista web, invece, i rispondenti compilano il questionario con estrema libertà e comodità e hanno il tempo di raccogliere tutte le informazioni necessarie (anche consultando documenti).

Infine, dopo aver richiamato alla mente le informazioni richieste, il rispondente decide quale strategia di risposta adottare. La decisione di sottostimare volontariamente le proprie fonti di reddito o la propria ricchezza è probabilmente la principale fonte di errore in questa fase. La presenza di un intervistatore può avere effetti sia positivi che negativi. L'intervistatore può infatti tranquillizzare il rispondente spiegando le finalità dell'indagine e rassicurandolo sulla confidenzialità delle risposte fornite, creando dunque un clima di intervista ottimale. D'altro canto, parlando della propria situazione economica, la presenza di un intervistatore potrebbe generare il cosiddetto fenomeno della "desiderabilità sociale" (Bagozzi, 1994, Pitrone, 2009)⁶.

Nel caso dell'indagine WEBIT la presenza di errori (non campionari) risulta evidente confrontando la composizione del campione teorico (ossia quello inizialmente selezionato) con quello finale effettivamente realizzato. Se non vi fossero errori rilevanti, le caratteristiche

⁶ Ad esempio, se il rispondente appartiene a una famiglia molto ricca, può decidere di dichiarare meno reddito a causa del desiderio di conformità sociale con l'intervistatore o per paura che queste informazioni non vengano tenute riservate. In teoria, potrebbe anche verificarsi il comportamento opposto in cui il soggetto che vuole impressionare l'intervistatore, sovrastima la propria condizione economica.

socio-demografiche dei due campioni non risulterebbero significativamente diverse fra loro. Il campione intervistato risulta invece composto con maggior frequenza da famiglie con intestatario della scheda anagrafica con età compresa fra i 35 e i 64 anni, da uomini, da famiglie con 3 o 4 componenti, con figli e che vivono in prevalenza al Nord. Si tratta inoltre di famiglie che presentano in genere una condizione economica più agiata delle altre (tavola 1).

Gli strumenti utilizzati per valutare l'indagine WEBIT sono essenzialmente tre. I problemi di copertura sono valutati attraverso l'indagine IBF-I che è stata condotta sull'intera popolazione e include domande sull'uso di Internet (par. 4.1). La distorsione da mancata risposta (par. 4.2) è analizzata utilizzando le informazioni demografiche e amministrative disponibili sul campione web dei non intervistati e quelle dell'indagine IBF-I (che permette di stimare la condizione di eleggibilità delle famiglie WEBIT). Gli errori di misura sono invece stimati confrontando le distribuzioni delle risposte con quelle fornite da rispondenti simili che hanno partecipato ad IBF-I, utilizzando i dati amministrativi della Banca Dati Reddittuale (BDR) ed alcune sperimentazioni condotte nell'indagine WEBIT (par. 4.3). Infine nel par. 5 sono descritti i risultati di alcune misure introdotte per limitare il problema della non risposta.

4.1. Gli errori di copertura

La popolazione di riferimento dell'indagine WEBIT è l'intera popolazione italiana. Lo stimatore di Horvitz-Thompson (\bar{y}_I) calcolato su un campione probabilistico della popolazione con accesso ad Internet ha una distorsione pari a:

$$[2] \quad B_{COP} = E(\bar{y}_I) - \bar{Y} = \frac{N_{NI}}{N} (\bar{Y}_I - \bar{Y}_{NI})$$

L'entità della distorsione dipende da due fattori:

- a) la proporzione di famiglie che non hanno accesso ad Internet;
- b) la differenza fra i valori medi di Y delle due sotto-popolazioni.

La valutazione del primo aspetto è resa possibile dall'indagine IBF-I che è condotta sull'intera popolazione. Secondo tale indagine la quota di famiglie che ha accesso ad Internet (tramite *pc*, *smartphone* o *tablet*) era pari a circa il 70 per cento nel 2015. Tale valore risulta in linea con quello che emerge dall'indagine annuale dell'Istat sull'uso delle tecnologie dell'informazione e della comunicazione (ICT) da parte di cittadini e imprese (Istat, 2016)⁷. La diffusione di Internet in Italia risulta lievemente più bassa della media europea (figura 1). Negli ultimi anni comunque la percentuale di famiglie con accesso ad Internet è in netta crescita (nel 2008 era pari al 47 per cento, secondo l'indagine Istat).

La diffusione di Internet non è uniforme tra i vari sotto gruppi socio-demografici della popolazione (tavola 2). L'età è un primo fattore discriminante, con le famiglie più anziane più difficilmente raggiungibili tramite questo strumento. Inoltre, l'accesso al web è maggiormente diffuso tra le famiglie più istruite o appartenenti alle fasce di reddito e di ricchezza più elevate. Infine, questo strumento è maggiormente diffuso al Centro-Nord e nelle grandi città⁸. In

⁷ L'indagine ICT dell'Istat stima che nel 2016 circa il 70 per cento delle famiglie aveva accesso ad Internet da casa. Considerando anche coloro che accedono da altri luoghi tale percentuale salirebbe al 72 per cento. La principale ragione per cui la famiglia non possiede accesso ad Internet (57 per cento dei casi) è che nessun componente sa usare Internet. Il secondo motivo (24 per cento) è che la famiglia non lo ritiene interessante o utile. I fattori economici incidono in meno del 10 per cento dei casi.

⁸ Risultati del tutto analoghi emergono a livello europeo dove circa il 96 per cento degli individui fra 16 e 24 anni sono utilizzatori regolari di Internet, contro il 57 per cento di coloro che hanno fra 55 e 74 anni. Inoltre, gli individui con un alto livello di istruzione sono quasi tutti utilizzatori regolari di Internet contro circa il 60 per cento di quelli che hanno un basso di livello di istruzione. Le statistiche a livello europeo sono disponibili nel sito *statistics explained* di Eurostat <http://ec.europa.eu/eurostat/statistics->

generale, l'unico gruppo che potrebbe risultare difficile da raggiungere è quello delle famiglie in cui il capofamiglia non ha alcun titolo di studio: solo nel 6 per cento dei casi queste hanno infatti accesso ad Internet. Tali famiglie comunque rappresentano meno del 4 per cento dell'intera popolazione (Banca d'Italia 2015).

L'entità del secondo fattore di distorsione (ossia la differenza fra i valori medi) può essere stimata sia tramite l'indagine WEBIT sia attraverso IBF-I. Nel primo caso si possono usare le informazioni di natura fiscale provenienti dalla Base dati reddituali disponibili per tutto il campione inizialmente selezionato di 10.030 famiglie. Occorre però imputare la condizione di eleggibilità delle famiglie in quanto tale informazione non è disponibile. L'imputazione è avvenuta in due fasi. Prima è stato stimato nell'indagine IBF-I un modello logistico sulla probabilità di accedere ad Internet in funzione delle seguenti variabili: genere, classe di età e nazionalità del capofamiglia, dummy sulla presenza di un giovane in famiglia (definito come un componente di età inferiore a 40 anni) e sulla presenza di un coniuge, numero di componenti, classe di reddito familiare, possesso di immobili diversi dall'abitazione di residenza (utilizzata come *proxy* della ricchezza), area geografica e la dimensione del comune di residenza. I coefficienti stimati sono in linea con i risultati dell'analisi descrittiva e inoltre mostrano che la presenza di un giovane in famiglia aumenta significativamente la probabilità di usare il web (tavola 3).

I coefficienti stimati sono stati quindi usati per stimare la probabilità di essere eleggibile nel campione estratto a partecipare all'indagine WEBIT. L'imputazione non ha riguardato le famiglie per cui è stato possibile accertare la condizione di eleggibilità in altro modo (attraverso ricontatti telefonici)⁹.

Utilizzando le informazioni reddituali presenti nella BDR, il reddito medio netto familiare risulta pari a circa 18.300 euro per le famiglie stimate come ineleggibili nel modello (\bar{Y}_{NI}), contro un valore di circa 31.800 per quelle eleggibili (\bar{Y}_I). Anche considerando tutti i decili della distribuzione del reddito familiare risulta evidente come le famiglie eleggibili presentino una situazione economica più agiata delle altre (figura 2). Utilizzando l'equazione [2], la distorsione che si otterrebbe utilizzando i dati dell'indagine WEB per la stima del reddito (senza sfruttare le informazioni ausiliarie della BDR) comporterebbe una sovrastima pari a circa il 15 per cento del valore medio del reddito (4.000 euro circa su un valore medio di 27.700 euro).

Risultati analoghi si ottengono con l'indagine IBF-I. Considerando il reddito netto familiare, che presenta comunque alcune differenze rispetto a quello fiscale, la distorsione che si avrebbe per problemi di copertura risulterebbe sempre nell'ordine del 15 per cento (del reddito medio). Considerando invece la ricchezza netta familiare, tale distorsione sarebbe sempre positiva e pari a circa il 10 per cento del valore medio osservato.

4.2. Gli errori di mancata risposta

Le indagini web hanno generalmente tassi di partecipazione sensibilmente più bassi rispetto

explained/index.php/Internet_access_and_use_statistics_households_and_individuals#Internet_activity_by_age_group.

⁹ Nel complesso, la percentuale stimata di famiglie eleggibili nel campione WEBIT è pari a circa il 70 per cento. Va comunque sottolineato che il processo di stima della condizione di eleggibilità nell'indagine web è soggetto ad alcune approssimazioni. In primo luogo la fonte dei dati è diversa, nel primo caso si tratta di informazioni raccolte tramite indagine, nel secondo di dati desumibili dalle liste anagrafiche (per quanto riguarda le informazioni socio-demografiche) e dall'aggancio con la base dati reddituale per quanto riguarda le informazioni economiche. In particolare, per stimare il possesso di altri immobili oltre all'abitazione di residenza si è utilizzata l'informazione sulla presenza di una tassazione IMU. Infine va ricordato che anche la definizione di capofamiglia è diversa, i dati dell'indagine CAPI si riferiscono infatti al maggior percettore di reddito, mentre nel caso del campione WEB ci riferiamo all'intestatario del foglio di famiglia anagrafico.

ad indagini analoghe condotte con modalità tradizionali (es. CAPI o indagini telefoniche). Uno studio, che ha analizzato oltre 40 indagini sperimentali, ha mostrato come in media i tassi di risposta delle indagini web risultino inferiori di circa 11 punti percentuali rispetto a quelli di indagini con modalità alternative (Manfreda et al., 2008). Jäckle et al. (2015) trovano risultati analoghi anche nel caso di indagini *panel* per le quali si potrebbe pensare che non esistano problemi di eleggibilità e che siano tipicamente caratterizzate da una maggiore propensione degli individui a partecipare.

Bassi tassi di partecipazione non implicano necessariamente una bassa qualità dei risultati (Groves e Peytcheva, 2008). Ipotizzando che ciascuna famiglia abbia una propensione ignota a rispondere $p_k = E(r_k)$, dove r_k è una variabile indicatrice che assume valore 1 se la famiglia k ha risposto e 0 altrimenti, uno stimatore naturale della media campionaria della popolazione con accesso ad internet (\bar{Y}_I) è quello calcolato sulla base dei soli rispondenti $\bar{y}_{I,r} = \frac{1}{n_r} \sum_{k=1}^N r_k Y_k$ dove n_r è la numerosità del campione dei rispondenti. Bethlehem (1988, 2002) ha mostrato come tale stimatore sia distorto (anche in assenza di errori di misura) e come tale distorsione sia pari a:

$$[3] \quad B_{NR} = E(\bar{y}_{I,r}) - \bar{Y}_I \approx \frac{\text{Cov}(p, Y)}{\bar{p}} \quad \text{Cov}(p, Y) = \frac{1}{N} \sum_{k=1}^N (p_k - \bar{p})(y_k - \bar{y})$$

L'espressione mostra in primo luogo come non sia corretto parlare di distorsione senza far riferimento ad una specifica variabile di interesse (Y): una stessa indagine può dunque produrre risultati distorti per alcune variabili e corretti per altri. La distorsione dipende inoltre dall'ammontare della mancata partecipazione ($1-\bar{p}$) e da quanto questa risulta associata alla variabile di interesse.

L'indagine WEBIT presenta, come previsto, un basso tasso di risposta. Una volta esclusi dal conteggio coloro che non hanno accesso ad Internet, il tasso di risposta è pari a circa il 12 per cento (contro un tasso del 37 per cento per la componente non panel dell'indagine sui bilanci delle famiglie italiane, Banca d'Italia 2015). Questo valore è probabilmente sottostimato in quanto per l'indagine WEB non è nota la quota di famiglie ineleggibili dovuta alle imprecisioni delle informazioni anagrafiche o ad altri motivi (indirizzi inesistenti, famiglie trasferite, decesso di tutti i componenti). Sulla base dell'indagine sui bilanci delle famiglie italiane nel 2014 tale quota è stimabile in circa il 5 per cento del totale campione selezionato. Anche considerando tale aspetto il tasso di risposta dell'indagine WEB rimarrebbe ad un valore pari all'incirca al 13 per cento.

La seconda componente della distorsione è legata all'associazione fra probabilità di risposta e variabile di interesse. La probabilità di risposta di ciascun individuo p_k non è nota e viene in genere stimata come funzione di determinate caratteristiche osservabili, sotto l'assunzione che queste contengano tutta l'informazione necessaria (la cosiddetta ipotesi di *missing at random*, MAR).

Le tavole 4 e 5 mostrano le stime dei parametri di due modelli logistici relativi:

$$\text{logit}(r_k) = \alpha + \beta X_k$$

dove $r_k = 1$ rispettivamente se il rispondente ha completato interamente l'intervista (prima specificazione del modello) o se il rispondente ha partecipato all'indagine indipendentemente dal fatto di averla completata o meno (seconda specificazione). X_k è un insieme di regressori di natura demografica (sesso, classe di età, nazionalità, stato civile della persona di riferimento, presenza di un giovane in famiglia, numero di componenti, area geografica e dimensione del comune di residenza) ed economica (classe di reddito familiare e possesso di altri immobili).

La partecipazione all'indagine WEBIT risulta più elevata al Nord Italia e per le famiglie più piccole. La probabilità di completare l'indagine è inoltre maggiore se l'intestatario della scheda anagrafica ha cittadinanza italiana o se è coniugata/o. Inoltre, anche a parità di altre caratteristiche osservabili, la probabilità cresce al crescere del reddito familiare.

L'entità della distorsione non dipende solo da quanto la variabile di interesse è associata direttamente alla propensione a rispondere, ma anche da quanto le variabili socio-demografiche (utilizzate per stimare la probabilità di risposta) sono associate con essa. Sempre nell'ipotesi che l'interesse sia la stima del reddito medio, la tavola 6 mostra come questo sia in effetti legato a fattori socio-demografici quali l'età, il sesso o l'area geografica di residenza, la cittadinanza e la presenza di un giovane in famiglia, caratteristiche che influenzano anche la probabilità di partecipazione.

La figura 3 mostra in sintesi come le probabilità di risposta stimate per i partecipanti all'indagine WEBIT risultino associate al reddito familiare di fonte fiscale. La probabilità media (e anche mediana) di partecipare è di circa il 7 per cento per la classe di reddito più bassa ed aumenta fino al 18 per cento per quella più alta. Il coefficiente di correlazione lineare fra le due variabili è pari a circa il 43 per cento. Utilizzando la formula [3] per calcolare la distorsione nella stima del reddito netto medio familiare che si avrebbe per la mancata risposta, si otterrebbe un valore di circa 4.500 euro pari a circa il 17 per cento del reddito medio.

È interessante notare come questi risultati vadano nella direzione opposta a quanto comunemente trovato per le indagini condotte con intervistatore. In questi casi infatti sono le famiglie più abbienti ad essere meno disponibili all'intervista. Ad esempio, Cannari e D'Alessio (1992) usando la componente panel dell'indagine sui bilanci delle famiglie italiane (SHIW) della Banca d'Italia hanno mostrato come le famiglie con redditi più elevati abbiano una minore propensione a partecipare alla successiva rilevazione. Analogamente, D'Alessio e Faiella, (2002) utilizzando un campione di individui estratti tra i clienti di una importante banca commerciale (e per i quali erano disponibili informazioni amministrative), hanno trovato una correlazione negativa fra ricchezza e propensione a partecipare ad una indagine.

Una spiegazione di questi risultati è da ricercarsi nel fatto che la modalità di svolgimento di una indagine può modificare la propensione alla partecipazione. In particolare, le famiglie più agiate, essendo in media più istruite delle altre, hanno una maggiore facilità a rispondere alle indagini web (che comunque richiedono un minimo livello di competenze informatiche). Inoltre, per tali famiglie è in genere più difficile trovare il tempo da dedicare in una sola giornata per rispondere ad un'indagine condotta con modalità *face to face*¹⁰. Sotto questo punto di vista il web rappresenterebbe una soluzione più adatta alle loro esigenze in quanto permetterebbe di rispondere nei momenti liberi e di compilare il questionario anche nell'arco di più giorni.

Le indagini web soffrono invece di una forma di mancata partecipazione che non è presente nelle indagini CAPI, ossia l'interruzione dell'intervista una volta iniziata. Nell'indagine WEBIT questo problema non è risultato molto significativo: meno del 4 per cento dei rispondenti ha interrotto l'indagine. Le interruzioni sono avvenute nella maggior parte dei casi nella sezione del questionario riguardante i redditi (figura 4), contribuendo ad aumentare la distorsione delle stime finali.

Un risultato interessante è che gli errori legati alla non risposta agiscono nella stessa direzione di quelli della mancata copertura, ampliando in questo modo la distorsione finale.

¹⁰ Kennickell (2009) mostra come nell'indagine SCF (Survey of Consumer Finance), il maggior ostacolo ad intervistare le famiglie estremamente ricche sia da parte degli intervistatori quello di riuscire a contattarle e da parte dei rispondenti quello di disporre del tempo necessario per la partecipazione all'intervista. Una volta contattate e risolto il problema del tempo la propensione a rispondere di queste famiglie non è inferiore a quella delle altre.

Considerando la stima del reddito medio familiare, la distorsione risulterebbe quindi nel complesso pari a circa 8.500 euro (di cui 4.500 dovuti alla non risposta e 4.000 a problemi di copertura), ossia pari al 31 per cento del valore medio. Questo risultato dipende in primo luogo dal fatto che le fasce di popolazione più agiate hanno una maggiore probabilità sia di avere accesso ad internet sia di partecipare all'indagine. Vi sono però altri fattori non osservabili che sembrano andare nella stessa direzione. La tavola 7 riporta i risultati della stima di un modello bivariato per la probabilità congiunta di essere eleggibile e di partecipare all'indagine in funzione di una serie di variabili osservabili sia di natura demografica (sesso, età, stato civile, cittadinanza dell'intestatario della scheda anagrafica, numero di componenti della famiglia, presenza di un giovane, macro area geografica e dimensione del comune di residenza), sia economica (classe di reddito familiare e possesso di altri immobili). I termini di errore dei due modelli presentano una significativa correlazione positiva, anche dopo aver controllato per i fattori sopra ricordati. Questo implica che le fasce della popolazione in cui l'utilizzo di Internet è più diffuso, come le famiglie più istruite sono anche quelle che partecipano con maggior probabilità (Roberts, 2007). Se queste variabili non osservabili sono correlate con l'oggetto della rilevazione, la distorsione dei risultati che ne consegue non può essere corretta con metodi ex-post (in quanto si tratta di variabili non osservabili).

4.3. Gli errori di misura

L'ultima componente dell'errore totale è quello di misura, definito come la differenza fra il valore riportato nell'indagine e quello vero che è ignoto. In indagini su reddito e ricchezza questa componente può risultare più rilevante di altre (si veda ad esempio Neri e Ranalli, 2011 e D'Alessio, 2017).

Nel caso in cui non vi sia il problema di mancata risposta selettiva descritta nel precedente paragrafo, la distorsione dovuta all'errore di misura può essere scritta come:

$$[3] \quad B_{MIS} = E(\bar{y}_r) - \bar{Y}_{I,R}$$

dove $\bar{Y}_{I,R}$ rappresenta il valore vero della media di Y nella popolazione con accesso ad Internet.

La situazione ottimale per poter stimare l'effetto modalità di intervista sarebbe quella in cui all'interno della stessa indagine, lo strumento di rilevazione fosse randomizzato fra le famiglie rispondenti (e che il tasso di non risposta fosse trascurabile). Nel nostro caso non è stato possibile realizzare questo esperimento per motivi di natura operativa. Pur trattandosi di due indagini indipendenti, i campioni sono comunque stati estratti casualmente dalle liste anagrafiche negli stessi comuni. L'indagine WEBIT è stata condotta inoltre il più possibile in parallelo a quella CAPI. Di conseguenza i motivi che possono spiegare le differenze nei risultati sono essenzialmente due: un effetto di non risposta selettiva o un effetto tecnica di intervista (o l'azione congiunta di entrambi).

Per analizzare il più possibile l'effetto modalità d'intervista, abbiamo in primo luogo usato nel confronto il sotto-campione dell'indagine IBF-I costituito dai soli intervistati che hanno dichiarato di utilizzare Internet. Inoltre, i pesi campionari per l'indagine WEBIT sono stati allineati per quanto possibile a quelli dell'indagine IBF-I in modo che le distribuzioni dei rispondenti in termini di età, sesso, area geografica, titolo di studio ed occupazione fra le due indagini siano le stesse (si veda l'appendice B: i pesi campionari). Per quanto riguarda il confronto fra i valori del reddito dichiarati nelle due indagini, la robustezza dei risultati è stata anche testata conducendo le analisi su sottogruppi omogenei della popolazione individuati attraverso tecniche di *propensity score matching* (Rubin, 1974). Infine, un'ulteriore strategia

utilizzata è quella di confrontare a livello di singolo rispondente le informazioni raccolte in WEBIT con quelle contenute nella BDR.

4.3.1. La misurazione dei redditi da lavoro e da pensione

In entrambe le indagini sono state chieste informazioni riguardo alcune tipologie di redditi percepiti, distinti in reddito da lavoro dipendente, da pensione, da lavoro indipendente come, ad esempio, libero professionista o lavoratore autonomo. Anche in questo caso l'uso del web ha prodotto risultati differenti da quelli ottenuti con le interviste di persona.

La metodologia CAPI sembra fare emergere un numero maggiore di percettori di reddito da lavoro o da pensione (fig. 5 e 6). In particolare, circa il 38 per cento delle famiglie intervistate in IBF-I e la cui persona di riferimento¹¹ è in pensione, dichiara di avere almeno due percettori in famiglia di redditi da pensione. Questa percentuale scende al 18 per cento nell'indagine WEBIT. Un risultato analogo si ottiene per i percettori di reddito da lavoro dipendente: in IBF-I il 38 per cento delle famiglie CAPI (con persona di riferimento occupata) dichiara almeno due percettori contro una percentuale del 28 per cento nell'indagine CAWI. Una possibile spiegazione è che la presenza dell'intervistatore aiuti e incentivi il rispondente (che generalmente fornisce le risposte anche per tutti gli altri componenti della famiglia) a ricordare tutte le fonti di reddito percepite. Nell'indagine CAWI, senza l'insistenza dell'intervistatore, il rispondente probabilmente tende a dichiarare solo le fonti di reddito più importanti.

Considerando invece il valore delle fonti di reddito, la metodologia CAWI produce risultati sistematicamente più elevati di quella CAPI per quanto riguarda i redditi da lavoro dipendente e da pensione (fig. 7 e 9). Per i redditi da lavoro autonomo invece la tecnica CAPI produce valori più elevati (fig. 8). In quest'ultimo caso, rispetto alle altre fonti reddito, la rilevazione risulta comunque più complessa in quanto i redditi autonomi sono maggiormente soggetti ad evasione fiscale. Il fatto che i rispondenti all'indagine IBF-I abbiano già partecipato a qualche edizione precedente, insieme alla presenza dell'intervistatore che è istruito per tranquillizzare il rispondente sulla riservatezza dei dati forniti, contribuisce a spiegare i più alti valori ottenuti con il metodo CAPI.

Come analisi di robustezza sono state utilizzate tecniche di *matching* statistico¹² per eliminare il più possibile l'effetto della mancata risposta. Le tavole 8 e 9 mostrano i risultati che si ottengono confrontando l'indagine WEBIT con dei sotto-campioni dell'indagine IBF-I formati da rispondenti che hanno caratteristiche socio-demografiche (età, sesso, area geografica e titolo di studio, occupazione, tipologia familiare e titolo di godimento dell'abitazione di residenza) molto simili a quelli dell'indagine WEBIT. Per i redditi da lavoro dipendente il valore medio dichiarato in WEBIT risulta significativamente superiore a quello di IBF-I, mentre per i redditi autonomi si osserva la situazione opposta. Infine, per i redditi da pensione risulta confermato che il metodo CAPI consenta di ottenere un maggior numero di percettori.

¹¹ In entrambe le indagini la persona di riferimento è definita come la persona maggiormente informata sull'economia familiare.

¹² Il *matching* statistico si basa sull'idea di abbinare a ciascun soggetto "trattato" un soggetto non trattato tendenzialmente equivalente, ovvero il più simile possibile. Questo avviene selezionando, attraverso una funzione di distanza prescelta, le unità che hanno caratteristiche osservabili il più simile possibile a quelle di interesse. In questo modo si crea un gruppo di controllo in un contesto che ricrea, almeno nelle ipotesi di lavoro, la situazione sperimentale.

Una ulteriore valutazione dell'indagine CAWI è stata ottenuta integrandola con la Base dati reddituale (BDR), che contiene le informazioni di fonte fiscale dei modelli di dichiarazione d'imposta (modello Unico, 730, 770). L'aggancio è stato realizzato solo per l'intestatario della scheda familiare (per gli altri componenti dichiarati nella famiglia non si è potuto effettuare un abbinamento univoco con gli altri individui presenti nelle anagrafi) ed è stato possibile per il 95 per cento dei rispondenti.

L'analisi sul numero di percettori mostra come le due fonti non risultino del tutto allineate. In particolare, le percentuali di individui che risultano avere un reddito nella BDR ma che non lo hanno dichiarato nell'indagine web oscilla intorno all'11 per cento per i redditi da lavoro dipendente e da pensione e al 35 per cento per i redditi da lavoro autonomo (tav. 10). Queste differenze possono dipendere, oltre che da fenomeni di under-reporting, anche da differenze definitorie (soprattutto nel caso degli autonomi) e dalla presenza di alcune categorie come i lavori atipici che nell'indagine web non sono facilmente individuabili (e classificabili come autonomi in coerenza con la definizione fiscale). D'altro lato, la quota di percettori presenti nell'indagine ma non nella BDR è pari a quasi il 2 per cento per i redditi da pensione, al 5 per cento per i redditi da lavoro dipendente e a quasi il 48 per cento per i redditi da lavoro autonomo.

Per poter confrontare i valori puntuali dei redditi percepiti, quelli presenti nella BDR sono stati trasformati in redditi netti ripartendo l'imposta complessiva proporzionalmente al valore di tali redditi. Considerando anche il margine di errore dovuto a questa approssimazione, i valori delle due fonti risultano in linea per quanto riguarda i redditi da lavoro dipendente e quelli da pensione (tav. 10). Le differenze oscillano intorno al 5 per cento per i redditi da lavoro dipendente e al 2 per cento per quelli da pensione. Si confermano differenze rilevanti invece per i redditi da lavoro indipendente, che probabilmente risentono in modo rilevante delle diverse definizioni adottate nelle due fonti. Questo risultato deve comunque essere interpretato con cautela data la bassa numerosità del campione.

In termini di variabilità delle stime dei redditi, le stime basate sull'indagine CAWI mostrano coefficienti di variazione superiori rispetto a quelli relativi all'indagine CAPI per tutte le tipologie di reddito quando si considerano le stime non pesate. Il risultato non è invece univoco quando si introducono i pesi: in questo caso la variabilità delle stime CAWI è inferiore per i redditi da lavoro autonomo (tav. 11).

4.3.2. Le valutazioni qualitative sulle condizioni economiche della famiglia

Una seconda serie di confronti riguarda le domande sulla autovalutazione (qualitativa) della propria condizione economica.

La figura 10 mostra la distribuzione delle risposte alla domanda sulle aspettative future sul reddito complessivo familiare: "Pensando a tutte le fonti di reddito (da lavoro e da pensione, affitti, rendite da capitale ecc.), nel 2016 prevede che il reddito complessivo della sua famiglia sarà più alto, uguale o più basso rispetto a quello del 2015?". Nell'indagine IBF-I, l'81% dei rispondenti utilizza la modalità di risposta più «neutra», dichiarando che il reddito rimarrà sostanzialmente invariato. Nell'indagine WEBIT tale percentuale scende al 64,5%. In modalità web, inoltre, molte più famiglie dichiarano di prevedere un peggioramento del proprio reddito (quasi il 18% rispetto all'8,4% del CAPI), forse sentendosi maggiormente rassicurate dal non dover dare questa risposta di fronte ad un estraneo. Bisogna tuttavia notare che nel web si riscontra una più elevata non risposta, pari al 10% di "Non so/Non risponde" rispetto al 3,5% nel CAPI, che verosimilmente viene limitata dalla presenza dell'intervistatore.

Risultati analoghi emergono per la domanda sui risparmi o i debiti della famiglia nel corso del 2015: “Considerando tutte le spese sostenute e tutte le fonti di reddito, la sua famiglia ha speso meno (risparmiando), tutto (senza riuscire a risparmiare nulla) o più del reddito annuo (indebitandosi)?”. Anche in questo caso, nelle interviste CAPI le famiglie hanno utilizzato con maggior frequenza l’opzione di risposta più neutrale (il 63,6% dichiara che le spese sono state uguali alle entrate, rispetto al 46,9% del CAWI). Compilando il questionario online, invece, le famiglie sembrano «aprirsi» di più, dichiarando con maggior frequenza di aver speso più del reddito annuo (14,0% contro il 6,5%) o anche di essere riuscite a mettere da parte qualche risparmio (39,1% contro il 30%) (fig. 11).

Analogamente, nel confronto delle spese correnti rispetto a quelle dell’anno precedente, una percentuale molto alta dei rispondenti CAPI ha dichiarato che non ci sono state variazioni (82% rispetto al 38,5%), mentre con il CAWI molte più famiglie hanno dichiarato un aumento delle spese (48,3% contro il 15,2%) (fig. 12). Si conferma nel CAWI una maggiore frequenza del Non so/non risponde (7,7% contro l’1,0%).

Anche l’informazione sugli aiuti economici (e non) ricevuti sembrerebbe misurata in modo più efficace con la modalità web: nel CAPI la quota di chi ha ricevuto aiuti negli ultimi 3 anni è meno della metà di quella registrata per il web (6,5% contro il 17%, fig. 13). La differenza fra i due metodi emerge anche per i debiti nei confronti di banche: se con il CAPI li dichiara il 23% degli intervistati, con il CAWI la stessa percentuale sale a quasi il 48% (fig. 14).

In conclusione, una prima differenza rilevante fra i due metodi è che nell’indagine WEB i rispondenti tendono a fornire con minore frequenza risposte «neutrali». Questo risultato potrebbe essere attribuibile al fatto che lo strumento web consentirebbe ai rispondenti di dichiarare più liberamente le proprie valutazioni, anche riportando situazioni economiche più difficili per la famiglia, in quanto l’assenza dell’intervistatore permette di evitare fenomeni di *misreporting* legati al *social desirability bias*. La presenza di un intervistatore, infatti, pur garantendo una minor presenza di mancate risposte, sembra però favorire l’aumento della scelta delle modalità centrali.

4.3.3. L’effetto delle modalità di risposta “Non so” / “Non intendo rispondere”

Nell’indagine è stato anche sperimentato l’uso delle modalità di risposta “Non so” / “Non intendo rispondere” per valutarne gli effetti sulla qualità dei dati. La scelta se sia meglio offrire o meno tali opzioni è ampiamente discussa in letteratura. Da un lato, “obbligare” la famiglia a dare una risposta per poter continuare l’indagine ha il vantaggio di spingere alcuni individui a fare uno sforzo per recuperare almeno nella propria memoria le informazioni necessarie per rispondere. Anche se poi venisse fornita una risposta imprecisa, questa sarà più informativa di qualsiasi modello di imputazione impiegato ex-post. Lo svantaggio è che alcuni rispondenti potrebbero sentirsi “incentivati/autorizzati” a fornire delle risposte totalmente inventate. Dare la possibilità di dire “Non so” potrebbe dunque aiutare ad ottenere risposte più sincere. Il rovescio della medaglia è che questa opzione di risposta potrebbe però rappresentare una facile scorciatoia sia per il rispondente sia per l’intervistatore per ridurre il tempo dell’intervista minimizzando il proprio onere di risposta. La prevalenza degli studi disponibili in letteratura sembrano concordi nell’indicare che non sia consigliabile includere la modalità “Non so”/“Non risponde” (per una rassegna di questi studi si veda Krosnick e Presser 2010).

L’effetto della presenza delle modalità “Non so” e “Non risponde” è stata testata in due domande qualitative, relative all’andamento del reddito familiare e al valore dell’abitazione di residenza.

Per quanto riguarda la domanda relativa al reddito, nella versione del questionario in cui era obbligatorio scegliere una delle tre modalità possibili, quasi il 70 per cento ha risposto che il reddito è rimasto sostanzialmente uguale, a differenza dell'altro questionario dove la stessa percentuale scende al 62,3 per cento (fig. 15). Si noti che in questa seconda versione, vi è anche un 4,5 per cento di rispondenti che dichiarano di non sapere in che modo sia variato il proprio reddito. Le modalità "Più alto" e "Più basso" mostrano invece frequenze più simili nei due questionari. Sembrerebbe dunque che quando non vi sia la possibilità di indicare "non so", venga scelta la modalità più "neutrale", in questo caso dichiarando che il reddito è stato uguale a quello dell'anno precedente.

La domanda sull'andamento del valore della propria abitazione mostra differenze più rilevanti fra le due versioni. La percentuale di famiglie che ha dichiarato che il valore rimarrà invariato è pari, infatti, al 75,4% quando non viene data l'opzione di risposta "Non so", contro il 55,1% dell'altra versione, dove si riscontra una quota di "non so" che potrebbe spiegare la differenza osservata (pari al 16%) (fig. 16).

I risultati della sperimentazione indicano in conclusione, che sarebbe opportuno inserire fra le opzioni di risposta il "Non so", almeno nei quesiti dove rappresenta una modalità plausibile, come nei casi in esame, dove non è sempre facile e immediato confrontare il proprio reddito con quello dell'anno precedente o dare un giudizio sul valore futuro della propria abitazione. Per queste tipologie di domande, escludendo questa opzione di risposta, potrebbe essere sovrastimata la modalità di risposta più "neutrale".

5. Le misure adottate per aumentare il tasso di partecipazione

L'uso di incentivi per aumentare i tassi di risposta rappresenta una strategia comunemente adottata nelle indagini web. Vari studi hanno mostrato come gli incentivi abbiano un effetto positivo sulla partecipazione e che tale efficacia aumenti per quelli monetari e quelli offerti in anticipo, indipendentemente dalla partecipazione all'indagine (per una raccolta degli studi relativi alle indagini web si vedano Göritz 2006 e 2010). Questi risultati sono sostanzialmente gli stessi trovati anche per indagini condotte con modalità diversa (si veda ad esempio la rassegna di Singer e Ye, 2012). Ciò nonostante, gli incentivi monetari e anticipati non sono molto frequenti nella pratica. Una possibile ragione è che i tassi di partecipazione che normalmente caratterizzano le indagini web sono talmente contenuti che i benefici che possono ottenersi da un loro incremento vengono in genere ritenuti inferiori ai costi (Tourangeau et. al. 2013).

Sulla base di queste considerazioni e trattandosi comunque di una indagine sperimentale si è deciso per l'indagine WEBIT di fornire un incentivo, atto a scongiurare tassi di risposta troppo bassi, pur prestando attenzione agli aspetti economici.

L'incentivo ha prodotto modesti effetti positivi: nel complesso, l'incremento della probabilità di partecipazione è di circa 0,5 punti percentuali, mentre la riduzione della probabilità di interruzione è di circa un punto percentuale (fig. 17); entrambi i valori non risultano però significativamente diversi da zero. L'incentivo è risultato comunque più efficace per alcuni sottogruppi della popolazione. Effetti significativi sono ad esempio emersi per le famiglie che vivono nel Nord Ovest, nelle Isole e per quelle che vivono in piccoli comuni. Inoltre, l'incentivo è risultato più efficace per segmenti della popolazione potenzialmente più interessati a strumenti tecnologici (come i maschi) o che hanno probabilmente pensato all'incentivo come ad un regalo per un parente al di fuori della famiglia, come ad esempio può essere accaduto per i single con oltre 65 anni o per le famiglie senza figli (fig. 18).

L'altra strategia adottata per sensibilizzare le famiglie è stata quella di usare dei solleciti sia telefonici (possibili per il 30 per cento circa delle famiglie campione, che disponevano di un recapito telefonico) sia per email (per circa l'11 per cento delle famiglie che si sono registrate al sito dell'indagine). Inoltre, a fine maggio, è stata inviata una seconda lettera a circa 1.000 famiglie scelte in modo casuale fra i non rispondenti.

Le telefonate sono state svolte da personale Istat, preventivamente formato sulla fase di sollecito e sui principali contenuti dell'indagine, utilizzando sia il laboratorio dell'Istituto deputato ai ricontatti telefonici che le singole postazioni di lavoro collegate in remoto al server del laboratorio. I solleciti telefonici sono necessariamente avvenuti in orario di ufficio. Tale vincolo organizzativo ha ovviamente diminuito l'efficacia della procedura di sollecito che ha registrato un numero elevato di telefoni liberi nonché di rifiuti per impedimenti a collaborare (mancanza di Internet a casa) soprattutto da parte di persone anziane maggiormente reperibili a casa negli orari diurni e pomeridiani.

I solleciti telefonici hanno raggiunto il 65,3% delle famiglie con telefono, ma solo con poco più della metà di queste (53,2%) è stato possibile entrare in contatto. Gli esiti 'Libero/Non risponde nessuno', 'Rifiuto' e 'Non ha internet' mostrano le percentuali più elevate (rispettivamente 34, 23 e 17 per cento). Il principale effetto dei solleciti telefonici è stato quello di indurre le persone a registrarsi al sito della rilevazione (circa il 9% dei contattati telefonicamente). Allo 'stimolo' a registrarsi non ha fatto seguito però un altrettanto forte impulso alla compilazione: il 4,7% dei sollecitati, infatti, ha inviato il questionario solo parzialmente compilato. I solleciti tramite email hanno raggiunto circa un terzo (32,1%) dei registrati al sito dell'indagine. L'invito a completare il questionario è stato accolto da circa il 15% dei sollecitati.

Grazie all'azione congiunta di tutte le forme di sollecito, inclusa la lettera di preavviso, il trend di compilazione dei questionari web è aumentato nel tempo (fig. 19) con una velocità di crescita che ha subito un naturale rallentamento dopo circa un mese e mezzo dall'apertura della rilevazione. Un lieve impulso alla compilazione si registra intorno alla data di arrivo della lettera di sollecito (fine maggio 2016.)

6. Conclusioni

Il lavoro ha esplorato i vantaggi e gli svantaggi dell'uso di Internet come strumento di rilevazione nelle indagini sul reddito e la ricchezza. L'analisi si è concentrata sui principali aspetti problematici del metodo ossia quelli di copertura, di mancata partecipazione e di misurazione.

Attualmente, lo strumento web esclude dal campo di rilevazione il 30 per cento circa delle famiglie italiane (pari a 7,7 milioni); la probabilità di usare Internet cresce all'aumentare del titolo di studio e del reddito. Il problema di mancata copertura non sembra comunque rappresentare un ostacolo insormontabile, soprattutto in chiave prospettica. L'uso di Internet in Italia risulta infatti in forte crescita. Inoltre, già ad oggi non esistono intere tipologie di famiglie (definite in base a caratteristiche socio-demografiche) totalmente irraggiungibili. È infatti sufficiente la presenza di un componente della famiglia che sappia usare Internet per poter partecipare all'indagine. Infine, disponendo di informazioni ausiliarie, come quelle di fonte fiscale e quelle sociodemografiche, è comunque possibile ridurre a posteriori gli effetti distorsivi della mancata copertura.

I problemi principali dello strumento sono la difficoltà di ottenere soddisfacenti tassi di partecipazione e il fatto che questi dipendano anche dalla condizione economica delle famiglie. L'uso di un concorso a premi per la vincita di 5 Ipad è risultato solo marginalmente

di aiuto ed esclusivamente per determinati sottogruppi della popolazione. Anche l'attività di sollecito telefonico ha dato un contributo modesto, probabilmente perché realizzata solo in orario di ufficio.

Un aspetto positivo che comunque è emerso dall'indagine è che, contrariamente a quanto avviene in presenza di un intervistatore, lo strumento web sembra adatto ad intervistare le famiglie economicamente più agiate. Una prima ragione potrebbe dipendere dal loro livello di istruzione, che è in genere più elevato della media, e dal fatto che possiedono anche le competenze informatiche necessarie per partecipare ad una indagine web. Inoltre, la possibilità di poter compilare il questionario nei momenti desiderati, senza alcuna pressione e anche nell'arco di più giorni, può far venire meno la resistenza di coloro che non riescono a trovare il tempo da dedicare con continuità ad un intervistatore. Il rovescio della medaglia è naturalmente rappresentato dalla necessità di dover disporre di informazioni ausiliarie, come ad esempio quelle di fonte fiscale, per bilanciare la tendenza del campione a sovra-rappresentare le famiglie economicamente più agiate.

Per quanto riguarda gli errori di misurazione, il comportamento dei rispondenti nell'indagine web sembra contribuire a ridurre l'errore totale delle stime per alcune tipologie di domande. In particolare, i valori pro-capite dei redditi dichiarati nella web risultano sistematicamente più elevati di quelli dell'indagine CAPI e più vicini a quelli di fonte fiscale, anche dopo aver controllato per la maggior probabilità delle famiglie abbienti di partecipare all'indagine web. Una possibile spiegazione è che questo strumento permetta ai rispondenti di recuperare l'informazione richiesta con tranquillità sia attraverso documenti (come ad esempio la dichiarazione dei redditi) sia coinvolgendo altri componenti della famiglia. I migliori risultati dello strumento web sono resi robusti anche dal fatto che l'indagine CAPI è condotta su famiglie panel, tipicamente caratterizzate da maggior fiducia verso gli intervistatori e che dunque non dovrebbero sottostimare in modo rilevante i redditi effettivamente percepiti. Inoltre, l'assenza di un intervistatore sembra mettere in grado gli individui di rispondere più liberamente a domande qualitative: in particolare, i rispondenti all'indagine WEBIT tendono sistematicamente ad utilizzare meno le opzioni di risposta più neutrali alle domande sulle proprie condizioni economiche, segnalando inoltre con maggior frequenza situazioni di difficoltà.

Il web sembra invece meno adatto per coloro che percepiscono redditi "più sensibili" come quelli da lavoro autonomo. Questo risultato, anche se va interpretato con cautela in quanto basato su poche osservazioni, potrebbe essere dovuto alla capacità degli intervistatori di rassicurare sulla riservatezza delle risposte e alla familiarità con l'indagine (trattandosi di famiglie panel). Inoltre, la presenza fisica dell'intervistatore presso l'abitazione dell'intervistato può rendere più difficile nascondere il proprio tenore di vita e quindi dichiarare redditi troppo lontani da quelli effettivamente percepiti. Inoltre, occorre anche considerare che nell'indagine WEBIT, a differenza dell'indagine IBF-I, i rispondenti hanno ricevuto una lettera di invito personale (con il loro nome e cognome) e ciò potrebbe aver creato qualche preoccupazione sulla riservatezza delle loro risposte (nonostante le rassicurazioni fornite).

Un altro limite della tecnica web è che il numero di percettori dichiarati appare essere inferiore a quello dell'indagine CAPI. Questo risultato, che appare robusto soprattutto per le pensioni, potrebbe dipendere dal fatto che, in assenza di intervistatore che "incalzi" i rispondenti ricordandogli di dichiarare ogni (seppur minima) fonte di reddito, questi tendano a limitarsi a riportare solamente quelle principali, nel tentativo di concludere l'indagine il prima possibile.

In generale quindi, anche se l'utilizzo degli intervistatori implica maggiori costi e la possibilità di introdurre errori, il loro contributo nel convincere le famiglie selezionate a partecipare all'indagine, nel persuadere, chiarire e motivare gli intervistati durante l'intervista, risulta ad oggi difficilmente sostituibile. È invece immaginabile uno sviluppo di tecniche miste di rilevazione che permettano di sfruttare i punti di forza di entrambe. Ad esempio, nel caso dell'indagine sui bilanci delle famiglie che viene svolta ogni due anni dalla Banca d'Italia, lo strumento web potrebbe essere impiegato negli anni intermedi per svolgere indagini qualitative su famiglie che hanno già partecipato alla rilevazione principale. Tali famiglie hanno in genere una maggiore propensione a collaborare e, inoltre, la disponibilità di informazioni ausiliarie presenti nell'indagine principale permetterebbe di tenere sotto controllo i fenomeni di mancata copertura e di non risposta.

Bibliografia

Bagozzi, R.P. (1994). Measurement in marketing research: Basic principles of questionnaire design. *Principles of marketing research*, 1, 1-49.

Banca d'Italia (2015), I bilanci delle famiglie italiane nell'anno 2014, Supplementi al Bollettino Statistico, 64.

Bethlehem, J.G., (1988), Reduction of the nonresponse bias through regression estimation. *J. Official Statist.*, 4, 251–260.

Bethlehem, J.G., (2002), Weighting nonresponse adjustments based on auxiliary information. In *Survey Nonresponse*, Eds. R.M. Groves, D.A. Dillman, J.L. Eltinge & R.J.A. Little. New York: Wiley & Sons.

Bethlehem, J., (2010), Selection Bias in Web Surveys, *International Statistical Review*, 78, issue 2, p. 161-188.

Couper, M.P., (2000), Web surveys: A review of issues and approaches. *The Public Opinion Quarterly*, 64(4), 464-494.

Couper, M.P., (2011), The Future of Modes of Data Collection, *Public Opinion Quarterly*, Volume 75, Issue 5, 889–908, <https://doi.org/10.1093/poq/nfr046>.

D'Alessio, G., (2017), Measurement errors in consumption surveys and the estimation of poverty and inequality indices, Temi di discussione (Working paper), Bank of Italy, 1116.

D'Alessio, G., Iezzi, S., (2015), How the time of interviews affects estimates of income and wealth, *Questioni di Economia e Finanza*, Bank of Italy, 273.

Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., Messer, B. L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social science research*, 38(1), 1-18.

Fricker, S., Galesic, M., Tourangeau, R., Yan, T. (2005). An experimental comparison of web and telephone surveys. *Public Opinion Quarterly*, 69(3), 370-392.

Görizt, A. S. (2006). Incentives in web studies: Methodological issues and a review. *International Journal of Internet Science*, 1(1), 58-70.

Görizt, A. S. (2010). Using lotteries, loyalty points, and other incentives to increase participant response and completion.

Groves, R.M., (2004), *Survey errors and survey costs* (Vol. 536). John Wiley & Sons.

Groves, R.M., (2004), *Survey Methodology*, John Wiley & Sons.

Groves, R.M., Peytcheva, E., (2008), The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis, *Public Opinion Quarterly*, Volume 72, Issue 2, 1 January 2008, Pages 167–189.

Heckathorn, D.D., (2002), Respondent-Driven Sampling II: Deriving Valid Estimates from Chain-Referral Samples of Hidden Populations, in *Social Problems*. 49 (1): 11–34. doi:10.1525/sp.2002.49.1.11.

Istat (2016), Cittadini, imprese e ICT, Statistiche <https://www.istat.it/it/archivio/194611>.

Jäckle, A., Lynn, P., Burton, J. (2015). Going online with a face-to-face household panel: Effects of a mixed mode design on item and unit non-response. In *Survey Research Methods* (Vol. 9, No. 1, pp. 57-70).

Kennickell, A.B., (2009), Getting to the Top: Reaching Wealthy Respondents in the SCF, <https://www.federalreserve.gov/econresdata/scf/files/ASA200911.pdf>

Kreuter, F., Presser, S., Tourangeau, R. (2008). Social Desirability Bias in CATI, IVR, and Web Surveys The Effects of Mode and Question Sensitivity. *Public opinion quarterly*, 72(5), 847-865.

Krosnick, J.A, Presser, S., (2010), Question and questionnaire design, in Marsden, P.V., J.D. Wright. 2010. *Handbook of survey research*. Bingley, UK: Emerald.

Martin, P., Lynn, P. (2011). *The effects of mixed mode survey designs on simple and complex analyses* (No. 2011-28). ISER Working Paper Series.

Manfreda, K.L., Berzelak, J., Vehovar, V., Bosnjak, M., Haas, I. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *International journal of market research*, 50(1), 79-104.

Neri, A., Rannalli, M.G., (2011), To misreport or not to report? The case of the Italian Survey on Household Income and Wealth. *Statistics in Transition new series*, 12(2), 281-300.

Pitrone, M.C., (2009), Sondaggi e interviste. Lo studio dell'opinione pubblica nella ricerca sociale, Franco Angeli, Metodologia delle scienze umane.

Poynter, R., (2010), *The Handbook of Online and Social Media Research*, Amsterdam: ESOMAR

Pudney, S., (2008), Heaping and leaping: Survey response behavior and the dynamics of self-reported consumption expenditure, Institute for Social and Economic Research, ISER working papers.

Roberts, C., (2007). *Mixing modes of data collection in surveys: A methodological review*, Southampton: ESRC National Centre for Research Methods. NCRM Methods Review Papers, NCRM/008.

Rubin, D. (1974), Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701.

Salganik, M.J., Heckathorn D.D., (2004). Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling, *Sociological Methodology*. 34 (1): 193–239. doi:10.1111/j.0081-1750.2004.00152.x.

Simon, H.A., (1957), *Models of Man*, New York: Wiley.

Singer, E., Ye, C. (2013). The use and effects of incentives in surveys. *The ANNALS of the American Academy of Political and Social Science*, 645(1), 112-141.

Teppa F. (2012), The CentERpanel and the DNB Household Survey: Methodological Aspects, *DNB Occasional Studies*, vol.10, No 4
https://www.dnb.nl/binaries/DNB_OS_1004_BIN_WEB_tcm46-277691.pdf

Tourangeau, R., Conrad, F. and Couper, M., (2013), *The Science of Web Surveys*, OUP Oxford.

Tourangeau, R., Rips, J.L., Rasinski, K., (2000), *The Psychology of Survey Response*, Cambridge University Press.

Tourangeau, R., Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859–883.

Zhang, C., (2013), *Satisficing in Web Surveys: Implications for Data Quality and Strategies for Reduction*, Unpublished doctoral dissertation, University of Michigan, Ann Arbor, MI.

Appendice A - Tavole e figure

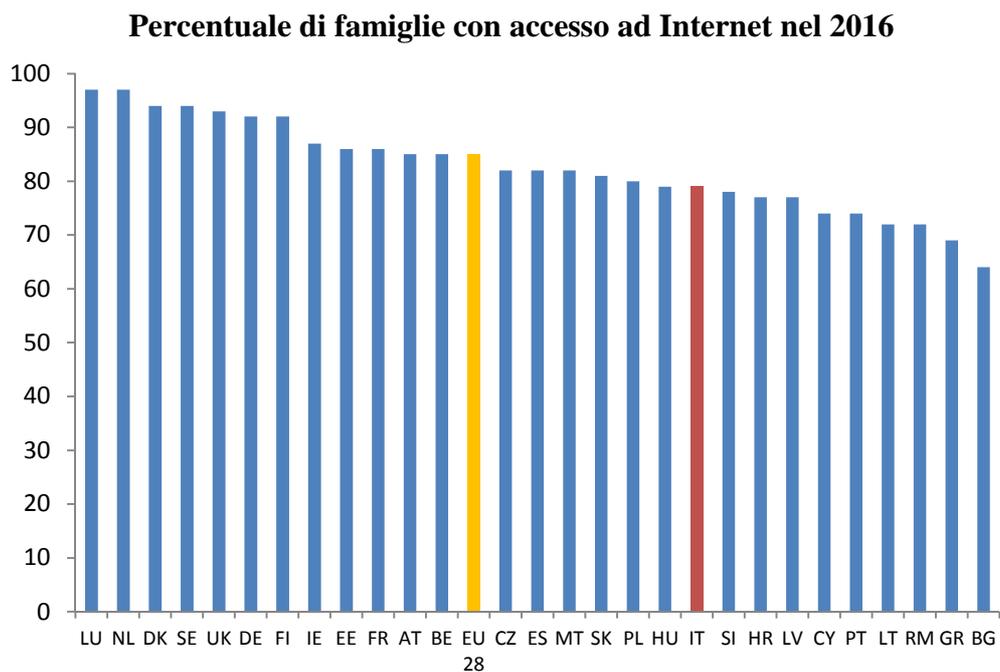
Tavola 1

Composizione del campione teorico e di quello realizzato per l'indagine web (Valori percentuali, statistiche non pesate)

		Campione	
		Teorico	Realizzato
Sesso*	Donna	35,99	27,90
	Uomo	64,01	72,10
Età*	Fino a 34 anni	7,11	6,97
	35-44 anni	15,84	17,14
	45-54 anni	20,52	22,93
	55-64 anni	17,98	21,87
	Oltre 64 anni	38,55	31,09
Giovane in famiglia**		51,44	60,28
Componenti famiglia	1	33,38	27,30
	2	26,33	24,70
	3	19,22	24,35
	4	14,99	18,91
	5 e più	6,08	4,73
Figli in famiglia		44,92	52,48
Coniugato*		47,84	56,97
Quintili di reddito (fiscale) familiare	Non in BDR***	11,34	5,44
	Primo	17,74	8,04
	Secondo	17,75	10,99
	Terzo	17,72	16,90
	Quarto	17,74	24,82
	Quinto	17,73	33,81
Possesso altri immobili		35,62	45,15
Area geografica	Nord	42,42	51,89
	Centro	21,59	21,87
	Sud e Isole	35,99	26,24
Ampiezza comune (n. abitanti)	fino a 20000	10,77	10,99
	20000-40000	16,55	15,84
	40000-500000	63,71	62,65
	oltre 500000	8,97	10,52
Cittadinanza*	Italiana	92,46	99,17
Totale		100,00	100,00

* Caratteristiche riferite all'intestatario della scheda anagrafica. ** Famiglia con almeno un componente con meno di 40 anni. *** Banca Dati Reddituale, ossia il registro di dati di natura fiscale con cui sono stati confrontati i dati dell'indagine.

Figura 1



Fonte: Eurostat, indagini annuali sull'uso di Information and Communication Technology in famiglia. Famiglie che contengono almeno una persona fra 16 e 74 anni.

Diffusione di Internet nella popolazione
(Valori percentuali)

Caratteristiche*		Famiglie che utilizzano Internet	Famiglie che non utilizzano Internet
		Percentuali	
Sesso	Uomo	73,3	26,7
	Donna	60,9	39,1
Età	fino a 34 anni	94,2	5,8
	da 35 a 44 anni	90,5	9,6
	da 45 a 54 anni	91,0	9,0
	da 55 a 64 anni	78,0	22,0
	oltre 64 anni	34,4	65,7
Condizione professionale	lavoratore dipendente	91,1	8,9
	lavoratore autonomo	92,7	7,3
	in condizione non professionale	40,9	59,1
Titolo di studio	Nessuno	5,7	94,3
	Licenza Elementare	18,0	82,0
	Licenza Media Inferiore	74,8	25,2
	Diploma Scuola Superiore	88,0	12,0
	Università	95,7	4,3
Quintili di reddito familiare	Primo	40,6	59,4
	Secondo	58,5	41,6
	Terzo	71,2	28,8
	Quarto	85,8	14,2
	Quinto	93,8	6,3
Quintili di ricchezza familiare	Primo	61,7	38,3
	Secondo	66,9	33,2
	Terzo	63,1	36,9
	Quarto	77,8	22,2
	Quinto	77,5	22,5
Area Geografica	Nord	75,2	24,8
	Centro	70,6	29,4
	Sud e Isole	60,1	39,9
Ampiezza Comunale	fino a 20000 abitanti	63,0	37,0
	da 20000 a 40000 abitanti	71,3	28,7
	da 40000 a 500000 abitanti	73,5	26,5
	oltre 500000 abitanti	79,9	20,1
Totale		69,4	30,6

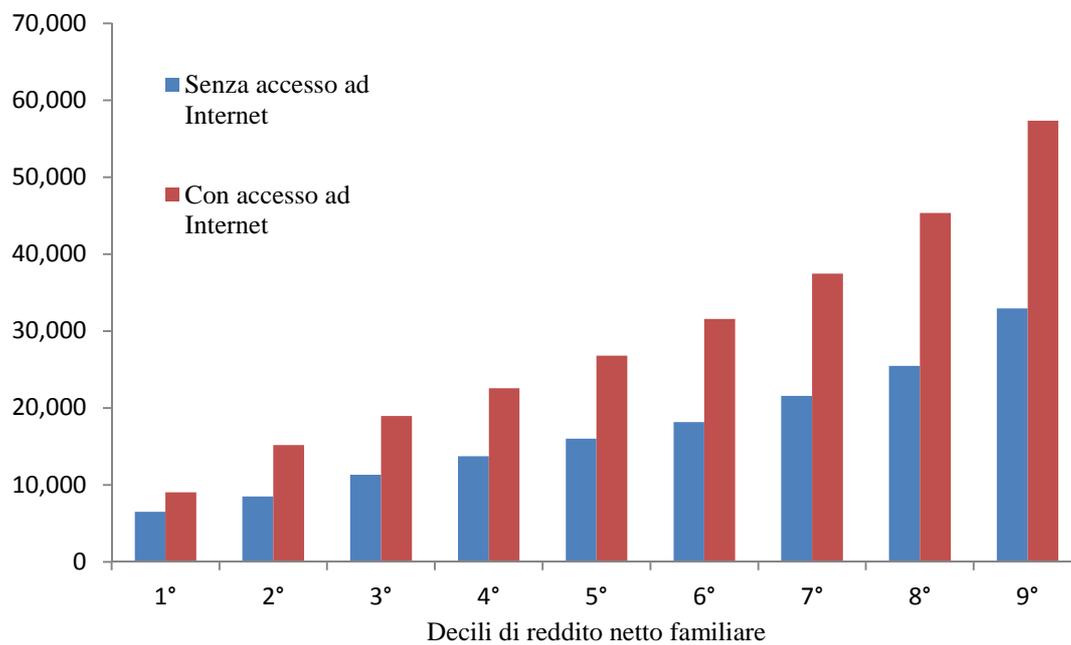
Fonte: Indagine IBF-I. Le caratteristiche individuali sono riferite al capofamiglia, definito come il maggiore percettore di reddito.

Probabilità di avere accesso ad Internet
(*modello logistico*)

Caratteristiche*		Parametro	P-value
Intercetta		0,694	0,3562
Sesso*	Uomo	0,141	0,3923
	Donna	--	--
Età*	Fino a 34 anni	1,881	<,0001
	35-44 anni	2,021	<,0001
	45-54 anni	2,503	<,0001
	55-64 anni	1,492	<,0001
	Oltre 64 anni	--	--
Giovane in famiglia		1,790	<,0001
Componenti famiglia	1	0,656	0,2515
	2	0,559	0,2273
	3	0,089	0,8216
	4	0,323	0,4399
	5 e più	--	--
Figli in famiglia		0,315	0,3305
Coniugato		0,000	0,9989
Quintili di reddito (fiscale) familiare	Primo	-3,489	<,0001
	Secondo	-2,530	<,0001
	Terzo	-1,946	<,0001
	Quarto	-1,203	<,0001
	Quinto	--	--
Possesso altri immobili		0,421	0,0167
Area geografica	Nord	0,398	0,017
	Centro	0,178	0,3518
	Sud e Isole	--	--
Ampiezza comune (n. abitanti)	fino a 20000	-1,175	0,0004
	20000-40000	-0,867	0,0048
	40000-500000	-0,803	0,003
	oltre 500000	--	--
Cittadinanza*	Italiana	0,301	0,3653

Fonte: Indagine IBF-I. Le caratteristiche individuali sono riferite al capofamiglia, definito come il maggiore percettore di reddito. Pseudo $R^2=0,35$

Decili di reddito netto familiare per condizione di accesso ad Internet
(euro)



Probabilità di completare l'indagine web
(*modello logistico*)

Caratteristiche*		Parametro	Pr > ChiSq
Intercetta		-4,621	<,0001
Sesso*	Uomo	0,023	0,8177
	Donna	--	--
Età*	Fino a 34 anni	-0,074	0,7123
	35-44 anni	-0,163	0,2252
	45-54 anni	-0,265	0,0187
	55-64 anni	-0,142	0,2001
	Oltre 64 anni	--	--
Giovane in famiglia		-0,033	0,8171
Componenti famiglia	1	1,075	0,0003
	2	0,528	0,0255
	3	0,459	0,0129
	4	0,368	0,0497
	5 e più	--	--
Figli in famiglia		0,153	0,3885
Coniugato		0,351	0,0073
Quintili di reddito (fiscale) familiare	Primo	-0,698	<,0001
	Secondo	-0,611	<,0001
	Terzo	-0,443	0,0002
	Quarto	-0,186	0,0682
	Quinto	--	--
Possesso altri immobili		-0,054	0,5088
Area geografica	Nord	0,390	<,0001
	Centro	0,205	0,0591
	Sud e Isole	--	--
Ampiezza comune (n. abitanti)	fino a 20000	-0,014	0,934
	20000-40000	-0,098	0,5111
	40000-500000	-0,043	0,7335
	oltre 500000	--	--
Cittadinanza*	Italiana	2,177	<,0001

Le caratteristiche individuali sono riferite alla persona di riferimento, definita come la persona più informata sull'economia familiare. Le interruzioni sono considerate come mancate risposte. Il modello è stimato sugli elegibili.

Probabilità di partecipare* all'indagine web
(modello logistico)

Caratteristiche*		Parametro	Pr > ChiSq
Intercetta		-3,447	<,0001
Sesso*	Uomo	-0,071	0,4365
	Donna	--	--
Età*	Fino a 34 anni	0,250	0,1883
	35-44 anni	0,003	0,9791
	45-54 anni	0,030	0,7713
	55-64 anni	0,027	0,7941
	Oltre 64 anni	--	--
Giovane in famiglia		-0,099	0,456
Componenti famiglia	1	0,738	0,0061
	2	0,470	0,0283
	3	0,523	0,002
	4	0,369	0,0321
	5 e più	--	--
Figli in famiglia		-0,009	0,9539
Coniugato		0,253	0,0299
Quintili di reddito (fiscale) familiare	Primo	-0,737	<,0001
	Secondo	-0,791	<,0001
	Terzo	-0,534	<,0001
	Quarto	-0,308	0,0009
	Quinto	--	--
Possesso altri immobili		0,032	0,6704
Area geografica	Nord	0,322	0,0002
	Centro	0,143	0,1537
	Sud e Isole	--	--
Ampiezza comune (n. abitanti)	fino a 20000	-0,226	0,1505
	20000-40000	-0,284	0,0436
	40000-500000	-0,081	0,4877
	oltre 500000	--	--
Cittadinanza*	Italiana	1,614	<,0001

* Le interruzioni sono considerate come partecipazione all'indagine. Le caratteristiche individuali sono riferite alla persona di riferimento, , definita come la persona maggiormente informata sull'economia familiare. Il modello è stimato sugli elegibili.

Associazione fra reddito netto familiare e caratteristiche socio-demografiche
(variabile dipendente: logaritmo del reddito netto)

Caratteristiche*		Parametro	Pr > ChiSq
Intercetta		10,26	<.0001
Sesso*	Donna	-0,134	<.0001
	Uomo	--	.
Età*	Fino a 34 anni	-0,588	<.0001
	35-44 anni	-0,365	<.0001
	45-54 anni	-0,235	<.0001
	55-64 anni	-0,044	0.1694
	Oltre 64 anni	--	--
Giovane in famiglia		-0,167	<.0001
Componenti famiglia	1	-0,612	<.0001
	2	-0,294	<.0001
	3	-0,025	0,56
	4	0,038	0,39
	5 e più	--	--
Senza figli		0,066	0,14
Non coniugato		-0,146	<.0001
Area geografica	Nord	0,390	<.0001
	Centro	0,300	<.0001
	Sud e Isole	--	--
Ampiezza comune (n. abitanti)	fino a 20000	0,075	0,12
	20000-40000	0,060	0,16
	40000-500000	0,052	0,15
	oltre 500000	--	--
Cittadinanza*		0,762	<.0001
	Italiana		

Le caratteristiche individuali sono riferite all'intestatario della scheda anagrafica. Il modello è stimato sugli individui presenti nella BDR (n=8.823).
Pseudo R²=0,21.

Probabilità stimata di partecipare all'indagine per decili di reddito netto familiare

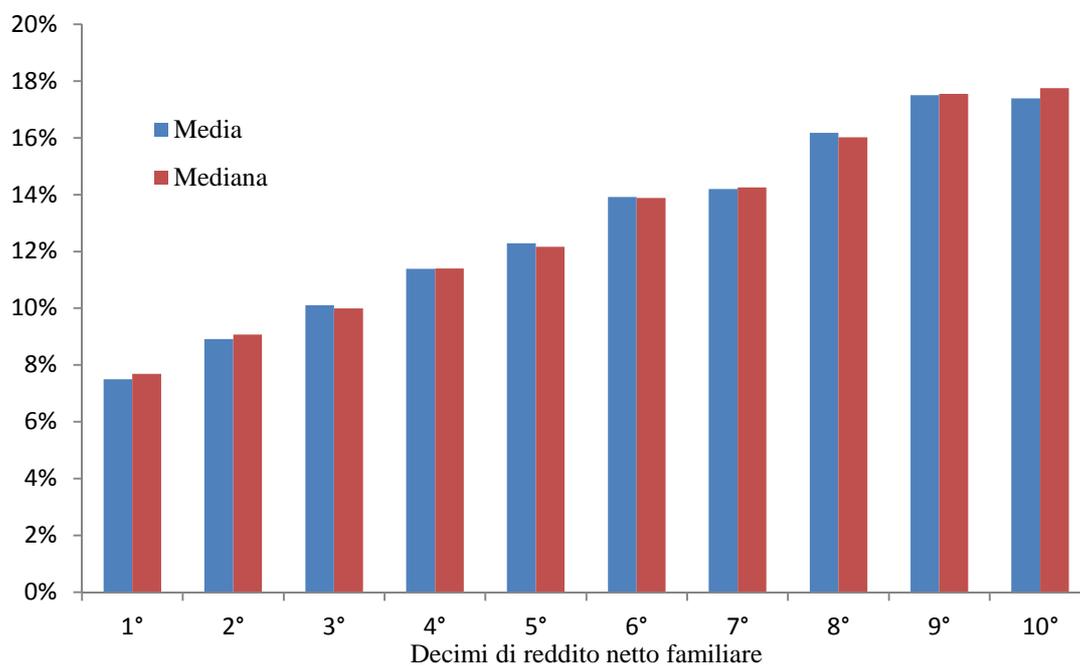
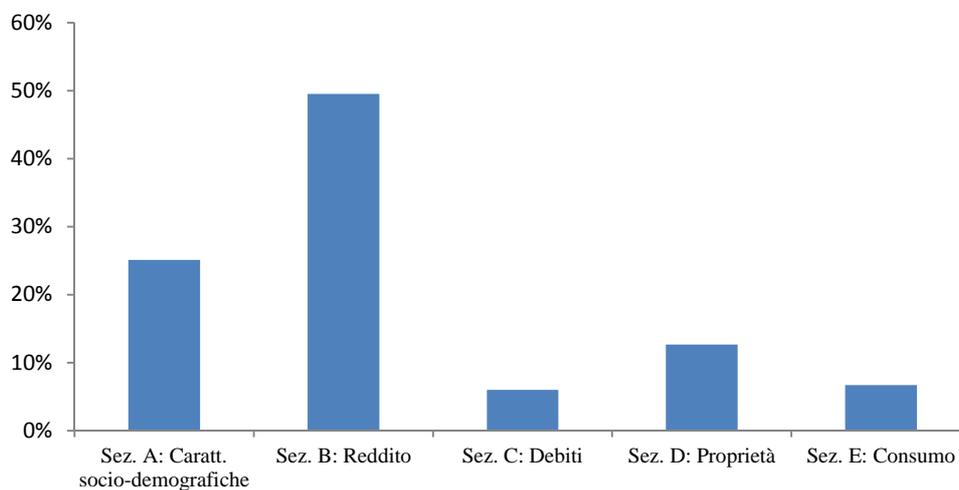


Figura 4

Distribuzione del numero di interruzioni per sezione del questionario



Probabilità congiunta di partecipare* all'indagine web e di essere eleggibile
(modello probit bivariato)

Caratteristiche*		Prob. eleggibilità		Prob. partecipazione	
		Parametro	Pr > ChiSq	Parametro	Pr > ChiSq
Intercetta		-0,521	0,000	-2,353	0,000
Sesso*	Uomo	0,133	0,001	0,051	0,321
	Donna	--	--	--	--
Età*	Fino a 34 anni	0,033	0,672	-0,070	0,437
	35-44 anni	0,288	0,000	-0,107	0,262
	45-54 anni	-0,287	0,000	-0,112	0,254
	55-64 anni	-1,130	0,000	-0,228	0,043
	Oltre 64 anni	--	--	--	--
Giovane in famiglia		0,966	0,000	0,145	0,058
Componenti famiglia	1	-0,051	0,363	-0,241	0,001
	2	-0,354	0,000	-0,314	0,011
	3	-0,361	0,001	-0,384	0,004
	4	-0,464	0,000	-0,563	0,000
	5 e più	--	--	--	--
Figli in famiglia		0,210	0,001	0,097	0,277
Coniugato		-0,027	0,617	0,142	0,033
Quintili di reddito (fiscale) familiare	Primo	-1,645	0,000	-0,628	0,000
	Secondo	-1,175	0,000	-0,511	0,000
	Terzo	-0,906	0,000	-0,355	0,000
	Quarto	-0,504	0,000	-0,160	0,004
	Quinto	--	--	--	--
Possesso altri immobili		0,254	0,000	0,021	0,629
Area geografica	Nord	0,279	0,000	0,228	0,000
	Centro	0,169	0,000	0,120	0,028
	Sud e Isole	--	--	--	--
Ampiezza comune (n. abitanti)	fino a 20000	-0,619	0,000	-0,098	0,260
	20000-40000	-0,393	0,000	-0,109	0,165
	40000-500000	-0,383	0,000	-0,081	0,226
	oltre 500000	--	--	--	--
Cittadinanza*	Italiana	0,254	0,000	0,924	0,000
Rho 0,68 (0,0001)					

* Le interruzioni sono considerate come partecipazione all'indagine. Le caratteristiche individuali sono riferite alla persona di riferimento,, definita come la persona maggiormente informata sull'economia familiare..

Figura 5

Distribuzione del numero di percettori di reddito dal lavoro
(famiglie con persona di riferimento occupata)

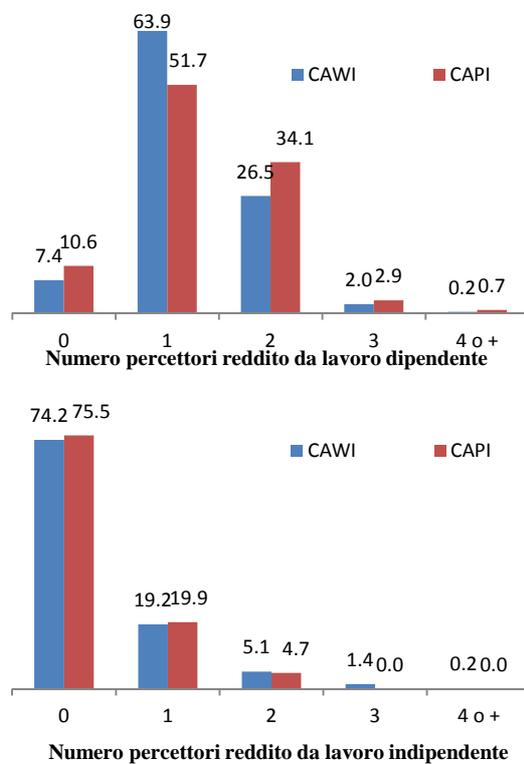


Figura 6

Distribuzione del numero di percettori di reddito da pensione
(famiglie con persona di riferimento in pensione)

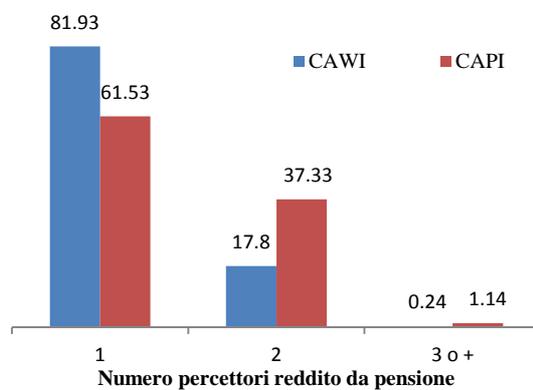
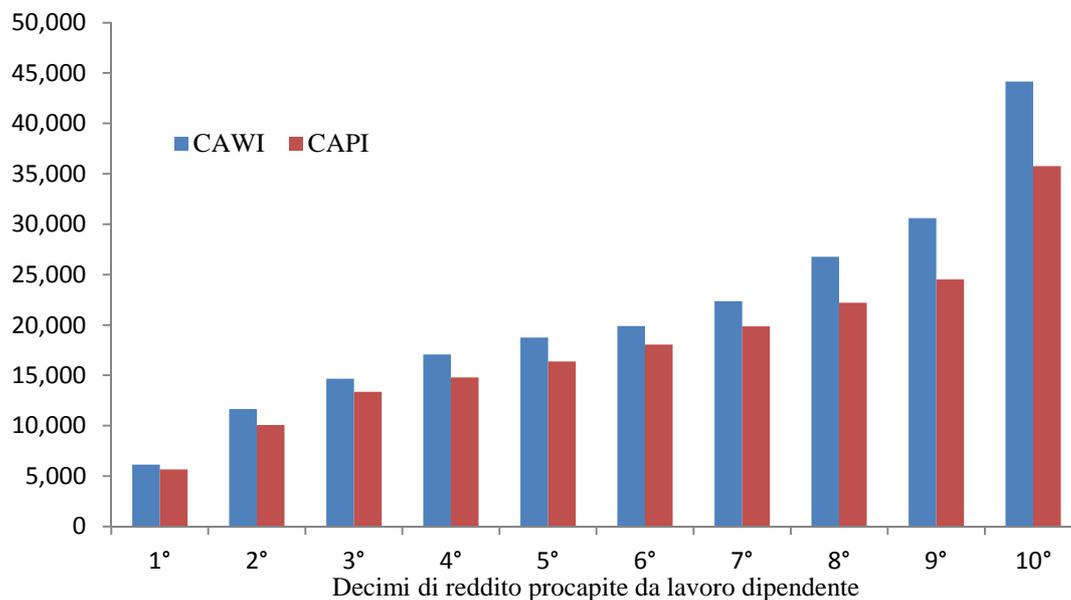


Figura 7

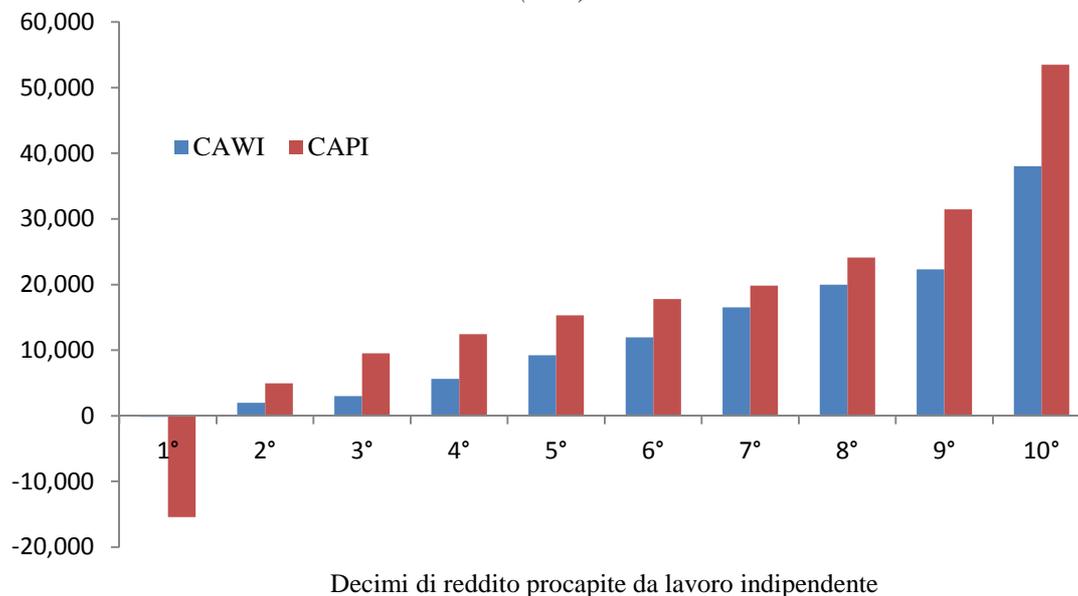
Valori medi di reddito da lavoro dipendente per classi di reddito
(euro)



Famiglie con persona di riferimento lavoratore dipendente. I decili sono calcolati sulla distribuzione a livello familiare del reddito pro-capite da lavoro dipendente. Il reddito pro-capite è calcolato per ciascuna famiglia come rapporto fra il totale dei redditi da lavoro dipendente e il numero di percettori.

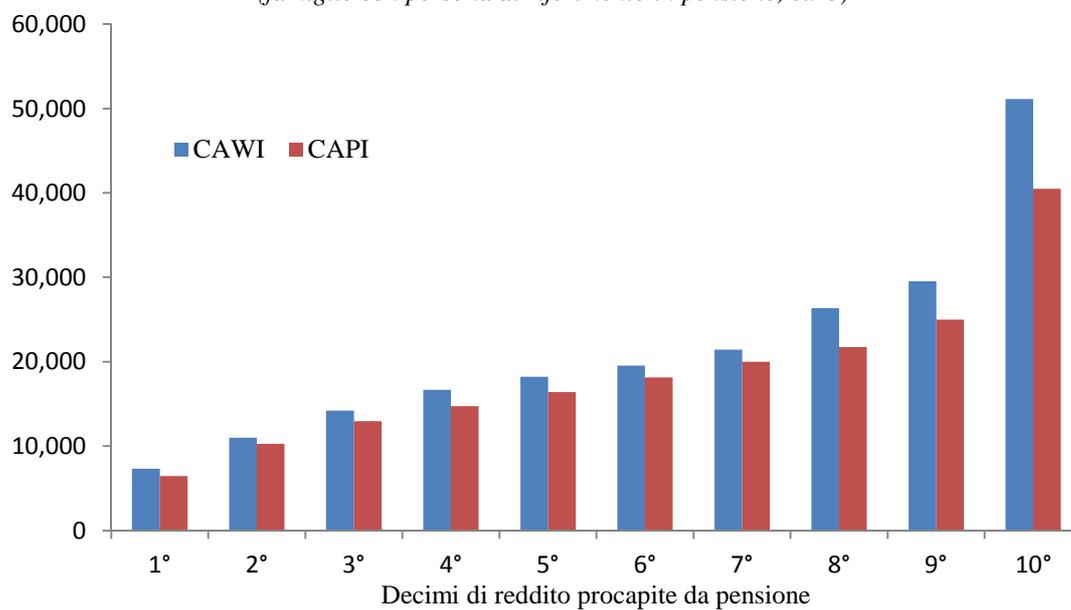
Figura 8

Valori medi di reddito da lavoro indipendente
(euro)



Famiglie con persona di riferimento lavoratore indipendente. I decili sono calcolati sulla distribuzione a livello familiare del reddito pro-capite da lavoro indipendente. Il reddito pro-capite è calcolato per ciascuna famiglia come rapporto fra il totale dei redditi da lavoro indipendente e il numero di percettori.

Valori medi del reddito da pensione per classi di reddito
(famiglie con persona di riferimento in pensione, euro)



Famiglie con persona di riferimento in pensione. I decili sono calcolati sulla distribuzione a livello familiare del reddito pro-capite da pensione. Il reddito pro-capite è calcolato per ciascuna famiglia come rapporto fra il totale dei redditi da pensione e il numero di percettori.

Bilanciamento dei campioni CAWI e CAPI selezionati con *propensity score matching*
(esempio relativo all'analisi del numero di percettori di reddito da lavoro dipendente)

Caratteristiche*	Valori medi		Significatività diff. fra medie
	CAWI	CAPI	
Donna	0,282	0,268	0,525
35-44 anni	0,178	0,174	0,818
45-54 anni	0,230	0,231	0,957
55-64 anni	0,234	0,244	0,618
Oltre 64 anni	0,293	0,288	0,827
Laurea	0,284	0,276	0,719
Occupato/a	0,354	0,344	0,688
Centro	0,217	0,215	0,911
Sud e Isole	0,263	0,267	0,854
Singoli sotto i 65 anni	0,181	0,188	0,739
Coppia senza figli	0,206	0,211	0,820
Coppia con figli	0,390	0,389	0,961
Monogenitore con figli	0,069	0,072	0,774
Altra tipologia familiare	0,064	0,060	0,731
In affitto	0,128	0,128	0,996

*Le caratteristiche individuali sono relative alla persona di riferimento.

Redditi da lavoro e da pensione nel confronto fra IBF-I e WEBIT
(numero di percettori e valori medi in euro)

	N. osservazioni		Valori medi		P-value diff.
	CAPI	CAWI	CAPI	CAWI	
Reddito da lavoro dipendente					
Numero medio di percettori per famiglia	1.272	843	0,9878	0,998	0,74
Valore medio per percettore	907	562	18.600	19.614	0,06
Reddito da lavoro autonomo					
Numero medio di percettori per famiglia	1.347	843	0,232	0,287	0,02
Valore medio per percettore	251	184	18.650	10.704	0,00
Reddito da pensione					
Numero medio di percettori per famiglia	1.347	845	0,437	0,331	0,00
Valore medio per percettore	503	314	17.255	17.249	0,90

Il metodo utilizzato per *matching* statistico è il metodo radius, caliper 0,03.

Confronto fra CAWI e BDR
(perceutori e valori medi in euro)

	n. oss.	CAWI	BDR
Reddito da lavoro dipendente	441		
Quota di perceutori non presenti in BDR		5,4	-
Quota di perceutori non presenti nella CAWI		-	11,3
Valore medio		24.907	26.326
Valore mediano		21.937	23.135
Reddito da lavoro autonomo	99		
Quota di perceutori non presenti in BDR		47,9	-
Quota di perceutori non presenti nella CAWI		-	35,1
Valore medio		25.754	23.635
Valore mediano		19.500	14.419
Reddito da pensione	261		
Quota di perceutori non presenti in BDR		1,9	-
Quota di perceutori non presenti nella CAWI		-	11,5
Valore medio		21.014	21.635
Valore mediano		19.123	19.416

I valori medi e mediani sono calcolati sugli individui presenti in entrambe le fonti.

Variabilità dei redditi da lavoro e da pensione nel confronto fra IBF-I e WEBIT
(coefficienti di variazione. rapporti interdecilici)

	Stime pesate		Stime non pesate	
	CAWI	CAPI	CAWI	CAPI
Reddito da lavoro dipendente				
Coefficiente di variazione	69,0	60,1	65,7	58,2
P95/P5	7,1	8,2	9,3	7,4
Reddito da lavoro autonomo				
Coefficiente di variazione	122,0	141,5	159,6	120,0
P90/P10*	135,0	22,7	178,0	16,7
Reddito da pensione				
Coefficiente di variazione	56,2	49,0	59,1	53,9
P95/P5	6,8	7,1	7,2	6,1

*Per questa tipologia di reddito P5 risulta pari a zero.

Pensi al reddito complessivo della Sua famiglia nel 2016. Lei prevede che rispetto a quello del 2015 sarà...
(valori percentuali)

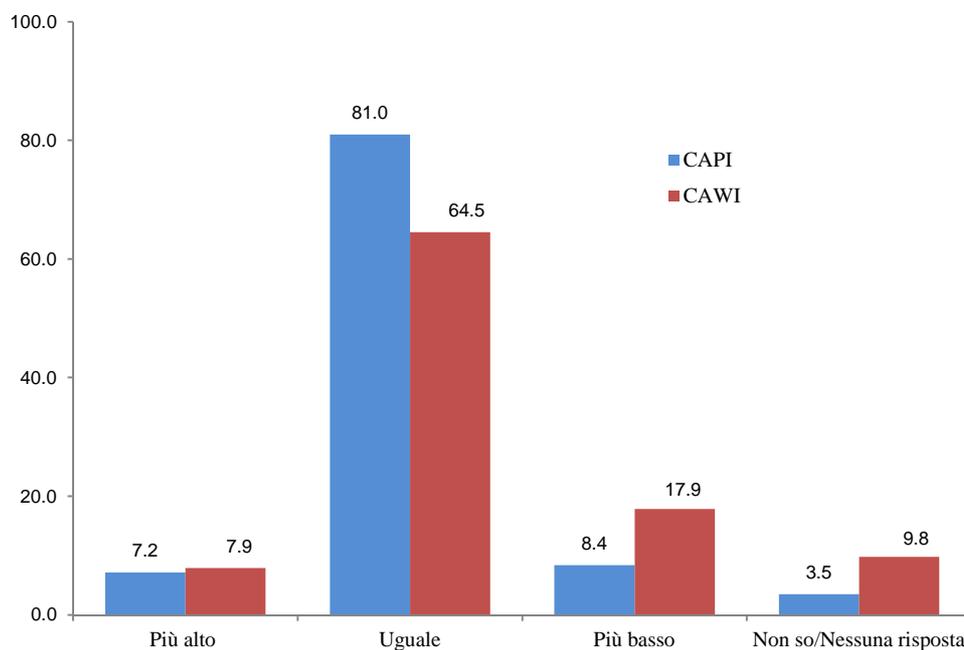
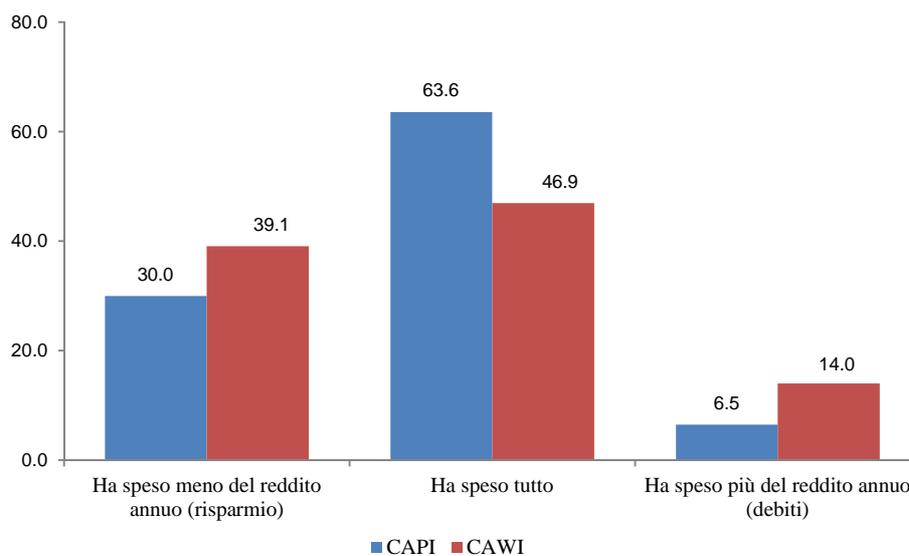


Figura 11

Considerando tutte le spese sostenute e tutte le fonti di reddito nel 2015, la sua famiglia ha speso meno (risparmiando), tutto (senza riuscire a risparmiare nulla) o più del reddito annuo (indebitandosi)?
(valori percentuali)



La spesa sostenuta mediamente dalla sua famiglia per tutti i consumi nel 2015 è stata più alta, più bassa o uguale rispetto a quanto ha speso nel 2014?
(valori percentuali)

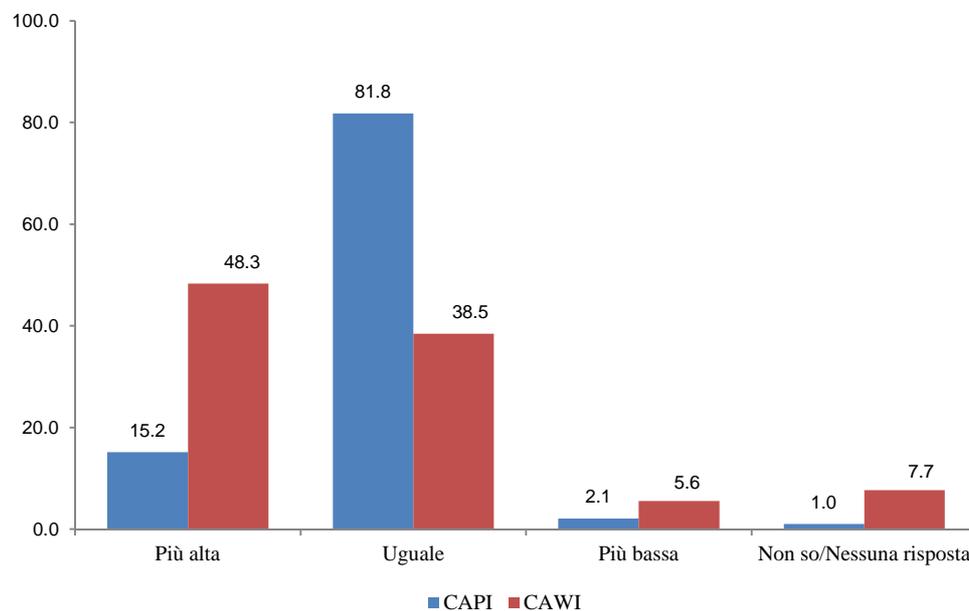
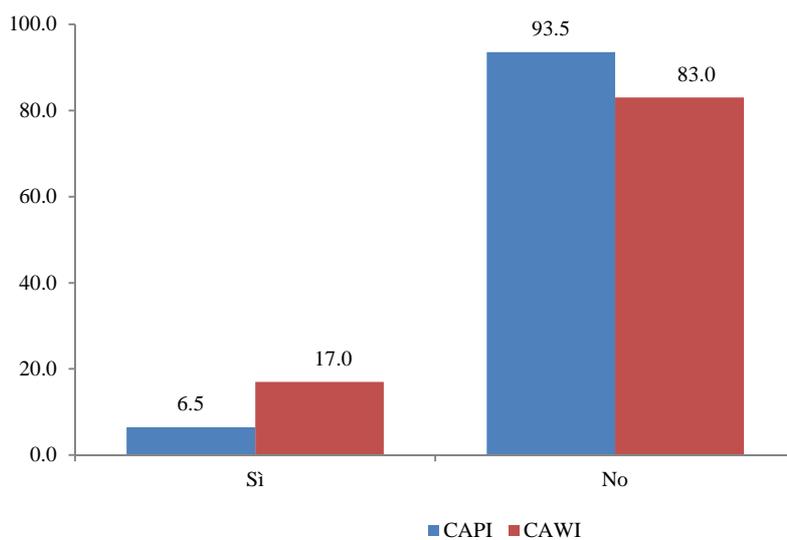
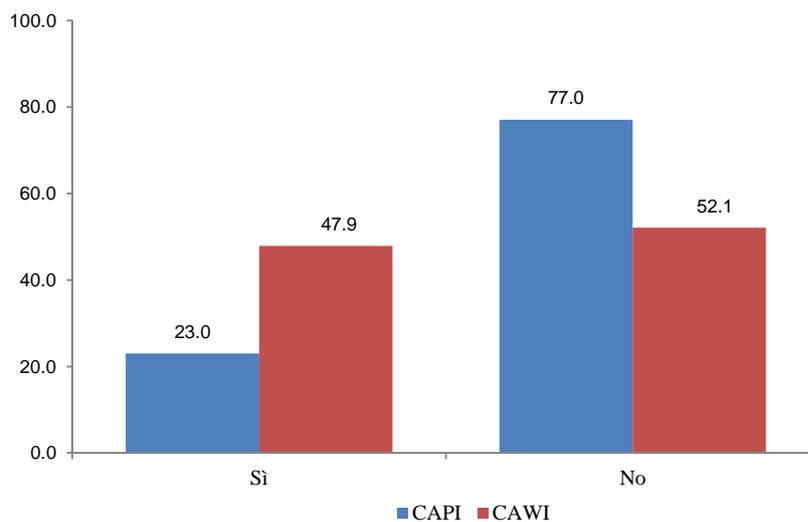


Figura 13

Nel corso degli ultimi 3 anni la sua famiglia ha ricevuto da parenti o amici non conviventi assistenza o aiuti non economici?
(valori percentuali)



Alla fine del 2015 la sua famiglia aveva debiti verso banche e società finanziarie?
(valori percentuali)



Pensando al reddito complessivo della sua famiglia nel 2015, Lei direbbe che a fine 2015 era più alto, più basso o sostanzialmente uguale a quello della fine del 2014?
(valori percentuali)

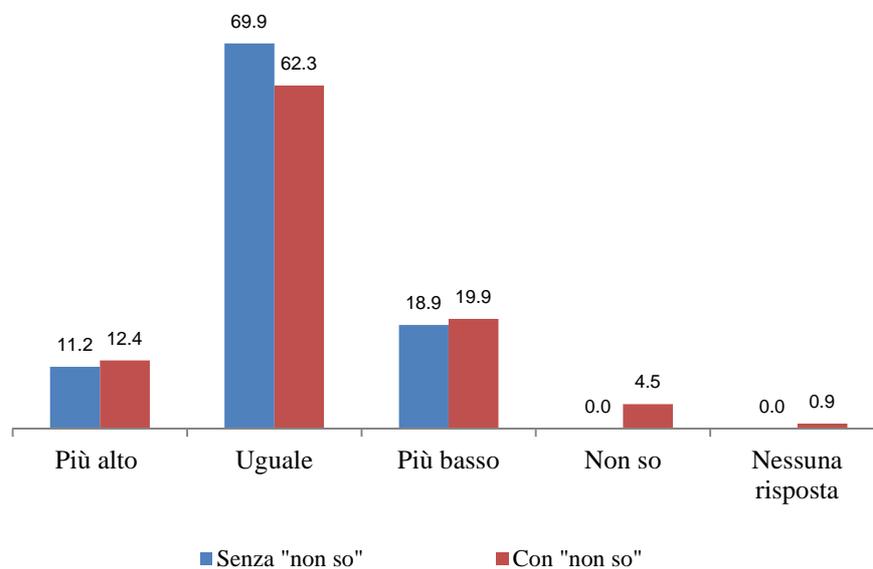


Figura 16

A Suo giudizio, alla fine del 2016 il valore della Sua abitazione principale sarà:.....?
(valori percentuali)

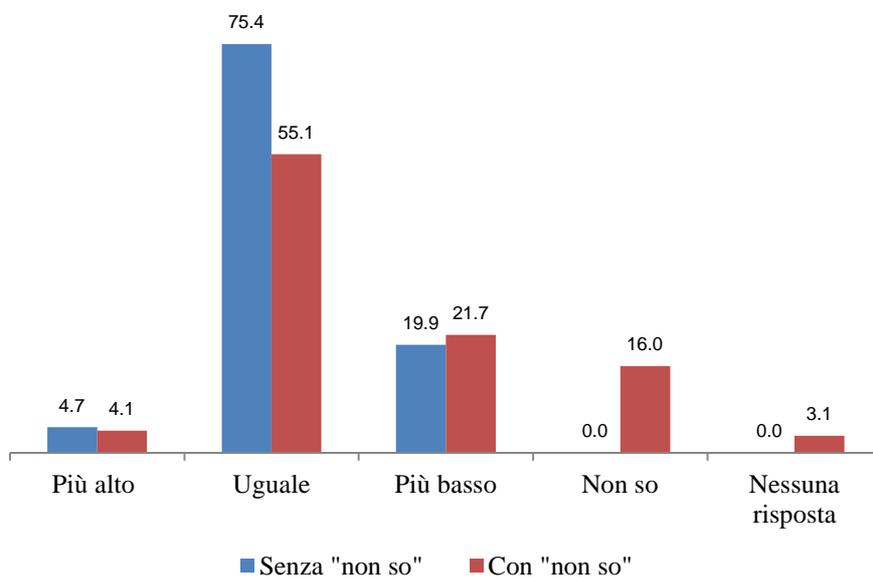
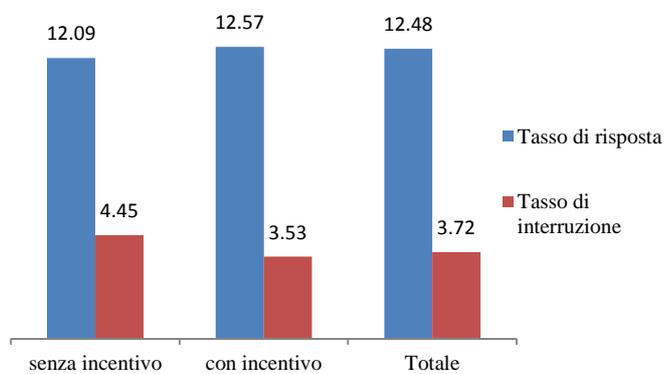


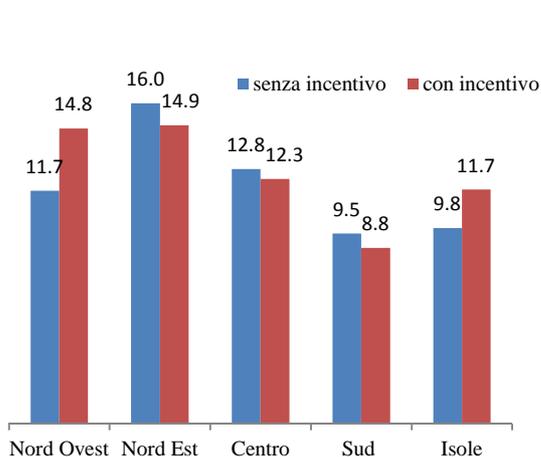
Figura 17

Tassi di risposta e di interruzione

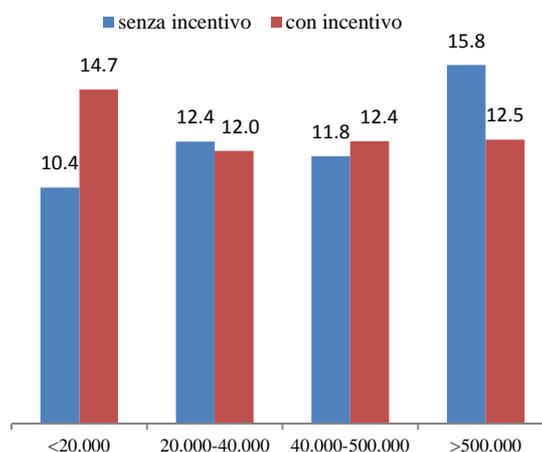


Tassi di risposta per:

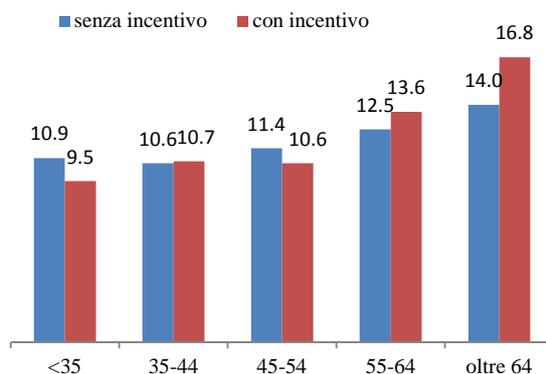
**Area geografica
(numero di abitanti)**



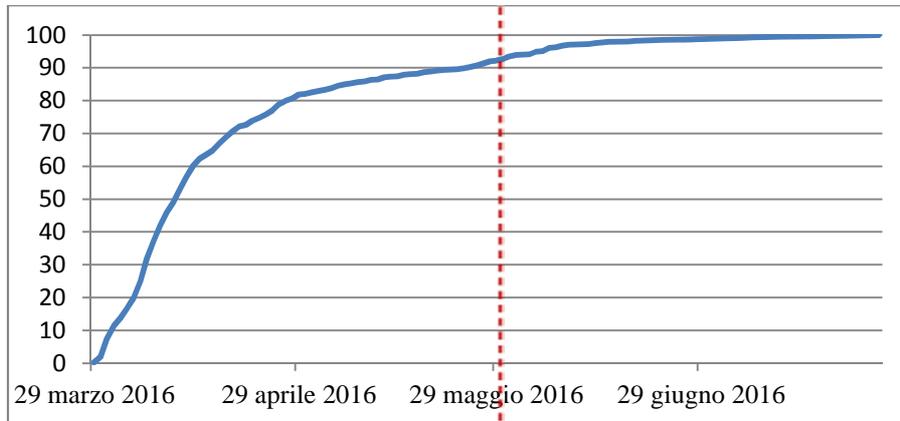
Dimensione di comune di residenza



Classe di età



Andamento della compilazione dei questionari web



Appendice B - I pesi campionari

Il peso di ciascuna unità campionaria è stato ottenuto attraverso le seguenti fasi:

- a) inizialmente viene calcolato il *peso da disegno* come reciproco della probabilità di inclusione delle unità ($w^{(0)}$);
- b) si ottiene poi un peso corretto per la mancata risposta totale ($w^{(1)}$) moltiplicando $w^{(0)}$ per un fattore di aggiustamento legato alla probabilità di partecipazione;
- c) il peso viene aggiustato sulla base di informazioni ausiliarie (*post-stratificazione*), giungendo ad un peso finale $w^{(2)}$

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

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

dove P_h e m_h sono rispettivamente la popolazione residente e il numero di comuni campione nell'*h-esimo* strato, e P_{hi} e n'_{hi} rispettivamente la popolazione e il numero di componenti da intervistare (teorici) nell'*i-esimo* comune dell'*h-esimo* strato¹³.

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

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

dove n_{hi} rappresenta il numero di componenti effettivamente intervistati nell'*i-esimo* comune dell'*h-esimo* strato. Questa correzione ipotizza che la probabilità di risposta sia uguale per tutte le unità che appartengono allo stesso comune

Come analisi di robustezza l'aggiustamento per la non risposta è stato anche effettuato sulla base del modello descritto nel precedente paragrafo. Il fattore di aggiustamento in questo caso è rappresentato dalla probabilità di risposta stimata per ciascuna famiglia (che varia dunque anche all'interno di ciascun comune):

$$(2a) \quad w_j^{(1)} = w_j^{(0)} \alpha_j$$

dove α_j rappresenta l'inverso della probabilità di risposta stimata per la famiglia j . Il presupposto per poter utilizzare questo approccio è che la stima del modello deve essere fatta sulle famiglie eleggibili ossia su coloro che hanno la possibilità di utilizzare il web. Come

¹³ Ciascun comune *auto-rappresentativo* è inserito in uno strato a se stante ed incluso con certezza nel campione. In tale caso la probabilità di estrazione di una famiglia è approssimativamente pari a n_{hi}/P_{hi} . I comuni *non auto-rappresentativi* sono selezionati con probabilità proporzionale alla dimensione (*probability proportional to size - PPS*); la probabilità di selezione dell'*i-esimo* comune nell'*h-esimo* strato è pertanto pari a $m_h P_{hi}/P_h$. La probabilità di inclusione di una famiglia nel campione può quindi essere scritta come: $m_h n_{hi}/P_h$.

descritto nel precedente paragrafo, questa informazione non è disponibile nel nostro caso in quanto il campione è stato selezionato a partire dalle liste anagrafiche che non hanno alcuna informazione sull'uso del web. Per ovviare a tale inconveniente è stato quindi necessario stimare per ciascuna famiglia del campione la sua eleggibilità utilizzando un modello stimato sull'indagine CAPI.

Infine, i pesi vengono calibrati a livello familiare utilizzando informazioni esterne provenienti dall'indagine CAPI¹⁴. La correzione viene condotta imponendo al campione finale l'allineamento alle caratteristiche alla popolazione web per quanto riguarda il sesso, la classe di età (meno di 34 anni, da 35 a 44, da 45 a 54, da 55 a 64, oltre 65 anni), l'area geografica (Nord, Centro, Sud e Isole) e la dimensione del comune di residenza (fino a 20.000 abitanti, da 20.000 a 40.000, da 40.000 a 500.000, oltre 500.000 abitanti) e titolo di studio (fino alla scuola secondaria di primo grado, diplomati, laureati o più)¹⁵.

I pesi finali sono dunque ottenuti come:

$$(3) \quad w_c^{(2)} = w_c^{(1)} \beta_c$$

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

Le caratteristiche della popolazione *web* sono stimate a partire dall'indagine CAPI, selezionando le famiglie in cui il rispondente ha accesso ad internet tramite pc o altro strumento. Le caratteristiche socio-demografiche sono naturalmente relative alla persona di riferimento (il responsabile dell'economia familiare) che risponde al questionario per la famiglia.

La finalità di questo aggiustamento è quella di allineare la composizione dei due campioni (WEB e CAPI) in modo che le differenze nei risultati non dipendano da caratteristiche socio-economiche osservabili.

¹⁴ La calibrazione consente nel caso in cui le variabili utilizzate siano correlate con le variabili oggetto di indagine, di migliorarne l'accuratezza delle stime e limitare ulteriormente le distorsioni legate alla mancata risposta.

¹⁵ La tecnica utilizzata consente di allineare in modo simultaneo i pesi campionari alla distribuzione di alcune caratteristiche note da fonti esterne considerando esclusivamente le probabilità marginali. Tale metodo è noto come *Iterative Proportional Fitting* (o *Raking*). Si veda G. Kalton e I. Flores Cervantes, *Weighting Methods*, in *Journal of Official Statistics*, Vol.19, No.2, 2003, pp. 81-97.