

Atti del 6° Censimento Generale dell'Agricoltura

LA VALUTAZIONE DELLA QUALITÀ

5



6° Censimento Generale
dell'Agricoltura

 Istat
Istituto Nazionale
di Statistica

RACCOGLIAMO RISPOSTE, SEMINIAMO FUTURO.

Atti del 6° Censimento Generale dell'Agricoltura

LA VALUTAZIONE DELLA QUALITÀ

Atti a cura di: Giampaola Bellini
Fascicolo 5 a cura di: Matteo Mazziotta

Atti del 6° Censimento Generale dell'Agricoltura
LA VALUTAZIONE DELLA QUALITÀ

© 2013
Istituto nazionale di statistica
Via Cesare Balbo, 16 - Roma

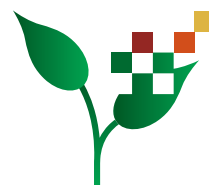
Salvo diversa indicazione la riproduzione è libera,
a condizione che venga citata la fonte.

Immagini, loghi (compreso il logo dell'Istat),
marchi registrati e altri contenuti di proprietà di terzi
appartengono ai rispettivi proprietari e
non possono essere riprodotti senza il loro consenso.

DISTRIBUITO DA
STEALTH
BY SIMPLICISSIMUS BOOK FARM

Indice

Presentazione dell'opera	5
Il 6° Censimento generale dell'agricoltura	7
Introduzione	11
Capitolo 1 - L'indagine di misura	13
1.1 Gli obiettivi.....	13
1.2 Il disegno dell'indagine	13
1.3 Organizzazione e calendario delle attività	15
1.4 Il piano di campionamento.....	17
1.5 Rilevazione e monitoraggio.....	19
1.6 Il controllo e la correzione dei dati	24
1.7 Modello di stima	25
1.8 I risultati	34
Appendice 1.A - Risultati delle componenti di varianza di risposta	43
Appendice 1.B - La variabilità delle stime	49
Capitolo 2 - L'indagine di copertura	55
2.1 Gli obiettivi.....	55
2.2 Il disegno dell'indagine	55
2.3 L'organizzazione.....	61
2.4 Il piano di campionamento, i metodi di stima e la variabilità	65
2.5 La rilevazione	82
2.6 Il controllo e la correzione dei dati	89
2.7 Le procedure di abbinamento.....	93
2.8 I risultati	103
Appendice 2.A - Variabilità delle stime di copertura.....	112
Appendice 2.B - Linearizzazione degli stimatori ${}_d\tilde{N}$, ${}_d\tilde{x}$, ${}_d\tilde{y}$ e ${}_d\tilde{s}$	117
Appendice 2.C - Varianze e stime delle varianze degli stimatori ${}_d\tilde{N}$, ${}_d\tilde{x}$, ${}_d\tilde{y}$ e ${}_d\tilde{s}$	119
Appendice 2.D - Numerosità di primo e secondo stadio teoriche e rispondenti per regione	123
Conclusioni	124
Riferimenti bibliografici	127
Allegati in formato digitale (cfr. cd-rom allegato al fascicolo <i>Il quadro normativo</i>)	
1.A Indagine di misura. Lettera informativa	
1.B Indagine di misura. Questionario di rilevazione	
2.A Indagine di copertura. I modelli per la rilevazione	





Presentazione dell'opera¹

Gli Atti del 6° Censimento generale dell'agricoltura, vista la rilevanza e la complessità degli argomenti trattati, sono stati strutturati in forma di cofanetto, articolato in cinque fascicoli che raccontano il Censimento attraverso l'iter normativo percorso, nonché le principali soluzioni organizzative, tecniche e metodologiche messe a punto e adottate per questa tornata censuaria. L'opera è stata realizzata in maniera unitaria, essendo evidenti le interconnessioni delle fasi e dei processi produttivi, ma ciascun fascicolo è autonomo nella trattazione tematica, ed è stato realizzato per affrontare in maniera esauriente il tema trattato. Il cofanetto è corredato di un cd-rom, collocato all'interno del primo fascicolo, contenente i documenti ufficiali del Censimento, in formato .pdf.

I cinque fascicoli che costituiscono il cofanetto sono i seguenti:

1. *Il quadro normativo;*
2. *L'organizzazione della rilevazione e le informazioni censuarie;*
3. *Il sistema informatico integrato;*
4. *Il Censimento sul territorio;*
5. *La valutazione della qualità.*

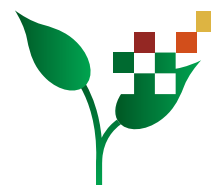
Il primo fascicolo *Il quadro normativo* contiene le norme internazionali e quelle nazionali sulla base delle quali si è progettato e realizzato il Censimento. Sono presentati e descritti nei loro aspetti principali la legge di indizione e finanziamento, il DPR relativo al regolamento di esecuzione, una serie di atti, che sono stati adottati per garantire un adeguato funzionamento della complessa macchina organizzativa, nonché le circolari operative emanate dall'Istat per definire e dare attuazione a tutte le fasi censuarie.

Il secondo fascicolo *L'organizzazione della rilevazione e le informazioni censuarie* descrive le linee guida adottate nell'organizzazione della rilevazione, la composizione e le caratteristiche della rete di rilevazione, nonché la modalità di definizione ed erogazione della formazione agli operatori censuari. Vengono altresì descritte le caratteristiche generali e tecniche della rilevazione, percorrendo le principali tappe che hanno portato dalla definizione della lista pre-censuaria di unità agricole e dei contenuti del questionario alla diffusione dei risultati.

Il terzo fascicolo *Il sistema informatico integrato* descrive l'architettura e le soluzioni informatiche adottate per un Censimento che per la prima volta utilizza la modalità di compilazione on-line e la registrazione dei questionari presso gli uffici di rilevazione. Tutti i processi adottati sono connessi e dialoganti tra loro garantendo coerenza logica, velocità e trasparenza al sistema e alle operazioni realizzate. Sono descritti il Sistema di Gestione della Rilevazione (SGR) e l'impianto informatico e gli strumenti di diffusione dei dati.

Il quarto fascicolo *Il Censimento sul territorio* descrive, da una parte, come la novità dell'organizzazione decentrata alle amministrazioni regionali delle operazioni censuarie, seppure sotto l'egida di norme e linee guida dettate dal centro, sia stata supportata dagli uffici territoriali Istat, dall'altra, come la realizzazione di un Censi-

¹ Si desidera ringraziare per l'attento lavoro di revisione e per i preziosi suggerimenti forniti Franco Lorenzini e Massimo Greco.



mento secondo una modalità organizzativa più tradizionale e più centralizzata per una parte delle regioni italiane sia stata gestita con un forte coordinamento dell'Istat a livello locale.

Il quinto fascicolo *La valutazione della qualità* descrive le metodologie adottate e i risultati ottenuti dalle due indagini di qualità: 1) l'indagine di misura è volta alla stima della distorsione e delle varianze di risposta dovute al complesso processo di rilevazione; 2) l'indagine di copertura è volta alla stima del tasso di sotto-copertura del 6° Censimento generale dell'agricoltura.

Il 6° Censimento generale dell'agricoltura

a cura di **Andrea Mancini**

Direttore del Dipartimento per i censimenti e gli archivi amministrativi e statistici dell'Istat

Come avviene tradizionalmente in occasione di ogni censimento, l'Istituto nazionale di statistica redige gli "Atti" del censimento, pubblicazione mediante la quale descrive le caratteristiche tecniche, organizzative e metodologiche delle attività svolte. In occasione del 6° Censimento generale dell'agricoltura si è scelto di pubblicare un cofanetto di cinque fascicoli tematici, corredato da un supporto informatico, collocato all'interno del primo fascicolo *Il quadro normativo*, contenente copia delle norme di riferimento e di altra documentazione prodotta dall'Istat. Gli "Atti del 6° Censimento generale dell'agricoltura" rappresentano lo strumento che garantisce la "memoria storica" del Censimento, punto di riferimento essenziale per chi è interessato a conoscere come è stata progettata ed eseguita una tra le rilevazioni più importanti del Paese. In questa pubblicazione sono descritte tutte le fasi dell'indagine censuaria in modo da consentire al lettore di conoscere le motivazioni delle strategie adottate, le soluzioni organizzative e tecniche, il processo di produzione dei dati.

Il 6° Censimento generale dell'agricoltura ha avuto luogo con riferimento alla data del 24 ottobre 2010 e ha aperto la tornata censuaria del 2010-2011^{II} contraddistinta da un ampio uso di dati amministrativi e da una forte attenzione all'innovazione. Esso è stato svolto con l'obiettivo principale di delineare un quadro informativo statistico, aggiornato e dettagliato della struttura delle aziende agricole e zootecniche operanti in Italia e dei metodi di produzione agricola adottati. La fotografia dell'agricoltura italiana emersa dai dati del censimento ha mostrato le trasformazioni che hanno interessato il comparto nel decennio 2000-2010, caratterizzato dalla crisi economica, dalla riforma della Politica agricola comunitaria (PAC) e dall'enfasi con cui è stata promossa l'adozione di metodi di produzione rispettosi dell'ambiente. Inoltre, il Censimento ha assolto gli obblighi internazionali di produrre statistiche comunitarie comparabili sui fenomeni agricoli, sulle condizioni di sviluppo rurale e di sostenibilità ambientale^{III} - necessarie ai fini di programmazione e valutazione della Politica agricola comunitaria - sulle superfici viticole^{IV}, utilizzando definizioni, concetti e metodi armonizzati a livello europeo. Inoltre, l'estensione delle analisi ai confronti storici fino al Censimento agricolo del 1982 ha permesso di delineare il quadro delle trasformazioni strutturali dell'agricoltura nazionale nell'ultimo trentennio.

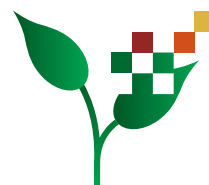
Dal punto di vista metodologico e tecnologico sono state introdotte rilevanti innovazioni, in parte riutilizzate con adattamenti per il 15° Censimento generale della popolazione e delle abitazioni e il 9° Censimento generale dell'industria e dei servizi, condotti nel 2011 e 2012.

La principale innovazione è stata l'ampio uso di dati amministrativi sia per la pre-

II 6° Censimento generale dell'agricoltura, 15° Censimento generale della popolazione e delle abitazioni, 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit.

III *Regolamento (CE) n. 1166/2008 del Parlamento europeo e del Consiglio del 19 novembre 2008* avente per oggetto l'istituzione di un quadro di riferimento per la produzione di statistiche comunitarie comparabili sulla struttura delle aziende agricole e per un'indagine sui metodi di produzione agricola.

IV *Regolamento (CEE) n. 357/79 del Consiglio del 5 febbraio 1979* e successive modificazioni, concernente le indagini statistiche sulle superfici viticole.



disposizione della lista pre-censuaria sia per il controllo - anche contestuale alla rilevazione - e la correzione dei dati raccolti. Gli obiettivi raggiunti sono stati la maggiore tempestività nel rilascio dei risultati, il contenimento del carico statistico sui rispondenti, il miglioramento della qualità dei dati prodotti. La lista è stata predisposta mediante integrazione di fonti amministrative a carattere generale e di fonti amministrative e statistiche specifiche del settore tra le quali l'archivio dei Fascicoli aziendali, contenuti nel Sistema informativo agricolo nazionale (Sian), e l'archivio delle Anagrafi zootecniche. Nel complesso sono state utilizzate 17 fonti amministrative o statistiche, dopo aver eseguito nel 2008 la rilevazione sperimentale e di controllo del prototipo di lista pre-censuaria, svolta in collaborazione con gli uffici di statistica delle Regioni e Province autonome.

Salvaguardando la continuità storica rispetto ai Censimenti precedenti, altre significative innovazioni sono state apportate attraverso:

- la nuova rilevazione sui metodi di produzione agricola adottati dalle aziende agricole, svolta dall'Italia in forma censuaria, sebbene per la Unione Europea fosse sufficiente l'esecuzione in forma campionaria;
- il campo di osservazione individuato dal solo Universo Ue e non anche l'Universo Italia come avvenuto nei precedenti censimenti;
- l'individuazione ex ante delle micro unità con segnali amministrativi di possibile attività di coltivazione o allevamento da escludere dalla rilevazione;
- la definizione del metodo per identificare ex ante l'unitarietà tecnico-economica dell'azienda agricola;
- l'intervista effettuata, a cura del rilevatore, presso la residenza o sede legale del conduttore anziché presso il centro aziendale;
- l'individuazione ex ante della possibile localizzazione del centro aziendale elaborando i dati di archivio;
- la pubblicazione di dati riferiti al comune di localizzazione dei terreni/allevamenti^v in aggiunta alla tradizionale pubblicazione di dati riferiti al centro aziendale;
- la richiesta di informazioni volte a consentire la georeferenziazione del centro aziendale e la stima dei volumi d'acqua aziendali consumati per l'irrigazione, nonché notizie relative alla cittadinanza della manodopera, alla presenza di impianti per la produzione di energie rinnovabili, alle superfici dedicate a colture energetiche;
- l'esclusione dal campo di osservazione delle unità esclusivamente forestali.

Ulteriori innovazioni tecniche hanno riguardato:

- la tecnica di rilevazione multicanale (*mixed mode*), con facoltà per le aziende agricole di compilare il questionario direttamente via internet;
- l'architettura informatica interamente on-line, che ha riguardato tutte le macro funzioni del processo produttivo;
- il ricorso, per la prima volta in Italia, ai social network come ausilio alle attività degli addetti al censimento;
- l'utilizzo di strumenti a distanza di e-learning a supporto delle attività formative degli addetti al Censimento;
- un nuovo impianto metodologico del sistema di controllo e correzione dei dati, assai accurato e supportato dal confronto con i dati amministrativi;

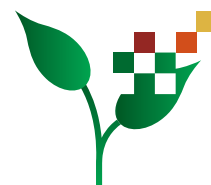
^v Riprendendo quanto pubblicato in occasione del Censimento dell'agricoltura del 1990, non replicato in quello del 2000.

- la diffusione assai tempestiva di dati provvisori di ampio contenuto informativo, con disaggregazione dell'informazione senza precedenti rispetto ai passati censimenti;
- la diffusione dei risultati definitivi mediante il sistema di data warehousing dell'Istat, I.Stat.

Il Sistema di Gestione della Rilevazione (SGR), nucleo portante dell'architettura informatica del Censimento, ha consentito di seguire l'andamento di tutte le fasi del processo di produzione per singolo organo di censimento, rilevatore o coordinatore, grazie alla realizzazione di oltre cinquanta funzioni componenti del sistema. SGR è stato il Centro di controllo dell'organizzazione ed è stato utilizzato autonomamente da ciascun organo di censimento, rilevatore o coordinatore, che attraverso proprie *userid* e *password*, ha avuto accesso alle funzioni e alle tabelle di pertinenza.

In generale, la realizzazione dei sistemi informatici a supporto della fase di acquisizione, monitoraggio, controllo, correzione e diffusione dei dati del 6° Censimento generale dell'agricoltura 2010 è stata condotta utilizzando tecniche di progettazione, modellazione e sviluppo software in un'ottica di standardizzazione complessiva. I sistemi realizzati, pur non essendo "generalizzati", sono stati progettati in modo da consentire un elevato grado di riuso dei metodi e del software, specialmente per quanto riguarda i sistemi di gestione della rilevazione e i sistemi di implementazione e presentazione dei questionari elettronici, tanto da essere stati di fatto riutilizzati, con adattamenti, nel 15° Censimento della popolazione e delle abitazioni e nel 9° Censimento dell'industria e dei servizi. Nondimeno SGR ha presentato alcune criticità connesse alla complessità della rete e dell'organizzazione del lavoro. All'avvio del sistema di compilazione on-line da parte delle aziende agricole, il consistente numero di accessi contemporanei ha messo in evidenza un comportamento anomalo dell'applicazione, non evidenziato in fase di test. È stato necessario, a poche ore dalla partenza, sospendere il servizio e sottoporre l'applicazione a test più approfonditi. SGR è stato riaperto alla rete territoriale di rilevazione dopo 48 ore, mentre il sistema di acquisizione on-line è stato riaperto alle aziende dopo due settimane, con conseguente abbassamento delle risposte via web da parte delle aziende agricole rispetto alle aspettative iniziali. Superate le difficoltà iniziali, il sistema ha permesso di gestire una rete articolata composta da più di 20.000 operatori. I questionari relativi alle oltre 2 milioni di unità di rilevazione sono stati registrati on-line, in tutto o in parte: circa 1.352.000 questionari sono stati inseriti via web dagli operatori di censimento delle regioni ad Alta partecipazione (66,1 per cento); circa 636.000 notizie di base riferite alle variabili primarie sono state inserite dagli operatori degli Uffici Comunali di Censimento (UCC) delle regioni a Partecipazione integrativa attraverso l'apposita funzionalità a disposizione in SGR (31,0 per cento) e 61.000 questionari sono stati compilati via web direttamente dalle aziende agricole (2,9 per cento).

Sul piano dell'organizzazione il Censimento ha adottato schemi a "geometria variabile", coinvolgendo assai più delle precedenti edizioni gli uffici di statistica e le direzioni di agricoltura delle Regioni e Province autonome. La flessibilità organizzativa ha trovato il suo compimento nella definizione di due modelli organizzativi alternativi, uno ad "Alta partecipazione" e uno a "Partecipazione integrativa", che la singola Regione o Provincia autonoma poteva adottare. Nel primo modello la Regione o Provincia autonoma ha definito in autonomia la struttura della propria rete territoriale. Nel secondo modello il ruolo della Regione è stato minore e analogo



a quello già espletato nel precedente Censimento in quanto l'organizzazione generale è stata a carico dell'Istat ed è spettato ai Comuni il compito di eseguire la rilevazione sul campo.

A conclusione di queste brevi note si vuole sottolineare che la trasmissione ad Eurostat delle informazioni di interesse comunitario raccolte dal Censimento è avvenuta in anticipo rispetto ai termini fissati per l'Italia dal Regolamento, che il rapporto di qualità trasmesso dall'Istituto a Eurostat è stato da questo vagliato positivamente, che i numerosi dati provvisori sono stati diffusi a meno di 9 mesi dalla data di riferimento e a 5 mesi dalla data di chiusura delle operazioni sul campo prefissata dal calendario censuario, che la diffusione dei dati definitivi è iniziata a luglio 2012 e terminata a dicembre 2012. Inoltre, le indagini di controllo della qualità e della copertura del Censimento hanno restituito il quadro della qualità statistica dei dati prodotti in modo trasparente e disponibile nel quinto fascicolo di questi "Atti". I risultati sono stati ottenuti grazie al lavoro svolto da tutte le componenti, interne ed esterne all'Istituto, che hanno operato nella rilevazione, fin dalle attività svolte dal "Comitato consultivo per la preparazione a livello regionale del 6° Censimento generale dell'agricoltura", dove la condivisione delle scelte e la trasparenza delle decisioni ha rappresentato un rilevante fattore di efficacia per il funzionamento della macchina censuaria.

Introduzione

L'impostazione metodologica del 6° Censimento generale dell'agricoltura è stata caratterizzata da numerose innovazioni nelle diverse fasi dell'indagine, a partire dalla lista pre-censuaria che è stata prodotta mediante l'integrazione di numerosi archivi amministrativi provenienti da fonti informative differenti.

La lista pre-censuaria è stata costruita con un duplice obiettivo: i) individuare le unità da sottoporre a rilevazione censuaria; ii) valutare, anche attraverso il Censimento, la qualità delle diverse fonti amministrative al fine di realizzare un registro statistico delle aziende agricole da aggiornare annualmente. Questo secondo obiettivo è coerente con le indicazioni emerse a livello nazionale e internazionale circa l'utilizzo di fonti amministrative per ridurre i costi e il fastidio statistico a carico dei rispondenti.

La bontà di tale approccio, anche convalidata dai risultati delle indagini di qualità descritte nel presente volume, può aprire nuovi scenari nella conduzione delle indagini in campo agricolo, in un'ottica di realizzazione del *Farm Register*, il registro statistico delle aziende agricole. La disponibilità di tale registro, aggiornabile annualmente, determinerà nel tempo un miglioramento della qualità di tutto il sistema di produzione delle statistiche agricole, consentendo di coprire esigenze informative attualmente soddisfatte solo una volta ogni dieci anni, in occasione della rilevazione a carattere censuario. Vale qui ricordare che la costruzione di un registro statistico di aziende agricole, a partire dalla base informativa del Censimento agricoltura 2010 e attraverso l'integrazione di fonti amministrative, implica lo svolgimento di verifiche sul campo dei risultati riscontrati e delle metodologie applicate. La qualità dell'informazione statistica, fornita dal 6° Censimento generale dell'agricoltura, è stata misurata e documentata dall'Istat attraverso due rilevazioni campionarie post-censuarie mirate a valutare il grado di attendibilità delle singole risposte fornite dai conduttori di azienda e il grado di copertura rispetto al numero di aziende agricole rilevate dal Censimento. Le due indagini di controllo presentano una dimensione campionaria che permette di fornire stime sulla precisione del Censimento a livello regionale.

Tali indagini sono:

- l'indagine di misura, finalizzata a fornire valutazioni dell'errore di misurazione dovuto al processo di risposta e condotta mediante re-intervista telefonica realizzata su un campione di circa 50.000 aziende già rilevate al Censimento;
- l'indagine di copertura, finalizzata a fornire una misurazione del grado di copertura del Censimento rispetto alla popolazione delle aziende agricole mediante un campione areale di circa 1.500 fogli di mappa estratti dal Catasto dei terreni.

Le due rilevazioni post-censuarie sono state eseguite con le medesime caratteristiche di quelle condotte in occasione del Censimento del 2000*.

Precedentemente la valutazione della qualità dei risultati del Censimento agricolo era realizzata con tecniche di analisi dei dati per la stima dell'errore di misura e delle sue principali componenti. Durante l'arco di 50 anni le tecniche delle indagini di misura si sono evolute: si è passati dalla re-intervista diretta, utilizzando gli stessi rilevatori del Censimento, alla re-intervista con tecnica CATI (*Computer Assisted Telephone Interviewing*) svolta da una società esterna, riducendo notevolmente i

* La prima esperienza di indagini post-censuarie così strutturate è stata effettuata in occasione del 5° Censimento generale dell'agricoltura del 2000.



tempi di raccolta dati. Inoltre, si è passati dal confronto manuale dei modelli cartacei compilati nelle due interviste (Censimento e indagine post-censuaria) a una procedura informatizzata di riconciliazione tra il dato fornito all'intervista telefonica e quello rilevato al Censimento.

Per quanto riguarda invece l'errore di copertura, solo in occasione del 5° Censimento generale dell'agricoltura del 2000, l'Istat ha applicato una metodologia già utilizzata per il Censimento della popolazione in Istat e in altri paesi europei, capace di fornire una stima diretta del grado di copertura facendo ricorso a un'apposita indagine campionaria areale.

Il presente fascicolo, presenta le due indagini di qualità condotte immediatamente dopo il Censimento del 2010 esponendo le loro caratteristiche principali: dalla fase di progettazione alla conduzione delle rilevazioni, dall'analisi dei dati al *record linkage*, dai risultati all'individuazione degli elementi utili per la prossima progettazione del *Farm Register*.

Il Capitolo 1 descrive tutte le fasi dell'indagine di misura e i risultati che hanno portato a misurare il livello di errore di risposta del 6° Censimento generale dell'agricoltura.

In particolare, nel Paragrafo 1.1 sono descritti gli obiettivi della rilevazione post censuaria, nel Paragrafo 1.2 il disegno dell'indagine e nel Paragrafo 1.3 l'organizzazione e il calendario delle attività. Il Paragrafo 1.4 si occupa degli aspetti legati al disegno di campionamento, alla definizione della numerosità campionaria e alla selezione delle aziende agricole dalla lista censuaria. Il Paragrafo 1.5 descrive tutte le fasi dell'indagine legate alla rilevazione e al monitoraggio (centrale e decentrato) della stessa. Il Paragrafo 1.6 esamina le procedure di controllo e correzione dei dati rilevati dall'indagine telefonica con particolare attenzione all'analisi dei valori anomali. Il Paragrafo 1.7 presenta le procedure metodologiche di stima dell'errore di misura e della varianza di risposta, con particolare attenzione alla variabilità delle suddette stime. Infine, nel Paragrafo 1.8 sono pubblicati i risultati dell'indagine di misura (distorsione e varianza di risposta) a livello regionale, di ripartizione e nazionale.

Il Capitolo 2 descrive tutte le fasi dell'indagine di copertura (PES - *Post Enumeration Survey*) e i risultati che hanno portato a misurare il livello di copertura del 6° Censimento generale dell'agricoltura.

In particolare, nel Paragrafo 2.1 sono descritti gli obiettivi della rilevazione post censuaria, nel Paragrafo 2.2 il disegno dell'indagine e nel 2.3 l'organizzazione delle attività. Il Paragrafo 2.4 si occupa degli aspetti legati al disegno di campionamento e ai metodi di stima dell'errore di sotto-copertura, nonché lo studio sulla variabilità delle stime stesse. Il Paragrafo 2.5 descrive tutte le fasi dell'indagine legate alla raccolta dei dati e al monitoraggio (centrale e decentrato) della stessa. Il Paragrafo 2.6 esamina le procedure di controllo e correzione dei dati, con particolare attenzione alla valutazione dell'eleggibilità dell'unità agricola ad azienda. Il Paragrafo 2.7 presenta le procedure metodologiche di "*record linkage*" tra il Censimento e l'indagine di copertura. Il Paragrafo 2.8, infine, presenta i risultati dell'indagine di copertura a livello regionale, di ripartizione geografica e nazionale.

Capitolo 1

L'indagine di misura¹

1.1 Gli obiettivi

Molti Istituti di statistica utilizzano l'indagine di misura come strumento per la valutazione dell'errore di risposta nelle indagini esaustive (censimenti) o campionarie. L'errore di rilevazione, o di risposta, è la parte di errore extra-campionario dovuta a difformità tra il valore vero e il valore rilevato presso le unità statistiche osservate. Esso è imputabile agli errori commessi nella fase di rilevazione dei dati e causati dall'azione di rispondenti, di rilevatori e supervisori.

Poiché è particolarmente difficile evitare questi errori non campionari, in letteratura sono state proposte varie tecniche (Forsman e Schreiner, 1991) al fine di misurare le componenti dell'errore di risposta. Tra le tecniche proposte, quella più utilizzata, soprattutto nelle indagini agricole (sia campionarie che censuarie), è la tecnica della re-intervista, da effettuarsi su un campione di aziende agricole già rilevate all'indagine principale (campionaria o censuaria) al quale è sottoposto tutto o una parte del questionario utilizzato nella prima indagine. Nel caso in cui le risposte fornite dalla stessa unità per il medesimo quesito nelle due indagini risulti essere differente, viene applicata una procedura di riconciliazione tra i due valori discordanti.

Ciò premesso, l'indagine post censuaria è finalizzata a fornire stime sufficientemente affidabili dell'errore di misura e delle sue principali componenti (distorsione e varianza di risposta) relativamente ad alcune variabili ritenute significative e rilevate al Censimento. L'indagine è stata condotta mediante re-interviste effettuate con tecnica telefonica (CATI) da una ditta esterna su un campione di circa 50.000 aziende, già rilevate al Censimento ed estratte con un disegno di campionamento a uno stadio stratificato dalla lista censuaria. Le variabili oggetto di re-intervista hanno riguardato le superfici dell'azienda e la consistenza dei principali allevamenti. I domini territoriali di riferimento delle stime sono l'intero territorio nazionale, le cinque ripartizioni geografiche e le 21 regioni e province autonome².

Nella progettazione e nella realizzazione di tale indagine di misura si è tenuto conto dell'esperienza maturata con le indagini di qualità effettuate sia per il 5° Censimento generale dell'agricoltura del 2000 sia per le tre indagini su Struttura e produzione delle aziende agricole degli anni 2003, 2005 e 2007.

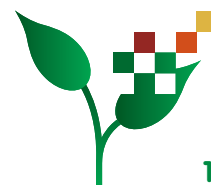
1.2 Il disegno dell'indagine

L'indagine di misura, a carattere nazionale, è di interesse pubblico ed è inserita nel Programma Statistico Nazionale 2011-2013 (codice IST-02440) approvato con DPCM del 31 marzo 2011. La popolazione di riferimento dell'indagine campionaria, è rappresentata dalle aziende agricole presenti sul territorio italiano e rilevate al 6° Censimento generale dell'agricoltura³. L'unità di rilevazione è l'azienda agricola o zootecnica, definita ai fini censuari come "l'unità tecnico-economica, costituita da terreni, anche in appezzamenti non contigui, ed eventualmente da impianti e attrezzature varie, in cui

1 Autori: A. Bernardini §§ 1.2, 1.5.2, 1.6, 1.A; L. De Gaetano §§ 1.3, 1.5.4; C. De Vitiis § 1.7.1; C. De Vitiis e M. Russo §§ 1.4.1, 1.4.3, 1.7.3; M. Mazziotta § 1.1, 1.7.4, 1.8; M. Russo §§ 1.7.2, 1.B; L. Soriani §§ 1.5.1, 1.5.3.

2 Le Province autonome di Trento e Bolzano sono state assimilate a singole regioni.

3 L'universo di riferimento è complessivamente di 1.654.954 aziende di cui 1.620.884 attive e 34.070 temporaneamente inattive.



si attua, in via principale o secondaria, l'attività agricola e zootecnica a opera di un conduttore - persona fisica, società, o ente - che ne sopporta il rischio sia da solo, come conduttore coltivatore o conduttore con salariati e/o compartecipanti, sia in forma associata".

Si tratta quindi di aziende con produzione agricola o zootecnica, indipendentemente dalla loro entità quantitativa (possono essere anche senza produzione nell'anno di riferimento e quindi definite come temporaneamente inattive) e dalla presenza di vendite (aziende con solo autoconsumo). Anche l'esistenza di terreno agrario non costituisce carattere distintivo, pertanto sono state considerate unità di rilevazione anche le aziende esclusivamente zootecniche, prive di terreno agrario.

L'Istat ha estratto dalla lista censuaria un campione di 50.000 aziende agricole⁴ (campione base), rappresentative dell'universo di riferimento, le quali hanno nuovamente fornito i dati per alcuni quesiti del questionario censuario che avevano precedentemente compilato.

Al momento dell'estrazione del campione sono state acquisite tutte le informazioni anagrafiche dell'azienda (denominazione, localizzazione del centro aziendale) e del conduttore (nome, cognome, codice fiscale, residenza, recapito telefonico). Successivamente la rilevazione si è articolata in due distinte fasi:

il reperimento dei questionari censuari già compilati dalle aziende agricole selezionate nel campione;

la rilevazione delle medesime attraverso un'intervista telefonica.

Il reperimento dei questionari censuari è stato condotto secondo due diverse modalità: *i*) nelle regioni ad Alta partecipazione, poiché la registrazione dei dati era avvenuta a cura degli Uffici Regionali di Censimento (URC) in tempo reale durante la fase di raccolta dei dati, da parte delle regioni stesse, i questionari censuari sono stati acquisiti da ciascun URC estraendo direttamente dal Sistema di Gestione della Rilevazione (SGR) informatizzato i dati relativi alle variabili precedentemente selezionate da sottoporre a verifica; *ii*) nelle regioni a Partecipazione integrativa, poiché la registrazione dei questionari è avvenuta a cura di una ditta esterna incaricata dall'Istat in tempo differito rispetto al termine della fase di raccolta dei dati, l'acquisizione è avvenuta dal data base dei questionari registrati dalla ditta. Al fine di garantire la disponibilità dei medesimi in tempo utile per l'effettuazione della rilevazione di misura, sono stati segnalati alla Società di registrazione i codici identificativi delle aziende appartenenti al campione, al fine di anteporre la registrazione dei loro questionari a quella delle altre. L'Istat, inoltre, ha effettuato il monitoraggio dello stato di avanzamento della loro registrazione, nonché lo scarico delle variabili utili allo svolgimento dell'indagine di misura, attraverso una particolare applicazione informatica sviluppata internamente all'Istituto e fornita alla società menzionata.

La rilevazione si è basata su una re-intervista al campione di aziende agricole selezionato, effettuata mediante tecnica telefonica, meglio nota come *Computer-Assisted Telephone Interviewing* (CATI), da una società esterna incaricata dall'Istat. La scelta di utilizzare una tecnica di rilevazione differente da quella utilizzata nella rilevazione del 6° Censimento generale dell'agricoltura si è resa necessaria per rispettare i vincoli di tempo e di budget. Infatti re-intervistare un campione di aziende distribuito sull'intero territorio nazionale con la stessa tecnica censuaria avrebbe comportato tempi molto lunghi e costi eccessivi. L'adozione della tecnica CATI ha consentito, invece, di re-intervistare un elevato numero di aziende agricole e soprattutto di rilevare l'informazione su tutte le principali variabili contenute nel questionario censuario.

⁴ È stato estratto, inoltre, un elenco suppletivo di 50.000 aziende da utilizzare per sostituire le unità non rispondenti.

Per la riuscita delle re-interviste sono state fondamentali: *i)* le operazioni di selezione dei rilevatori, realizzata secondo criteri di elevata professionalità da parte della Società incaricata; *ii)* la formazione tecnica e teorica tenuta direttamente dal personale Istat ai rilevatori selezionati; *iii)* l'informativa inviata dall'Istat alle aziende del campione e *iv)* la sensibilizzazione delle regioni e delle associazioni di categoria al fine di favorire la collaborazione delle aziende a partecipare all'indagine.

1.3 Organizzazione e calendario delle attività

Per la progettazione e la successiva implementazione dell'indagine, l'Istat ha costituito al suo interno Gruppi di lavoro⁵ dedicati, mentre la realizzazione della rilevazione è stata affidata a una struttura dell'Istituto⁶ autonoma dalla Direzione centrale per i censimenti generali responsabile del 6° Censimento dell'agricoltura.

Per effettuare l'intervista è stato progettato e disegnato un questionario (Allegato elettronico 1.B), la cui complessa articolazione si è resa indispensabile per consentire la riconciliazione del dato rilevato. Ciò ha imposto delle scelte sulla tipologia e sul numero dei quesiti presenti nel questionario di azienda del Censimento dell'agricoltura da inserire nell'indagine di qualità. In particolare, si è ritenuto opportuno inserire soltanto i quesiti che si riferivano ai totali delle superfici adibite alle coltivazioni di più larga diffusione (tenendo conto di alcune specificità territoriali), alla consistenza delle principali specie di bestiame e altre informazioni ritenute importanti per la descrizione della struttura aziendale. Oltre al modello di rilevazione l'Istat ha predisposto anche lo schema di intervista, mentre la società incaricata di effettuare l'indagine CATI ha sviluppato il software per la rilevazione.

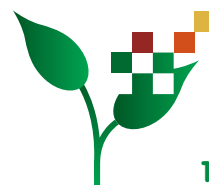
Le operazioni preliminari alla raccolta dei dati hanno riguardato:

- l'estrazione delle unità campionarie, mediante campionamento a uno stadio stratificato (Paragrafo 1.4);
- l'invio, alla ditta incaricata della registrazione dei questionari cartacei del 6° Censimento dell'agricoltura, dei codici identificativi delle aziende appartenenti al campione di aziende, per le regioni a Partecipazione integrativa;
- l'estrazione delle variabili oggetto di re-intervista relative ai questionari delle aziende appartenenti al campione delle regioni ad Alta partecipazione;
- l'assegnazione dell'incarico alla ditta esterna per l'indagine di misura mediante espletamento di gara d'appalto;
- l'esecuzione dei test del software realizzato dalla ditta incaricata della rilevazione;
- la formazione degli operatori telefonici (rilevatori)⁷;
- l'invio, a partire dal mese di Maggio 2011 e a tutte le aziende del campione, della lettera informativa contenente la descrizione degli obiettivi, delle tematiche trattate durante l'intervista, delle modalità con cui le aziende sarebbero state contattate, in-

5 La progettazione dell'indagine di misura è stata affidata nel mese di giugno 2010 a un Gruppo di lavoro interno all'Istat avente il compito di predisporre i metodi, le tecniche e i disegni organizzativi delle rilevazioni dell'errore di misura e della copertura del 6° Censimento generale dell'agricoltura (Del. DPTS n. 92 del 27 maggio 2010). L'esecuzione dell'indagine è stata affidata a una apposita struttura dell'Istat, coadiuvata da un gruppo di lavoro (Del. DPTS n. 209 del 14 dicembre 2010), costituito da diversi esperti appartenenti alle varie direzioni dell'Istituto, con il compito di fornire il supporto organizzativo, tecnico e metodologico e di sovrintendere a tutte le fasi delle indagini post-censuarie del 6° Censimento dell'agricoltura (misura e copertura).

6 La struttura competente preposta alle due rilevazioni post-censuarie (MSS1 - Indagini post censuarie dei censimenti della popolazione e dell'agricoltura del Dipartimento per l'integrazione, la qualità e lo sviluppo delle reti di produzione e di ricerca) si è avvalsa delle riunioni del Comitato consultivo per la preparazione a livello regionale del 6° Censimento generale dell'agricoltura, che hanno rappresentato occasioni di discussione sullo stato dei lavori e dell'intero progetto di realizzazione dell'indagine.

7 In considerazione della tipologia di rilevazione, l'Istat ha dedicato particolare attenzione alla formazione degli operatori (circa 70) utilizzando molteplici strumenti: accanto alla tradizionale formazione in aula con supporti didattici, si è fatto ricorso anche a test/esercizi per la valutazione del grado di apprendimento degli intervistatori e a momenti di focalizzazione su casi particolari. Formazione ad hoc è stata fatta ai rilevatori operanti nella provincia di Bolzano, le cui interviste telefoniche sono state effettuate in lingua tedesca.



clusi i giorni e gli orari nei quali avrebbero ricevuto la telefonata (cfr. lettera informativa Allegato elettronico 1.A).

- allestimento di un servizio di assistenza telefonica e per mezzo di casella di posta elettronica, al quale le unità del campione potevano rivolgersi per avere chiarimenti e assistenza sulla rilevazione⁸.

Il servizio di assistenza è stato attivato presso l'Istat a metà Aprile 2011 ed è rimasto attivo per tutto il periodo della rilevazione. Esso è risultato di particolare efficacia nel tranquillizzare le aziende circa l'ufficialità dell'indagine condotta dall'Istat e nell'offrire le garanzie che le stesse esigevano in termini di riservatezza dei dati loro richiesti. Il servizio è stato usato anche per migliorare la gestione delle interviste e degli appuntamenti telefonici.

La rilevazione si è svolta nell'arco di otto mesi, dal 20 maggio 2011 al 31 gennaio 2012, data di chiusura delle interviste telefoniche da parte della ditta esterna, per un totale di 164 giorni di attivazione. Al termine della raccolta dei dati, la ditta ha provveduto all'elaborazione degli output e all'invio dei file completi all'Istat per i successivi controlli di qualità e analisi, per il calcolo delle stime e la diffusione dei dati, secondo il calendario prefissato (Prospetto 1.1).

Prospetto 1.1 - Calendario delle operazioni

ATTIVITÀ/OPERAZIONI	Periodo effettivo
Estrazione delle unità campionarie	Marzo - Aprile 2011
Assegnazione dell'incarico alla Società di rilevazione telefonica	Aprile - Maggio 2011
Realizzazione e verifica del <i>software</i> CATI e definizione del sistema di <i>output</i>	Aprile - Maggio 2011
Istruzione ai rilevatori e supporto alle operazioni (<i>briefing</i> teorico e <i>briefing</i> pratico)	Maggio 2011
Invio della lettera informativa alle aziende campione degli elenchi base e suppletivo	Maggio 2011
Inizio interviste	20 Maggio 2011
Monitoraggio delle interviste	Tutto il periodo di rilevazione
Termine delle interviste	31 Gennaio 2012
Elaborazione degli output e invio all'Istat, da parte della Società CATI, dei file completi	Febbraio 2012
Controllo e correzione dei dati	Marzo 2012 - Giugno 2012
Analisi dei dati e calcolo delle stime	Settembre 2012 - Giugno 2013
Diffusione dei dati definitivi	Dicembre 2013

Nel complesso le operazioni si sono svolte all'interno delle scadenze fissate, sebbene la fase di raccolta dei dati abbia registrato qualche slittamento della chiusura, a causa di ritardi imputabili alle attività di registrazione dei questionari cartacei del Censimento e relativi alle regioni a partecipazione integrativa.

⁸ Per la provincia di Bolzano il servizio di assistenza telefonica è stato curato dall'ufficio di statistica di Bolzano, così come la lettera informativa che è stata predisposta, in collaborazione con l'Istat, in lingua tedesca.

1.4 Il piano di campionamento

1.4.1 Il disegno di campionamento

Il disegno di campionamento è stato studiato, usando le informazioni contenute nella lista pre-censuaria, già disponibili prima dell'inizio delle operazioni censuarie, sia per la stratificazione delle unità della popolazione di interesse sia per lo studio dell'allocatione del numero di aziende agricole campione negli strati.

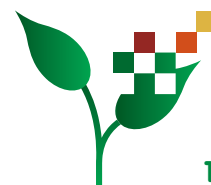
La selezione del campione è stata realizzata sulle unità della lista censuaria, in quanto questa fase si è svolta quando la rilevazione censuaria era terminata, nonostante fossero disponibili solo le informazioni censuarie relative alle variabili principali necessarie per la selezione. È stata altresì effettuata una verifica sulle distribuzioni delle unità tra gli strati, dalla quale è risultato che non vi erano differenze significative tra le due liste e che, pertanto, lo studio dell'allocatione sulla lista pre-censuaria si poteva ritenere ancora valido. I domini di riferimento delle stime sono l'intero territorio nazionale, le cinque ripartizioni geografiche (Nord-ovest, Nord-est, Centro, Sud, Isole), le 21 regioni e province autonome e le seguenti classi di Superficie agricola utilizzata (SAU):

- 0, aziende con SAU uguale a 0 ettari;
- 1, aziende con SAU compresa in (0 - 0,99) ettari;
- 2, aziende con SAU compresa in (1 - 1,99) ettari;
- 3, aziende con SAU compresa in (2 - 4,99) ettari;
- 4, aziende con SAU compresa in (5 - 9,99) ettari;
- 5, aziende con SAU compresa in (10 - 19,99) ettari;
- 6, aziende con SAU compresa in (20 - 49,99) ettari;
- 7, aziende con SAU compresa in (50 - 99,99) ettari;
- 8, aziende con SAU compresa in (100 - ∞) ettari.

Il disegno di campionamento adottato è di tipo probabilistico a uno stadio stratificato, comunemente utilizzato per le indagini sulle aziende agricole condotte dall'Istat. In generale, l'obiettivo della stratificazione è quello di migliorare l'efficienza delle stime rispetto a un campione non stratificato della stessa dimensione. È chiaro che una tale finalità può essere conseguita solo stratificando sulla base di una o più variabili correlate con le variabili di interesse. Tale obiettivo è stato reso possibile dalla disponibilità di una lista contenente, per tutte le unità della popolazione d'interesse, oltre alle informazioni identificative, ulteriori variabili caratterizzanti l'azienda agricola e correlate con le variabili oggetto di studio. Si è ritenuto opportuno esaminare le variabili di tipo territoriale, quelle relative alla dimensione dell'azienda agricola, le variabili relative al tipo di produzione e all'affidabilità del dato presente nella lista pre-censuaria⁹.

Le variabili territoriali considerate sono la regione e la provincia. In base alla prima sono stati definiti i domini pianificati - in altre parole le regioni geografiche - ossia gli ambiti territoriali di stima con dimensioni campionarie atte a garantire prefissati livelli di precisione delle corrispondenti stime campionarie. Relativamente alla seconda variabile territoriale, poiché le risorse disponibili non consentivano di avere una numerosità campionaria adeguata a garantire una sufficiente precisione campionaria delle stime a livello sub-regionale, essa è stata utilizzata solamente al fine di assicurare la copertura di tutte le province e migliorare eventuali stime a livello provinciale, nel caso in cui, in un prossimo futuro, si dovesse avere necessità di ottenere informazioni ad un livello di dettaglio maggiore, ricorrendo però a metodi di stima indiretti, come, ad esempio, quelli basati sugli stimatori per piccole aree.

⁹ Le variabili relative alla produzione e all'affidabilità del dato, dopo uno studio avente la finalità di stabilire l'opportunità o meno del loro impiego, sono state ignorate per le motivazioni illustrate nel prosieguo.



Per ciò che concerne la stratificazione provinciale, si è valutata, anche, la possibilità di tener conto, nell'allocazione della dimensione campionaria di ciascuna regione tra le province appartenenti a essa, di un'eventuale loro specializzazione in alcune colture tipiche (ad esempio, la produzione di fiori nella Provincia di Imperia). Non è stato però possibile stratificare le aziende secondo la specializzazione, in quanto tale informazione non era presente nella lista pre-censuaria né nei dati delle fonti amministrative di riferimento.

Era, inoltre, opportuno, per ragioni di efficienza del disegno, poter stratificare le aziende rispetto a una misura della loro dimensione. Tra le variabili presenti nella lista quelle che meglio esprimevano una misura della dimensione aziendale sono la SAU e la Superficie totale (ST). Tuttavia, il loro utilizzo non era possibile poiché tale informazione non era disponibile per tutte le unità contenute nella lista pre-censuaria; inoltre, la SAU non è comprensiva della superficie eventualmente destinata al ricovero per la stabulazione dell'allevamento, quindi può non essere indicata per le aziende zootecniche, in particolare per quelle di tipo intensivo.

L'unica variabile presente in archivio per tutte le unità e che poteva essere utile per i fini in discussione era la "eleggibilità", che presenta modalità da 1 a 10 ed esprime "l'intensità del segnale per l'inserimento dell'unità nella lista pre-censuaria"; tale variabile fornisce da un lato una misura della rilevanza economica dell'azienda, dall'altro esprime l'affidabilità dell'inclusione dell'unità statistica nella lista.

In ogni caso, la scelta finale delle variabili di stratificazione è stata effettuata anche sulla base di una valutazione del numero di strati che ogni combinazione delle modalità delle variabili considerate avrebbe comportato, in quanto è opportuno tenere sempre sotto controllo il numero di strati.

Da qui la necessità - a partire dalla SAU - di creare una variabile, di seguito indicata, per semplicità espositiva, con SAUM (dove la "M" sta per "modificata"), le cui modalità sono più aggregate di quelle precedentemente individuate e così definite:

- 1, aziende con SAU compresa in (0 - 1) ettari;
- 2, aziende con SAU compresa in (1 - 2) ettari;
- 3, aziende con SAU compresa in (2 - 5) ettari;
- 4, aziende con SAU compresa in (5 - 10) ettari;
- 5, aziende con SAU compresa in (10 - ∞) ettari;
- 6, aziende con SAU non valorizzata, con o senza allevamenti.

Una volta definite la stratificazione e la numerosità campionaria di ogni strato, la selezione delle unità è stata effettuata da ogni strato elementare senza re-immissione e con probabilità uguali. Il campione di aziende agricole è stato selezionato al termine della rilevazione del Censimento a partire dalla lista censuaria (si veda in proposito il Paragrafo 1.2).

1.4.2 La determinazione della numerosità campionaria e sua allocazione tra gli strati

La dimensione del campione di aziende agricole è stata definita con l'obiettivo di garantire una prefissata affidabilità a livello dei domini di stima definiti in fase di progettazione del disegno campionario; tenendo conto sia dei vincoli di disponibilità economica, sia dell'esperienza acquisita con le indagini sulla *Struttura e la produzione delle aziende agricole* (SPA)¹⁰ realizzate dall'Istat nel decennio precedente il 2010.

¹⁰ La numerosità del campione in tali indagini è di circa 50.000 aziende agricole.

Ai fini del calcolo della numerosità campionaria e della relativa allocazione tra i suddetti domini, non disponendo di informazioni sulla variabilità dei fenomeni di interesse derivanti da precedenti indagini¹¹, la procedura di allocazione è stata impostata considerando come parametro di stima una frequenza relativa del 10 per cento in tutti i domini di interesse; in tal modo il disegno campionario è tale da garantire una sufficiente precisione per stime corrispondenti almeno al 10 per cento a livello nazionale, di ripartizione e di regione.

Il metodo adottato è un'estensione dell'allocazione ottimale di Bethel (1989) al caso multi-dominio (De Vitiis *et al.*, 1998). Tale metodo, che è stato implementato nel software *Multivariate Allocation of Units in Sampling Surveys - MAUSS* (utilizzato per questa indagine), consente di definire le numerosità campionarie negli strati in modo tale da rispettare i vincoli sugli errori attesi di campionamento (espressi in termini di coefficienti di variazione) delle stime "tipiche", prefissati a livello dei differenti domini pianificati.

I vincoli suddetti sono stati fissati tutti sotto al 10 per cento, tranne per i domini le cui modalità sono date dall'incrocio tra le classi di SAU e le ripartizioni geografiche o le regioni e province autonome.

I risultati ottenuti hanno portato alla definizione di una dimensione campionaria di circa 50.000 unità e alla loro conseguente distribuzione tra le regioni. Per far fronte a eventuali mancate risposte totali all'indagine, è stato previsto un elenco suppletivo di unità campionarie da utilizzare per sostituire le unità non rispondenti. Tale elenco è costituito, per ogni strato, da un numero di unità pari a quello del campione base, tranne per gli strati incrocio con la modalità 6 - "aziende con SAU non valorizzata, con o senza allevamenti" della variabile SAUM, per i quali è stato selezionato un campione con numerosità doppia appartenente all'elenco base.

1.5 Rilevazione e monitoraggio

1.5.1 Il questionario d'indagine

Tutte le operazioni di organizzazione e realizzazione dell'indagine di misura sono state concepite per assicurare il più alto grado di qualità all'intero processo, a partire dalla progettazione del questionario.

Il modello di rilevazione (Mod. Istat CATI/2011) è lo schema di intervista messo a punto per l'indagine telefonica, sulla base del quale la ditta esterna - con la collaborazione dell'Istat - ha predisposto il software per la rilevazione.

Per la progettazione del questionario di intervista sono state operate delle scelte sulla tipologia e sul numero dei quesiti da selezionare tra quelli presenti nel questionario di azienda del 6° Censimento generale dell'agricoltura 2010.

In particolare, si è ritenuto opportuno prendere in esame soltanto i quesiti che si riferiscono ai totali delle superfici adibite alle coltivazioni ritenute a più larga diffusione (pur tenendo conto di alcune specificità territoriali), alle superfici irrigate, alla consistenza delle principali specie di bestiame in allevamento al 24 ottobre 2010.

Il modello di rilevazione (Allegato elettronico 1.B) è costituito dalle sezioni come di seguito indicato:

¹¹ In realtà, ai fini dello studio dell'allocazione della numerosità campionaria tra i domini di stima, si era pensato di sfruttare i dati a disposizione dell'indagine di controllo condotta nel 2008 dell'indagine SPA (in quanto, tra le diverse indagini realizzate a tale scopo, quella del 2008 risulta essere maggiormente affidabile e si basa su una metodologia di rilevazione e validazione più consolidata), il cui campione è costituito da circa 3.500 aziende. Prendendo in considerazione le principali variabili relative a superfici, colture, allevamenti e altre importanti caratteristiche caratterizzanti l'azienda agricola. Questa possibilità, però, è decaduta in quanto non erano disponibili misure di variabilità per le variabili che si volevano considerare ai fini dell'allocazione del campione.



a) Sezione I: Informazioni sull'intervista di base.

Questa prima sezione è stata strutturata in maniera da individuare il rispondente al questionario di azienda del Censimento dell'agricoltura (Mod. Istat CEAGR), cui sottoporre l'intervista telefonica; tale operazione risponde all'esigenza di garantire l'omogeneità intertemporale dei rispondenti e quindi la coerenza delle risposte alle due indagini prese in esame. Poiché per gli obiettivi dell'indagine era importante ottenere le informazioni dalla medesima persona che aveva partecipato all'indagine censuaria, il questionario di intervista prevedeva una serie di domande filtro volte a identificare in maniera esatta la persona che rispondeva al questionario della CATI. A tale scopo, questa sezione è suddivisa in due parti (riquadro A e riquadro B) che permettono di differenziare l'intervista a seconda che l'intervistato sia o meno la stessa persona a cui è stato somministrato il questionario censuario.

b) Sezione II: Utilizzazione dei terreni nell'annata agraria 2009-2010.

La seconda sezione prende in esame alcune caratteristiche dell'azienda relative alla sua dimensione in termini di superfici (ST e SAU), all'utilizzazione dei terreni (per gruppi di coltivazioni e per colture principali) e ad alcuni metodi di produzione aziendale nell'annata agraria 1 novembre 2009 - 31 ottobre 2010. Il questionario dell'indagine di misura è stato progettato in modo da rispettare il seguente rapporto gerarchico: *Colture principali ≤ Gruppi di coltivazioni ≤ Superfici totali*. La somma delle colture principali deve essere coerente (inferiore o uguale) con il rispettivo gruppo di coltivazione; la somma dei gruppi di coltivazione deve essere coerente (inferiore o uguale) con la SAU; la risposta relativa alla SAU stessa deve essere coerente con quella fornita per la ST (SAU≤ST).

Le colture principali selezionate sono:

- Frumento;
- Mais da granella;
- Terreni a riposo soggetti a regime di aiuto;
- Vite;
- Olivo;
- Agrumi;
- Melo.

I gruppi di coltivazioni sono:

- Seminativi;
- Coltivazioni legnose agrarie;
- Prati permanenti e pascoli.

Questi tre gruppi, insieme agli orti familiari, compongono la SAU.

I totali di superficie sono due:

- Superficie agricola utilizzata;
- Superficie totale.

Inoltre, in merito alle informazioni relative ai metodi di produzione agricola richieste dal Regolamento (CE) 1166/2008, è stata introdotta nel questionario la variabile Superficie irrigata;

c) Sezione III: Notizie sulla consistenza degli allevamenti al 24 ottobre 2010.

Nella terza sezione sono stati inseriti i quesiti volti a misurare la qualità dei dati censuari relativi alla consistenza di cinque tipologie di bestiame in allevamento. Le specie prese in esame sono:

- Bovini;
- Bufalini;
- Ovini;

- Caprini;
- Suini.

Per ovini, caprini e suini sono richieste informazioni solamente relative ai capi, o i cui prodotti, siano destinati alla vendita (coerentemente al questionario di rilevazione del Censimento).

1.5.2 Il processo di riconciliazione

Il questionario, redatto dall'Istat in forma cartacea, è stato implementato dalla ditta esterna, in collaborazione con l'Istat, nel software di gestione della rilevazione CATI. Il disegno dell'indagine ha richiesto l'upload dei dati censuari relativi alle aziende del campione selezionato, poiché il software prevedeva l'eventuale attivazione del processo di riconciliazione. Infatti, poiché uno degli obiettivi dell'indagine era la stima della distorsione, la re-intervista alle aziende è stata predisposta allo scopo di determinare il valore vero della variabile di interesse mediante l'applicazione del processo di riconciliazione. Per questo motivo, durante la re-intervista il rilevatore ha avuto a disposizione le risposte originarie cosicché in caso di discordanza del dato dichiarato al censimento con il nuovo dato dichiarato durante la re-intervista, poteva appurare, con l'aiuto del rispondente, quale fosse la risposta vera. Per rispettare le condizioni di indipendenza tra la rilevazione del Censimento e l'indagine telefonica (CATI), i rilevatori non conoscevano, per ciascuna variabile, il valore indicato nel questionario censuario. Questo valore, essendo pre-caricato nel sistema per la gestione delle re-interviste, veniva utilizzato in automatico per il controllo di coerenza con il dato fornito alla re-intervista telefonica. Per la riconciliazione è stato fissato un campo di variazione al di fuori del quale non venivano tollerate le discrepanze tra i due dati: se la differenza tra i valori indicati per il medesimo quesito alla CATI e al Censimento, differivano di ± 5 per cento per l'estensione delle colture principali e di ± 10 per cento per il numero dei capi delle principali specie animali, il software sottoponeva i due valori a riconciliazione. Durante questa fase l'operatore, poteva visualizzare i due valori e chiedeva all'intervistato, senza alcun condizionamento, quale valore doveva essere considerato il valore "vero".

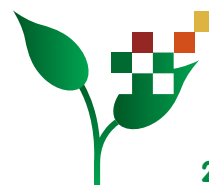
Alla fine della formazione sono state predisposte delle interviste di prova, per testare il software CATI implementato sulla base del questionario elettronico, e per valutare la professionalità e il grado di apprendimento degli operatori sul campo. Le interviste di prova hanno permesso di implementare delle soluzioni migliorative del questionario elettronico e correggere dei quesiti, che potevano portare a errate interpretazioni da parte degli intervistati.

1.5.3 Contatto ed esito delle interviste

L'intervistatore telefonico, nel prendere contatto con le aziende agricole del campione, doveva precisare all'intervistato che l'Istat era il titolare della rilevazione e che le informazioni fornite erano tutelate dal segreto statistico e sottoposte alle regole sulla riservatezza stabilite dal d.lgs. 30 giugno 2003, n.196.

Visti gli obiettivi dell'indagine era opportuno cercare di acquisire le informazioni richieste con l'intervista telefonica dalla stessa persona (conduttore¹² o altra persona) che aveva risposto alle domande del questionario del Censimento 2010. Le interviste erano effettuate dal lunedì al venerdì, nell'intervallo orario 8.30-20.30.

12 Responsabile giuridico ed economico che sopporta il rischio della gestione aziendale sia da solo (conduttore coltivatore e conduttore esclusivamente con salariati e/o compartecipanti), sia in associazione. Il conduttore può essere una persona fisica, una società o un ente.



Il processo che regolava i tentativi di contatto doveva uniformarsi al seguente schema:

1. i contatti telefonici e le eventuali sostituzioni dovevano essere gestiti con procedure automatiche messe a punto dalla ditta esterna;
2. la procedura doveva prevedere diversi tentativi di contatto telefonico per intervistare i nominativi campione; in particolare la ditta incaricata doveva ripetere la chiamata almeno cinque volte in giorni diversi prima di sostituirla con un'unità appartenente all'elenco sostitutivo e facente parte dello stesso strato campionario. A tale riguardo la ditta doveva registrare l'ora e il giorno in cui erano avvenuti i diversi contatti con la stessa unità.

La sostituzione di un nominativo avveniva solo se si verificavano le seguenti condizioni:

- rifiuto dell'intervista;
- numero telefonico non reperibile¹³;
- irreperibilità prolungata del nominativo.

Riguardo quest'ultima eventualità, un nominativo veniva considerato irreperibile quando "non rispondeva nessuno" o rispondeva una "segreteria telefonica" per almeno cinque tentativi di contatto da effettuare in orari e giorni differenti, all'interno della fascia oraria di riferimento. Nel caso in cui il numero telefonico dell'interessato risultasse "occupato" il software predisposto dalla ditta doveva prevedere in automatico tentativi di contatto ogni 15 minuti per almeno quattro volte consecutive, da ripetere in giorni ed orari differenti per almeno cinque volte. Solo nel caso in cui i tentativi non avevano dato esito, il nominativo non era più contattato e si provvedeva alla sua sostituzione.

Inoltre, in caso di un nominativo trasferito (nel medesimo comune o in altro comune) la ditta doveva reperire il nuovo numero telefonico e provvedere successivamente ad effettuare l'intervista secondo le modalità di rilevazione definite.

Per quanto concerne la gestione degli appuntamenti successivi (nel caso in cui, cioè, si aveva avuto più di un contatto telefonico conclusosi con esito provvisorio "appuntamento"), l'unità doveva essere sostituita dopo cinque appuntamenti che non avevano dato luogo ad intervista.

Le sostituzioni delle unità dell'elenco sono avvenute con una rigida procedura stabilita dall'Istat, che forniva dei criteri di selezione dall'elenco sostitutivo in funzione delle caratteristiche del nominativo da sostituire. Il software realizzato per l'intervista doveva, inoltre, creare un collegamento tra il nominativo sostituito e quello selezionato per la sostituzione (mediante, ad esempio, la segnalazione del codice azienda dell'azienda sostitutiva). Tutte le informazioni, relative ai contatti, agli esiti e alle informazioni statistiche acquisite tramite questionario, dovevano essere quotidianamente registrate su supporto digitale da parte della Società, in modo da risultare immediatamente disponibili ed inviate giornalmente all'Ufficio Istat responsabile dell'indagine di misura (entro le ore 11.00 del giorno successivo), che si riservava eventuali interventi.

1.5.4 Il sistema di monitoraggio

All'organizzazione della rete di rilevazione CATI e al forte investimento sulla formazione effettuata dall'Istat, si è affiancato il continuo e capillare monitoraggio del lavoro svolto dagli intervistatori telefonici. Sono stati predisposti indicatori di sintesi che si riferivano al campione complessivo e altri riferiti ai due elenchi (base e suppletivo).

Il flusso informativo ha permesso, sulla base dei contatti presi e delle interviste svolte, di produrre e analizzare un vasto set di indicatori di "*performance della rilevazione*".

¹³ La ditta doveva provvedere alla ricerca del numero di telefono ove mancante in lista o non corretto.

In particolare, il sistema ha permesso di seguire il lavoro del rilevatore per tutti i contatti con le aziende agricole. Venivano pertanto elaborati indicatori relativi: i) all'avvenuta ricezione della lettera informativa; ii) al numero complessivo di contatti con le aziende agricole; iii) alla gestione degli appuntamenti; iv) alla durata dell'intervista; v) alle risposte *proxy*; vi) agli esiti definitivi (interviste complete, mancate risposte totali e parziali, motivazione delle mancate risposte, sostituzioni).

In presenza di valori anomali, si sono adottate tempestivamente le misure correttive idonee a risolvere le situazioni critiche, intervenendo sia sugli operatori sia sulle generiche problematiche legate all'indagine.

Alla Società, come già detto precedentemente, è stato fornito un elenco, con relativo tracciato record, delle aziende appartenenti al campione base (50.000 aziende) ed un elenco suppletivo (circa 50.000 aziende) da utilizzare per le sostituzioni di unità, nel caso di contatti non utili. Le integrazioni, in questo ultimo caso, sono state effettuate secondo uno schema di riferimento fornito dall'Istat e indicato nei paragrafi precedenti.

La funzione di monitoraggio dell'andamento delle diverse fasi della rilevazione si alimentava automaticamente attraverso l'esecuzione delle varie funzioni previste nel software di registrazione (intervista completa, parziale, occupato, non risponde, appuntamento, eccetera) generando una serie di rapporti riepilogativi.

I file di reportistica, prodotti giornalmente, si distinguevano in:

- “*Report sullo sfruttamento del file nominativi*”; per ogni strato erano riportate le seguenti situazioni: totale aziende, da contattare, successi, non chiamate, fallite, appese (da richiamare), mancanza di linea, occupato, non risponde, dialogo impossibile, segreteria telefonica, persona irraggiungibile, rifiuto, numero errato, appuntamento, quota raggiunta, fax o modem, numero doppio, cliente disconnesso, errore sconosciuto, no contatto, questionario compilato da associazione di categoria, eccetera;
- “*Report intervistatore per giorno*”; per ogni rilevatore, veniva indicato nome e cognome, codice identificativo univoco ed esito del contatto (successo, occupato, non risponde, dialogo impossibile, segreteria telefonica, rifiuto, numero errato, appuntamento, fax o modem, cliente disconnesso, questionario compilato da associazione di categoria) avuto nello svolgimento dell'intervista;
- “*Report riempimento nominativi*”; per ogni strato veniva indicato lo stato dell'intervista completa e parziale, con la *componente* dell'esito, ovvero per l'intervista completa: attiva per conduttore oppure attiva per altra persona diversa dal conduttore. Per le interviste parziali veniva indicato uno dei seguenti stati: cessata; temporaneamente inattiva, fuori campo di osservazione. Nel report oltre alla situazione per strato era riportato il totale complessivo degli esiti delle interviste.

I report riepilogativi, alimentati automaticamente con il progredire della rilevazione, hanno rappresentato uno strumento prezioso per l'Istat. Tramite essi l'Istat ha potuto approntare opportuni interventi al verificarsi di situazioni critiche. Attraverso funzioni download messe a punto dalla società, le informazioni contenute nei riepiloghi e il data base di lavorazione sono stati resi costantemente disponibili all'Istat in file di formato e tracciato predefinito Excel.

Giornalmente l'ufficio competente dell'Istat, sulla base della reportistica inviata, provvedeva a calcolare alcuni indicatori elementari, come ad esempio il *tasso di completezza*, inteso come rapporto tra le aziende intervistate e quelle che occorreva intervistare. Sono stati calcolati inoltre altri indicatori della performance qualitativa, ad esempio il *tasso di fedeltà* al campione: considerata la possibilità di sostituire, dopo cinque tentativi, l'azienda del campione base, l'indicatore è stato definito come rap-



porto tra il numero delle aziende del campione base con intervista completa e il numero totale delle aziende del campione base.

Per una valutazione del lavoro svolto è stato utile anche disporre della distribuzione delle *mancate risposte totali* per motivo della non risposta. La caduta campionaria è riconducibile a due motivi: assenza del numero di telefono fisso e irreperibilità delle aziende da intervistare.

È da segnalare che 34.995 (81,5 per cento) delle interviste complete sono state somministrate al conduttore di azienda agricola (Prospetto 1.2) .

Prospetto 1.2 - Distribuzione delle interviste complete e parziali per esito

Attiva, intervista al conduttore	Attiva intervista ad altra persona	Totale attiva	Cessata	Fuori campo di osservazione	Temporaneamente inattiva	Totale interviste parziali (cessata, fuori campo, temp. inattiva)	Totale complessivo
34.995	7.932	42.927	1.885	1.749	829	4.463	47.390

1.6 Il controllo e la correzione dei dati

Nella procedura di Controllo e Correzione (C&C) dei risultati dell'indagine sono state integrate diverse metodologie e tecniche per il trattamento delle diverse tipologie di errore presenti nei dati (errori influenti, valori anomali, errori casuali, mancate risposte parziali).

In particolare, la procedura di C&C si è articolata nelle fasi descritte di seguito:

1. un approccio deterministico tradizionale è stato adottato per la localizzazione e la correzione degli errori di tipo sistematico;
2. l'individuazione degli errori influenti è stata realizzata attraverso l'approccio dell'*editing selettivo* (Latouche *et al.*, 1992), mentre l'individuazione dei valori anomali è avvenuta principalmente attraverso l'analisi di grafici e matrici di transizione in grado di evidenziare le variazioni (rispetto ai dati, rilevati al Censimento dell'agricoltura) delle principali caratteristiche strutturali e produttive di ciascuna unità e la loro influenza sul livello delle stime finali;
3. l'individuazione ed eliminazione delle incoerenze non influenti attribuibili a fattori casuali è stata effettuata mediante la metodologia probabilistica nota come *algoritmo di Fellegi e Holt* (Fellegi, Holt, 1976);
4. validazione dei dati.

Le diverse fasi del processo di C&C sono state coadiuvate dal software, implementato in Istat, *Selective Editing via Mixtures* (Selemix).

La particolare natura dell'indagine, ossia la re-intervista di aziende già, censite, la tecnica utilizzata per la costruzione del questionario elettronico e la presenza di un operatore qualificato che ha supportato l'intervista telefonica sono stati fattori fondamentali nel determinare l'assenza di mancate risposte parziali; pertanto, nel processo di C&C; non è stato necessario intervenire su tale tipologia di errore.

1.7 Modello di stima

1.7.1 La calibrazione dei pesi coinvolti nella determinazione delle stime

Il principio su cui è basato ogni metodo di stima campionaria consiste nella condizione che le unità incluse nel campione rappresentino anche le restanti unità della popolazione oggetto di studio. Questo principio viene realizzato attribuendo a ogni unità campionaria un *peso*, che indica il numero di unità della popolazione rappresentate dall'unità medesima.

Pertanto, per l'indagine in discussione, all'insieme delle unità rispondenti è stato assegnato un sistema di pesi determinato con la tecnica nota in letteratura con il nome di *calibration* (Deville, Särndal, 1992; Singh, Mohl, 1996), che costituisce, peraltro, il metodo di stima standard per la maggior parte delle indagini campionarie condotte dall'Istat.

Al fine di rendere più chiara la successiva esposizione, introduciamo la seguente simbologia. Sia:

- k indice di azienda agricola
- d indice di dominio di stima ($d = 1, \dots, D$)
- h indice di strato ($h = 1, \dots, H_d$)
- H_d numero di strati inerenti al dominio d
- N_h numero di aziende agricole dello strato h
- U popolazione oggetto di studio di dimensione $N \left(N = \sum_{d=1}^D \sum_{h=1}^{H_d} N_h \right)$
- n_h numero di aziende agricole campione dello strato h
- s campione estratto da U di dimensione $n \left(n = \sum_{d=1}^D \sum_{h=1}^{H_d} n_h \right)$
- y generica variabile di interesse
- y_{hk} valore della variabile y rilevato sull'unità k dello strato h , ipotizziamo di voler stimare, con riferimento ad un generico dominio d , il totale della variabile y oggetto di indagine, espresso dalla seguente relazione:

$$y_d = \sum_{h=1}^{H_d} \sum_{k=1}^{N_h} y_{hk} \quad (1)$$

La stima del totale (1) è data da:

$$\hat{y}_d = \sum_{h=1}^{H_d} \hat{y}_h = \sum_{h=1}^{H_d} \sum_{k=1}^{n_h} w_{hk} y_{hk} \quad (2)$$

in cui w_{hk} è il *peso finale* da attribuire alle n_h aziende dello strato h .

La procedura di calibrazione, in generale, viene utilizzata al fine di: *i*) correggere l'effetto distorsivo della mancata risposta totale; nonché *ii*) tenere conto della conoscenza di totali noti di rilevanti variabili ausiliarie (disponibili da fonti esterne all'indagine).

In particolare, la procedura per l'ottenimento dei pesi finali in oggetto è articolata nelle seguenti fasi:



1. si calcolano i pesi base come reciproco della probabilità di inclusione nel campione delle unità costituenti la popolazione d'interesse;
2. si costruiscono i fattori correttivi che consentono di soddisfare le condizioni di uguaglianza tra i totali noti delle variabili ausiliarie e le corrispondenti stime campionarie;
3. si calcolano, infine, i pesi finali come prodotto dei pesi base, di cui al passo 1, per i fattori correttivi ottenuti al passo 2.

I fattori correttivi, sopra citati, sono ottenuti dalla risoluzione di un problema di minimo vincolato, in cui la funzione da minimizzare è una funzione di distanza (opportuna-mente prescelta) tra i pesi base e i pesi finali, mentre i vincoli sono definiti dalla condizione di uguaglianza tra stime campionarie dei totali noti di popolazione e valori noti degli stessi. La funzione di distanza generalmente utilizzata è la funzione logaritmica troncata; l'adozione di tale funzione garantisce che i pesi finali siano positivi e contenuti in un predeterminato intervallo di valori possibili, eliminando in tal modo i pesi positivi estremi (troppo grandi o troppo piccoli).

Un importante stimatore appartenente alla classe dei *calibration estimators*, che si ottiene utilizzando la funzione di distanza euclidea, è lo stimatore di regressione generalizzata. Tale stimatore riveste un ruolo centrale perché è possibile dimostrare che tutti i *calibration estimators* convergono asintoticamente, all'aumentare della numerosità campionaria, allo stimatore di regressione generalizzata (Deville, Särndal, 1992). Nell'indagine in oggetto, si è deciso¹⁴ di utilizzare come totale noto esterno, ai fini della calibrazione, il numero di aziende censite a livello regionale allo scopo di garantire la coerenza tra la popolazione censita e quella stimata dall'indagine.

Indicando, quindi, con x_j ($j = 1, \dots, J$) il totale noto della variabile ausiliaria j per la generica regione geografica e con x_{hkj} il valore assunto dalla variabile ausiliaria j sull'azienda rispondente k dello strato h , la condizione sopra descritta è espressa dalla seguente uguaglianza:

$$x_j = \hat{X}_j = \sum_{h=1}^H \sum_{k=1}^{n_h} w_{hkj} x_{hkj} \quad (j = 1, \dots, J) \quad (3)$$

in cui H indica il numero complessivo di strati definiti nella regione.

È bene sottolineare, infine, che i *calibration estimators* sono approssimativamente corretti per n sufficientemente grande.

14 Tale decisione è scaturita da uno studio volto a valutare l'impatto di tre differenti sistemi di calibrazione dei pesi di riporto all'universo sui coefficienti di variazione di alcune tra le più importanti variabili di interesse per l'indagine, tra le quali la SAU. L'analisi è stata condotta sia sulla stima del totale che sulla differenza relativa tra la stima del totale censuario e quella del totale all'indagine di misura. In particolare, sono stati posti a confronto i coefficienti di variazione relativi a stime i cui pesi erano ottenuti: i) come inverso della probabilità di inclusione (pertanto, senza calibrazione); ii) calibrando i pesi base con il numero di aziende censite a livello regionale; iii) calibrando i pesi base con il totale SAU a livello regionale; iv) calibrando i pesi base con il numero di aziende censite ed il totale SAU, a livello regionale.

Dai risultati di suddetta analisi si evince che sia nel caso in cui il parametro d'interesse è un totale sia se è una differenza relativa di stime di totali, non ci sono significativi guadagni in efficienza. Si è deciso di calibrare solo rispetto al numero di aziende censite a livello regionale, poiché:

a) la lista censuaria disponibile al momento della selezione del campione non coincideva con l'insieme definitivo delle aziende censite quindi i pesi campionari senza calibrazione non avrebbero riportato le stime al numero di aziende agricole per regione; b) i totali di calibrazione dovrebbero possedere il requisito di essere vicini al valore vero o almeno approssimativamente tali (stati- sticamente sarebbe auspicabile che le corrispondenti variabili fossero legate con quelle d'interesse). Si possono verificare le due seguenti situazioni: (i) il totale SAU è un valore vero (o approssimativamente tale), in tal caso potrebbe rivestire il ruolo di totale di calibrazione, conseguentemente, non avrebbe senso valutarne il relativo errore di misura; (ii) non c'è alcuna certezza sul grado di vicinanza rispetto al valore vero (che rappresenta la situazione in questione), in questo caso è sconsigliabile utilizzarlo nella calibrazione, ma è invece opportuno determinarne il relativo errore di misura. Per quanto appena spiegato e poiché la sua introduzione tra i totali di calibrazione non apporta un incremento nella efficienza delle stime, si è stabilito di non prenderlo in considerazione.

1.7.2 Lo stimatore dell'errore di misura

Per errore di misura (o errore di risposta) si intende qualsiasi errore che si manifesta nella fase della raccolta dei dati e per effetto del quale si osserva un valore diverso da quello *vero*. Tale tipo di errore - che può essere dovuto al rispondente, all'intervistatore, al criterio di raccolta delle informazioni, alla predisposizione del questionario, eccetera - introduce, nelle stime delle costanti caratteristiche inerenti la popolazione d'interesse, sia una componente distorsiva sia un aumento di variabilità. In questa indagine dell'errore di misura si stimerà l'entità della componente distorsiva mediante il procedimento di stima di seguito illustrato in modo dettagliato.

Il presente paragrafo è strutturato in due parti: nella prima è descritta la formalizzazione del modello matematico cui si è fatto riferimento, che, opportunamente adattato, costituisce una ragionevole approssimazione della realtà e che, sotto opportune condizioni, consente di misurare le componenti dell'errore di misura; nella seconda parte è descritto lo stimatore adottato per la stima della distorsione di risposta e sono, altresì, illustrate le sue proprietà rispetto al modello e al disegno di campionamento.

Nel prosieguo della trattazione si supporrà che le condizioni per la misurazione nelle due occasioni (indagine primaria e re-intervista) siano identiche o le più simili possibili e che la rilevazione censuaria non sia affetta da errori di copertura.

Per poter valutare l'errore di misura occorre introdurre dei modelli probabilistici, caratterizzati da ipotesi più o meno stringenti sul meccanismo che genera il valore risultante al termine delle operazioni di misurazione. Il modello adottato nella presente indagine - che evidenzia l'impatto dell'errore di misura sulle stime di interesse e ne quantifica l'effetto - fu introdotto da Hansen, Hurwitz e Bershada (1961) e applicato da diversi statistici (Cochran, 1977; Fellegi, 1963, 1964 e 1974) ed Istituti di Statistica internazionali, quali il U.S. Bureau of the Census.

Si denoti con μ_k ($k = 1, \dots, N$) il valore vero della variabile y sull'unità k ($k \in U$); tale valore è una caratteristica propria dell'unità ed è quindi indipendente dalle condizioni che influenzano la risposta individuale. Si supponga, inoltre, che si voglia stimare il totale di popolazione di questi valori *veri*, espresso da:

$$\mu = \sum_{k=1}^N \mu_k \quad (4)$$

La variabile y può essere rilevata su ogni unità k della popolazione infinite volte, ognuna indipendente dall'altra, e, in ciascuna occasione di rilevazione, può assumere un valore diverso dagli altri e dal valore vero; pertanto, i valori di y assunti su ogni unità rilevata al Censimento o all'indagine di misura sono due sue possibili realizzazioni.

Sia y_{tk} il valore osservato sulla k -esima unità nella t -esima ($t = 1, 2, \dots$) misurazione, la differenza $y_{tk} - \mu_k$ è chiamato *errore di misura individuale* per l'unità k .

Indicando rispettivamente con C e I i valori di t per il Censimento e per l'indagine di misura, il modello m che formalizza il processo di misurazione e che descrive la distribuzione congiunta delle $N + n$ misurazioni $\{y_{ck} : k \in U\}$ e $\{y_{ik} : k \in s\}$, in cui U e s ($s \subseteq U$) sono dati, sotto l'ipotesi che le y_{ck} siano incorrelate con le y_{ik} , può essere specificato nel modo di seguito illustrato:



$$E_m(y_{Ck}|U, s) = m_k \quad k \in U \quad (5)$$

$$E_m(y_{Ik}|U, s) = E_m(y_{Ck}|U, s) = m_k \quad k \in s \subseteq U \quad (6)$$

$$E_m[(y_{Ck} - m_k)^2|U, s] = \sigma_k^2 \quad k \in U \quad (7)$$

$$E_m[(y_{Ik} - m_k)^2|U, s] = E_m[(y_{Ck} - m_k)^2|U, s] = \sigma_k^2 \quad k \in s \subseteq U \quad (8)$$

$$E_m[(y_{Ck} - m_k)(y_{Cl} - m_l)|U, s] = \sigma_{kl} \quad k, l \in U \quad (9)$$

$$E_m[(y_{Ik} - m_k)(y_{Il} - m_l)|U, s] = \sigma_{kl} \quad k, l \in s \subseteq U \quad (10)$$

$$E_m[(y_{Ck} - m_k)(y_{Il} - m_l)|U, s] = 0 \quad k \in U \text{ e } l \in s \subseteq U \quad (11)$$

in cui con $E_m(\cdot | U, s)$ si è indicato il valore atteso rispetto al modello m condizionato ad U ed s ($s \subseteq U$).

Dalle (5) e (6) si evince che il valore atteso individuale $E_m(y_{tk}|U, s)$ (in cui, nel caso specifico, $t = C, I$) non coincide con il valore vero μ_k ; la differenza tra i due valori è detta *distorsione di risposta individuale* dell'unità k . Si denoti, ora, con:

$$\hat{y}_t = \sum_{k=1}^n w_k y_{tk} \quad (12)$$

uno stimatore del totale dato dalla (4); ipotizzando che le misurazioni della variabile y seguano il modello m su specificato, l'errore $\hat{y}_t - \mu$ è una variabile casuale la cui distribuzione di probabilità è determinata congiuntamente dal disegno di campionamento e dal modello.

Una misura dell'accuratezza dello stimatore \hat{y}_t è fornita dal valore atteso - rispetto al disegno ed al modello - del quadrato di questo errore, ossia in formula

$$MSE_{dm}(\hat{y}_t) = E_{dm}[(\hat{y}_t - \mu)^2] \quad (13)$$

in cui con il pedice si è indicato il valore atteso rispetto al disegno di campionamento. Si vedano ora gli effetti¹⁵ degli errori di misura sull'accuratezza di \hat{y}_t , dividendo l' $MSE_{dm}(\hat{y}_t)$ nelle sue diverse componenti.

La (13) può essere scritta come somma della varianza totale e del quadrato della distorsione:

$$MSE_{dm}(\hat{y}_t) = V_{dm}(\hat{y}_t) + [B_{dm}(\hat{y}_t)]^2 \quad (14)$$

dove la varianza totale è data da:

$$V_{dm}(\hat{y}_t) = E_{dm}\{[\hat{y}_t - E_{dm}(\hat{y}_t)]^2\} \quad (15)$$

15 L'analisi è incentrata sulla distorsione indotta dall'errore di misura, in quanto essa è l'obiettivo della presente indagine.

mentre la distorsione è:

$$B_{dm}(\hat{y}_t) = E_{dm}(\hat{y}_t) - \mu \quad (16)$$

in cui:

$$\begin{aligned} E_{dm}(\hat{y}_t) &= E_d[E_m(\hat{y}_t|U, s)] = E_d\left[E_m\left(\sum_{k=1}^n w_k y_{tk} \middle| U, s\right)\right] = E_d\left[\sum_{k=1}^n w_k E_m(y_{tk}|U, s)\right] = \\ &= E_d\left(\sum_{k=1}^n w_k m_k\right) = E_d(\hat{M}) = \sum_{k=1}^N m_k = M \end{aligned} \quad (17)$$

e, quindi,

$$B_{dm}(\hat{y}_t) = E_{dm}(\hat{y}_t) - \mu = M - \mu \quad (18)$$

Si indichi, ora, con y_c il totale di y calcolato al termine delle operazioni di misurazione censuaria:

$$y_c = \sum_{k=1}^N y_{ck} \quad (19)$$

in base a quanto ora illustrato, e considerando il fatto che il valore atteso totale coincide in questo caso con il valore atteso rispetto al modello, segue che la distorsione di y_c è data da:

$$\begin{aligned} B_{dm}(y_c) &= B_m(y_c) = E_m(y_c) - \mu = E_m(y_c|U, s) - \mu = E_m\left(\sum_{k=1}^N y_{ck} \middle| U, s\right) - \mu = \\ &= \sum_{k=1}^N E_m(y_{ck}|U, s) - \mu = \sum_{k=1}^N m_k - \mu = M - \mu = B \end{aligned} \quad (20)$$

Un metodo per stimare la distorsione è la re-intervista delle unità rilevate durante l'indagine censuaria, ossia la replicazione della rilevazione sulle unità della popolazione o su una parte di esse. Una condizione fondamentale per stimare la distorsione è che la seconda indagine (indagine di controllo) sia effettuata con un processo di misurazione più preciso dell'indagine originaria, in quanto lo scopo è quello di individuare il valore *vero*. A tal fine, si è replicata l'indagine con riconciliazione delle risposte: in tal modo l'intervistatore, avendo a disposizione le risposte originarie, in caso di discordanza tra la risposta fornita al Censimento e quella data durante l'intervista corrente, può appurare con l'aiuto del rispondente quale sia la risposta vera. Sia, pertanto, y_{lk}^* il valore riconciliato per l'unità k che si ipotizza coincida con il valore *vero*, segue che:

$$y_{lk}^* = \mu_k \quad k \in s \quad (21)$$

Indichiamo con \hat{y}_c ed \hat{y}_l^* gli stimatori del totale della variabile y del tipo dato dalla (12), i cui pesi sono determinati così come descritto nel Paragrafo 1.7.1, calcolati sfrut-



tando i valori rispettivamente rilevati al Censimento e riconciliati (così come su descritto) sulle unità rispondenti all'indagine di misura.

Dalla (21) segue che:

$$E_{dm}(\hat{y}_I^*) = E_d(\hat{y}_I^*) = E_d(\hat{\mu}) = \mu \quad k \in s \quad (22)$$

$$V_{dm}(\hat{y}_I^*) = V_d(\hat{y}_I^*) = V_d(\hat{\mu}) \quad k \in s \quad (23)$$

in cui

$$\hat{\mu} = \sum_{k=1}^n w_k \mu_k \quad (24)$$

Ciò premesso, uno stimatore corretto B di μ è fornito dalla seguente espressione:

$$\hat{B} = \hat{y}_C - \hat{y}_I^* \quad (25)$$

Con semplici passaggi, infatti, è possibile dimostrare che:

$$\begin{aligned} E_{dm}(\hat{B}) &= E_{dm}(\hat{y}_C - \hat{y}_I^*) = E_{dm}(\hat{y}_C) - E_{dm}(\hat{y}_I^*) = E_d \left[E_m \left(\sum_{k=1}^n w_k y_{Ck} \middle| U, s \right) \right] - \mu = \\ &= E_d \left[\sum_{k=1}^n w_k E_m(y_{Ck} | U, s) \right] - \mu = E_d \left(\sum_{k=1}^n w_k m_k \right) - \mu = M - \mu = B \end{aligned} \quad (26)$$

È importante sottolineare che lo stimatore espresso dalla (25) è stato preferito allo stimatore:

$$\tilde{B} = y_C - \hat{y}_I^* \quad (27)$$

in quanto \hat{B} sfrutta la correlazione tra le osservazioni nelle due occasioni di indagine, risultando pertanto più efficiente.

In questa indagine, in realtà, si è preferito valutare la distorsione relativa di y_C indotta dagli errori di misura, in formula

$$\begin{aligned} RB_{dm}(y_C) &= RB_m(y_C) = \frac{E_m(y_C) - \mu}{\mu} = \frac{E_m(y_C | U, s) - \mu}{\mu} = \frac{E_m \left(\sum_{k=1}^N y_{Ck} \middle| U, s \right) - \mu}{\mu} = \\ &= \frac{\sum_{k=1}^N E_m(y_{Ck} | U, s) - \mu}{\mu} = \frac{\sum_{k=1}^N m_k - \mu}{\mu} = \frac{M - \mu}{\mu} = RB \end{aligned} \quad (28)$$

che, rispetto ad una misura assoluta, è di più semplice lettura ed interpretazione.

Uno stimatore della (28) è fornito dalla seguente espressione:

$$\hat{RB} = \frac{\hat{y}_C - \hat{y}_I^*}{\hat{y}_I^*} = \frac{\hat{y}_C}{\hat{y}_I^*} - 1 \quad (29)$$

Essendo uno stimatore non lineare delle osservazioni campionarie, al fine di valutarne le proprietà, si ricorre al *metodo di linearizzazione di Woodruff* (Woodruff 1971; Cicchitelli *et al.* 1992) basato sull'espansione in serie di Taylor (Särndal *et al.* 1989; Deville e Särndal 1992). In particolare, si è valutata la distorsione e la varianza dell'approssimazione arrestata ai termini di ordine rispettivamente quadratico e lineare nell'intorno del punto:

$$\{E_{dm}(\hat{y}_C) = M; E_{dm}(\hat{y}_I^*) = \mu\}$$

Tenendo presenti le ipotesi (21), (22) e (23), analogamente a quanto visto per lo stimatore espresso dalla (12), una misura che quantifica l'attendibilità dello stimatore proposto per il parametro RB è dato dal suo errore quadratico medio calcolato sia rispetto al disegno sia al modello, in formula:

$$MSE_{dm}(\hat{RB}) = V_{dm}(\hat{RB}) + [B_{dm}(\hat{RB})]^2 \quad (30)$$

Il primo addendo della (30), esprime la varianza dello stimatore \hat{RB} , si decompone secondo la ben nota relazione:

$$V_{dm}(\hat{RB}) = E_d[V_m(\hat{RB})] + V_d[E_m(\hat{RB})] \quad (31)$$

in cui è possibile dimostrare che:

$$E_d[V_m(\hat{RB})] = \frac{1}{\mu^2} \left[\sum_{k=1}^N \sum_{l=1}^N w_k w_l \pi_{kl} \sigma_{kl} + \sum_{k=1}^N \sum_{l=1}^N \sigma_{kl} - \sum_{k=1}^N w_k \pi_k \sum_{l=1}^N \sigma_{kl} \right] \quad (32)$$

$$V_d[E_m(\hat{RB})] = \frac{1}{\mu^2} V_d(\hat{M}) + \frac{M^2}{\mu^4} V_d(\hat{\mu}) - 2 \frac{M}{\mu^3} C_d(\hat{M}, \hat{\mu}) \quad (33)$$

dove: π_{kl} indica la probabilità congiunta che le unità k ed l siano incluse nel campione s , π_k è la probabilità che l'unità k sia inclusa nel campione s , \hat{M} indica lo stimatore di M coinvolto nella formula (17), C_d è l'operatore di covarianza rispetto al disegno di campionamento.



Il secondo termine della (30) esprime, invece, il quadrato della distorsione di \hat{RB} , data da:

$$B_{dm}(\hat{RB}) = \frac{M}{\mu^3} V_d(\hat{\mu}) - \frac{C_d(\hat{\mu}, \hat{M})}{\mu^2} \quad (34)$$

da cui si evince che lo stimatore (29) è distorto.

1.7.3 L'attendibilità della stima dell'errore di misura

Una valutazione dell'attendibilità delle stime prodotte da un'indagine è fornita dalla stima della varianza campionaria. Nel Paragrafo 1.7.2 è stata riportata l'espressione analitica della varianza totale dell'espressione linearizzata dello stimatore adottato. Poiché la varianza da modello è solitamente di entità trascurabile rispetto alla variabilità indotta dal disegno campionario, si è proceduto alla valutazione solo di quest'ultima componente. Per il calcolo della varianza di campionamento è stato utilizzato *il package R ReGenesis* (Zardetto, 2012), sviluppato in Istat, che consente di ottenere le stime della varianza di campionamento anche per parametri di natura non lineare, come quelli prodotti dall'indagine in esame.

Indicando con $\hat{V}_{dm}(\hat{RB}_d)$ la stima della varianza di \hat{RB} per il generico dominio di stima d ($d = 1, \dots, D$), la stima dell'errore di campionamento assoluto di \hat{RB} si ottiene mediante la seguente espressione:

$$\hat{\sigma}_{dm}(\hat{RB}_d) = \sqrt{\hat{V}_{dm}(\hat{RB}_d)} \quad (35)$$

A partire dall'errore assoluto è possibile costruire un *intervallo di confidenza* per ciascuna stima di interesse. Tale intervallo contiene il valore *vero* del parametro oggetto di stima con un livello di fiducia pari a P ed è ottenuto come:

$$\left\{ \hat{RB}_d - k_p \hat{\sigma}_{dm}(\hat{RB}_d) \leq RB_d \leq \hat{RB}_d + k_p \hat{\sigma}_{dm}(\hat{RB}_d) \right\} \quad (36)$$

Nella (36) il valore di k_p dipende dal valore fissato per il livello di fiducia P , ad esempio, per $P = 0.95$ si ha $k_{0.95} = 1.96$.

Data la natura del parametro di interesse, definito come differenza relativa, gli errori di campionamento delle stime dell'errore di misura sono presentati in termini di intervalli di confidenza ed è stato scelto come livello di fiducia $P = 95\%$. Tali intervalli, per le principali stime dell'indagine, sono riportati nei prospetti dell'Appendice 1.B.

1.7.4 La varianza di risposta

L'errore di risposta è l'errore non campionario dovuto alla differenza tra la risposta data e il valore vero.

La varianza di risposta totale costituisce la parte di varianza di stima dovuta agli errori di rilevazione di una indagine statistica. Le componenti si identificano nelle varianze dei rispondenti, dei rilevatori e degli eventuali supervisori. Mentre la varianza dei rispondenti è di tipo incorrelato, quelle dei rilevatori e dei supervisori sono di tipo correlato. La varianza di risposta rappresenta la parte più cospicua della varianza extra-campionaria.

La varianza del rispondente misura la variabilità indotta sulle stime dall'errato operato dei rispondenti che può essere identificato in errori di memoria, approssimazioni e confusione accidentale nel riportare i dati. Per misurare la varianza del rispondente è necessario disporre di almeno due misure dello stesso valore (indagine primaria e re-intervista).

La varianza del rilevatore misura la variabilità indotta sulle stime dall'azione dei rilevatori nella fase di raccolta dei dati; di conseguenza cresce sia a causa della non adeguata qualità del lavoro degli intervistatori, sia a causa del carico medio di lavoro.

Sia $y_{i_{CENS}}$ il generico valore osservato al 6° Censimento generale dell'agricoltura (indagine primaria) e si indichi con \bar{y}_{CENS} la media degli $y_{i_{CENS}}$; sia $y_{i_{CATI}}$ il generico valore osservato all'indagine di misura svolta con tecnica CATI (re-intervista) e si indichi con \bar{y}_{CATI} la media degli $y_{i_{CATI}}$, dove $i=1...n$.

Si consideri, quindi, una misura relativa percentuale della varianza di risposta totale (VRTR) la seguente espressione:

$$VRTR = \left\{ \frac{\left[\frac{1}{2} (\bar{y}_{CENS} - \bar{y}_{CATI})^2 \right]^{\frac{1}{2}}}{\bar{y}_{CATI}} \right\} \cdot 100$$

La varianza elementare, o semplice, di risposta è la varianza di una stima affetta da errori (variabili) di risposta tra loro indipendenti. Dato un insieme di unità statistiche, la varianza elementare di risposta è il valore medio degli scarti quadratici dell'errore di risposta atteso per le singole unità.

Si consideri, quindi, una misura relativa percentuale della varianza di risposta semplice (VRSR) la seguente espressione:

$$VRSR = \left\{ \frac{\left[\frac{1}{2n^2} \sum_i (y_{i_{CENS}} - y_{i_{CATI}})^2 \right]^{\frac{1}{2}}}{\bar{y}_{CATI}} \right\} \cdot 100$$



1.8 I risultati

I risultati dell'indagine si riferiscono alla stima della distorsione del dato censuario (Paragrafo 1.8.1) per il livello nazionale, per le 5 ripartizioni geografiche e le 21 regioni e province autonome. È utile sottolineare che per la prima volta l'Istat fornisce dati sull'errore di risposta del Censimento dell'agricoltura con tale dettaglio territoriale poiché, i risultati dell'analoga indagine eseguita in occasione del 5° Censimento generale dell'agricoltura del 2000 si riferivano alle sole ripartizioni geografiche e all'Italia.

Al fine di garantire una corretta lettura dei risultati, si precisa che il segno negativo indica una sottostima del Censimento, ossia un errore per difetto del dato censuario. Per esempio, se una variabile in una data area geografica presenta un valore percentuale di distorsione pari a -1, significa che il valore vero è pari al dato ufficiale risultato al Censimento più l'1 per cento del dato ufficiale stesso. Viceversa, un segno positivo della distorsione indica una sovrastima del Censimento. A testimonianza della variabilità delle stime delle distorsioni, nell'Appendice 1.B sono riportati gli intervalli di confidenza.

Nel Paragrafo 1.8.2 sono presentati i risultati della varianza di risposta semplice e totale calcolati considerando i valori 'grezzi' (dati pre-check) delle risposte fornite al Censimento e all'indagine di misura. La stima di questi parametri è fondamentale per comprendere l'origine dell'errore statistico e, quindi, aumentare la qualità di future rilevazioni, agendo sugli aspetti anche organizzativi delle fasi dell'indagine (la tecnica di rilevazione, il questionario, la formazione dei rilevatori, eccetera).

1.8.1 La distorsione di risposta

Le distorsioni, espresse in percentuale, delle variabili seminativi, coltivazioni legnose agrarie, prati permanenti e pascoli utilizzati, SAU (Superficie Agricola Utilizzata), ST (Superficie Totale), superficie totale irrigata, per regione, ripartizione geografica e Italia sono tutte molto contenute e denotano un elevato livello di precisione del Censimento.

Tavola 1.1 - Distorsioni percentuali della SAU (e sue componenti), ST e Superficie irrigata

REGIONI	Seminativi	Coltivazioni legnose agrarie	Prati permanenti e pascoli utilizzati	SAU	ST	Superficie irrigata
Piemonte	-0,2	-2,5	0,1	1,1	0,8	0,8
Valle d'Aosta/Vallée d'Aoste	-0,8	-0,7	-0,6	-0,6	1,5	-4,4
Liguria	2,3	-0,8	-3,8	-1,2	-4,2	-7,9
Lombardia	0,2	5,4	-3,1	-5,6	-10,7	-1,4
Bolzano/Bozen	4,0	-0,4	-1,7	-1,5	-2,5	-6,2
Trento	-4,8	2,1	-3,6	-2,6	-1,5	-2,0
Veneto	-0,2	-1,7	-0,2	-0,3	-2,7	-1,0
Friuli-Venezia Giulia	-0,3	1,8	7,0	1,0	-1,8	-4,8
Emilia-Romagna	1,1	-0,9	-8,1	0,6	-0,8	-2,6
Toscana	2,0	0,4	-5,4	0,8	-3,5	-1,0
Umbria	1,7	-2,7	-6,9	0,5	-0,8	-4,8
Marche	-3,6	1,5	-3,4	-3,0	-3,8	-1,6
Lazio	-0,9	-4,4	-4,4	-2,6	-7,1	-8,1

continua alla pagina successiva

Tavola 1.1 - Distorsioni percentuali della SAU (e sue componenti), ST e Superficie irrigata

REGIONI	Seminativi	Coltivazioni legnose agrarie	Prati permanenti e pascoli utilizzati	SAU	ST	Superficie irrigata
Abruzzo	-3,2	1,3	-1,3	-1,0	-0,4	-5,1
Molise	1,1	1,6	-2,4	3,3	1,6	-5,7
Campania	0,8	-1,5	-6,1	-1,2	-1,2	-6,7
Puglia	-2,4	-0,3	3,7	-1,4	-2,0	-1,0
Basilicata	0,0	-2,3	-3,3	-0,2	3,6	0,1
Calabria	-4,6	-6,1	-6,2	-5,6	-9,2	-2,6
Sicilia	1,5	-0,9	-8,5	-2,7	-2,9	-13,6
Sardegna	-1,0	-1,0	-1,7	-0,6	-2,2	-7,8
Nord-ovest	0,1	-0,5	-0,9	-2,0	-4,5	-0,7
Nord-est	0,4	-0,8	-2,0	-0,1	-1,9	-2,4
Centro	-0,2	-1,4	-5,0	-1,1	-4,1	-5,2
Sud	-1,5	-1,9	-3,3	-1,5	-1,8	-2,6
Isole	0,6	-0,9	-4,0	-1,8	-2,6	-11,9
Italia	-0,2	-1,4	-2,9	-1,3	-2,9	-2,8

(a) Il segno negativo indica una sottostima del Censimento

A livello Italia, la SAU presenta una distorsione pari a -1,3 per cento, ossia una sottostima del Censimento pari a tale entità. Tra le ripartizioni geografiche, le distorsioni variano dal -2,0 per cento del Nord-ovest al -0,1 per cento del Nord-est. A livello regionale, i risultati presentano risultati più diversificati, ma complessivamente concentrati in un intorno ristretto del valore nullo: si passa dal circa -5,5 per cento di Lombardia e Calabria al +3,3 per cento del Molise.

I risultati delle componenti della SAU rispecchiano naturalmente gli andamenti appena descritti. La superficie totale dei seminativi presenta a livello Italia il risultato qualitativamente più vicino al Censimento, con una lieve distorsione pari al -0,2 per cento; ossia, si può affermare che il valore vero è molto prossimo al dato ufficiale pubblicato dal Censimento. All'interno delle ripartizioni geografiche le distorsioni variano dal -1,5 per cento del Sud al +0,6 per cento delle Isole. A livello regionale si notano dei valori prossimi allo zero per Piemonte, Valle d'Aosta, Lombardia, Veneto, Friuli-Venezia Giulia, Lazio, Campania, Basilicata e Sardegna.

I dati sulla qualità delle coltivazioni legnose agrarie presentano dei risultati decisamente lusinghieri per il Censimento. A livello Italia è stata misurata una distorsione pari a -1,4 per cento e, a livello di ripartizione geografica, una complessiva sottostima del dato censuario che varia dal -1,9 per cento nel Sud al -0,5 per cento nel Nord-ovest. A livello regionale, numerosi sono i valori compresi tra il -1 per cento e +1 per cento (Valle d'Aosta, Liguria, Emilia-Romagna, Toscana, Puglia, Sicilia, Provincia autonoma di Bolzano). Si devono però notare sottostime dei valori censuari meno contenute nel Lazio (-4,4 per cento) e in Calabria (-6,1 per cento) e la sovrastima di circa il 5 per cento in Lombardia.

La distorsione della variabile riferita a prati permanenti e pascoli utilizzati è pari a circa -3 per cento a livello nazionale; nelle ripartizioni geografiche, le distorsioni, tutte indicanti una sottostima del Censimento, sono comprese tra -5 per cento del Centro e -0,9 per cento del Nord-ovest. Nei dati regionali sono presenti principalmente segni negativi; fanno eccezione il Friuli-Venezia Giulia con circa +7 per cento e la Puglia con +3,7 per cento. Si segnalano il Piemonte, la Valle d'Aosta e il Veneto che presentano valori prossimi allo zero.



La ST del Censimento è stata sottostimata del 2,9 per cento a livello nazionale; di segno negativo risultano anche i valori delle ripartizioni geografiche che variano dal -4,5 per cento del Nord-ovest al -1,8 per cento del Sud. A livello regionale i dati risultano molto variabili poiché si passa dal -10,7 per cento della Lombardia al +3,6 per cento della Basilicata. Il Censimento presenta dei risultati di ST molto precisi in Piemonte, Emilia-Romagna, Umbria e Abruzzo con valori rispettivamente pari a +0,8 per cento, -0,9 per cento, -0,8 per cento e -0,4 per cento.

La superficie totale irrigata presenta una distorsione a livello nazionale pari a -2,8 per cento. Risultano di segno negativo anche i risultati delle ripartizioni geografiche: il Nord-ovest si caratterizza per un valore particolarmente basso di distorsione (-0,7 per cento) mentre le Isole presentano un valore medio-alto (-11,9 per cento).

L'indagine ha previsto la re-intervista anche per alcuni utilizzi del terreno inclusi tra i seminativi: frumento, mais, terreni a riposo soggetti a regime di aiuto. Per una corretta lettura dei risultati, si precisa che, ove siano assenti i valori (e quindi siano presenti due trattini, '--'), la stima della distorsione non è stata calcolata perché ritenuta non attendibile. La motivazione di tale esito è dovuta sia all'esigua numerosità campionaria della cella sia alla contenuta incidenza della variabile nel panorama agricolo italiano. Anche in questa circostanza, le distorsioni molto contenute denotano un'elevata qualità dei risultati censuari e confermano la bontà del processo di produzione del dato.

Tavola 1.2 - Distorsioni percentuali della superficie di alcune componenti dei seminativi per regione (a, b)

REGIONI	Frumento	Mais	Terreni a riposo soggetti a regime di aiuto
Piemonte	0,1	--	-9,8
Valle d'Aosta/Vallée d'Aoste	--	-3,0	--
Liguria	--	-7,4	--
Lombardia	-2,4	-8,7	5,6
Bolzano/Bozen	--	--	--
Trento	--	--	--
Veneto	-2,0	2,5	-4,2
Friuli-Venezia Giulia	-1,2	-2,5	4,6
Emilia-Romagna	0,1	-1,8	-5,4
Toscana	7,8	-7,7	2,1
Umbria	-0,5	4,4	-1,8
Marche	2,2	-5,8	-5,4
Lazio	-4,5	-7,5	-7,4
Abruzzo	-3,5	-8,9	-5,6
Molise	2,2	4,1	5,1
Campania	-3,0	-9,8	-6,3
Puglia	-0,5	--	-5,8
Basilicata	0,6	--	0,0
Calabria	0,5	--	-8,1
Sicilia	5,2	--	-7,8
Sardegna	-5,3	--	-0,6
Nord-ovest	0,1	-5,9	-5,7
Nord-est	0,4	0,7	-3,0
Centro	-0,2	-4,2	-0,7

continua alla pagina successiva

Tavola 1.2 - Distorsioni percentuali della superficie di alcune componenti dei seminativi per regione (a, b)

REGIONI	Frumento	Mais	Terreni a riposo soggetti a regime di aiuto
Sud	-1,5	-8,6	-2,3
Isole	0,6	--	-6,4
Italia	0,6	-2,9	-3,0

(a) Il simbolo (--) indica che la numerosità campionaria e la contenuta incidenza della variabile non consentono di fornire stime attendibili

(b) Il segno negativo indica una sottostima del Censimento

La variabile frumento è presentata come aggregato delle variabili frumento duro e frumento tenero e spelta. La distorsione, a livello Italia, è pari a +0,6 per cento; le ripartizioni geografiche presentano risultati molto prossimi allo zero, tra i quali spicca il +0,1 per cento del Nord-ovest. A livello regionale, Piemonte, Emilia-Romagna, Umbria, Puglia, Basilicata e Calabria hanno distorsioni nell'intorno del valore nullo, indicando, quindi, un'elevata qualità del dato censuario.

La variabile mais per la produzione di granella, calcolata principalmente nelle regioni del Nord e del Centro, presenta una distorsione a livello Italia pari a -2,9 per cento. Particolarmente contenuta è la distorsione per la ripartizione Nord-est (+0,7 per cento) mentre al Nord-ovest e al Centro si registrano rispettivamente il -5,9 per cento e il -4,2 per cento. A livello regionale, le distorsioni passano dal -9,9 per cento della Campania al +4,4 per cento dell'Umbria.

La variabile terreni a riposo soggetti a regime di aiuto è caratterizzata da distorsioni di segno negativo. L'Italia presenta un valore pari a -3,0 per cento. Le distorsioni delle ripartizioni geografiche variano da -6,4 per cento delle Isole a -0,7 per cento del Centro. A livello regionale, si segnalano i risultati, assolutamente in linea con il Censimento, di Basilicata e Sardegna; altre regioni presentano sottostime superiori al 4 per cento.

La rilevazione post censuaria è stata progettata ed eseguita per misurare la qualità anche di alcune componenti delle coltivazioni legnose agrarie. Come è noto, in tale categoria, rientrano le variabili vite e olivo che destano molta attenzione poiché caratterizzano l'agricoltura del nostro paese.

Tavola 1.3 - Distorsioni percentuali della superficie di alcune componenti delle coltivazioni legnose agrarie (a,b)

REGIONI	Vite	Olivo	Agrumi	Melo
Piemonte	-4,6	--	--	0,5
Valle d'Aosta/Vallée d'Aoste	2,6	--	--	-8,1
Liguria	0,3	0,3	--	--
Lombardia	3,6	-2,9	--	1,0
Bolzano/Bozen	1,0	--	--	-1,2
Trento	5,9	--	--	-0,3
Veneto	4,1	-9,7	--	-1,4
Friuli-Venezia Giulia	1,9	-6,2	--	-2,4
Emilia-Romagna	0,7	-5,2	--	4,1
Toscana	4,1	-0,1	--	--
Umbria	3,1	-1,4	--	--
Marche	8,9	-2,8	--	--

continua alla pagina successiva



Tavola 1.3 - Distorsioni percentuali della superficie di alcune componenti delle coltivazioni legnose agrarie (a,b)

REGIONI	Vite	Olivo	Agrumi	Melo
Lazio	0,9	-1,4	--	--
Abruzzo	10,9	-1,1	--	--
Molise	7,6	1,9	--	--
Campania	3,7	-4,4	-3,7	0,0
Puglia	10,6	-1,8	2,0	--
Basilicata	-7,0	1,3	-0,2	--
Calabria	-6,5	-5,3	-6,6	--
Sicilia	5,2	-3,0	-1,3	--
Sardegna	10,0	-4,9	-5,5	--
Nord-ovest	-2,0	-0,6	-7,6	0,2
Nord-est	2,6	-7,3	--	-0,5
Centro	4,3	-0,9	-4,1	0,5
Sud	7,0	-2,8	-4,0	0,5
Isole	5,9	-3,4	-1,5	--
Italia	4,0	-2,6	-2,6	-0,4

(a) Il simbolo (--) indica che la numerosità campionaria e la contenuta incidenza della variabile non consentono di fornire stime attendibili

(b) Il segno negativo indica una sottostima del Censimento

La variabile di superficie investita a vite, contrariamente a gran parte delle altre variabili sottoposte a re-intervista, è caratterizzata da distorsioni di segno positivo che denotano una sovrastima del dato censuario. La distorsione per i dati relativi all'Italia è pari a circa +4 per cento; tra le ripartizioni, solo il Nord-ovest presenta un valore negativo (circa il -2 per cento), mentre tutte le altre aree geografiche registrano distorsioni oscillanti tra il +2,6 per cento del Nord-est e il +7 per cento del Sud. A livello regionale, la Liguria, l'Emilia-Romagna, il Lazio e la Provincia autonoma di Bolzano presentano distorsioni molto contenute e prossime allo zero. Le Marche, l'Abruzzo, il Molise, la Puglia e la Sardegna registrano distorsioni positive superiori al 7 per cento, mentre la Basilicata e la Calabria hanno distorsioni negative inferiori al 6 per cento.

La variabile olivo non è stata calcolata per il Piemonte, la Valle d'Aosta e le Province autonome di Trento e Bolzano. A livello nazionale e di ripartizione, le distorsioni per questa variabile sono tutte negative: l'Italia presenta un valore pari a -2,6 per cento; i risultati del Nord-ovest e del Centro denotano una elevata qualità del dato censuario con valori prossimi allo zero (-0,6 per cento e -0,9 per cento). Il Sud registra -2,8 per cento e le Isole -3,4 per cento. A livello regionale prevalgono i segni negativi, tra i quali si segnalano il -0,1 per cento della Toscana (come valore più vicino allo zero) e il -9,7 per cento del Veneto (come valore più distante dallo zero). I segni positivi sono tutti molto contenuti e sono presenti in Liguria (0,3 per cento), Molise (1,9 per cento) e Basilicata (1,3 per cento).

La variabile agrumi è stata calcolata solo in alcune regioni del Sud e delle Isole; tuttavia è stato possibile fornire una stima affidabile a livello di ripartizione per il Nord-ovest e il Centro. La distorsione a livello nazionale è pari a -2,6 per cento. Le Isole presentano una distorsione pari a -1,5 per cento, mentre il Sud e il Centro registrano un valore prossimo a -4 per cento. A livello regionale, si segnalano i risultati della Calabria (-0,2 per cento) e della Sardegna (-1,3 per cento); l'unica distorsione con segno positivo si registra in Puglia (+2 per cento).

La variabile melo presenta risultati esclusivamente per alcune regioni settentrionali e

la Campania. Le distorsioni, tra le più basse dell'indagine, sono, a livello nazionale e di ripartizione, prossime allo zero, dimostrando una elevata qualità del dato prodotto dal Censimento.

Nella tavola 1.4 sono riportati i risultati inerenti le principali specie di bestiame. Solo per i bovini è stato possibile presentare i dati per tutte le aree geografiche considerate; le variabili bufalini, ovini, caprini e suini, essendo particolarmente caratterizzate dalla localizzazione geografica, sono presenti in quantità significative solo in alcune regioni.

A livello nazionale le distorsioni sono negative per tutte le specie di bestiame, e i dati censuari sono sottostimati, in media, di circa il 5 per cento.

Tavola 1.4 - Distorsioni percentuali del numero di capi delle principali specie di bestiame (a, b)

REGIONI	Bovini	Bufalini	Ovini	Caprini	Suini
Piemonte	0,3	--	--	-8,8	-3,0
Valle d'Aosta/Vallée d'Aoste	-1,1	--	--	--	--
Liguria	-1,5	--	--	--	--
Lombardia	-6,6	--	--	-14,1	-9,9
Bolzano/Bozen	-1,8	--	--	--	--
Trento	-0,9	--	--	--	--
Veneto	-8,4	--	--	--	0,0
Friuli-Venezia Giulia	-0,3	--	--	--	--
Emilia-Romagna	-0,2	--	--	--	-1,0
Toscana	-11,3	--	-4,0	--	--
Umbria	-4,1	--	--	--	--
Marche	-9,2	--	--	--	--
Lazio	-3,9	--	-4,6	--	--
Abruzzo	-8,0	--	4,4	--	--
Molise	-4,0	--	--	--	--
Campania	-10,5	-5,8	--	--	--
Puglia	-5,9	--	--	--	--
Basilicata	-8,3	--	-2,7	-4,8	--
Calabria	-19,3	--	-14,5	-12,4	--
Sicilia	-6,9	--	-9,2	-7,5	--
Sardegna	-6,6	--	-1,5	-3,5	--
Nord-ovest	-4,4	--	--	-11,3	-8,6
Nord-est	-5,1	--	--	-4,8	0,1
Centro	-5,9	--	-3,0	-1,6	--
Sud	-9,5	-6,5	-2,6	-9,3	--
Isole	-6,8	--	-3,2	-4,8	--
Italia	-5,5	-5,0	-3,3	-6,6	-5,6

(a) Il simbolo (--) indica che la numerosità campionaria e la contenuta incidenza della variabile non consentono di fornire stime attendibili

(b) Il segno negativo indica una sottostima del Censimento

La distorsione della variabile bovini, a livello nazionale, è pari a -5,5 per cento; anche le ripartizioni geografiche sono caratterizzate da distorsioni negative che variano dal -9,6 per cento del Sud al -4,4 per cento del Nord-ovest. A livello regionale, si registrano risultati decisamente soddisfacenti per il Piemonte (+0,3 per cento), la Valle d'Aosta (-1,1 per cento), il Friuli-Venezia Giulia (-0,3 per cento), la Liguria (-1,5 per cento), l'Emilia-Romagna (-0,2 per cento), la Provincia autonoma di Bolzano



(-1,8 per cento) e la Provincia autonoma di Trento (-0,9 per cento). In Toscana, Campania e Calabria le distorsioni sono negative e inferiori al 10 per cento.

Vista la concentrazione geografica degli allevamenti bufalini la stima della distorsione della variabile è risultata significativa esclusivamente in Campania dove si registra un dato censuario sottostimato del 5,8 per cento. A livello nazionale, la distorsione è pari a circa -5 per cento.

La variabile ovini presenta una distorsione, a livello nazionale, pari a -3,3 per cento; le ripartizioni geografiche in cui la variabile è risultata significativa presentano un dato prossimo a circa -3 per cento. I risultati regionali sono particolarmente oscillanti e passano da -14,5 per cento della Calabria a +4,4 per cento del Lazio.

La variabile caprini presenta distorsioni negative per l'Italia (-6,7 per cento) e tutte le ripartizioni geografiche: -11,3 per cento il Nord-ovest, -4,8 per cento il Nord-est, -1,6 per cento il Centro, -9,3 per cento il Sud e -4,9 per cento le Isole. Le distorsioni calcolate per le regioni sono tutte negative e oscillano tra -14,1 per cento della Lombardia e -3,5 per cento della Sardegna.

La distorsione, a livello nazionale, della variabile suini risulta pari a -5,6 per cento; il Nord-est e, in particolare, il Veneto presentano delle distorsioni pressoché nulle. Si evidenziano, inoltre, i risultati del Piemonte e della Liguria che registrano distorsioni pari, rispettivamente, a -3,0 per cento e -1,1 per cento.

1.8.2 Le varianze di risposta

Come già accennato nel Paragrafo 1.8, la varianza di risposta totale e la varianza di risposta semplice sono calcolate sui dati grezzi del Censimento e dell'indagine di misura, ossia non è considerato il dato modificato dalle procedure di controllo e correzione.

Nella formulazione adottata si è preferito calcolare le varianze di risposta attraverso una misura relativa che sia identificabile come un coefficiente di variazione percentuale; questa soluzione è stata adottata al fine di consentire una migliore lettura dei risultati. Inoltre, i coefficienti di variazione percentuali delle varianze di risposta sono stati raccolti in classi: A=[0-10); B=[10-20); C=[20-30); D=[30-40); E=[40-50); F>50. In questo paragrafo sono presentati i risultati della varianza di risposta semplice e della varianza di risposta totale per le principali variabili rilevate all'indagine di misura: seminativi, coltivazioni legnose agrarie, prati permanenti e pascoli. I risultati di tutte le altre variabili sono presentati nell'Appendice 1.A.

La tavola 1.5 mostra i risultati della varianza di risposta semplice a livello nazionale, di ripartizione e regionale. Al fine di agevolare la lettura dei risultati, si ricorda che una classe bassa (dalla A alla C) deve essere interpretata come una lieve differenza tra il dato grezzo rilevato al Censimento e il dato rilevato all'indagine di misura (non riconciliato). Ovviamente, una classe alta (dalla D alla F) indica una forte discrepanza tra i due valori rilevati sulla stessa azienda in momenti differenti. Come si evince dalla tavola 1.5 la varianza semplice di risposta presenta molti valori nelle classi alte di coefficiente di variazione. Ciò indica che le differenze tra dato censuario e dato dichiarato in prima battuta al rilevatore dell'indagine di controllo sono numerose ed elevate e che solo con la domanda di riconciliazione effettuata dal rilevatore della seconda indagine il dichiarante ha confermato il dato del censimento. Infatti, come si è detto in precedenza, la distorsione presenta valori contenuti. Questo fenomeno si spiega con la diversità di tecniche di rilevazione adottate nelle due occasioni di rilevazione: con intervista diretta il censimento e con intervista telefonica l'indagine di misura.

Tavola 1.5 - Varianza di risposta semplice (a) della superficie di alcune componenti della SAU e di ST

REGIONI	Seminativi	Coltivazioni legnose agrarie	Prati permanenti e pascoli utilizzati	SAU	ST
Piemonte	F	F	F	F	F
Valle d'Aosta/Vallée d'Aoste	F	F	F	A	F
Liguria	F	F	F	D	A
Lombardia	F	F	F	F	F
Bolzano/Bozen	A	F	F	D	F
Trento	E	D	E	B	F
Veneto	F	F	F	F	F
Friuli-Venezia Giulia	F	D	F	C	A
Emilia-Romagna	F	F	F	F	F
Toscana	F	F	F	A	F
Umbria	F	A	C	E	E
Marche	F	F	F	C	E
Lazio	F	F	F	E	D
Abruzzo	F	D	F	B	B
Molise	F	F	C	F	F
Campania	F	F	F	F	F
Puglia	F	F	F	B	F
Basilicata	F	F	F	F	F
Calabria	F	F	F	F	F
Sicilia	F	F	F	B	A
Sardegna	F	F	F	E	C
Nord-ovest	F	F	F	F	F
Nord-est	F	F	F	F	F
Centro	F	F	F	A	F
Sud	F	F	F	C	C
Isole	F	F	F	B	B
Italia	F	F	F	F	F

(a) Classi di ampiezza: A=0%-10%; B=10%-20%; C=20%-30%; D=30%-40%; E=40%-50%; F=>50%

Nella tavola 1.6 sono presentati i risultati della varianza di risposta totale che, come descritto sopra per la varianza di risposta semplice, è stata calcolata attraverso il coefficiente di variazione percentuale raccolto in classi di ampiezza. I risultati mostrano che, in questa circostanza, le classi maggiormente presenti sono quelle basse (dalla A alla C). La motivazione deve essere anche ricercata nella formula stessa della varianza totale poiché non sono calcolate le differenze sulle singole osservazioni (come per la varianza semplice), bensì sono calcolate le differenze dei valori medi che, per definizione, hanno un effetto calmierante sulle differenze stesse.

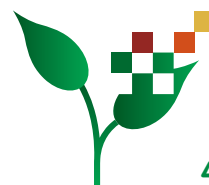


Tavola 1.6 - Varianza di risposta totale (a) della superficie di alcune componenti della SAU e di ST

REGIONI	Seminativi	Coltivazioni legnose agrarie	Prati permanenti e pascoli utilizzati	SAU	ST
Piemonte	A	A	C	A	A
Valle d'Aosta/Vallée d'Aoste	C	B	C	A	C
Liguria	A	E	B	A	A
Lombardia	A	B	A	A	A
Bolzano/Bozen	A	A	C	A	A
Trento	A	A	B	A	B
Veneto	A	A	D	A	A
Friuli-Venezia Giulia	A	A	A	A	B
Emilia-Romagna	B	F	A	F	F
Toscana	B	A	A	A	A
Umbria	B	A	A	A	A
Marche	A	A	A	A	A
Lazio	A	A	A	A	A
Abruzzo	A	A	A	A	A
Molise	B	A	A	A	A
Campania	A	A	B	A	A
Puglia	A	A	B	A	A
Basilicata	B	A	A	A	A
Calabria	A	A	A	A	A
Sicilia	A	A	A	A	A
Sardegna	A	A	A	A	A
Nord-ovest	A	A	B	A	A
Nord-est	A	F	C	C	C
Centro	A	A	A	A	A
Sud	A	A	A	A	A
Isole	A	A	A	A	A
Italia	A	C	A	A	A

(a) Classi di ampiezza: A=0%-10%; B=10%-20%; C=20%-30%; D=30%-40%; E=40%-50%; F=>50%

Appendice 1.A - Risultati delle componenti di varianza di risposta

Tavola 1.A.1 - Varianza di risposta semplice (a) della superficie di alcune componenti dei seminativi

REGIONI	Frumento	Mais	Terreni a riposo soggetti a regime di aiuto
Piemonte	F	--	F
Valle d'Aosta/Vallée d'Aoste	--	F	--
Liguria	--	F	--
Lombardia	D	F	F
Bolzano/Bozen	--	--	--
Trento	--	--	--
Veneto	D	F	F
Friuli-Venezia Giulia	F	D	C
Emilia-Romagna	F	F	B
Toscana	A	F	F
Umbria	F	F	F
Marche	F	F	A
Lazio	F	F	F
Abruzzo	F	F	F
Molise	F	F	F
Campania	F	F	F
Puglia	F	--	F
Basilicata	F	--	A
Calabria	F	--	F
Sicilia	F	--	F
Sardegna	F	--	F
Nord-ovest	F	F	F
Nord-est	F	F	C
Centro	F	F	F
Sud	A	F	F
Isole	F	--	F
Italia	C	F	F

(a) Classi di ampiezza: A=0%-10%; B=10%-20%; C=20%-30%; D=30%-40%; E=40%-50%; F=>50%



Tavola 1.A.2 - Varianza di risposta totale (a) della superficie di alcune componenti dei seminativi

REGIONI	Frumento	Mais	Terreni a riposo soggetti a regime di aiuto
Piemonte	A	--	B
Valle d'Aosta/Vallée d'Aoste	--	C	--
Liguria	--	A	--
Lombardia	A	B	A
Bolzano/Bozen	--	--	--
Trento	--	--	--
Veneto	A	A	A
Friuli-Venezia Giulia	A	A	A
Emilia-Romagna	A	A	A
Toscana	A	A	A
Umbria	A	A	A
Marche	B	B	A
Lazio	A	B	B
Abruzzo	A	B	B
Molise	A	A	A
Campania	A	C	A
Puglia	A	--	A
Basilicata	A	--	A
Calabria	A	--	D
Sicilia	A	--	A
Sardegna	A	--	B
Nord-ovest	A	A	A
Nord-est	A	A	A
Centro	A	B	A
Sud	A	B	A
Isole	A	--	A
Italia	A	A	A

(a) Classi di ampiezza: A=0%-10%; B=10%-20%; C=20%-30%; D=30%-40%; E=40%-50%; F=>50%

Tavola 1.A.3 - Varianza di risposta semplice (a) della superficie di alcune componenti delle coltivazioni legnose agrarie

REGIONI	Vite	Olivo	Agrumi	Melo
Piemonte	F	--	--	F
Valle d'Aosta/Vallée d'Aoste	A	--	--	C
Liguria	F	F	--	--
Lombardia	E	A	--	F
Bolzano/Bozen	F	--	--	A
Trento	F	--	--	A
Veneto	F	F	--	F
Friuli-Venezia Giulia	B	F	--	F
Emilia-Romagna	F	F	--	F
Toscana	F	C	--	--
Umbria	F	A	--	--
Marche	F	F	--	--
Lazio	F	C	--	--
Abruzzo	F	F	--	--
Molise	F	B	--	--
Campania	F	F	F	F
Puglia	F	F	F	--
Basilicata	F	F	E	--
Calabria	F	F	F	--
Sicilia	F	C	F	--
Sardegna	F	F	F	--
Nord-ovest	F	C	F	F
Nord-est	F	F	--	F
Centro	F	F	F	F
Sud	F	F	F	F
Isole	F	E	F	--
Italia	F	F	F	F

(a) Classi di ampiezza: A=0%-10%; B=10%-20%; C=20%-30%; D=30%-40%; E=40%-50%; F=>50%



Tavola 1.A.4 - Varianza di risposta totale (a) della superficie di alcune componenti delle coltivazioni legnose agrarie

REGIONI	Vite	Olivo	Agrumi	Melo
Piemonte	A	--	--	A
Valle d'Aosta/Vallée d'Aoste	A	--	--	A
Liguria	D	A	--	--
Lombardia	A	A	--	A
Bolzano/Bozen	B	--	--	A
Trento	E	--	--	A
Veneto	C	F	--	A
Friuli-Venezia Giulia	A	A	--	B
Emilia-Romagna	F	A	--	A
Toscana	A	A	--	--
Umbria	A	A	--	--
Marche	B	F	--	--
Lazio	A	A	--	--
Abruzzo	C	A	--	--
Molise	B	A	--	--
Campania	A	A	B	B
Puglia	B	A	B	--
Basilicata	B	B	A	--
Calabria	B	A	A	--
Sicilia	A	A	A	--
Sardegna	B	A	A	--
Nord-ovest	A	A	C	A
Nord-est	F	C	--	A
Centro	A	F	A	B
Sud	A	A	A	A
Isole	A	A	A	
Italia	F	B	A	A

(a) Classi di ampiezza: A=0%-10%; B=10%-20%; C=20%-30%; D=30%-40%; E=40%-50%; F=>50%

Tavola 1.A.5 - Varianza di risposta semplice (a) del numero di capi di alcune specie di bestiame

REGIONI	Bovini	Bufalini	Ovini	Caprini	Suini
Piemonte	F	--	--	F	F
Valle d'Aosta/Vallée d'Aoste	A	--	--	--	--
Liguria	A	--	--	--	--
Lombardia	F	--	--	F	F
Bolzano/Bozen	F	--	--	--	--
Trento	A	--	--	--	--
Veneto	F	--	--	--	F
Friuli-Venezia Giulia	F	--	--	--	--
Emilia-Romagna	F	--	--	--	F
Toscana	F	--	A	--	--
Umbria	F	--	--	--	--
Marche	C	--	--	--	--
Lazio	F	--	F	--	--
Abruzzo	F	--	C	--	--
Molise	B	--	--	--	--
Campania	F	F	--	--	--
Puglia	F	--	--	--	--
Basilicata	F	--	F	E	--
Calabria	F	--	F	F	--
Sicilia	F	--	C	F	--
Sardegna	D	--	F	F	--
Nord-ovest	A	--	--	F	B
Nord-est	E	--	--	F	F
Centro	E	--	D	F	--
Sud	F	F	F	F	--
Isole	F	--	F	D	--
Italia	F	F	F	F	F

(a) Classi di ampiezza: A=0%-10%; B=10%-20%; C=20%-30%; D=30%-40%; E=40%-50%; F=>50%



Tavola 1.A.6 - Varianza di risposta totale (a) del numero di capi di alcune specie di bestiame

REGIONI	Bovini	Bufalini	Ovini	Caprini	Suini
Piemonte	A	--	--	B	A
Valle d'Aosta/Vallée d'Aoste	A	--	--	--	--
Liguria	A	--	--	--	--
Lombardia	A	--	--	C	A
Bolzano/Bozen	A	--	--	--	--
Trento	A	--	--	--	--
Veneto	A	--	--	--	B
Friuli-Venezia Giulia	B	--	--	--	--
Emilia-Romagna	A	--	--	--	A
Toscana	A	--	A	--	--
Umbria	A	--	--	--	--
Marche	A	--	--	--	--
Lazio	A	--	A	--	--
Abruzzo	C	--	A	--	--
Molise	A	--	--	--	--
Campania	B	A	--	--	--
Puglia	B	--	--	--	--
Basilicata	B	--	A	A	--
Calabria	B	--	A	A	--
Sicilia	A	--	A	A	--
Sardegna	A	--	A	A	--
Nord-ovest	A	--	--	B	A
Nord-est	A	--	--	B	A
Centro	A	--	A	B	--
Sud	A	B	C	A	--
Isole	A	--	A	A	--
Italia	A	B	A	A	A

(a) Classi di ampiezza: A=0%-10%; B=10%-20%; C=20%-30%; D=30%-40%; E=40%-50%; F=>50%

Appendice I.B - La variabilità delle stime

Tavola 1.B.1 - Intervalli di confidenza delle stime delle distorsioni della superficie delle principali variabili aggregate per regione

REGIONI	Seminativi		Coltivazioni legnose agrarie		Prati permanenti e pascoli utilizzati	
	Inf	Sup	Inf	Sup	Inf	Sup
Piemonte	-0,0104	0,0069	-0,0571	0,0071	-0,0647	0,0671
Valle d'Aosta/ Vallée d'Aoste	-0,1746	0,0807	-0,0419	0,0260	-0,0197	0,0070
Liguria	-0,0038	0,0559	-0,0294	0,0121	-0,1619	0,0865
Lombardia	-0,0225	0,0269	-0,0206	0,1283	-0,0811	0,0182
Bolzano/Bozen	-0,0154	0,0964	-0,0183	0,0101	-0,1006	0,0662
Trento	-0,0856	-0,0030	0,0099	0,0324	-0,0683	-0,0032
Veneto	-0,0114	0,0073	-0,0466	0,0122	-0,0237	0,0202
Friuli-Venezia Giulia	-0,0090	0,0021	-0,0049	0,0399	-0,1359	0,2765
Emilia-Romagna	-0,0118	0,0340	-0,0322	0,0133	-0,2572	0,0957
Toscana	0,0048	0,0344	-0,0234	0,0307	-0,1370	0,0300
Umbria	-0,0007	0,0346	-0,0470	-0,0061	-0,2468	0,1095
Marche	-0,1228	0,0500	-0,0883	0,1180	-0,1580	0,0895
Lazio	-0,0475	0,0289	-0,0893	0,0010	-0,1048	0,0166
Abruzzo	-0,0844	0,0206	-0,0082	0,0336	-0,0725	0,0473
Molise	-0,0038	0,0260	-0,0129	0,0443	-0,0719	0,0249
Campania	-0,0363	0,0516	-0,0477	0,0181	-0,1666	0,0454
Puglia	-0,0570	0,0082	-0,0250	0,0180	-0,0364	0,1229
Basilicata	-0,0332	0,0330	-0,0418	-0,0022	-0,1682	0,0979
Calabria	-0,0958	0,0046	-0,1142	-0,0074	-0,1747	0,0506
Sicilia	-0,0186	0,0484	-0,0351	0,0168	-0,1666	-0,0028
Sardegna	-0,0401	0,0211	-0,0797	0,0591	-0,0678	0,0340
Nord-ovest	-0,0140	0,0152	-0,0331	0,0223	-0,0537	0,0363
Nord-est	-0,0074	0,0156	-0,0221	0,0068	-0,0576	0,0178
Centro	-0,0274	0,0228	-0,0361	0,0076	-0,0965	-0,0021
Sud	-0,0321	0,0028	-0,0364	-0,0012	-0,0765	0,0113
Isole	-0,0181	0,0297	-0,0337	0,0151	-0,0833	0,0028
Italia	-0,0105	0,0062	-0,0240	-0,0036	-0,0497	-0,0086

continua alla pagina successiva



Tavola 1.B.1 - Intervalli di confidenza delle stime delle distorsioni della superficie delle principali variabili aggregate per regione

REGIONI	SAU		ST		Superficie irrigata	
	Inf	Sup	Inf	Sup	Inf	Sup
Piemonte	-0,0094	0,0318	-0,0097	0,0268	-0,0123	0,0278
Valle d'Aosta/ Vallée d'Aoste	-0,0197	0,0066	-0,0259	0,0568	-0,1310	0,0430
Liguria	-0,0745	0,0528	-0,0795	-0,0036	-0,1289	-0,0300
Lombardia	-0,2084	0,0968	-0,3320	0,1175	-0,0473	0,0189
Bolzano/Bozen	-0,0869	0,0564	-0,1050	0,0542	-0,1317	0,0078
Trento	-0,0518	0,0004	-0,0460	0,0151	-0,0376	-0,0030
Veneto	-0,0120	0,0057	-0,0691	0,0148	-0,0551	0,0375
Friuli-Venezia Giulia	-0,0179	0,0380	-0,0669	0,0300	-0,0822	-0,0134
Emilia-Romagna	-0,0136	0,0255	-0,0309	0,0140	-0,0593	0,0038
Toscana	-0,0068	0,0225	-0,0592	-0,0116	-0,1847	0,1650
Umbria	-0,0247	0,0341	-0,0300	0,0137	-0,1278	0,0308
Marche	-0,1131	0,0528	-0,1066	0,0304	-0,1179	0,0849
Lazio	-0,0611	0,0090	-0,1703	0,0277	-0,1368	-0,0250
Abruzzo	-0,0432	0,0222	-0,0413	0,0337	-0,1398	0,0385
Molise	-0,0141	0,0800	-0,0178	0,0488	-0,2335	0,1202
Campania	-0,0455	0,0208	-0,0392	0,0153	-0,1136	-0,0209
Puglia	-0,0340	0,0063	-0,0387	-0,0017	-0,0869	0,0659
Basilicata	-0,0433	0,0378	-0,0571	0,1264	-0,0715	0,0763
Calabria	-0,0926	-0,0186	-0,1422	-0,0422	-0,2277	0,1748
Sicilia	-0,0502	-0,0038	-0,0513	-0,0073	-0,2144	-0,0577
Sardegna	-0,0404	0,0281	-0,0439	-0,0003	-0,1442	-0,0110
Nord-ovest	-0,0928	0,0526	-0,1518	0,0628	-0,0283	0,0147
Nord-est	-0,0111	0,0101	-0,0383	0,0012	-0,0479	0,0008
Centro	-0,0316	0,0105	-0,0725	-0,0103	-0,0993	-0,0050
Sud	-0,0283	-0,0020	-0,0363	0,0004	-0,0760	0,0232
Isole	-0,0377	0,0021	-0,0414	-0,0104	-0,1774	-0,0597
Italia	-0,0276	0,0015	-0,0487	-0,0085	-0,0434	-0,0121

Tavola 1.B.2 - Intervalli di confidenza delle stime delle distorsioni della superficie di alcune componenti dei seminativi per regione

REGIONI	Frumento		Mais		Terreni a riposo soggetti a regime di aiuto	
	Inf	Sup	Inf	Sup	Inf	Sup
Piemonte	-0,0224	0,0248	--	--	-0,2336	0,0373
Valle d'Aosta/ Vallée d'Aoste	--	--	-0,0599	0,0009	--	--
Liguria	--	--	-0,1688	0,0214	--	--
Lombardia	-0,0483	0,0006	-0,1290	-0,0456	-0,4079	0,5198
Bolzano/Bozen	--	--	--	--	--	--
Trento	--	--	--	--	--	--
Veneto	-0,0648	0,0240	-0,0041	0,0543	-0,1279	0,0441
Friuli-Venezia Giulia	-0,0537	0,0289	-0,0499	-0,0002	-0,0498	0,1419
Emilia-Romagna	-0,0218	0,0247	-0,0545	0,0188	-0,2670	0,1586
Toscana	-0,0061	0,1609	-0,1625	0,0089	-0,1331	0,1738
Umbria	-0,0431	0,0330	-0,0069	0,0958	-0,1113	0,0753
Marche	-0,0213	0,0586	-0,1930	0,0770	-0,1361	0,0806
Lazio	-0,0869	-0,0024	-0,1640	0,0138	-0,3711	0,2225
Abruzzo	-0,0841	0,0140	-0,2198	0,0415	-0,2271	0,1143
Molise	-0,0131	0,0570	-0,0351	0,1162	-0,0244	0,1272
Campania	-0,0803	0,0195	-0,1811	-0,0159	-0,3177	0,1910
Puglia	-0,0535	0,0439	--	--	-0,2656	0,1497
Basilicata	-0,0231	0,0357	--	--	-0,1343	0,1339
Calabria	-0,0718	0,0814	--	--	-0,2715	0,1093
Sicilia	0,0016	0,1049	--	--	-0,3025	0,1325
Sardegna	-0,1266	0,0204	--	--	-0,2576	0,2447
Nord-ovest	-0,0267	0,0071	-0,0877	-0,0298	-0,2138	0,0990
Nord-est	-0,0282	0,0146	-0,0131	0,0262	-0,1454	0,0870
Centro	-0,0115	0,0482	-0,0870	0,0036	-0,0910	0,0876
Sud	-0,0311	0,0232	-0,1430	-0,0283	-0,1106	0,0643
Isole	-0,0092	0,0793	--	--	-0,2546	0,1146
Italia	-0,0082	0,0208	-0,0453	-0,0122	-0,0908	0,0300



Tavola 1.B.3 - Intervalli di confidenza delle stime delle distorsioni della superficie di alcune componenti delle coltivazioni legnose agrarie per regione

REGIONI	Vite		Olivo		Agrumi		Melo	
	Inf	Sup	Inf	Sup	Inf	Sup	Inf	Sup
Piemonte	-0,1010	0,0097	--	--	--	--	-0,0146	0,0241
Valle d'Aosta/ Vallée d'Aoste	-0,0203	0,0715	--	--	--	--	-0,1421	-0,0201
Liguria	-0,0439	0,0497	-0,0171	0,0232	--	--	--	--
Lombardia	0,0077	0,0650	-0,0696	0,0125	--	--	-0,0200	0,0395
Bolzano/Bozen	-0,0106	0,0311	--	--	--	--	-0,0318	0,0070
Trento	0,0403	0,0770	--	--	--	--	-0,0171	0,0103
Veneto	-0,0195	0,1009	-0,2127	-0,0190	--	--	-0,0426	0,0143
Friuli-Venezia Giulia	0,0094	0,0279	-0,1285	0,0045	--	--	-0,0726	0,0252
Emilia-Romagna	-0,0165	0,0314	-0,0960	-0,0089	--	--	-0,0431	0,1246
Toscana	-0,0328	0,1155	-0,0260	0,0235	--	--	--	--
Umbria	0,0034	0,0578	-0,0374	0,0101	--	--	--	--
Marche	-0,0761	0,2545	-0,0722	0,0078	--	--	--	--
Lazio	-0,0809	0,0990	-0,0492	0,0223	--	--	--	--
Abruzzo	0,0702	0,1474	-0,0335	0,0113	--	--	--	--
Molise	-0,0068	0,1595	-0,0227	0,0606	--	--	--	--
Campania	-0,0330	0,1069	-0,1111	0,0224	-0,1553	0,0816	-0,0420	0,0410
Puglia	0,0389	0,1712	-0,0432	0,0076	-0,0354	0,0745	--	--
Basilicata	-0,1421	0,0162	-0,0079	0,0335	-0,0389	0,0349	--	--
Calabria	-0,2537	0,1233	-0,1125	0,0066	-0,1196	-0,0122	--	--
Sicilia	0,0146	0,0886	-0,0567	-0,0037	-0,0539	0,0285	--	--
Sardegna	-0,1340	0,3340	-0,0880	-0,0089	-0,2329	0,1220	--	--
Nord-ovest	-0,0592	0,0189	-0,0221	0,0121	-0,2012	0,0496	-0,0137	0,0177
Nord-est	-0,0015	0,0543	-0,1297	-0,0281	-	-	-0,0187	0,0079
Centro	-0,0087	0,0950	-0,0259	0,0078	-0,1629	0,0813	-0,0452	0,0560
Sud	0,0297	0,1097	-0,0495	-0,0056	-0,0774	-0,0031	-0,0321	0,0430
Isole	0,0112	0,1069	-0,0565	-0,0114	-0,0548	0,0255	--	--
Italia	0,0222	0,0586	-0,0404	-0,0109	-0,0536	0,0021	-0,0157	0,0085

Tavola 1.B.4 - Intervalli di confidenza delle stime delle distorsioni del numero di capi di alcune specie di bestiame per regione

REGIONI	Bovini		Bufalini		Ovini	
	Inf	Sup	Inf	Sup	Inf	Sup
Piemonte	-0,0351	0,0415	--	--	--	--
Valle d'Aosta/ Vallée d'Aoste	-0,0211	-0,0011	--	--	--	--
Liguria	-0,0586	0,0281	--	--	--	--
Lombardia	-0,1178	-0,0158	--	--	--	--
Bolzano/Bozen	-0,0317	-0,0052	--	--	--	--
Trento	-0,0211	0,0022	--	--	--	--
Veneto	-0,1946	0,0305	--	--	--	--
Friuli-Venezia Giulia	-0,0061	-0,0006	--	--	--	--
Emilia-Romagna	-0,1041	0,1005	--	--	--	--
Toscana	-0,2303	0,0039	--	--	-0,2420	0,1615
Umbria	-0,0793	-0,0027	--	--	--	--
Marche	-0,2144	0,0306	--	--	--	--
Lazio	-0,0664	-0,0118	--	--	-0,1084	0,0167
Abruzzo	-0,1835	0,0232	--	--	-0,1702	0,2576
Molise	-0,0858	0,0063	--	--	--	--
Campania	-0,1906	-0,0195	-0,1393	0,0235	--	--
Puglia	-0,1376	0,0195	--	--	--	--
Basilicata	-0,1573	-0,0093	--	--	-0,0698	0,0158
Calabria	-0,2981	-0,0870	--	--	-0,2886	-0,0020
Sicilia	-0,1315	-0,0055	--	--	-0,1757	-0,0089
Sardegna	-0,1277	-0,0052	--	--	-0,0477	0,0168
Nord-ovest	-0,0803	-0,0079	--	--	--	--
Nord-est	-0,1210	0,0198	--	--	--	--
Centro	-0,0904	-0,0272	--	--	-0,1081	0,0486
Sud	-0,1361	-0,0547	-0,1432	0,0123	-0,0880	0,0360
Isole	-0,1124	-0,0230	--	--	-0,0632	-0,0012
Italia	-0,0832	-0,0263	-0,1084	0,0090	-0,0592	-0,0068

continua alla pagina successiva



Tavola 1.B.4 - Intervalli di confidenza delle stime delle distorsioni del numero di capi di alcune specie di bestiame per regione

REGIONI	Caprini		Suini	
	Inf	Sup	Inf	Sup
Piemonte	-0,1934	0,0171	-0,0950	0,0343
Valle d'Aosta/ Vallée d'Aoste	--	--	--	--
Liguria	--	--	--	--
Lombardia	-0,2217	-0,0605	-0,2115	0,0137
Bolzano/Bozen	--	--	--	--
Trento	--	--	--	--
Veneto	--	--	-0,0008	0,0012
Friuli-Venezia Giulia	--	--	--	--
Emilia-Romagna	--	--	-0,3361	0,3152
Toscana	--	--	--	--
Umbria	--	--	--	--
Marche	--	--	--	--
Lazio	--	--	--	--
Abruzzo	--	--	--	--
Molise	--	--	--	--
Campania	--	--	--	--
Puglia	--	--	--	--
Basilicata	-0,1132	0,0170	--	--
Calabria	-0,2422	-0,0068	--	--
Sicilia	-0,1849	0,0340	--	--
Sardegna	-0,1298	0,0602	--	--
Nord-ovest	-0,1716	-0,0537	-0,1798	0,0075
Nord-est	-0,0926	-0,0039	-0,0844	0,0862
Centro	-0,2219	0,1895	--	--
Sud	-0,1581	-0,0271	--	--
Isole	-0,1203	0,0234	--	--
Italia	-0,1090	-0,0239	-0,1138	0,0022

Capitolo 2

L'indagine di copertura¹⁶

2.1 Gli obiettivi

L'indagine di copertura è la rilevazione post censuaria volta a ottenere una stima del grado di copertura del 6° Censimento dell'agricoltura (Mazziotta, 2011).

L'errore di copertura può verificarsi quando:

- i. alcune unità facenti parte del campo di osservazione sfuggono alla rilevazione;
- ii. alcune unità facenti parte del campo di osservazione vengono rilevate più di una volta;
- iii. alcune unità, che non hanno le caratteristiche necessarie per fare parte della popolazione di interesse, sono erroneamente incluse nel campo di osservazione.

Ciascuna delle situazioni precedenti è la causa di una specifica componente dell'errore di copertura. La componente di errore derivante dalla situazione (i) è quella predominante; le altre due - (ii) e (iii) - presentano in genere un'influenza di tipo residuale, si manifestano durante la fase di raccolta dei dati e nella maggior parte dei casi sono riconosciute e corrette nelle successive fasi di correzione e controllo.

Al fine di ottenere una valutazione dell'errore di copertura, nei mesi di giugno-luglio 2011 (per il Veneto) e ottobre-novembre 2011 (per tutte le altre regioni e province autonome), è stata condotta l'indagine di copertura post-censuaria avente lo scopo di stimare per l'intero territorio nazionale, per le cinque ripartizioni geografiche e per le regioni italiane il numero di aziende agricole realmente esistenti alla data di riferimento del 6° Censimento generale dell'agricoltura e il suo tasso di copertura, definito come rapporto tra il numero di aziende rilevato al Censimento e il numero di aziende realmente esistenti.

L'indagine di copertura è campionaria e si basa su uno schema di tipo areale in cui le unità di campionamento sono i fogli di mappa del Catasto dei terreni¹⁷.

Come accennato in premessa, già nel 2000, in occasione del 5° Censimento generale dell'agricoltura, l'Istat ha valutato l'errore di copertura attraverso un'indagine *ad hoc* che potesse fornire una stima diretta del grado di copertura, adottando metodologie innovative e applicate per la prima volta in Istituto a indagini di carattere economico¹⁸. L'impostazione metodologica dell'indagine condotta per il 5° Censimento generale dell'agricoltura ha rappresentato la base di partenza per la progettazione e la conduzione dell'indagine di qualità del 6° Censimento generale dell'agricoltura. Così come il Censimento è stato caratterizzato da numerose innovazioni, anche l'indagine di copertura, presenta diversi aspetti innovativi, sia metodologici sia organizzativi. Tra i secondi si segnala, in particolare, la maggiore dimensione campionaria che permette di fornire stime a livello regionale.

2.2 Il disegno dell'indagine

La tecnica di rilevazione adottata garantisce l'indipendenza tra i due processi di individuazione delle aziende, quello censuario e quello dell'indagine campionaria. Infatti,

¹⁶ Autori: A. Bernardini §§ 2.2, 2.2.1, 2.2.2, 2.5, 2.5.1, 2.5.2, 2.6, 2.7.5; N. Cibella § 2.7.3; L. De Gaetano §§ 2.2.3, 2.3, 2.3.1, 2.3.2, 2.3.3, 2.3.4; C. De Vitiis §§ 2.4.3.1, 2.4.3.4, 2.4.4; C. De Vitiis e M. Russo § 2.4.1; M. Mazziotta §§ 2.1, 2.8; M. Russo §§ 2.4.2, 2.4.3.2, 2.4.3.3, 2.4.5, 2.A, 2.B, 2.C, 2.D; L. Soriani §§ 2.3.5, 2.5.3; L. Tosco § 2.7.4; T. Tuoto § 2.7.1; L. Valentino §§ 2.7.2, 2.7.6.

¹⁷ Foglio di mappa catastale: unità territoriale in cui ogni comune è suddiviso. Il foglio a sua volta è suddiviso in particelle continue, disegnate dal Catasto dei terreni.

¹⁸ Nei Censimenti dell'agricoltura dal 1961 al 1990 la valutazione del grado di copertura, ebbe l'obiettivo di stimare la superficie agricola non censita a diversi livelli territoriali e, solo indirettamente, l'esattività raggiunta nell'individuazione di tutte le aziende agricole esistenti alla data di riferimento.



per poter applicare i metodi di stima basati sul modello di Petersen (modello duale o cattura e ricattura) è necessario mantenere l'indipendenza tra le due occasioni di rilevazione. Per questo motivo nell'indagine di copertura non è stato possibile utilizzare le banche dati e la lista di aziende agricole - da esse derivata - preparata per il Censimento. Per tale motivo si è partiti direttamente dal territorio, basando l'indagine su uno schema campionario di tipo areale. Sono stati, pertanto, selezionati circa 1.500 fogli di mappa del Catasto dei terreni localizzati sul territorio nazionale ad eccezione delle Province autonome di Trento e Bolzano¹⁹, ove sono state estratte rispettivamente 1.641 e 776 particelle catastali²⁰ che hanno costituito le unità finali di campionamento, come sarà descritto ampiamente nel Paragrafo 2.4.

2.2.1 L'unità di rilevazione

L'unità di rilevazione dell'indagine campionaria è l'azienda agricola e zootecnica, la cui definizione è la stessa adottata al Censimento²¹. Di seguito si riporta una sintesi delle informazioni fornite ai rilevatori impegnati nell'indagine di copertura per la corretta individuazione delle aziende agricole sul territorio.

In base alla definizione, sono caratteri distintivi fondamentali dell'azienda agricola:

- l'unità tecnico-economica nella conduzione dei terreni agricoli e degli allevamenti;
- l'utilizzazione dei terreni per la produzione agricola e/o zootecnica;
- la gestione unitaria ad opera di un conduttore;
- lo svolgimento di una o più delle attività economiche specificate dal Regolamento (CE) n. 1166/2008, con riferimento alla Classificazione europea delle attività economiche (Nace Rev.2).

In relazione al primo carattere distintivo di azienda agricola, l'unità tecnico-economica è individuata dall'uso comune della forza lavoro (manodopera aziendale) e dei mezzi di produzione (fabbricati rurali, macchinari agricoli e terreni).

In relazione al secondo carattere distintivo di azienda agricola, i terreni per la produzione agricola e/o zootecnica, possono essere costituiti da una o più particelle (contigue o non), situate nello stesso comune oppure in comuni diversi della stessa regione (criterio geografico - amministrativo applicato dall'Istat per l'individuazione della singola azienda) o di province confinanti di regioni diverse.

Con riferimento al terzo carattere distintivo, la gestione unitaria è esercitata dal "conduttore", inteso come il responsabile giuridico ed economico, che sopporta il rischio della gestione aziendale sia da solo (conduttore coltivatore e conduttore esclusivamente con salariati e/o compartecipanti) sia in associazione. Il conduttore può essere una persona fisica, una società o un ente.

Infine, con riferimento alle attività economiche considerate di tipo agricolo e/o zootecnico, si riporta di seguito l'elenco contenuto nell'Allegato I del Regolamento (Ce) n. 1166/2008 (Prospetto 2.1) e fornito ai rilevatori ai fini dell'individuazione dell'azienda agricola.

19 Nelle Province autonome di Trento e Bolzano l'unità catastale che può essere identificata in maniera univoca è la particella catastale, quindi le unità campionate in questi territori sono le particelle catastali.

20 Particella catastale: poligono chiuso di una porzione di territorio situata in uno stesso Comune, intestato ad una o più persone o a società, avente la stessa qualità o classe e la stessa destinazione d'uso. Le particelle catastali sono individuate dal numero di foglio di mappa catastale in cui è inserita e dal numero (o lettera) di particella. Nel tipo di catasto in vigore nelle province autonome di Trento e Bolzano il foglio di mappa non concorre all'individuazione della particella sul terreno.

21 La definizione di azienda stabilita dall'art. 3 del Regolamento di esecuzione in conformità a quanto disposto dal Regolamento (Ce) n. 1166/2008 è la seguente: unità tecnico-economica, costituita da terreni, anche in appezzamenti non contigui, ed eventualmente da impianti e attrezzature varie, in cui si attua, in via principale o secondaria, l'attività agricola e zootecnica ad opera di un conduttore - persona fisica, società, ente - che ne sopporta il rischio sia da solo, come conduttore coltivatore o conduttore con salariati e/o compartecipanti, sia in forma associata.

Prospetto 2.1 - Elenco delle attività agricole richiamate nella definizione di azienda agricola (Gruppi di attività economiche della classificazione Nace Rev. 2). Allegato I del Regolamento (Ce) n. 1166/2008

DESCRIZIONE DELL'ATTIVITÀ	Codice NACE Rev. 2	Note aggiuntive sulle attività incluse nella definizione di attività agricola o da essa escluse
Coltivazione di colture agricole non permanenti	01.1	
Coltivazione di colture permanenti da uve o da olive	01.2	Sono incluse le attività di produzione di vino o di olio d'oliva di produzione propria
Riproduzione delle piante	01.3	
Allevamento di animali	01.4	Sono escluse tutte le attività classificate nella classe 01.49 della Nace Rev. 2 (allevamento di altri animali), tranne: i) l'allevamento e la riproduzione di struzzi, emù e conigli; ii) l'apicoltura e la produzione di miele e di cera d'api.
Attività mista (coltivazioni agricole associate all'allevamento di animali)	01.5	
Attività di supporto all'agricoltura e attività successive alla raccolta	01.6	Sono escluse tutte le attività del gruppo 01.6 della Nace Rev. 2, laddove tali attività abbiano carattere esclusivo. Sono, invece, incluse le attività della classe 01.61 della Nace Rev. 2 limitatamente a: - attività di conservazione del territorio agricolo al fine di mantenerlo in buone condizioni agricole ed ecologiche; - manutenzione del terreno al fine di mantenerlo in buone condizioni ambientali per uso agricolo (rinverdimento, nuove coltivazioni, miglioramento dei terreni, zone di ritenzione, bacini antiallagamento, eccetera).

Il riferimento alla classificazione statistica europea delle attività economiche Nace Rev. 2 consente di adeguare la definizione statistica di azienda agricola alla nuova Politica Agricola Comune (PAC), che prevede tra le attività agricole aziendali anche il mantenimento delle superfici agricole in buone condizioni agronomiche ed ambientali.

Secondo la definizione ulteriori caratteri, distintivi ma non fondamentali, di azienda agricola sono rappresentati dall'eventuale esistenza di mezzi meccanici, di impianti per la lavorazione e la trasformazione dei prodotti (cantine, frantoi, eccetera), di fabbricati rurali ed abitazioni.

Non costituisce invece carattere distintivo la destinazione della produzione aziendale (autoconsumo o vendita).

Tra le aziende agricole sono comprese anche le aziende zootecniche prive di terreno agrario. Si tratta di aziende esclusivamente zootecniche che:

- praticano allevamenti intensivi (ad esempio di bovini o di avi-cunicoli²²) o allevamenti di suini annessi a caseifici industriali;
- allevano il bestiame utilizzando terreni pascolativi appartenenti a Comuni, ad altri Enti pubblici o a privati, senza che i terreni possano configurarsi come elementi costitutivi di dette aziende.

Per terreno agrario si intende la superficie dell'unità agricola destinata alla pratica delle

²² Allevamento di polli e/o conigli.



varie colture o che potrebbe essere ad esse destinata mediante l'impiego di mezzi normalmente disponibili presso l'azienda agricola. Non è terreno agrario la superficie costituita da aree occupate da fabbricati, cortili, strade poderali, eccetera ("altra superficie").

Il campo di osservazione dell'indagine campionaria per la verifica della copertura del 6° Censimento dell'agricoltura comprende tutte le aziende agricole e zootecniche, da chiunque condotte e le cui dimensioni in termini di superficie o di consistenza del bestiame allevato siano uguali o superiori alle soglie minime fissate dall'Istat nel rispetto di quanto stabilito dal Regolamento (CE) n. 1166/2008. Rientrano nel campo di osservazione, purché aventi i requisiti di azienda agricola, anche:

- le aziende agricole gestite da istituzioni pubbliche e istituzioni non profit, ad esempio le aziende agricole degli istituti di ricerca, degli ospedali, delle cliniche, delle comunità religiose, delle scuole, degli istituti penitenziari;
- le aziende agricole gestite da imprese industriali, commerciali e dei servizi;
- gli allevamenti di tori, verri, montoni e becchi per la riproduzione, gli allevamenti di cavalli (esclusa la gestione di scuderie di cavalli da corsa e le scuole di equitazione), gli impianti di incubazione per pollame;
- le aziende zootecniche che praticano esclusivamente allevamento del bestiame, anche se prive di terreno agrario (ad es. allevamenti intensivi di bovini o avicunicoli, allevamenti di suini annessi a caseifici industriali);
- le aziende zootecniche che utilizzano terreni pascolativi che non si configurano come elementi costitutivi di dette aziende agricole (ad esempio terreni appartenenti a Comuni, ad altri Enti pubblici o a privati);
- le proprietà collettive a uso agricolo ("*common land*") costituite da terreni di proprietà privata o pubblica su cui gravano forme di diritto (usi civici)²³;
- le aziende che svolgono attività appartenenti alla classe 01.61 della Nace Rev.2 (Attività di supporto alla produzione vegetale) limitatamente ad:
 - attività di conservazione del terreno agricolo al fine di mantenerlo in buone condizioni agricole ed ecologiche;
 - attività di manutenzione del terreno al fine di mantenerlo in buone condizioni ambientali per uso agricolo (rinverdimento, nuove coltivazioni, miglioramento dei terreni, zone di ritenzione, bacini anti-allagamento, eccetera).

2.2.2 La tecnica di rilevazione

La procedura per l'individuazione delle aziende agricole si è basata essenzialmente sulle informazioni contenute negli archivi catastali. In particolare per arrivare a intervistare l'unità finale di rilevazione si è passati per gli intestatari di particella catastale dei fogli di mappa campione. Ciascun foglio di mappa (o particella catastale per le Province autonome di Trento e Bolzano) selezionato nel campione è stato sottoposto a rilevazione esaustiva, al fine di enumerare le aziende agricole che hanno almeno una particella di terreno nel territorio delimitato dal foglio stesso.

Questo approccio ha previsto una rigorosa procedura di individuazione delle aziende agricole a partire dalle informazioni presenti nell'archivio del Catasto dei terreni.

La rilevazione sul campo è stata realizzata dai rilevatori selezionati dalle regioni. Le informazioni e i materiali forniti al rilevatore sono stati:

²³ L'uso civico è definito dalla Legge n.1766 del 16 giugno 1927 come il "diritto che gruppi di persone quali collettività di abitanti di un Comune o di una sua frazione o, anche, di una sola parte di essi o di una collettività di altro tipo (associazione agraria, università, comunanza, partecipazione, dominio collettivo, vicinia, regola, eccetera) esercitano su terreni appartenenti a privati, oppure ad Enti territoriali od associativi di vario genere". Gli usi civici più frequenti sono quelli di legnatico (raccolta della legna), pascolatico od erratico (pascolo degli animali), fungatico (raccolta di funghi).

- la cartografia del foglio di mappa (o della particella) campione, con la componente dei confini perimetrali del foglio, le strade, i corsi d'acqua, i fabbricati, eccetera, in cui sono circoscritte le particelle appartenenti al foglio;
- l'elenco dei nominativi degli intestatari delle particelle catastali facenti parte del foglio di mappa campione o campionate direttamente (per Trento e Bolzano); per ogni nominativo è stato riportato il codice fiscale, l'indirizzo del domicilio, il numero di telefono dell'intestatario della particella catastale²⁴ e altre informazioni identificative di supporto alla rilevazione.

A partire da queste informazioni, riportate nel primo modello a disposizione del rilevatore (modello Istat CensAgrCop1), è avvenuta l'individuazione delle aziende agricole e l'intervista alle medesime, seguendo due distinti passi: individuazione delle aziende agricole e intervista al conduttore.

Passo 1) individuazione delle aziende agricole

L'obiettivo di questo passo è stato quello di raccogliere tutte le informazioni utili ad associare a ciascuna particella catastale (campionata o facente parte del foglio di mappa campione) le eventuali aziende agricole corrispondenti, tentando di acquisire tutti gli elementi utili a stabilire un successivo contatto con i conduttori delle aziende stesse. In tale passo il rilevatore:

- prendeva contatto con gli intestatari delle particelle al fine di conoscere se, alla data del Censimento, il terreno di tale particella fosse condotto da un'azienda agricola;
- individuava tutte le aziende agricole (con centro aziendale²⁵ all'interno o all'esterno del foglio di mappa o della particella campione) i cui terreni o allevamenti, in data 24 ottobre 2010, ricadevano in tutto o in parte nel foglio di mappa (o nella particella campione per le Province autonome di Trento e Bolzano);
- acquisiva il nome del conduttore dell'azienda individuata.

Nel fare ciò, il rilevatore doveva compilare il secondo modello che costituisce l'elenco delle aziende agricole (modello Istat CensAgrCop2).

Al termine di questo passo il rilevatore ha prodotto una lista di conduttori di aziende agricole per poi procedere con il passo 2.

Passo 2) intervista al conduttore

In questo passo il rilevatore prendeva contatto con ciascuno dei conduttori delle aziende individuate al passo precedente e, dopo aver verificato di essere in presenza di un'azienda agricola o zootecnica, doveva rilevare le principali caratteristiche dell'azienda (SAU, superficie totale, eccetera) con riferimento alla data del Censimento, attraverso la compilazione del terzo modello che rappresenta il questionario di rilevazione vero e proprio (modello Istat CensAgrCop3):

Al termine di questo passo il rilevatore procedeva anche a completare la compilazione del secondo modello in base alle informazioni contenute nel terzo modello già compilato, integrando e correggendo, dove necessario, le informazioni preliminarmente ricevute dagli intestatari delle particelle catastali.

Al termine di questi due passi, il rilevatore aveva definito l'elenco delle aziende agricole individuate sul territorio con le loro principali caratteristiche.

²⁴ In taluni casi, non è stato possibile reperire le informazioni relative all'intestatario, pertanto, il rilevatore per poterlo rintracciare poteva avvalersi solo della localizzazione della particella e recarsi direttamente sul campo.

²⁵ Il Centro aziendale è identificato dal fabbricato, o dal complesso dei fabbricati, connesso all'attività aziendale e situato entro il perimetro dei terreni aziendali. In assenza di fabbricati, il centro aziendale si identifica con la porzione più estesa dei terreni aziendali.



2.2.3 Tempi e fasi della rilevazione

Le attività preliminari di preparazione dell'indagine di copertura hanno incluso la costituzione - nel mese di maggio 2010 - di un Gruppo di lavoro²⁶ *ad hoc* per la predisposizione del documento, che ha incluso la metodologia, l'organizzazione, nonché la definizione dei tempi e delle fasi della rilevazione stessa.

A partire dalle linee guida e azioni definite in tale documento sono state realizzate le attività e le fasi che caratterizzano questa rilevazione, dalla sua preparazione fino alla conclusione, come di seguito descritto.

In particolare, la data di riferimento dell'indagine di copertura del 6° Censimento dell'agricoltura è il 24 ottobre 2010, in linea con il riferimento temporale del 6° Censimento dell'agricoltura.

Nel periodo dicembre 2010 - marzo 2011 sono stati definiti gli aspetti metodologici e organizzativi, mentre, entro marzo 2011, si è provveduto all'estrazione dei fogli di mappa (e delle particelle) nei comuni campione dall'archivio del Catasto dei terreni. Individuati i comuni e i fogli di mappa (e le particelle) campione, sono state predisposte le mappe (cartografie) aggiornate. Tali mappe sono state inserite nel Sistema di Gestione della Rilevazione denominato SGR COP²⁷, al fine di fornire un valido ausilio ai rilevatori durante la raccolta dei dati, unitamente all'elenco degli intestatari di particella catastale.

Le operazioni di rilevazione hanno preso avvio ufficialmente con l'emanazione della circolare Istat, nella quale sono stati definiti tutti gli aspetti organizzativi ed esecutivi della rilevazione stessa (cfr. circolare n. 13 del 29.07.2011, Prot.gen. n. 6403²⁸).

Entro luglio 2011 l'Istat ha rilasciato tutte le funzionalità di SGR COP (funzioni di gestione della rilevazione, utilità, monitoraggio) e del sistema di acquisizione dei questionari in data entry controllato che ha permesso ai rilevatori una piena operatività all'avvio della rilevazione.

Nei mesi di giugno - luglio 2011 è stata realizzata l'indagine nella sola Regione Veneto anche al fine di testare la rete di rilevazione, il funzionamento di SGR COP, la modulistica adottata²⁹ (elenco degli intestatari, questionario, documentazione a supporto) per la rilevazione, la gestione degli aspetti critici non considerati in fase di progettazione.

Al fine di garantire l'immediato funzionamento degli Uffici Regionali di Censimento (URC), costituiti presso le regioni, l'Istat ha erogato entro novembre 2011 il contributo forfetario fisso, ripartito in funzione della numerosità campionaria delle unità di rilevazione presenti nei fogli di mappa (o particelle) campione³⁰.

Entro settembre 2011, si è svolta a cura degli URC la fase di reclutamento dei rilevatori e degli eventuali coordinatori.

Entro ottobre 2011, gli URC hanno inviato all'Istat l'elenco e l'anagrafica dei rilevatori ed eventuali coordinatori che dovevano operare all'interno di SGR COP; quindi l'Istat ha provveduto all'inserimento di tale elenco nel sistema di rilevazione (SGR COP) e al

26 Gruppo di Lavoro avente il compito di coadiuvare la struttura preposta all'esecuzione della rilevazione della sotto-copertura di aziende agricole e relative superfici relative al 6° Censimento generale dell'agricoltura (delibera n. 209/DPTS/14-12-2010).

27 Per la rilevazione è stato predisposto, a cura dell'Istat il sistema informatico di gestione della rilevazione SGR COP. Tale rilascio è stato effettuato, in versione test, nel maggio 2011 per la sola Regione Veneto, al fine di verificarne le varie funzionalità.

28 Una prima circolare è stata emanata solo per i Comuni del Veneto, al fine di disciplinare l'organizzazione e l'esecuzione dell'indagine all'interno della Regione (cfr. circolare n. 12 del 01.06.2011, Prot.gen. n. 4465).

29 Cfr. Paragrafo 2.5.1.

30 Tale numerosità è stata rilasciata in allegato alla circolare tecnica predisposta dall'Istat sull'organizzazione della rilevazione.

rilascio delle credenziali di accesso ai rilevatori e ai coordinatori per la gestione della rete e ha proceduto all'assegnazione, ai rilevatori, dei fogli di mappa (o particelle) da lavorare.

Nel periodo settembre - ottobre 2011, sono state organizzate giornate di formazione per i rilevatori e i coordinatori, curate direttamente dall'Istat centrale, su tutti gli aspetti tecnico-operativi riguardanti la rilevazione.

La raccolta dei dati sul campo si è svolta tra il 15 ottobre 2011 e il 31 gennaio 2012. In questo arco temporale i rilevatori hanno individuato le aziende agricole sul territorio e i coordinatori e i responsabili URC hanno organizzato il lavoro e garantito lo svolgimento dell'attività di registrazione dei questionari e di controllo e correzione dei dati, attraverso l'utilizzo di apposite funzioni di SGRCOP.

Entro la fine di marzo 2012, gli URC hanno confezionato il materiale cartaceo della rilevazione (modelli ausiliari e questionario), e restituito, secondo le indicazioni fornite da Istat, i plichi/pacchi così realizzati. Il materiale restituito all'Istat era stato preventivamente registrato sul sistema informatico.

A partire da giugno 2012, i dati raccolti sono stati sottoposti a operazioni di controllo e correzione - a cura dell'Istat - e al loro trattamento informatico ai fini del *record linkage* e delle analisi di sotto-copertura, le cui operazioni sono descritte in dettaglio nei paragrafi successivi.

2.3 L'organizzazione

L'Istat ha effettuato la rilevazione di copertura in collaborazione con gli Uffici di Censimento (URC) delle Regioni e Province autonome. Il Piano generale di censimento del 6° Censimento generale dell'agricoltura disciplina i compiti dei differenti attori che partecipano al Censimento e alla rilevazione della copertura del Censimento stesso. Le norme in esso contenute sono specificate e chiarite anche in successive circolari informative, emanate dall'Istat durante lo svolgimento del Censimento, prima, e successivamente all'indagine di copertura.

La rilevazione, prevista dal Programma Statistico Nazionale 2011-2013, approvato con Decreto del Consiglio dei ministri del 31 marzo 2011 e registrato alla Corte dei Conti il 28 giugno 2011, è compresa tra le rilevazioni che comportano l'obbligo di risposta per i soggetti privati (Codice IST-02357), così come disciplinato dagli artt. 7 e 11 del Decreto legislativo 6 settembre 1989, n. 322 e successive modifiche.

L'indagine di copertura, inoltre, è un'attività censuaria prevista nel Paragrafo 4.8 del Piano Generale di Censimento (PGC), che, ai sensi del Decreto legge n. 78 del 31 maggio 2010, convertito con modificazioni nella Legge 30 luglio 2010, n. 122, art. 50, comma 6, lettera a), stabilisce che: "l'Istat organizza le operazioni censuarie (...) secondo il Piano Generale di Censimento di cui al numero Istat SP/1275.2009 del 23 dicembre 2009 e relative circolari applicative che individuano anche gli enti e gli organismi pubblici impegnati nelle operazioni censuarie".

Nei paragrafi seguenti sono dettagliatamente descritti i compiti e i ruoli degli attori della rete di rilevazione della copertura.



2.3.1 La rete di rilevazione

La rete di rilevazione è costituita dall'Istat e dagli Uffici Regionali di Censimento. In tale contesto l'Istat è il titolare della rilevazione della copertura del 6° Censimento generale dell'agricoltura e ne definisce gli aspetti tecnici, organizzativi e metodologici.

A livello regionale operano gli Uffici Regionali di Censimento (URC) che sovrintendono al funzionamento della rete di rilevazione nell'ambito del territorio di competenza, controllando il regolare svolgimento delle varie operazioni realizzate. Nell'ambito di ciascun URC, un ruolo particolare è riservato alla figura del responsabile che coordina tutte le attività, comprese le comunicazioni con l'Istat.

Inoltre, in base a specifiche intese stipulate con Istat, per le Province autonome di Trento e Bolzano le funzioni di URC sono svolte rispettivamente dal Servizio di Statistica della Provincia autonoma e dall'Istituto Provinciale di Statistica (ASTAT)³¹.

All'interno degli URC hanno operato figure professionali di rilievo: i responsabili dell'Ufficio, i coordinatori e i rilevatori (con compiti e funzioni dettagliate dal PGC), nonché altro personale dipendente dell'Ente, che faceva parte dell'URC e svolgeva specifici compiti assegnati dal responsabile.

La creazione della rete di rilevazione, il suo funzionamento e il coordinamento delle attività di tutti gli attori coinvolti sono realizzati attraverso un sistema strutturato di comunicazioni, basato sull'emanazione di atti ufficiali (circolari) e lettere informative che l'Istat ha predisposto e inviato agli organi della rilevazione. Tutto il materiale relativo alla rilevazione (manuale, questionario della rilevazione, modelli ausiliari) è messo a disposizione di tutti gli attori della rete (URC e rilevatori) sul Sistema di Gestione della Rilevazione (SGRCOP) informatizzato, al quale si accede attraverso l'indirizzo URL <https://indata.istat.it/sgrcop>. Gli atti ufficiali (circolari), invece, sono resi disponibili sul portale web del Censimento agricoltura all'indirizzo: <https://Censimentoagricoltura.istat.it/rete> e consultabili da tutti gli attori della rete censuaria.

Le principali fasi delle attività della rilevazione che i suddetti soggetti sono stati chiamati a svolgere sono sintetizzate nel calendario delle attività già riportato e approfondito nel Paragrafo 2.2.3.

2.3.2 Il ruolo e i compiti dell'Istat

Per la progettazione e la successiva implementazione dell'indagine, come precedentemente anticipato, l'Istat ha costituito al suo interno appositi gruppi di lavoro, mentre la realizzazione della rilevazione è stata affidata a un'apposita struttura dell'Istituto.

A livello nazionale, l'Istat definisce gli aspetti organizzativi, tecnici e metodologici della rilevazione di copertura, sovrintende alle operazioni censuarie, ne assicura il monitoraggio e adotta le circolari, i provvedimenti e le misure necessarie che ne approfondiscono specifici aspetti, al fine di garantirne il buon andamento. Compito dell'Istat è quello di elaborare i dati, sia di fonte amministrativa sia raccolti mediante rilevazione diretta sulle unità di rilevazione. Per il raggiungimento di tale fine, l'Istat ha adottato appropriati metodi di controllo e di correzione, ha validato i dati raccolti dagli organi di rilevazione ed ha provveduto al calcolo delle stime campionarie. Sempre a cura dell'Istat risulta la diffusione e la comunicazione dei risultati della rilevazione di copertura del 6° Censimento generale dell'agricoltura.

³¹ Pertanto, l'ASTAT della Provincia autonoma di Bolzano e il Servizio Statistica della Provincia autonoma di Trento hanno provveduto alla rilevazione esercitando il ruolo e le funzioni equivalenti a quelle di un Ufficio Regionale di Censimento.

L'Istat ha predisposto il materiale della rilevazione e, in particolare, i questionari, il manuale della rilevazione e i modelli ausiliari, nonché il materiale per la formazione, che è stata organizzata dagli URC e svolta, agli organi della rete, direttamente dai ricercatori dell'Istat.

L'Istat, inoltre, ha assicurato - ai sensi della legge - il rispetto del segreto statistico e la tutela della riservatezza delle informazioni personali acquisite presso i rispondenti. In particolare, la normativa in materia prevede che i dati raccolti in occasione della rilevazione di copertura siano coperti dal segreto d'ufficio e dal segreto statistico, in conformità agli articoli 8 e 9 del Decreto legislativo 6 settembre 1989, n. 322 e successive modificazioni e integrazioni. Inoltre, le operazioni della rilevazione che richiedono il trattamento di dati personali sono svolte nel rispetto della disciplina dettata dal Decreto legislativo 30 giugno 2003 e successive modificazioni e integrazioni (c.d. Codice in materia di protezione dei dati personali) e dal Codice di deontologia e di buona condotta per il trattamento dei dati personali a scopi statistici e di ricerca scientifica effettuati nell'ambito del Sistema Statistico Nazionale. In base alla normativa richiamata, l'Istat è titolare del trattamento dei dati personali, mentre responsabili del trattamento dei dati personali, per le operazioni di rispettiva competenza, sono il Direttore del Dipartimento per i censimenti e gli archivi amministrativi e statistici dell'Istat, il Direttore del Dipartimento per l'integrazione, la qualità e lo sviluppo delle reti di produzione e di ricerca, i responsabili degli Uffici Regionali di Censimento, i coordinatori e i rilevatori.

L'Istat ha monitorato il funzionamento della rete territoriale mediante i propri Uffici Territoriali. Nell'ambito di ciascun Ufficio territoriale hanno operato i Responsabili Istat Territoriali (RIT), che svolgono nel contesto delle attività di rilevazione campionaria compiti di supporto e di assistenza (monitoraggio) agli URC.

2.3.3 Il ruolo e i compiti degli Uffici Regionali di Censimento (URC)

Per l'esecuzione della rilevazione della copertura del 6° Censimento generale dell'agricoltura l'Istat si è avvalso, come anticipato, quali organi intermedi, degli Uffici Regionali di Censimento (URC) delle Regioni e Province autonome costituiti per l'effettuazione del 6° Censimento generale dell'agricoltura con le modalità e i tempi stabiliti dalle apposite circolari tecniche predisposte dall'Istituto.

Tenuto conto anche delle indicazioni emerse dal Comitato consultivo per la preparazione a livello regionale del 6° Censimento generale dell'agricoltura, a ciascun Ufficio Regionale di Censimento (URC)³², nell'ambito della rilevazione di copertura, sono assegnati i seguenti compiti:

1. selezionare, nominare e assegnare i rilevatori ai fogli di mappa (o particelle) campionati per la rilevazione, rispettando i criteri di indipendenza dalla rilevazione censuaria;
2. nominare gli eventuali loro coordinatori (CoC);
3. comunicare via e-mail, alla casella di posta elettronica Istat dell'indagine: i) il nominativo del/i responsabile/i e il/i relativo/i recapito/i telefonico/i e indirizzo/i di posta elettronica al quale inviare tutte le comunicazioni di servizio relative alla rilevazione in oggetto, previa invio lettera di adesione alla realizzazione dell'indagine³³; ii) le informazioni anagrafiche dei rilevatori selezionati e dei coordinatori comunali, ove esistenti (seguendo lo schema fornito dall'Istat³⁴);

³² Per l'esecuzione dell'indagine di copertura l'Istat si è avvalso degli Uffici Regionali di Censimento, ad eccezione dell'URC Toscana dove la rilevazione è stata effettuata dagli Uffici Territoriali di Censimento di Grosseto, Livorno, Siena, Firenze, Pistoia, Lucca e dall'Ufficio di Censimento della Comunità Montana della Lunigiana. L'Ufficio Territoriale di Censimento di Pisa non ha effettuato la rilevazione. Il compito di URC è stato assegnato, con apposita lettera d'incarico, all'Ufficio territoriale Istat per la Toscana.

³³ L'Istat ha predisposto un format con il quale ciascun Ufficio Regionale di Censimento aderiva alla realizzazione dell'indagine.

³⁴ Il fac-simile dell'anagrafica del rilevatore è stato allegato alle circolari n. 12 (per la Regione Veneto) e n. 13 per le altre regioni.



4. organizzare in collaborazione con la struttura competente dell'Istat, la formazione dei coordinatori (ove esistenti) e dei rilevatori, nonché gestire la logistica per la formazione, predisponendo un'aula dotata di adeguati supporti informatici;
5. coordinare sotto il profilo tecnico-statistico e organizzativo i coordinatori (ove esistenti) e i rilevatori;
6. monitorare costantemente l'andamento della rilevazione e assicurare il buon andamento delle operazioni di rilevazione nel territorio di competenza;
7. curare il confezionamento e la trasmissione del materiale cartaceo a conclusione della rilevazione.

2.3.4 I compiti dei rilevatori

Le modalità di reclutamento dei coordinatori, dei loro eventuali responsabili e dei rilevatori sono regolate dal Titolo IV - Personale addetto al Censimento- artt. 21, 22, 23 e 24 del Decreto del Presidente della Repubblica 23 luglio 2010, n. 154 e dalla circolare dell'Istat n. 5, paragrafo 1, del 28 luglio 2010 (prot. n. 4857).

L'URC, poteva, se lo riteneva opportuno, utilizzare i rilevatori e i coordinatori già selezionati o utilizzati per il 6° Censimento generale dell'agricoltura, a condizione che operassero in aree diverse da quelle loro precedentemente assegnate, allo scopo di garantire le condizioni di indipendenza tra le due rilevazioni. I rilevatori sono i soggetti ai quali è stata affidata la rilevazione delle unità loro assegnate da parte dei responsabili URC o da eventuali coordinatori. Essi dovevano agire in modo tale da predisporre i soggetti sottoposti alla rilevazione alla massima collaborazione. Nel contattare le unità di rilevazione, essi erano tenuti a rendere nota la propria identità, la propria funzione e le finalità della raccolta, anche attraverso adeguata documentazione (cartellino di riconoscimento visibile), ai sensi dell'art. 10 del Codice di deontologia e di buona condotta per il trattamento di dati personali a scopi statistici e di ricerca scientifica. I rilevatori dovevano, inoltre, fornire tutti i chiarimenti che consentivano all'interessato di compilare in modo consapevole e adeguato il questionario della rilevazione.

In particolare, i rilevatori hanno eseguito i seguenti compiti:

- partecipato alle riunioni di formazione ed acquisito le necessarie conoscenze e abilità attraverso i diversi strumenti formativi messi a disposizione dall'Istat;
- gestito quotidianamente, mediante l'uso del sistema informatico di gestione della rilevazione SGRCOP, i fogli di mappa a loro assegnati;
- verificato la completezza e la correttezza dell'elenco dei nominativi degli intestatari delle particelle catastali campionate o facenti parte del foglio di mappa campione fornito dall'Istat;
- contattato gli intestatari delle particelle catastali ricadenti in ogni foglio di mappa loro assegnato;
- contattato i conduttori delle aziende agricole per la somministrazione del questionario di rilevazione, fornendo loro chiarimenti e supporto sulle finalità e la natura obbligatoria della rilevazione;
- revisionato i dati contenuti nei questionari e compilato i riquadri di propria pertinenza;
- sollecitato gli intestatari di particella non ancora rispondenti al rispetto dei tempi e delle modalità definite dal responsabile dell'URC.

Le operazioni sul campo hanno impegnato i rilevatori che, quotidianamente, dovevano contattare l'intestatario di particella catastale e da questi risalire al conduttore dell'azienda agricola, supportati dagli ausili forniti dall'Istat.

Per lo svolgimento dei compiti sopraelencati i rilevatori hanno avuto accesso al sistema

informatico di gestione della rilevazione (SGRCOP) tramite un'utenza personalizzata assegnata dall'Istat³⁵.

2.3.5 La formazione dei rilevatori

La formazione dei rilevatori è stata effettuata direttamente a cura dall'Istat ai rilevatori presso le sedi regionali, al fine di garantire uniformità nei contenuti trasmessi e un rapporto diretto Istat - rilevatori di particolare importanza per una rilevazione di qualità.

Le giornate formative, in collaborazione con ciascun URC, sono state organizzate, tra settembre e ottobre 2011, con tradizionale formazione in aula e ricorso anche a test/esercitazioni per valutare il grado di apprendimento.

Le tematiche trattate in occasione dei momenti formativi hanno principalmente riguardato i seguenti punti:

- la metodologia dell'indagine;
- l'organizzazione e i compiti della rete;
- il processo di rilevazione;
- l'analisi del questionario e le relative modalità di compilazione;
- il funzionamento di SGRCOP.

Oltre alle giornate formative, sono stati predisposti strumenti idonei a garantire all'intera rete di rilevazione una formazione continua e un'assistenza costante durante la rilevazione. Più nello specifico, la rete di rilevazione ha potuto trovare risposta alle domande relative agli aspetti gestionali e tecnici della rilevazione tramite l'attivazione, da parte di Istat, di una casella di posta elettronica dedicata e un *help desk* per soddisfare in tempo reale le richieste di chiarimenti e di approfondimento da parte dei rilevatori e dei responsabili URC.

Dalla medesima casella l'Istat ha inviato tutte le comunicazioni di servizio alla rete territoriale.

2.4 Il piano di campionamento, i metodi di stima e la variabilità

2.4.1 La popolazione obiettivo, i parametri di interesse e i domini di stima

Come detto, l'indagine è volta a produrre le stime dei tassi di copertura e di sotto-copertura del 6° Censimento generale dell'agricoltura. Per *tasso di copertura* (nell'ipotesi di assenza di *sovra-copertura*) si intende il rapporto tra il numero di aziende agricole rilevate con il Censimento e il numero di aziende che costituiscono la popolazione obiettivo del Censimento, corrispondente al numero di aziende agricole realmente esistenti sul territorio nazionale alla data di riferimento del Censimento stesso (24 ottobre 2010). Il *tasso di sotto-copertura*, invece, è espresso dal rapporto tra il numero di aziende agricole sfuggite all'enumerazione censuaria e la dimensione effettiva della popolazione su definita.

Come già descritto, l'indagine di copertura effettuata è areale e a due stadi di selezione (comuni e fogli di mappa - o particelle -, inclusi nel *Catasto dei terreni*) basata sulla rilevazione di tutte le aziende esistenti nei fogli di mappa (o particelle) campione.

Tuttavia, non essendo nota la dimensione effettiva della popolazione, per il calcolo dei tassi sopra definiti è necessario ricorrere all'utilizzo di appositi modelli probabilistici fondati sulla mancata enumerazione delle unità, che può verificarsi sia al Censimento che all'indagine di copertura.

³⁵ Per approfondimenti sul sistema informatico di gestione della rilevazione (SGRCOP) si rimanda al Paragrafo 2.5.2 del presente Volume.



In questa indagine è stato scelto il *modello di Petersen*³⁶ (o di *omogeneità entro le liste*) che - opportunamente adattato - costituisce una ragionevole approssimazione della realtà³⁷.

La finalità primaria dell'indagine in esame è l'ottenimento di stime *affidabili* dei tassi di copertura e di sotto-copertura inerenti ai seguenti domini di studio:

- l'intero territorio nazionale;
- le cinque ripartizioni geografiche (Nord-ovest, Nord-est, Centro, Sud, Isole);
- le diciannove regioni geografiche e le Province autonome di Trento e Bolzano;
- le classi di superficie agricola utilizzata (SAU).

I suddetti domini presentano, dunque, l'importante caratteristica di essere *domini pianificati*, essendo le loro dimensioni campionarie definite a priori.

2.4.2 Il modello di stima degli errori di copertura

Al fine di illustrare con chiarezza il *modello di Petersen*, si fa riferimento al caso ipotetico in cui anche l'indagine di copertura sia una rilevazione di tipo censuario e ripeta, quindi, le operazioni censuarie su tutti i fogli di mappa dell'intero territorio nazionale. In tal caso il censimento della popolazione di riferimento e l'indagine di copertura danno luogo a due liste distinte della popolazione, entrambe soggette a errore di copertura. Le assunzioni di seguito presentate risultano particolarmente restrittive, tali da presentare alcuni aspetti critici; la validità dell'applicazione del modello è connessa al rispetto delle assunzioni stesse.

Indichiamo con *C* la lista ottenuta con il Censimento e con *I* quella ricavata con l'indagine di copertura.

Per la valutazione della numerosità *N* della popolazione è possibile utilizzare le due liste integrando l'informazione fornita da entrambe per mezzo di un modello di rappresentazione dell'errore di copertura.

Il *modello di Petersen* è caratterizzato dalle seguenti assunzioni:

1. la popolazione di riferimento è chiusa e di dimensione fissata pari a *N*;
2. la probabilità che l'unità *j* appartenga o meno alla lista *C* e che appartenga o meno alla lista *I* è rappresentabile mediante una distribuzione multinomiale in cui le probabilità congiunte e quelle marginali sono riportate nello schema seguente:

		Lista <i>I</i>		
		Sì	No	
Lista <i>C</i>	Sì	p_{j11}	p_{j12}	p_{j1+}
	No	p_{j21}	p_{j22}	p_{j2+}
		p_{j+1}	p_{j+2}	1

36 Wolter, 1986.

37 Cfr. Paragrafo 2.4.2.

3. le due liste e C e I , sono considerate il risultato di N prove mutualmente indipendenti, usando le distribuzioni multinomiali descritte al punto 2. Per ogni singola unità j si definisce una variabile indicatrice x_{jab} pari a uno se l'unità j cade nella cella ab e 0 altrimenti ($a, b = 1, 2$). Per le unità della popolazione la situazione è descritta come segue

		Lista I		
		<i>Si</i>	<i>No</i>	
Lista C	<i>Si</i>	x_{11}	x_{12}	x_{1+}
	<i>No</i>	x_{21}	x_{22}	x_{2+}
		x_{+1}	x_{+2}	N

dove:

$$x_{ab} = \sum_{j=1}^N x_{jab} \text{ è il numero di unità nella cella } ab \text{ (} a, b = 1, 2 \text{), } x_{a+} = \sum_{j=1}^N \sum_{b=1}^2 x_{jab} \text{ (} a = 1, 2 \text{)}$$

$$\text{e } x_{+b} = \sum_{j=1}^N \sum_{a=1}^2 x_{jab} \text{ (} b = 1, 2 \text{)}.$$

Ovviamente, le quantità x_{22} e N non sono osservabili;

4. è possibile determinare senza errore quali unità registrate nella lista I sono presenti nella lista C e quali no (ossia, non sono presenti errori di abbinamento);
5. entrambe le liste sono depurate da errori di registrazione e duplicazione;
6. le probabilità che le unità siano incluse nella lista C , p_{j1+} ($j=1, \dots, N$), e le probabilità che le unità siano incluse nella lista I , p_{j+1} ($j=1, \dots, N$), sono costanti per ciascuna lista, ossia soddisfano le condizioni $p_{j1+} = p_{1+}$ e $p_{j+1} = p_{+1}$; d'altra parte le due probabilità p_{1+} e p_{+1} possono essere differenti.

La validità delle suddette assunzioni è un aspetto critico per l'applicabilità del modello, ma è necessaria per prevenire la presenza di distorsione nei risultati dell'indagine.

In particolare, l'ipotesi 1 assume che non ci siano fenomeni di entrata nella popolazione ("nascite", ossia, nel caso in esame, costituzione di nuove aziende agricole) o di uscite dalla stessa ("morti", ossia, nel presente contesto di indagine, cessazione di aziende agricole) in modo che la popolazione enumerata nelle due occasioni di rilevazione (Censimento e indagine di copertura) sia la stessa. Tale condizione è assicurata sia dalla circostanza che la data di riferimento è la stessa per le due rilevazioni (24 ottobre 2010) e sia dalla vicinanza tra il periodo in cui è eseguito il Censimento e quello in cui è effettuata l'indagine.

Per ciò che concerne invece l'ipotesi 4, le procedure impiegate devono essere tali da garantire il più possibile l'assenza di errori di abbinamento; in ogni caso, deve essere valutato l'effetto di eventuali errori di linkage sulle stime dei parametri di interesse. Inoltre, l'indagine prevede un'accurata fase di controllo e correzione dei dati, per cui si può assumere che l'ipotesi 5 sia verificata. Infine, una delle condizioni affinché la stima dell'ammontare ignoto della popolazione risulti non distorta, impone che la probabilità di cattura sia costante per ogni azienda agricola j in ciascuna delle due catture (ipotesi 6). Il modo più agevole di rispettare tale ipotesi è costituito dal ricorso a una stratificazione delle unità statistiche mediante l'utilizzo di un insieme di variabili scelte allo scopo



di assicurare la suddetta omogeneità delle probabilità di cattura. Ciò consentirà di calcolare la popolazione ignota di ogni dominio di studio come somma delle popolazioni stimate per ciascuno strato.

Sfruttando le notazioni simboliche precedentemente introdotte, le stime³⁸ dei tassi di copertura e sotto-copertura sono rispettivamente esprimibili nella forma:

$$\hat{\tau} = \frac{x_{1+}}{N}$$

$$\hat{\zeta} = 1 - \hat{\tau} = 1 - \frac{x_{1+}}{N} = \frac{x_{2+}}{N}$$

Le suddette formule non sono calcolabili in quanto è incognito. Tuttavia, avendo assunto l'indipendenza delle due rilevazioni, i tassi in questione possono essere stimati come segue:

$$\hat{\tau} = \frac{x_{11}}{x_{+1}} \quad (1)$$

$$\hat{\zeta} = 1 - \hat{\tau} = \frac{x_{21}}{x_{+1}} \quad (2)$$

mentre una stima della numerosità della popolazione è data da:

$$\hat{N} = x_{1+} + x_{1+} \frac{x_{21}}{x_{11}} = \frac{x_{1+} x_{+1}}{x_{11}} \quad (3)$$

Le stime dell'errore di copertura e di sotto-copertura nei domini stratificati

Poiché con l'indagine post-censuaria non si effettua una enumerazione completa su tutti i fogli di mappa ma solo su un campione di essi, le quantità x_{11} , x_{21} ed x_{+1} sono note solo per fogli di mappa inclusi nel campione. Pertanto, denotando con il generico foglio di mappa campione, le precedenti relazioni (1), (2) e (3) possono essere adottate a livello dei fogli di mappa sopra menzionati, per i quali le quantità x_{i11} , x_{i21} e x_{i+1} sono osservabili, essendo l'operazione di conteggio completa entro i fogli campione. Il tasso di copertura τ_i del foglio i è, dunque, stimato da $\hat{\tau}_i = (x_{i11} / x_{i+1})$, mentre quello di sotto-copertura - sempre relativo al foglio i - è dato da:

$$\hat{\zeta}_i = (x_{i21} / x_{i+1})$$

Si indichi ora con ($d = 1, \dots, D$) un generico dominio di interesse stratificato³⁹ (ad esempio, l'intero territorio nazionale, la ripartizione geografica o la regione), si possono stimare⁴⁰ le quantità a numeratore e denominatore della (3) sulla base delle osservazioni campionarie; pertanto, le stime dei tassi di copertura e di sotto-copertura riferiti a d sono date da:

38 Sono stime da modello e non da disegno, dato che si suppone che entrambe le rilevazioni siano esaustive.

39 Un dominio si dice *stratificato* (o stratificabile) se le sue unità sono raggruppate (o raggruppabili) in uno o più strati costituiti da quelle sole unità; il che significa che esso contiene tutte le unità della popolazione appartenenti a uno strato oppure ad aggregazioni di strati. Viceversa, un dominio *non stratificato* (o stratificabile) contiene solo una parte delle unità della popolazione appartenenti a uno o più strati del disegno.

40 In tal caso, le stime sono espresse in termini di variabilità sia rispetto al modello sia rispetto al disegno di campionamento, in quanto si suppone che l'indagine di copertura non riguardi l'intero territorio nazionale, ma un campione di fogli di mappa.

(4)

$${}_d \tilde{N} = \frac{{}_d \hat{X}_{11}}{{}_d \hat{X}_{+1}}$$

(5)

$$\text{in cui } {}_d \hat{X}_{11} = \sum_{i \in d} x_{i11} k_i, \quad {}_d \hat{X}_{21} = \sum_{i \in d} x_{i21} k_i \quad \text{e} \quad {}_d \hat{X}_{+1} = \sum_{i \in d} x_{i+1} k_i$$

sono rispettivamente le stime campionarie⁴¹ delle quantità x_{11} , x_{21} e x_{+1} calcolate nel dominio d , essendo k_i il peso campionario finale attribuito al foglio di mappa i incluso nel campione. Parimenti, la stima di N è fornita da:

$${}_d \tilde{N} = \frac{{}_d X_{1+} \quad {}_d \hat{X}_{+1}}{{}_d \hat{X}_{11}} \quad (6)$$

Le stime dell'errore di copertura e di sotto-coverage nei domini non stratificati

Un ulteriore aspetto da considerare è quello inerente le stime per classi di SAU. Poiché tali classi costituiscono domini non stratificati⁴², si è ricorsi a un artificio che permette di ricondurre tali tipi di stime a quelle relative a domini pianificati senza dover costruire un impianto metodologico apposito⁴³.

Sia d' il dominio trasversale rispetto al quale si vuole calcolare la stima (nel caso in esame la classe di SAU) e si indichi, inoltre, con x'_{ab} la variabile i cui valori nella popolazione sono:

$$x'_{jab} = \begin{cases} x_{jab} & \text{se } j \in d' \\ 0 & \text{altrimenti} \end{cases}$$

il totale di x'_{ab} nella popolazione equivale al totale di x_{ab} nel dominio d' , in simboli:

$$X'_{ab} = \sum_{j=1}^N x'_{jab} \equiv \sum_{j=1}^{N_{d'}} x_{jab} = {}_d X_{ab}$$

e, pertanto, stimare il totale di popolazione X'_{ab} equivale a stimare ${}_d X_{ab}$

La trasformazione di un parametro di dominio trasversale in parametro di popolazione permette di sfruttare le formule illustrate nel caso di stime riferite a domini stratificati; tuttavia, essa comporta un aumento nella variabilità delle stime riferite al dominio

41 È bene sottolineare che le espressioni degli stimatori qui riportati fanno riferimento ad un campione casuale semplice di fogli di mappa e a uno stimatore di Horvitz-Thompson. In realtà, sia il disegno di campionamento sia lo stimatore sono diversi da quelli qui presi in considerazione; la giustificazione a questa semplificazione è dovuta solo al fatto che in questo paragrafo si vuole focalizzare l'attenzione sul modello di stima adottato e quindi non si è voluto appesantire la simbologia relativa alla strategia campionaria. Per una descrizione dettagliata del disegno di campionamento si veda il Paragrafo 2.4.3; mentre per una trattazione completa del procedimento di stima campionaria si rimanda al Paragrafo 2.4.4.

42 Vedi nota 39.

43 Cicchitelli *et al.*, 1992.



stesso, e quindi una perdita di efficienza, dovuta alla circostanza che la dimensione campionaria nel dominio è una variabile casuale.

Le stime dell'errore di copertura e di sotto-copertura per componente

Oltre a fornire le stime dei tassi di copertura e di sotto-copertura per i domini di interesse, l'indagine di copertura in oggetto ha anche prodotto i suddetti tassi per *componente*⁴⁴ che ha generato i due fenomeni. In particolare, dopo la fase di *linkage*, ogni azienda agricola rilevata con l'indagine di copertura è caratterizzata dal possesso di una modalità della variabile presa in considerazione ed esplicativa del risultato del procedimento di abbinamento tra l'azienda stessa e la lista di aziende rilevate con il Censimento. Questa variabile, che per comodità è stata indicata con il nome "*componente*" e denotata con la lettera *Y*, presenta modalità che esprimono i motivi per cui una azienda è stata "coperta" o "sottocoperta" al Censimento, ossia:

$Y =$	{	$1C$ Azienda rilevata al censimento
		DE Demografia d'azienda agricola
		LM Azienda che presenta legami multipli con il censimento
		0 Azienda non presente in alcun archivio
		IL Azienda trovata in un archivio (esclusa la lista precensuaria)
		MR Azienda in lista precensuaria (non rispondente o non esistente al censimento)
		NU Azienda in lista precensuaria (non eleggibile al censimento)

In particolare, le prime tre modalità (*1C*, *DE*, *LM*) caratterizzano una azienda che è stata rilevata sia con l'indagine di copertura che con il Censimento; mentre le ultime quattro modalità (*0*, *IL*, *MR*, *NU*) contraddistinguono una azienda che è stata rilevata con l'indagine di copertura ma non con il Censimento.

Ciò premesso, con riferimento al generico dominio ($d=1, \dots, D$), le stime dei tassi di copertura e di sotto-copertura per *componente* si ricavano semplicemente scomponendo le espressioni (4) e (5) secondo le modalità della variabile medesima; in simboli, si ha:

$${}^d\tilde{t} = \frac{{}^d\hat{X}_{11}}{{}^d\hat{X}_{+1}} = \frac{\sum_Y {}^Y\hat{X}_{11}}{\sum_Y {}^Y\hat{X}_{+1}} = \frac{({}^1\hat{X}_{11} + {}^{DE}\hat{X}_{11} + {}^{LM}\hat{X}_{11})}{{}^d\hat{X}_{+1}} = \frac{{}^1\hat{X}_{11}}{{}^d\hat{X}_{+1}} + \frac{{}^{DE}\hat{X}_{11}}{{}^d\hat{X}_{+1}} + \frac{{}^{LM}\hat{X}_{11}}{{}^d\hat{X}_{+1}} = {}^1\tilde{t} + {}^{DE}\tilde{t} + {}^{LM}\tilde{t} = \sum_Y {}^Y\tilde{t} \quad (7)$$

$${}^d\tilde{s} = \frac{{}^d\hat{X}_{21}}{{}^d\hat{X}_{+1}} = \frac{\sum_Y {}^Y\hat{X}_{21}}{\sum_Y {}^Y\hat{X}_{+1}} = \frac{({}^0\hat{X}_{21} + {}^{IL}\hat{X}_{21} + {}^{MR}\hat{X}_{21} + {}^{NU}\hat{X}_{21})}{{}^d\hat{X}_{+1}} = \frac{{}^0\hat{X}_{21}}{{}^d\hat{X}_{+1}} + \frac{{}^{IL}\hat{X}_{21}}{{}^d\hat{X}_{+1}} + \frac{{}^{MR}\hat{X}_{21}}{{}^d\hat{X}_{+1}} + \frac{{}^{NU}\hat{X}_{21}}{{}^d\hat{X}_{+1}} = {}^0\tilde{s} + {}^{IL}\tilde{s} + {}^{MR}\tilde{s} + {}^{NU}\tilde{s} = \sum_Y {}^Y\tilde{s} \quad (8)$$

in cui $\frac{{}^1\hat{X}_{11}}{{}^d\hat{X}_{+1}} = \sum_{i \in d} {}^1X_{i1} k_i$, $\frac{{}^{DE}\hat{X}_{11}}{{}^d\hat{X}_{+1}} = \sum_{i \in d} {}^{DE}X_{i1} k_i$, $\frac{{}^{LM}\hat{X}_{11}}{{}^d\hat{X}_{+1}} = \sum_{i \in d} {}^{LM}X_{i1} k_i$, $\frac{{}^0\hat{X}_{21}}{{}^d\hat{X}_{+1}} = \sum_{i \in d} {}^0X_{i21} k_i$
 $\frac{{}^{IL}\hat{X}_{21}}{{}^d\hat{X}_{+1}} = \sum_{i \in d} {}^{IL}X_{i21} k_i$, $\frac{{}^{MR}\hat{X}_{21}}{{}^d\hat{X}_{+1}} = \sum_{i \in d} {}^{MR}X_{i21} k_i$ e $\frac{{}^{NU}\hat{X}_{21}}{{}^d\hat{X}_{+1}} = \sum_{i \in d} {}^{NU}X_{i21} k_i$

sono rispettivamente le stime campionarie⁴⁵ delle quantità:

44 Si veda il Paragrafo 2.7.5 per una trattazione più dettagliata.

45 Vale l'osservazione riportata nella nota 41 per quanto riguarda il motivo che ci ha fatto propendere per una semplificazione nella formulazione degli stimatori testé citati. Inoltre, occorre far presente che lo stimatore preso in considerazione per il computo delle stime per componente $\frac{{}^Y\hat{X}_{11}}{{}^d\hat{X}_{+1}}$, $\frac{{}^{DE}\hat{X}_{11}}{{}^d\hat{X}_{+1}}$, $\frac{{}^{LM}\hat{X}_{11}}{{}^d\hat{X}_{+1}}$, $\frac{{}^0\hat{X}_{21}}{{}^d\hat{X}_{+1}}$, $\frac{{}^{IL}\hat{X}_{21}}{{}^d\hat{X}_{+1}}$, $\frac{{}^{MR}\hat{X}_{21}}{{}^d\hat{X}_{+1}}$ e $\frac{{}^{NU}\hat{X}_{21}}{{}^d\hat{X}_{+1}}$ è dello stesso tipo di quello illustrato approfonditamente nel Paragrafo 2.4.4 con riferimento a stime non per componente.

${}^1 X_{11}, {}^{DE} X_{11}, {}^{LM} X_{11}, {}^0 X_{21}, {}^{IL} X_{21}, {}^{MR} X_{21}$

e ${}^{NU} X_{21}$ riferite al dominio d , essendo k_i il peso campionario finale attribuito al foglio i di mappa incluso nel campione e ${}^Y X_{ia1}$ ($a = 1, 2$) il numero di aziende agricole appartenenti alla cella a_i ($a = 1, 2$) caratterizzate da una certa modalità della variabile Y .

2.4.3 Il disegno di campionamento

2.4.3.1 Caratteristiche generali

Il disegno di campionamento scelto si inquadra nell'approccio generale del campionamento indiretto, cui si ricorre quando non esiste una lista riferita alle unità della popolazione di interesse e si utilizza pertanto una lista inerente a una popolazione a essa collegata⁴⁶. L'applicabilità di tale metodologia è basata sulla possibilità di conoscere i legami (o link) tra le due popolazioni, in modo da poter calcolare i pesi campionari utilizzando il metodo noto come *Weight Sharing Method* (WSM, in seguito)⁴⁷. Nello specifico contesto di una indagine campionaria su aziende agricole, i link tra le unità appartenenti all'archivio di selezione - costituito dalla lista dei fogli di mappa del Catasto dei terreni - e le unità della popolazione target - costituita dalle aziende agricole presenti sul territorio nazionale - sono rappresentati dalla presenza di ciascuna azienda (nella sua interezza o in parti della stessa) in uno o più fogli di mappa (o particelle). Per l'applicazione del WSM in modo corretto sarà, pertanto, necessario conoscere per ciascuna azienda rilevata il numero totale dei fogli di mappa (o particelle) in cui essa ricade. A tale scopo, questa informazione è stata rilevata per ciascuna azienda agricola intervistata.

Per la formazione del campione di fogli di mappa (o particelle) è stato adottato un disegno di campionamento di tipo probabilistico a due stadi di selezione, con stratificazione delle unità di primo stadio. Il primo stadio di campionamento è costituito dai comuni che vengono stratificati sulla base della provincia di appartenenza, mentre le unità di secondo stadio sono costituite dai fogli di mappa del Catasto dei terreni, che costituisce pertanto l'archivio di selezione del campione. Nella fase di rilevazione devono essere individuate le aziende agricole presenti nel territorio delimitato da ciascun foglio di mappa del campione selezionato.

Le ragioni che sottendono la scelta di questo disegno sono diverse: necessità, o opportunità, di tipo organizzativo; riduzione dei costi e tempi tecnici di esecuzione dell'indagine. Infatti, l'introduzione dei comuni come primo stadio di campionamento consente di tenere sotto controllo il numero dei comuni coinvolti nell'indagine; il disegno alternativo consistente nella formazione di un campione di fogli di mappa catastali avrebbe comportato un aumento del numero di comuni coinvolti e, di conseguenza, un aumento dei costi dell'indagine.

Fissata, sulla base essenzialmente di vincoli di costo, la dimensione campionaria in termini di numero di aziende da intervistare a livello nazionale e la sua allocazione tra le regioni e le province, il numero di comuni campione in ogni provincia è stato definito tenendo conto del numero medio di aziende agricole⁴⁸ per foglio di mappa all'interno della provincia.

⁴⁶ Lavallée, 2007; Lavallée, Rivest, 2012.

⁴⁷ Cfr. Paragrafo 2.4.4.

⁴⁸ Il numero medio è stato ottenuto sulla base delle numerosità dei fogli di mappa dell'archivio del Catasto dei terreni e del numero di unità agricole presenti in Agea 2009.



I comuni campione sono stati selezionati all'interno degli strati provinciali, con probabilità proporzionale alla loro dimensione in termini di numero di aziende presenti nella lista pre-censuaria e senza re-immissione.

Il numero di fogli di mappa da selezionare in ciascun comune campione è stato determinato in modo tale da garantire la condizione di auto-ponderazione del campione di aziende all'interno di ciascuna provincia (ossia garantire che i fogli di mappa campione abbiano la stessa probabilità d'inclusione all'interno di ciascuna provincia), la selezione, inoltre, è senza re-immissione.

2.4.3.2 La determinazione della numerosità campionaria e allocazione del campione tra gli strati

La dimensione del campione in termini di unità finali è stabilita a livello nazionale secondo un criterio di costo. Il numero totale di aziende da intervistare è stato determinato dividendo le risorse monetarie disponibili per la rilevazione sul campo per il costo medio previsto per ciascuna intervista. Il numero di fogli di mappa campione è stato quindi ottenuto dividendo il numero totale di aziende da intervistare per il numero medio di aziende per foglio di mappa.

In base al budget disponibile per l'indagine di copertura, il numero di aziende da intervistare è stato fissato pari a circa 27.900. Inoltre, dalle analisi effettuate sulla base dell'archivio del Catasto dei terreni e dell'Agea 2009 (Agenzia per le erogazioni in agricoltura), è risultato che a livello nazionale il numero medio teorico di aziende per foglio di mappa è pari a 18 e quindi la dimensione campionaria complessiva in termini di fogli di mappa è risultata approssimativamente pari a 1.500.

Per l'allocazione⁴⁹ della dimensione campionaria complessiva tra i domini di studio sub-nazionali (le 19 regioni e le due Province autonome di Trento e Bolzano) occorre osservare che l'indagine in esame ha il duplice obiettivo di produrre stime sia nazionali che regionali. In tali circostanze, per allocare il campione occorre trovare un compromesso ragionevole tra le due allocazioni ottimali corrispondenti ai due obiettivi separatamente considerati: da una parte, le stime nazionali e, dall'altra, le stime regionali. Infatti, se l'indagine avesse soltanto l'obiettivo di determinare una stima dei tassi a livello nazionale, sarebbe opportuno un criterio di allocazione che attribuisca alle regioni dimensioni campionarie approssimativamente proporzionali al numero di unità esistenti nelle regioni medesime. Se, invece, l'indagine fosse mirata alla sola produzione di stime regionali, l'allocazione preferibile, sotto il vincolo che le stime siano caratterizzate dallo stesso livello atteso di precisione, sarebbe quella di attribuire alle varie regioni dimensioni campionarie approssimativamente uguali.

Per l'indagine di copertura è stata trovata una soluzione di compromesso, in base alla quale le dimensioni campionarie espresse in termini di numero di aziende da intervistare a livello regionale sono state definite come combinazioni lineari convesse delle numerosità relative alle due allocazioni ottimali. Per la parte dell'allocazione proporzionale è stata utilizzata come variabile di dimensionamento il numero di aziende agricole presenti nelle regioni. Definito il numero di aziende da intervistare in ogni regione e provincia autonoma, si è proceduto con la sua allocazione tra le province in modo proporzionale al numero di aziende ivi presenti.

⁴⁹ Si è seguita la metodologia qui descritta poiché non si disponeva di informazioni circa la variabilità dei fenomeni oggetto di indagine utili per definire l'allocazione del campione complessivo tra i domini di stima sulla base degli errori campionari attesi.

2.4.3.3 La formazione del campione di primo stadio

Al fine di introdurre ulteriori guadagni di efficienza (a parte quelli dovuti all'utilizzo di probabilità di selezione variabili e di stimatori speciali) nelle stime uscenti dall'indagine in discussione, si è proceduto alla stratificazione delle unità primarie. Per quanto concerne, quindi, il primo stadio di campionamento, i comuni sono stati ripartiti in H ($h = 1, \dots, H$) strati definiti dalle province. All'interno di ogni provincia i comuni sono stati selezionati con probabilità proporzionale alla loro dimensione definita in termini di numero di aziende agricole ivi presenti. Da ciascun comune campione è stato selezionato un campione di fogli di mappa a probabilità uguali.

Al fine di conferire maggior chiarezza ai successivi sviluppi algebrici, conviene introdurre le seguenti notazioni simboliche.

Con riferimento al generico strato dell'indagine h ($h = 1, \dots, H$), si denotino con c e i rispettivamente gli indici di comune e foglio di mappa⁵⁰. Inoltre, N_{hci} è il numero di aziende agricole presenti nel generico foglio di mappa i del comune c , N_{hc} e L_{hc} indicano rispettivamente il numero di aziende agricole presenti e di fogli di mappa nel generico comune c , mentre N_h e L_h indicano le analoghe quantità marginali riferite ad h ; infine, I_{hc} rappresenta il numero di fogli di mappa campione nel comune campione c . Siano ancora m_h e n_h rispettivamente il numero di comuni campione e il numero pianificato di aziende agricole campione nello strato h , e:

$$n = \sum_{h=1}^H n_h$$

il numero pianificato di aziende agricole campione a livello totale.

Tenendo presente quanto sopra premesso, si riassume ora la procedura operativa di formazione del campione di primo stadio.

Determinate le numerosità n e n_h ($h = 1, \dots, H$), la procedura in questione si articola nei seguenti passi:

a) si definisce il numero di comuni campione, m_h , per ogni h ($h = 1, \dots, H$), in modo da rispettare la condizione che il valore atteso del numero di aziende agricole campione relativo ad h sia uguale al numero programmato; in simboli:

$$\sum_{c \in h} \sum_{i \in c} N_{hci} \pi_{hci} = n_h \quad (9)$$

in cui: $\pi_{hci} = (m_h N_{hc} / N_h) (I_{hc} / L_{hc})$ è la probabilità di inclusione del foglio di mappa i appartenente al comune c ; $I_{hc} = I_h = n_h^* / (N_h / L_h)$ è il numero minimo di fogli di mappa campione per comune campione⁵¹, determinato attraverso una procedura di tipo iterativo, facendo variare il numero minimo di aziende agricole da intervistare in ciascun comune campione n_h^* , in modo da rispettare il numero complessivo di comuni che si intende coinvolgere nell'indagine; N_h / L_h è il numero medio di aziende agricole per foglio di mappa riferito a h ; inoltre, si ha $\pi_{hci} = \pi_{hc}$, poiché tutti i fogli di mappa del comune c hanno la stessa probabilità di far parte del campione.

50 Al fine di non appesantire la notazione introdotta, si adotta qui e nel seguito la convenzione che laddove un indice è assente, la quantità deve essere intesa come marginale rispetto a quell'indice.

51 L'uguaglianza $I_{hc} = I_h$ è dovuta alla circostanza che il numero minimo di fogli di mappa campione per comune campione è fisso nello strato dell'indagine h .



Risolvendo l'equazione (9) rispetto all'incognita m_h , si ottiene:

$$m_h = \frac{n_h}{J_h} \left(\frac{1}{N_h} \sum_{c \in h} N_{hc} \frac{1}{L_{hc}} \sum_{i \in c} N_{hci} \right)^{-1}$$

Preme sottolineare che questo passo è effettuato con lo scopo di avere una prima valutazione sul numero di comuni coinvolti nell'indagine. Inoltre, è necessario specificare che il numero di aziende agricole per foglio di mappa, N_{hci} , si riferisce ad aziende univoche all'interno del foglio di mappa i , ma se sommato per comune può contenere duplicazioni nel caso in cui in quel comune ci siano aziende presenti su più fogli di mappa. Nel prosieguo, quando si farà riferimento a numerosità riferite ad aziende agricole non univoche (poiché ottenute sommando i totali per foglio di mappa) si apporrà una diresi sul simbolo; ad esempio, la somma del numero di aziende agricole dei fogli di mappa del comune c sarà indicato con il simbolo N_{hc} ;

b) si attua il processo di selezione dei comuni con probabilità proporzionale alla dimensione degli stessi espressa in termini di aziende agricole ivi presenti:

$$\pi_{hc} = m_h \frac{N_{hc}}{N_h}$$

I comuni con una probabilità di inclusione maggiore o uguale a uno sono definiti *Auto Rappresentativi* (o AR); i restanti comuni sono *Non Auto Rappresentativi* (o NAR);

c) definito il numero di comuni campione AR in ciascuno strato, ${}_{AR}m_h$, si selezionano⁵² senza reimmissione ${}_{NAR}m_h = m_h - {}_{AR}m_h$ comuni campione NAR con probabilità di selezione ${}_{NAR}z_{hc} = N_{hc} / {}_{NAR}N_h$, in cui ${}_{NAR}N_h$ è il numero di aziende agricole appartenenti ai comuni NAR dello strato. I comuni AR, invece, entrano con certezza⁵³ nel campione.

Dunque, alla luce di quanto sopra introdotto, si può concludere dicendo che:

- i) il disegno di campionamento inerente alla parte AR è del tipo ad uno stadio stratificato in cui le unità di campionamento sono costituite dai fogli di mappa catastali, che possono essere definiti come grappoli di aziende agricole;
- ii) il disegno di campionamento relativo alla parte NAR è del tipo a due stadi stratificato in cui le unità primarie sono costituite dai comuni e le unità secondarie dai fogli di mappa catastali.

2.4.3.4 La formazione del campione di secondo stadio

Come premesso nel Paragrafo 2.4.2, il secondo stadio di campionamento è costituito dai fogli di mappa del Catasto dei terreni che rappresentano, pertanto, le unità di secondo stadio.

La selezione dei fogli di mappa è stata effettuata a probabilità uguali e senza re-immisione; pertanto, la probabilità di inclusione del generico foglio di mappa i è data da:

⁵² Per l'estrazione delle unità primarie si è utilizzata la procedura di selezione sistematica, suggerita da Madow (1949) e Murthy (1967), che presenta le seguenti caratteristiche: (i) assegna una probabilità di inclusione espressa nella forma $\pi_{hc} = {}_{NAR}m_h / {}_{NAR}z_{hc}$; (ii) la sua implementazione è estremamente semplice; (iii) permette di ottenere stime generalmente più efficienti rispetto a quelle che si ottengono con altre procedure di selezione (Cicchitelli *et al.*, 1992; Fabbri, 1991).

⁵³ Negli strati AR si ha ${}_{AR}m_h$ e ${}_{AR}z_{hc} = 1$.

$$\pi_{hci} = I_{hc} \frac{1}{L_{hc}} \quad \text{per i comuni AR} \quad (10)$$

$$\pi_{i|hc} = I_{hc} \frac{1}{L_{hc}} \quad \text{per i comuni NAR.}$$

Per i comuni NAR la probabilità di inclusione del foglio di mappa i appartenente al comune c è condizionata alla circostanza che il comune c sia stato incluso nel campione di primo stadio.

Si indichi con $F_h = n_h / N_h$ la frazione di campionamento riferita al generico strato (la provincia) dell'indagine h ; per ognuno dei comuni campione si definisce il numero di fogli di mappa da selezionare al secondo stadio imponendo la condizione di autoponderazione a livello di strato⁵⁴ (la provincia):

$$\pi_{hci} = F_h$$

da cui, sostituendo π_{hci} e F_h con le rispettive espressioni, si ottiene

$$m_h \frac{N_{hc}}{N_h} \frac{I_{hc}}{L_{hc}} = \frac{n_h}{N_h} \quad \text{per i comuni NAR} \quad (11)$$

$$\frac{I_{hc}}{L_{hc}} = \frac{n_h}{N_h} \quad \text{per i comuni AR.} \quad (12)$$

Risolviendo la (11) e la (12) rispetto ad I_{hc} (che è incognita) si ricava

$$I_{hc} = \frac{n_h}{N_h} \frac{1}{m_h} \frac{N_h}{N_{hc}} L_{hc} \quad \text{per i comuni NAR}$$

$$I_{hc} = \frac{n_h}{N_h} L_{hc} \quad \text{per i comuni AR.}$$

La probabilità di inclusione finale, ossia comprensiva della probabilità di inclusione del comune c al quale afferisce il foglio di mappa i , è dunque uguale alla (10) per i fogli di mappa appartenenti ai comuni AR, mentre è pari a:

$$\pi_{hci} = \pi_{hc} \pi_{i|hc} = m_h \frac{N_{hc}}{N_h} \frac{I_{hc}}{L_{hc}} \quad (13)$$

per i fogli di mappa che ricadono in comuni NAR.

⁵⁴ Si sottolinea che, poiché l'allocazione del numero di aziende da intervistare per provincia è stata effettuata in modo proporzionale alla dimensione della provincia in termini di numero di aziende ivi presenti, ne consegue che la condizione di autoponderazione è verificata anche per regione geografica.



2.4.4 La procedura di stima campionaria

Il calcolo dei pesi campionari da assegnare alle aziende intervistate è stato effettuato seguendo l'approccio del WSM⁵⁵, che costituisce il metodo di stima utilizzato per attribuire i pesi alle unità di un campione selezionato mediante campionamento indiretto. Nell'indagine in esame, infatti, le unità campionarie selezionate sono i fogli di mappa e le interviste sono effettuate a tutte le aziende agricole che hanno parte dei loro terreni (o allevamenti) ricadente nel foglio di mappa campione. Per attribuire un peso campionario alle aziende intervistate (quindi raggiunte indirettamente attraverso i fogli di mappa campione) si è proceduto considerando le probabilità di inclusione dei fogli campione e sfruttando l'informazione relativa al numero complessivo di fogli di mappa in cui ricade ciascuna azienda intervistata, rilevato nel corso dell'intervista. Tale informazione è fondamentale per compiere l'operazione di frazionamento del peso dell'azienda, che consente di ottenere pesi campionari che producono una stima corretta del numero di aziende⁵⁶.

La scelta di seguire l'approccio del campionamento indiretto e del WSM è stata preferita alla strategia di stima alternativa, che avrebbe previsto di assegnare un peso campionario solamente alle aziende il cui centro aziendale ricade in un foglio di mappa campione, escludendo quindi tutte quelle aventi nel foglio di mappa una quota parte di terreni (particelle) ma non il centro aziendale. In tal caso la strategia campionaria seguita sarebbe consistita in un disegno campionario standard avente il centro aziendale come unità finale che riceve in modo diretto il peso campionario del foglio di mappa campione in cui ricade⁵⁷.

La scelta della procedura di stima basata sul WSM ha consentito di sfruttare al meglio le potenzialità del disegno campionario e della rilevazione effettuata, dal momento che tutte le interviste effettuate vengono sfruttate per stimare il numero di aziende.

Il peso attribuito a ciascuna azienda rispondente è stato calcolato in passi successivi. Inizialmente è stato calcolato il peso diretto teorico di ogni foglio di mappa campione come inverso della probabilità di inclusione (10) o (13) a seconda che il foglio di mappa ricada in un comune AR o NAR; in simboli: $k_{hci} = 1/\pi_{hci}$. Successivamente, poiché in fase di rilevazione si è verificata una mancata risposta di primo stadio da parte di alcuni comuni campione, è stato necessario correggere i pesi diretti teorici dei fogli di mappa per tenere conto di questa mancata rilevazione. È stato, pertanto, necessario calcolare un coefficiente correttivo della caduta comunale all'interno di ogni provincia, ottenuto come rapporto tra il numero di comuni selezionati nella provincia h , m_h , e il numero di comuni rispondenti sempre nella provincia h , m'_h . Tale fattore correttivo della mancata risposta di primo stadio è stato poi moltiplicato per il peso diretto teorico k_{hci} , ottenendo il peso campionario corretto per mancata risposta relativo al foglio i :

$$k'_{hci} = k_{hci} \frac{m_h}{m'_h} \quad (14)$$

Inoltre, poiché si è verificata una caduta di tutti i comuni campione della provincia di Pisa (che quindi non sono stati rilevati), solamente per la Regione Toscana è stata ef-

55 Lavallée, 2007.

56 Deville, Lavallée, 2006.

57 In realtà anche questo caso si sarebbe potuto inquadrare nell'ambito generale del campionamento indiretto, dal momento che si sarebbe utilizzata una lista di campionamento non coincidente con la popolazione di interesse, ma della quale ne avrebbe costituito il caso più elementare, in cui il numero di link tra le unità target e le unità della lista sarebbe stato al più pari a uno.

fettuata una post-stratificazione rispetto al totale dei fogli di mappa della regione presenti nell'archivio di selezione L_R . Moltiplicando tale fattore di post-stratificazione per il peso dato dalla (14), si è ottenuto il peso corretto post-stratificato:

$$k_{hci}^* = k_{hci}' \frac{L_R}{\sum_{h \in R} \sum_{c=1}^{m_h} \sum_{i=1}^{I_{hc}} k_{hci}'}$$

Per i fogli di mappa appartenenti a tutte le altre regioni si è posto $k_{hci}^* = k_{hci}'$. Ogni azienda j rilevata nel foglio di mappa campione i eredita il peso associato ad i ; poiché però il territorio di una azienda può ricadere in più fogli di mappa campione, e quindi essere intervistata più di una volta, il peso di ciascuna intervista relativa alla stessa azienda è stato frazionato per il numero di fogli di mappa in cui ricade parte dell'azienda stessa; il peso finale associato ad ogni intervista sarà pertanto:

$$k_{hcij} = \frac{k_{hci}^*}{L_j} \quad (15)$$

essendo $L_j = \sum_{i=1}^L I_{ij}$, in cui I_{ij} è una variabile indicatrice che rappresenta il legame tra il foglio di mappa i e l'azienda j ed è pari a 1 se l'azienda (nella sua interezza o in parte del suo territorio) j ricade nel foglio di mappa i e 0 altrimenti. Tale tecnica è nota in letteratura con il nome di *Weight Sharing Method* ed è usualmente adottata nel contesto del campionamento indiretto.

Il peso finale espresso dalla (15) è stato utilizzato per ottenere, in un'ottica di stima diretta del tipo Horvitz-Thompson, tutte le stime campionarie componenti le stime da modello dei tassi di copertura e sotto-copertura descritte nel Paragrafo 2.4.2.

2.4.5 L'attendibilità delle stime del tasso di copertura e di sotto-copertura

La valutazione degli errori relativi è stata effettuata, per i diversi domini di stima, relativamente alle stime del totale delle aziende agricole, del tasso di copertura e di sotto-copertura - questi ultimi due sia globalmente che distintamente per *componente* - espressi dalle (4), (5), (6), (7) e (8).

Al riguardo, si sottolinea che la variabilità delle suddette stime presenta due componenti: una generata dal modello e l'altra dovuta al disegno campionario. Infatti, adottando la ben nota scomposizione⁵⁸ della varianza, è possibile esprimere la varianza totale come somma di due addendi

$$V = E_p V_m + V_p E_m$$

in cui E_m e V_m denotano rispettivamente gli operatori di valore atteso e di varianza sotto il modello specificato nel Paragrafo 2.4.2, mentre E_p e V_p indicano i medesimi operatori sotto il disegno di campionamento adottato per l'indagine in oggetto. Poiché è possibile dimostrare che la varianza da modello di \tilde{N} è trascurabile rispetto

58 Cicchitelli et al., 1992.



alla sua varianza da disegno⁵⁹, in quanto x_{+I} e x_{I+} in un contesto censuario sono quantità molto grandi, l'analisi è stata eseguita con riferimento esclusivamente a quest'ultima. Data la natura non lineare degli stimatori dei parametri d'interesse, per stimare la varianza da disegno delle stime prodotte relativamente ai vari domini di stima su citati, è consuetudine calcolare una sua approssimazione. Essa è fornita dalla stima della varianza della forma linearizzata degli stimatori sopra definiti ricorrendo al *metodo della linearizzazione*, che consiste nell'applicare lo sviluppo in serie di Taylor degli stessi.

Per ciò che riguarda lo stimatore del numero totale di aziende agricole, sviluppando ${}_{d}\tilde{N} = {}_{d}\tilde{M}({}_{d}\hat{x}_{+1}, {}_{d}\hat{x}_{11})$ in serie di Taylor nell'intorno del punto di coordinate $({}_{d}x_{+1}, {}_{d}x_{11})$ e limitando lo sviluppo ai termini di ordine lineare, si ottiene⁶⁰:

$${}_{d}\tilde{N} \cong {}_{d}\hat{N} + \frac{{}_{d}x_{+1}}{{}_{d}x_{11}} {}_{d}\hat{x}_{+1} - \frac{{}_{d}\hat{N}}{{}_{d}x_{11}} {}_{d}\hat{x}_{11}.$$

Da cui, applicando l'operatore V_p , si ha⁶¹:

$$\begin{aligned} V_p\{ {}_{d}\tilde{N} \} &\cong V_p\{ {}_{d}\hat{N} \} + {}_{d}\hat{N}^2 \left(\frac{V_p\{ {}_{d}\hat{x}_{+1} \}}{x_{+1}^2} + \frac{V_p\{ {}_{d}\hat{x}_{11} \}}{x_{11}^2} - 2 \frac{C_p\{ {}_{d}\hat{x}_{+1}, {}_{d}\hat{x}_{11} \}}{{}_{d}x_{+1} {}_{d}x_{11}} \right) = \\ &= {}_{d}\hat{N}^2 \left(\frac{V_p\{ {}_{d}\hat{x}_{+1} \}}{x_{+1}^2} + \frac{V_p\{ {}_{d}\hat{x}_{11} \}}{x_{11}^2} - 2 \frac{C_p\{ {}_{d}\hat{x}_{+1}, {}_{d}\hat{x}_{11} \}}{{}_{d}x_{+1} {}_{d}x_{11}} \right), \end{aligned} \quad (16)$$

dove il simbolo C_p indica l'operatore di covarianza sotto il disegno di campionamento adottato e l'ultima uguaglianza è dovuta alla circostanza che $V_p\{ {}_{d}\hat{N} \} = 0$ poiché ${}_{d}\hat{N}$ non è affetto da errore campionario.

Una stima della (16) è fornita, pertanto, dalla relazione:

$$V_p\{ {}_{d}\tilde{N} \} \cong {}_{d}\tilde{N}^2 \left(\frac{V_p\{ {}_{d}\hat{x}_{+1} \}}{\hat{x}_{+1}^2} + \frac{V_p\{ {}_{d}\hat{x}_{11} \}}{\hat{x}_{11}^2} - 2 \frac{C_p\{ {}_{d}\hat{x}_{+1}, {}_{d}\hat{x}_{11} \}}{\hat{x}_{+1} \hat{x}_{11}} \right). \quad (17)$$

Ai fini del calcolo delle quantità $V_p\{ {}_{d}\hat{x}_{+1} \}$, $V_p\{ {}_{d}\hat{x}_{11} \}$ e $C_p\{ {}_{d}\hat{x}_{+1}, {}_{d}\hat{x}_{11} \}$ coinvolte nella (17), si è ricorso alle formule degli stimatori della varianza e covarianza campionaria relative al caso di un disegno di campionamento a due stadi, con stratificazione delle unità di primo stadio e selezione delle stesse con re-immissione e probabilità proporzionali, e selezione con probabilità uguali e senza re-immissione delle unità di secondo stadio.

59 Rao, 2003.

60 Per gli sviluppi si rimanda all'Appendice 2.B.

61 Per una trattazione più dettagliata della varianza e della stima della varianza si veda l'Appendice 2.C.

Poiché in realtà le unità primarie sono state estratte senza re-immissione, tale approssimazione comporta, chiaramente, una distorsione⁶² dello stimatore, con conseguente inflazione della variabilità. La principale giustificazione per tale scelta è la difficoltà di calcolo delle probabilità d'inclusione di secondo ordine delle unità primarie, richieste per definire gli stimatori corretti della varianza e della covarianza per un disegno con selezione senza re-immissione.

Ciò premesso, gli stimatori di $V_p\{\hat{x}_{+1}\}$, $V_p\{\hat{x}_{11}\}$ e $C_p\{\hat{x}_{+1}, \hat{x}_{11}\}$ e sono dati rispettivamente da:

$$V_p\{\hat{x}_{+1}\} = \sum_{h=1}^{H_d} \frac{1}{m_h(m_h-1)} \sum_{c=1}^{m_h} (T_{hc,+1} - \hat{x}_{+1})^2 \quad (18)$$

$$V_p\{\hat{x}_{11}\} = \sum_{h=1}^{H_d} \frac{1}{m_h(m_h-1)} \sum_{c=1}^{m_h} (T_{hc,11} - \hat{x}_{11})^2 \quad (19)$$

$$C_p\{\hat{x}_{+1}, \hat{x}_{11}\} = \sum_{h=1}^{H_d} \frac{1}{m_h(m_h-1)} \sum_{c=1}^{m_h} (T_{hc,+1} - \hat{x}_{+1})(T_{hc,11} - \hat{x}_{11}), \quad (20)$$

in cui $h = 1, \dots, H_d$ indica il generico strato appartenente al dominio d , H_d è il numero complessivo degli strati appartenenti al dominio d , mentre $T_{hc,+1}$ e $T_{hc,11}$ sono dati rispettivamente da:

$$T_{hc,+1} = \sum_{i=1}^{I_{hc}} \sum_{j=1}^{N_{hcj}} X_{hcij,+1} \frac{1}{\frac{N_{hc}}{N_h} \frac{I_{hc}}{L_{hc}} L_j} = \sum_{i=1}^{I_{hc}} \sum_{j=1}^{N_{hcj}} X_{hcij,+1} \frac{N_h}{N_{hc}} \frac{L_{hc}}{I_{hc}} \frac{1}{L_j}.$$

Analogamente si ottiene la varianza di $\tilde{d}\mathcal{T}$ e $\tilde{d}\mathcal{S}$. Infatti, sviluppando $\tilde{d}\mathcal{T} = \tilde{d}\mathcal{T}(\hat{x}_{+1}, \hat{x}_{11})$

in serie di Taylor nell'intorno del punto di coordinate $(\hat{x}_{+1}, \hat{x}_{11})$ e considerando solo i termini di ordine lineare, si ha⁶³:

$$\tilde{d}\mathcal{T} \cong \frac{\partial \mathcal{T}}{\partial x_{+1}} \hat{x}_{+1} - \frac{\partial \mathcal{T}}{\partial x_{11}^2} \hat{x}_{+1}^2 + \frac{\partial \mathcal{T}}{\partial x_{11}} \hat{x}_{11}$$

e, quindi, l'approssimazione lineare locale della varianza da disegno della (4) è data da⁶⁴:

$$V_p\{\tilde{d}\mathcal{T}\} \cong \frac{\partial \mathcal{T}}{\partial x_{+1}^2} V_p\{\hat{x}_{+1}\} + \frac{\partial \mathcal{T}}{\partial x_{11}^2} V_p\{\hat{x}_{11}\} - 2 \frac{\partial \mathcal{T}}{\partial x_{+1}} C_p\{\hat{x}_{+1}, \hat{x}_{11}\} \quad (21)$$

che può essere stimata dall'espressione:

$$V_p\{\tilde{d}\mathcal{T}\} \cong \frac{\partial \mathcal{T}}{\partial \hat{x}_{+1}^2} V_p\{\hat{x}_{+1}\} + \frac{\partial \mathcal{T}}{\partial \hat{x}_{11}^2} V_p\{\hat{x}_{11}\} - 2 \frac{\partial \mathcal{T}}{\partial \hat{x}_{+1}} C_p\{\hat{x}_{+1}, \hat{x}_{11}\}, \quad (22)$$

62 Ci preme sottolineare che la distorsione è trascurabile quando il tasso di campionamento all'interno degli strati è piccolo (Wolter, 1985).

63 Per gli sviluppi si rimanda all'Appendice 2.B.

64 Per una trattazione più dettagliata della varianza e della stima della varianza si veda l'Appendice 2.C.



in cui le quantità $V_p\{\hat{X}_{+1}\}$, $V_p\{\hat{X}_{11}\}$ e $C_p\{\hat{X}_{+1}, \hat{X}_{11}\}$ sono fornite dalle espressioni (18), (19) e (20).

Per ciò che concerne la varianza di ${}^Y\tilde{S}$, poiché il tasso di sotto-copertura è il complemento a 1 del tasso di copertura, ne consegue che la sua varianza è uguale a quella di ${}^Y\tilde{T}$, espressa dalla (21), in simboli:

$$V_p\{{}^Y\tilde{S}\} = V_p\{1 - {}^Y\tilde{T}\} = V_p\{{}^Y\tilde{T}\}.$$

Stesso discorso, chiaramente, vale per la sua stima, che quindi sarà fornita dalla (22). Seguendo lo stesso procedimento si ricava lo sviluppo in serie di Taylor limitato ai termini di ordine lineare dei tassi di copertura ${}^Y\tilde{T} = {}^Y\tilde{T}({}^Y\hat{X}_{+1}, {}^Y\hat{X}_{11})$ e sotto-copertura ${}^Y\tilde{S} = {}^Y\tilde{S}({}^Y\hat{X}_{+1}, {}^Y\hat{X}_{21})$ per *componente* nell'intorno dei punti di coordinate rispettivamente $({}^YX_{+1}, {}^YX_{11})$ e $({}^YX_{+1}, {}^YX_{21})$, in simboli⁶⁵:

$${}^Y\tilde{T} \cong \frac{{}^YX_{11}}{{}^YX_{+1}} - \frac{{}^YX_{11}^2}{{}^YX_{+1}^2} {}^Y\hat{X}_{+1} + \frac{{}^Y\hat{X}_{11}}{{}^YX_{+1}}$$

$${}^Y\tilde{S} \cong \frac{{}^YX_{21}}{{}^YX_{+1}} - \frac{{}^YX_{21}^2}{{}^YX_{+1}^2} {}^Y\hat{X}_{+1} + \frac{{}^Y\hat{X}_{21}}{{}^YX_{+1}}.$$

Da ciò segue che l'approssimazione lineare locale della varianza da disegno degli stimatori sopra citati è data da⁶⁶:

$$V_p\{{}^Y\tilde{T}\} \cong \frac{{}^Y\hat{X}_{11}^2}{{}^YX_{+1}^4} V_p\{{}^Y\hat{X}_{+1}\} + \frac{V_p\{{}^Y\hat{X}_{11}\}}{{}^YX_{+1}^2} - 2 \frac{{}^Y\hat{X}_{11}}{{}^YX_{+1}^3} C_p\{{}^Y\hat{X}_{+1}, {}^Y\hat{X}_{11}\} \quad (23)$$

$$V_p\{{}^Y\tilde{S}\} \cong \frac{{}^Y\hat{X}_{21}^2}{{}^YX_{+1}^4} V_p\{{}^Y\hat{X}_{+1}\} + \frac{V_p\{{}^Y\hat{X}_{21}\}}{{}^YX_{+1}^2} - 2 \frac{{}^Y\hat{X}_{21}}{{}^YX_{+1}^3} C_p\{{}^Y\hat{X}_{+1}, {}^Y\hat{X}_{21}\} \quad (24)$$

che possono essere stimate rispettivamente dalle espressioni:

$$V_p\{{}^Y\tilde{T}\} \cong \frac{{}^Y\hat{X}_{11}^2}{{}^Y\hat{X}_{+1}^4} V_p\{{}^Y\hat{X}_{+1}\} + \frac{V_p\{{}^Y\hat{X}_{11}\}}{{}^Y\hat{X}_{+1}^2} - 2 \frac{{}^Y\hat{X}_{11}}{{}^Y\hat{X}_{+1}^3} C_p\{{}^Y\hat{X}_{+1}, {}^Y\hat{X}_{11}\}$$

$$V_p\{{}^Y\tilde{S}\} \cong \frac{{}^Y\hat{X}_{21}^2}{{}^Y\hat{X}_{+1}^4} V_p\{{}^Y\hat{X}_{+1}\} + \frac{V_p\{{}^Y\hat{X}_{21}\}}{{}^Y\hat{X}_{+1}^2} - 2 \frac{{}^Y\hat{X}_{21}}{{}^Y\hat{X}_{+1}^3} C_p\{{}^Y\hat{X}_{+1}, {}^Y\hat{X}_{21}\},$$

65 Per gli sviluppi si rimanda all'Appendice 2.B.

66 Per una trattazione più dettagliata della varianza e della stima della varianza si veda l'Appendice 2.C.

in cui $V_p\{\hat{x}_{s1}\}$ è fornita dalla (18),

mentre le quantità $V_p\{\hat{x}_{11}\}$, $C_p\{\hat{x}_{s1}, \hat{x}_{11}\}$, $V_p\{\hat{x}_{21}\}$,

e $C_p\{\hat{x}_{s1}, \hat{x}_{21}\}$ sono date da:

$$V_p\{\hat{x}_{11}\} = \sum_{h=1}^{H_d} \frac{1}{m_h(m_h-1)} \sum_{c=1}^{m_h} ({}^Y T_{hc,11} - {}^Y \hat{x}_{11})^2$$

$$C_p\{\hat{x}_{s1}, \hat{x}_{11}\} = \sum_{h=1}^{H_d} \frac{1}{m_h(m_h-1)} \sum_{c=1}^{m_h} (T_{hc,s1} - {}^Y \hat{x}_{s1}) ({}^Y T_{hc,11} - {}^Y \hat{x}_{11})$$

$$V_p\{\hat{x}_{21}\} = \sum_{h=1}^{H_d} \frac{1}{m_h(m_h-1)} \sum_{c=1}^{m_h} ({}^Y T_{hc,21} - {}^Y \hat{x}_{21})^2$$

$$C_p\{\hat{x}_{s1}, \hat{x}_{21}\} = \sum_{h=1}^{H_d} \frac{1}{m_h(m_h-1)} \sum_{c=1}^{m_h} (T_{hc,s1} - {}^Y \hat{x}_{s1}) ({}^Y T_{hc,21} - {}^Y \hat{x}_{21}),$$

dove ${}^Y T_{hc,11}$ e ${}^Y T_{hc,21}$ sono dati rispettivamente da:

$${}^Y T_{hc,11} = \sum_{l=1}^{I_{hc}} \sum_{j=1}^{N_{hcl}} Y_{X_{hclj,11}} \frac{1}{\frac{N_{hc}}{N_h} \frac{I_{hc}}{L_{hc}} L_j} = \sum_{l=1}^{I_{hc}} \sum_{j=1}^{N_{hcl}} Y_{X_{hclj,11}} \frac{N_h}{N_{hc}} \frac{L_{hc}}{I_{hc}} \frac{1}{L_j}$$

$${}^Y T_{hc,21} = \sum_{l=1}^{I_{hc}} \sum_{j=1}^{N_{hcl}} Y_{X_{hclj,21}} \frac{1}{\frac{N_{hc}}{N_h} \frac{I_{hc}}{L_{hc}} L_j} = \sum_{l=1}^{I_{hc}} \sum_{j=1}^{N_{hcl}} Y_{X_{hclj,21}} \frac{N_h}{N_{hc}} \frac{L_{hc}}{I_{hc}} \frac{1}{L_j}$$

Gli errori di campionamento relativi inerenti alle stime del generico dominio d sono stati, quindi, calcolati secondo la ben nota formula:

$$CV({}_d Z) = \frac{\sqrt{V({}_d Z)}}{{}_d Z} \quad (25)$$

in cui ${}_d Z \equiv \{ {}_d \tilde{\tau}, {}_d \tilde{\zeta}, {}_d \tilde{\tau}, {}_d \tilde{\zeta} \}$. Inoltre, l'errore di campionamento assoluto, dato dal numeratore della (25), permette di costruire l'intervallo di confidenza sotto indicato, nel quale con un certo livello di fiducia è contenuto il parametro oggetto di stima:

$$[{}_d Z - \delta_{\alpha/2} \sqrt{V({}_d Z)}, {}_d Z + \delta_{\alpha/2} \sqrt{V({}_d Z)}]$$



dove $1 - \alpha$ (solitamente 0.95 o 0.99) è detto *coefficiente di confidenza* e $\delta_{\alpha/2}$ è la quantità tale che $\Pr(Z \geq \delta_{\alpha/2}) = \alpha/2$; per la variabile casuale normale se $\alpha = 0,05$ allora $\delta_{\alpha/2} = 1,96$.

2.5 La rilevazione

Ad ogni rilevatore è stato assegnato un numero di fogli di mappa campione da sottoporre a rilevazione. Il rilevatore aveva a disposizione quattro modelli e poteva condurre la rilevazione sul campo servendosi di un computer portatile con accesso a internet per compilarli direttamente attraverso il sistema SGR COP⁶⁷ oppure poteva stampare il modello da SGR COP (Allegato elettronico 2.A), compilarlo in forma cartacea e in seguito registrare i dati all'interno del sistema. I quattro modelli di rilevazione sono di seguito descritti.

2.5.1 I modelli per la rilevazione

Come già descritto nel Paragrafo 2.2.3, la rilevazione si è articolata in due passi principali, per ciascuno dei quali sono stati utilizzati dei modelli, come illustrato nel prospetto seguente.

Prospetto 2.2 - Schema delle operazioni

	Modelli coinvolti	Azioni	Soggetto attuatore
Passo 1	Istat CensAgrCop1 Istat CensAgrCop1_agg Istat CensAgrCop2	Individuazione delle aziende agricole	Rilevatore
Passo 2	Istat CensAgrCop2 Istat CensAgrCop3	Contatto e intervista al conduttore dell'azienda agricola	Rilevatore
	Monitoraggio di tutte le operazioni svolte dai rilevatori attraverso SGR COP		Istat

Nel Passo 1 - *Individuazione delle aziende agricole*, il rilevatore doveva contattare gli intestatari delle particelle incluse nel foglio di mappa campione e individuare i conduttori delle aziende agricole, attraverso l'utilizzo dei seguenti modelli: Istat CensAgrCop1, Istat CensAgrCop1_agg e Istat CensAgrCop2.

Modello Istat CensAgrCop1

Questo modello denominato anche "Elenco intestatari per particella catastale" contiene i nominativi di tutti gli intestatari di una stessa particella. Il corpo del modello è impostato secondo una struttura di righe e colonne. Ciascuna riga corrisponde a una particella e le colonne identificano sia le informazioni relative all'intestatario della particella sia le informazioni da acquisire durante la rilevazione. Il rilevatore doveva analizzare tutte le particelle elencate nel modello al fine di individuare il conduttore, nel caso in cui la particella fosse condotta da un'azienda agricola. Il modello è precompilato per quanto riguarda i codici delle particelle catastali e le informazioni anagrafiche relative all'intestatario di ciascuna particella catastale (cognome e nome; codice fiscale; comune, provincia e indirizzo; numero di telefono).

67 Cfr. Paragrafo 2.5.2.

Le informazioni da inserire nel modello sono:

- esito del contatto con l'intestatario della particella;
- presenza o assenza di un'azienda agricola sulla particella in esame;
- nominativo del conduttore dell'eventuale azienda esistente al 24 ottobre 2010.

Se il nominativo contattato corrispondeva effettivamente al conduttore dell'azienda agricola, allora il rilevatore procedeva alla trascrizione del nominativo nel modello Istat CensAgrCop2, altrimenti doveva chiedere all'intestatario della particella informazioni (cognome, nome, telefono e indirizzo) utili a contattare il presunto conduttore dell'azienda agricola, registrando queste informazioni nel modello di ausilio Istat CensAgrCop1_agg.

Modello **Istat CensAgrCop1_agg**

È un modello di lavoro ausiliario nel quale il rilevatore registrava tutte le operazioni necessarie per passare dal modello Istat CensAgrCop1 alla lista dei conduttori del modello Istat CensAgrCop2. I nominativi in esso elencati potevano essere: presunti conduttori di azienda agricola, nuovi intestatari di una particella, persone utili a fornire informazioni per arrivare al conduttore dell'azienda agricola. Per la sua compilazione, il rilevatore procedeva come per il modello Istat CensAgrCop1.

Modello **Istat CensAgrCop2**

Al termine della compilazione dei modelli precedenti questo modello Istat CensAgrCop2 risultava compilato automaticamente da SGR PES in ciascuna riga con l'elenco di tutti i conduttori di azienda agricola desunti dai contatti effettuati in questo Passo dai rilevatori. Analogamente al modello Istat CensAgrCop1, le informazioni contenute in ciascuna riga si riferiscono al:

- codice della particella presente nel modello Istat CensAgrCop1);
- conduttore (persona fisica, società o ente) che gestisce l'azienda (cognome e nome della persona fisica o denominazione della società o ente);
- codice fiscale;
- regione, provincia, comune e indirizzo di residenza del conduttore o della sede legale della società o ente);
- il numero telefonico.

Con questo modello il rilevatore contattava tutti i conduttori di azienda agricola in esso registrati per confermare ed eventualmente integrare le informazioni anagrafiche dei conduttori. Per ciascuno nominativo elencato la compilazione veniva completata dopo aver sottoposto il questionario di azienda agricola riportato nel modello Istat CensAgrCop3.

Il Passo 2 è caratterizzato dall'*Intervista ai conduttori*'. I modelli utilizzati in questo passo sono: Istat CensAgrCop2, già descritto sopra, e Istat CensAgrCop3.

Modello **Istat CensAgrCop 3**

Esso costituisce il *questionario d'azienda agricola*. Il rilevatore doveva compilare tale modello in ogni sua parte tramite intervista telefonica ai conduttori di azienda agricola, identificati nella fase precedente e riportati sul modello Istat CensAgrCop2 e facendo riferimento alla situazione in essere alla data del Censimento. In assenza del conduttore l'intervista poteva essere rivolta ad una persona informata sulle caratteristiche dell'azienda alla data censuaria. Il modello riportava in testata:

- il codice intervistatore, nonché le informazioni relative alla localizzazione dell'azienda (in termini di regione, provincia e comune in cui è situato il foglio di mappa campione);



- il foglio di mappa campione (identificato con il medesimo codice riportato sui modelli Istat CensAgrCop1 e Istat CensAgrCop2) o la particella (limitatamente ai comuni delle Province autonome di Trento e Bolzano) e il comune catastale.

Il concatenamento di tali informazioni, insieme alla particella e ad un progressivo all'interno della particella ha contribuito ad assegnare un codice, univoco, identificativo dell'azienda agricola rilevata. Il corpo del questionario è costituito invece da sette sezioni di seguito sinteticamente descritte.

Sezione 1 - *Caratteri identificativi della persona fisica (conduttore) o società o ente che gestisce l'azienda agricola*, in questa sezione vengono riportate tutte le informazioni relative al conduttore dell'azienda agricola già presenti nel modello Istat CensAgrCop2 eventualmente integrate e/o corrette.

Sezione 2 - *Localizzazione del centro aziendale*, contenente l'esatta ubicazione del centro aziendale, che può essere localizzato sia nel foglio di mappa campione sia altrove. In questa sezione il rilevatore riporta, oltre che l'indirizzo completo, anche il riferimento catastale, del centro aziendale.

Sezione 3a - *Codici delle particelle*, rileva tutte le particelle del foglio di mappa campione su cui insiste l'azienda agricola.

Sezione 3b - *Altri comuni*, rileva i comuni sui quali insiste l'azienda agricola. Il rilevatore deve indicare, in questa sezione, la/le provincia/e, il/i comune/i e il numero complessivo dei fogli di mappa su cui insiste l'azienda.

Sezione 4 - *Forma di conduzione*.

Sezione 5 - *Superficie totale e Superficie Agricola Utilizzata (SAU)*.

Quesito A - Rileva se l'azienda agricola è stata intervistata (*Si/No*) in occasione del 6° Censimento generale dell'agricoltura.

Sezione 6 - *Informazioni sul rispondente*, contiene informazioni sul rispondente all'intervista. Nel caso trattasi di persona diversa dal conduttore allora vengono rilevate anche le informazioni anagrafiche di colui che ha risposto alle domande del questionario.

Sezione 7 - *Altri componenti della famiglia*, rileva informazioni anagrafiche riguardo gli altri membri della famiglia che lavorano nell'azienda agricola.

Le informazioni rese alle sezioni da due a cinque si intendono riferite al 24 ottobre 2010, data del Censimento.

2.5.2 Il Sistema di Gestione della Rilevazione (SGRCOP)

Grazie a nuovi strumenti tecnologici sviluppati in Istituto, l'indagine di copertura si è basata su un Sistema di Gestione della Rilevazione (SGRCOP) informatizzato, al fine di semplificare la rilevazione e lo scambio di informazioni tra l'Istat ed i rilevatori. Attraverso il sistema, il rilevatore poteva accedere alle liste di partenza e alla cartografia, nonché compilare tutta la modulistica, compreso il questionario di indagine.

L'uso del sistema ha richiesto ai rilevatori selezionati competenze informatiche minime e i seguenti requisiti tecnologici:

- il browser, Internet Explorer versione 7 o superiore;
- una risoluzione minima 1024x768 (con i moderni schermi piatti LCD si è potuto avere accesso a risoluzioni più alte, fino a 1280x1024);
- l'attivazione di *Javascript*.

Ciascun rilevatore poteva accedere al sistema attraverso l'indirizzo URL <https://indata.istat.it/sgrcop> utilizzando una password univoca che gli consentiva di poter interrogare il sistema esclusivamente per le sezioni relative alle sue aree territoriali di competenza. I rilevatori, dopo aver effettuato l'accesso al sistema SGRCOP, avevano a disposizione il seguente materiale necessario allo svolgimento della rilevazione:

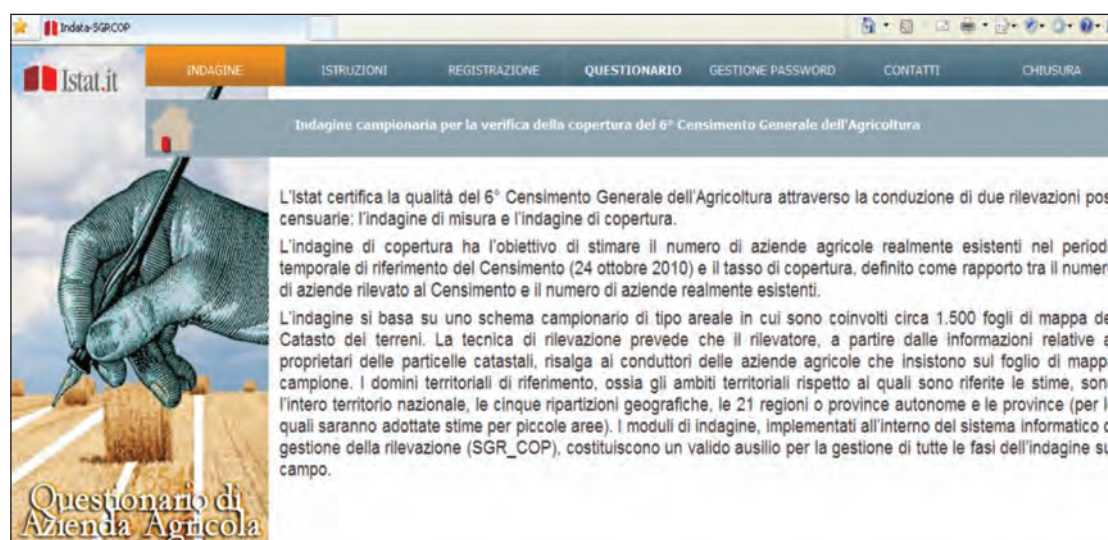
1. il manuale delle istruzioni;
2. i contatti informatici e telefonici dei referenti della rilevazione (mail: copertura@istat.it);
3. l'elenco dei fogli di mappa assegnati;
4. i modelli di rilevazione.

La struttura dell'indagine, da un punto di vista dell'acquisizione dati, è stata progettata e costruita secondo le tecniche CAWI (*Computer Assisted Web Interview*). L'immissione dei dati è ideata in modo tale che il rilevatore sia assistito in maniera flessibile e rapida. Il sistema prevede:

- l'upload, da parte dell'Istat, di tutto il materiale utile ai rilevatori per svolgere il proprio lavoro e il download degli stessi, da parte dei rilevatori;
- la compilazione, da parte dei rilevatori, dei modelli di rilevazione Istat CensAgrCop1, Istat CensAgrCop1_agg, Istat CensAgrCop2 e Istat CensAgrCop3;
- il download, da parte dell'Istat, dei modelli compilati;
- il monitoraggio, da parte dell'Istat, di tutte le fasi della rilevazione sul campo;
- il continuo scambio informativo tra l'Istat e i rilevatori.

Nella Figura 2.1 è mostrata l'homepage di SGR COP in cui la prima voce del menù, INDAGINE, mostra la pagina di presentazione in cui sono sintetizzati i contenuti e gli obiettivi dell'indagine.

Figura 2.1 - Pagina iniziale



La seconda voce del menù, ISTRUZIONI, mostrata in Figura 2.2, permette di accedere ad una pagina dove è possibile scaricare i documenti, utili alla rilevazione, in pdf e in Excel. In questa sezione è possibile visionare ed eventualmente scaricare in formato pdf il manuale delle istruzioni, la normativa riguardante il segreto statistico e la legenda della cartografia utile al rilevatore a leggere e interpretare le cartografie dei fogli di mappa campione. Inoltre in questa sezione sono stati collocati tutti i modelli di rilevazione in formato Excel. Ciò affinché il rilevatore possa stamparli e avere un supporto cartaceo per la rilevazione sul campo.



Figura 2.2 - Pagina istruzioni



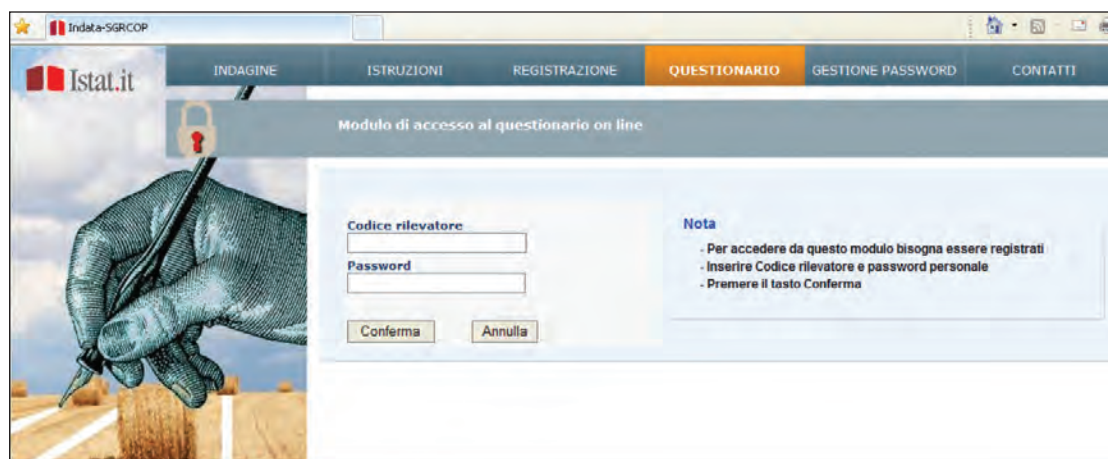
Un'operazione fondamentale è la registrazione del rilevatore, possibile nel menù REGISTRAZIONE (Figura 2.3). Questo elemento ha permesso la facile comunicazione dell'Istat con ogni singolo rilevatore e la possibilità di verificare in tempo reale il lavoro svolto e, quando richiesto, fornire supporto.

Figura 2.3 - Modulo di registrazione



Eseguita la procedura di registrazione, ogni accesso al sistema, successivo alla registrazione, avviene attraverso la sezione del menù QUESTIONARIO inserendo le proprie credenziali (Figura 2.4).

Figura 2.4 - Modulo di accesso



Il rilevatore è abilitato ed accede direttamente alla pagina che presenta la lista dei fogli di mappa assegnati (Figura 2.5). In questa pagina ogni rilevatore visualizza i propri dati anagrafici, inseriti al momento della registrazione e la lista di fogli di mappa che gli sono stati assegnati. Inoltre, ha la possibilità di avere in tempo reale: un consuntivo del lavoro svolto, infatti sulla riga di ogni foglio di mappa è possibile avere sempre aggiornati:

- il numero di intestatari per particella catastale da contattare;
- il numero dei conduttori da intervistare;
- il numero dei probabili conduttori da intervistare (Altri contatti);
- il numero dei questionari compilati.



Figura 2.5 - Esempio di Lista dei fogli mappa assegnati

INFORMAZIONI GENERALI SUL RILEVATORE										
Codice:		- Cod. Fiscale:		- Nome e Cognome:						
Indirizzo:		- CAP		- Comune:		- Provincia:				
Telefono:		- Fax:		- E-mail:						
  Modifiche										
Fogli mappa assegnati: 4										
N°	Foglio mappa	Elenco intestatari per particella catastale		Elenco conduttori da intervistare		Elenco conduttori da intervistare (Altri contatti)		Questionari compilati		Recupero
1	0002		n.146		n.1		n° 2		n° 1	
2	0017		n.119		n.4					
3	0047		n.583							
4	0048		n.184		n.1					

Per ciascun foglio di mappa assegnato l'intervistatore aveva a disposizione:

- il modello Istat CensAgrCop1 precompilato, con la lista degli intestatari di particella catastale;
- il modello Istat CensAgrCop1_agg, di ausilio per registrare informazioni utili ad individuare un conduttore di azienda agricola;
- il modello Istat CensAgrCop2, da compilare con la lista di conduttori di azienda agricola;
- il modello Istat CensAgrCop3, da compilare con le informazioni relative all'azienda agricola individuata;
- la cartografia del foglio di mappa, in cui sono disegnati i confini perimetrali del foglio, le strade, i corsi d'acqua, i fabbricati, eccetera e in cui sono circoscritte le particelle appartenenti al foglio;
- la cartografia del foglio di mappa in cui è stata sovrapposta l'ortofoto;
- la cartografia del comune in cui è stato evidenziato il foglio campione assegnato al rilevatore, al fine di individuare la sua collocazione nel territorio comunale.

All'interno del portale era anche reperibile una sezione "CONTATTI" dalla quale è stato possibile accedere al "Modulo di comunicazione", ossia un format mediante il quale i rilevatori potevano segnalare direttamente all'Istat i casi di criticità (sia informatici sia tecnici) e da questo ricevere assistenza e risolvere in tempo reale ogni criticità nei contatti e nell'identificazione delle aziende agricole.

2.5.3 Il monitoraggio

Tutta la rilevazione sul campo è stata attentamente monitorata dai referenti d'indagine Istat e dai responsabili regionali e provinciali usando SGR COP, al fine di garantire la massima qualità possibile dell'output per l'applicazione dei modelli di stima e il calcolo dei risultati. e, all'occorrenza, di intervenire sulle richieste di assi-

stenza inviate dai rilevatori attraverso i canali di comunicazione disponibili (e-mail, telefono).

Il sistema di monitoraggio permetteva ai referenti d'indagine Istat di accedere direttamente dal sito della rilevazione a qualsiasi utenza, semplicemente inserendo il codice utente del rilevatore e una password apposita. In questo modo si poteva visualizzare in tempo reale lo stato di lavorazione di ogni singola utenza e i verificare eventuali anomalie nella compilazione dei campi nei vari modelli di rilevazione. Inoltre, attraverso una pagina di gestione, l'Istat poteva intervenire nella correzione di record già lavorati a seguito di segnalazioni da parte dei rilevatori o dei referenti regionali/provinciali.

Anche ai responsabili regionali/provinciali sono state fornite delle utenze che permettevano di visionare l'elenco dei rilevatori di propria competenza e monitorare il lavoro di ciascuno di loro. Sempre dallo stesso sito della rilevazione era possibile, solo ai referenti d'indagine Istat e utilizzando un'utenza fittizia, entrare in un ambiente di lavoro del tutto simile a quello a disposizione dei rilevatori per effettuare prove di lavorazione. Questa opzione è stata utile soprattutto in fase di test del sistema e successivamente per verificare in prima persona eventuali malfunzionamenti segnalati dai rilevatori.

Al termine della rilevazione e concluse le attività di monitoraggio il sistema è stato chiuso e i dati scaricati su database per le successive attività di analisi ed elaborazioni dei risultati.

2.6 Il controllo e la correzione dei dati

Per la sua particolare natura l'indagine di copertura, ha richiesto che i casi di mancata risposta totale fossero risolti direttamente sul campo, cosicché le attività di controllo delle informazioni fornite dalle aziende intervistate si sono concentrate principalmente sugli aspetti riguardanti gli errori di misura e la mancata risposta parziale, limitatamente ai campi del questionario in cui non era stato posto un vincolo informatico. Alle attività di controllo e correzione sono state sottoposte tutte le unità rilevate all'indagine di copertura, che abbiano poi contribuito a determinare sia la 'copertura', sia la 'sotto-copertura' del Censimento.

Le procedure di controllo e correzione (C&C) sono state realizzate *ad hoc* per l'indagine in questione e possono essere sintetizzate in due passi distinti:

1. nel primo sono state eseguite le procedure finalizzate al controllo e alla correzione dei microdati, concentrando l'attenzione:
 - sulle variabili anagrafiche e di localizzazione dell'azienda e del conduttore, poiché sono informazioni rilevanti ai fini dell'eleggibilità e dell'abbinamento informatico con il Censimento;
 - sulle variabili quantitative come la "Superficie Agricola Utilizzata" e il "numero di fogli di mappa" in cui è distribuita l'azienda agricola, in quanto utili, rispettivamente, a stabilire l'eleggibilità dell'unità rilevata e ad applicare il *Weight Sharing Method*(WSM)⁶⁸;

2. nel secondo passo si è eseguita l'analisi dell'eleggibilità delle unità rilevate.

Dato il ruolo centrale che le operazioni di *record linkage*⁶⁹ hanno nella produzione delle stime, successivamente ai due passi descritti è stata effettuata la normalizzazione dell'indirizzo di residenza del conduttore e di quello relativo al centro aziendale.

68 Cfr. Paragrafo 2.4.4.

69 Cfr. Paragrafo 2.7.



2.6.1 Controllo e correzione dei microdati

Il primo degli obiettivi consiste quindi nel valutare se le informazioni riportate nei modelli CensAgrCop2 e CensAgrCop3 sono sufficienti a stabilire l'appartenenza dell'unità rilevata alla popolazione di interesse (le aziende agricole). A questo fine si è deciso di considerare eleggibile un'unità se, oltre a rispondere alla definizione censuaria di azienda agricola, presentava informazioni sufficienti per risalire all'identità del conduttore e a stabilire la localizzazione del centro aziendale.

In questa fase, quindi, sono stati oggetto d'esame principalmente i campi riportanti le variabili identificative della persona fisica (conduttore) o società o ente che gestisce l'azienda agricola e le variabili di localizzazione sia dell'azienda sia del conduttore stesso. È stata inoltre valutata la congruenza tra le informazioni riportate nel modello CensAgrCop2 e nel modello CensAgrCop3.

Per quanto riguarda le variabili "Cognome" e "Nome" del conduttore, grazie al flusso di dati che il sistema informatico ha permesso di trasferire da un modello all'altro, non ci sono state informazioni mancanti e spesso, per il rilevatore, non è stato necessario riscrivere il nome avendolo ereditato dal modello Istat CensAgrCop1 precompilato; tale accortezza ha fortemente limitato gli errori di digitazione (Figura 2.6). È stato necessario, infatti, intervenire in pochi casi per errori di scrittura, abbreviazioni, inversione nell'ordine delle variabili o differenze riscontrate tra i diversi modelli.

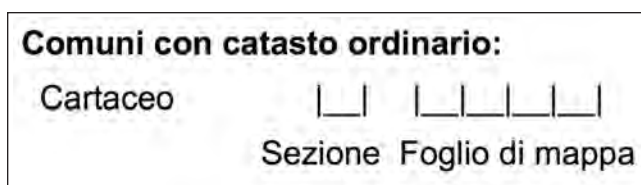
Figura 2.6 - Caratteri identificativi della persona fisica (conduttore) o società, ente che gestisce l'azienda agricola

sezione 1 - Caratteri identificativi della persona fisica(Conduttore) o società o ente che gestisce l'azienda agricola	
Anagrafica	
1.1 Cognome o ragione sociale:	<input type="text" value="ROSSI"/>
1.2 Nome della persona fisica:	<input type="text" value="MARIO"/>
1.3 Codice fiscale(CUAA):	<input type="text" value="RSSMRA00A01H501C"/>
Residenza	
1.4 Provincia:	<input type="text" value="Belluno"/>
1.5 Comune:	<input type="text" value="Auronzo di Cadore"/>
1.6 Indirizzo(via-piazza-località, numero civico):	<input type="text" value="VIA DELLA MARINA 15"/>
1.7 CAP:	<input type="text"/>

Anche i vincoli informatici inseriti nella sezione della localizzazione del centro aziendale sono stati appositamente molto rigidi, quindi non erano presenti valori mancanti; infatti non era possibile chiudere l'intervista se questa sezione non era compilata; tuttavia, in alcuni casi è stato necessario intervenire per correggere il riferimento catastale (se-

zione, foglio di mappa, particella) del centro aziendale perché scritto non correttamente. I rilevatori hanno trovato diverse difficoltà a rilevare questa variabile in quanto non sempre il conduttore dell'azienda agricola conosceva tale informazione, illustrata nel prospetto seguente⁷⁰:

Prospetto 2.3 - Riferimento catastale del centro aziendale



L'errore più frequente è stata l'assenza dell'identificativo della sezione oppure una non corretta scrittura nel formato richiesto, ma attraverso la revisione del modello cartaceo, dell'indirizzo rilevato e grazie alle informazioni contenute nell'archivio del Catasto dei terreni, si è proceduto alla ricostruzione della sezione legata al foglio di mappa indicato.

Per quanto riguarda la variabile relativa alla Superficie Agricola Utilizzata (SAU), nonostante l'indagine avesse la finalità di valutare l'errore di copertura e non la qualità dei valori censuari, è stato necessario analizzare la correttezza di tale variabile, sia per stabilire l'eleggibilità dell'azienda agricola, sia per preservare la correttezza delle stime e l'omogeneità delle stesse rispetto ai corrispondenti dati censuari. Complessivamente si sono registrati pochissimi valori mancanti, in questi casi si è proceduto ad un *recall* telefonico al fine di recuperare il dato e stabilire l'eleggibilità dell'unità in questione.

Per quanto riguarda gli errori di misura contenuti nei questionari di azienda agricola, si è proceduto al controllo di coerenza della SAU con la variabile superficie totale e in alcuni casi si è ricorsi alla revisione del cartaceo oppure al contatto con i rilevatori per risolvere alcune incongruenze.

Un'altra variabile su cui è stata posta l'attenzione in fase di controllo e correzione è stata quella che rilevava il 'Numero dei fogli di mappa': tale variabile non ammetteva valori mancanti in quanto per poter applicare correttamente il metodo del *Weight Sharing Method* era necessario conoscere per ciascuna azienda rilevata il numero totale dei fogli di mappa in cui essa ricade. In riferimento a questa variabile, è stata dapprima verificata una coerenza interna al questionario tra il valore indicato rispetto alla superficie totale dell'azienda, al numero di comuni in cui essa è dislocata e alla zona altimetrica del comune di localizzazione del centro aziendale. Vista l'importanza della variabile, prima di procedere al calcolo dei pesi, è stato effettuato un ulteriore controllo dei valori di questa variabile attraverso un confronto, per le aziende abbinate, con i dati contenuti nell'archivio amministrativo dell'AGEA.

2.6.2 Analisi dell'eleggibilità

Al termine della fase di rilevazione sul campo, le unità rilevate erano 22.513. L'analisi dell'eleggibilità si è svolta in due passi principali.

Il primo passo ha preso in considerazione la variabile SAU per verificare se fosse ri-

⁷⁰ Per un approfondimento si veda: M. Mazziotta, A. Bernardini, L. Soriani (2011). Indagine di copertura del 6° Censimento generale dell'agricoltura. Manuale delle istruzioni, Istat.



spettata la soglia regionale adottata al Censimento affinché l'unità rilevata potesse essere definita azienda agricola, coerentemente con la definizione adottata per il Censimento stesso.

A tale scopo in sede di formazione sono state fornite (riportate nella tavola 2.1) le soglie regionali di SAU al di sotto delle quali un'unità non può essere considerata azienda agricola e quindi non doveva essere rilevata.

Tavola 2.1 - Soglie di SAU per l'eleggibilità ad azienda agricola per Regione

REGIONI	Soglia di SAU (in ettari)
Piemonte	>= 0,3
Valle d'Aosta/Vallée d'Aoste	>= 0,4
Liguria	>= 0,3
Lombardia	>= 0,3
Bolzano/Bozen	>= 0,2
Trento	>= 0,2
Veneto	>= 0,3
Friuli-Venezia Giulia	>= 0,3
Emilia-Romagna	>= 0,3
Toscana	>= 0,3
Umbria	>= 0,3
Marche	>= 0,4
Lazio	>= 0,3
Abruzzo	>= 0,3
Molise	>= 0,3
Campania	>= 0,3
Puglia	>= 0,2
Basilicata	>= 0,3
Calabria	>= 0,3
Sicilia	>= 0,2
Sardegna	>= 0,2

Il secondo passo invece ha preso in considerazione le variabili anagrafiche del conduttore e quelle di localizzazione dell'unità rilevata per valutare l'univocità delle unità rilevate. Il sistema informatico non permetteva infatti a un rilevatore di registrare due volte la stessa azienda, ma essendo le aziende dislocate su più fogli di mappa dello stesso comune o di comuni diversi, la stessa unità poteva essere rilevata da più rilevatori, tante volte quanti erano i fogli di mappa campionati in cui ricadeva. Dopo un'attenta analisi sulle variabili anagrafiche, e soprattutto sulla variabile univoca di identificazione di azienda agricola - il codice fiscale della persona fisica che gestisce l'azienda (conduttore) o Partita Iva in caso di persona giuridica - i record duplicati sono risultati essere 925. Si è proceduto quindi, dopo aver confrontato le informazioni rilevate da ciascun rilevatore, all'eliminazione delle duplicazioni di record, mantenendo, a parità di contenuto, quelli con maggiori informazioni.

Al termine delle operazioni di controllo e correzione le aziende eleggibili rilevate dall'indagine di copertura sono 21.558, distribuite per regione come riportato nella Tavola 2.2.

Tavola 2.2 - Aziende intervistate per regione

REGIONI	Numero
Piemonte	1.013
Valle d'Aosta/Vallée d'Aoste	249
Liguria	213
Lombardia	822
Trento	454
Veneto	1.195
Friuli-Venezia Giulia	414
Emilia-Romagna	1.046
Toscana	1.007
Umbria	459
Marche	631
Lazio	1.337
Abruzzo	771
Molise	503
Campania	2.126
Puglia	3.540
Basilicata	1.361
Calabria	1.660
Sicilia	1.946
Sardegna	841
Italia	21.588

2.7 Le procedure di abbinamento

2.7.1 Strategia di abbinamento adottata

L'abbinamento esatto dei dati elementari del Censimento con quelli dell'indagine di copertura è un passaggio fondamentale per le stime del tasso di copertura del Censimento. Il termine "abbinamento esatto" (o "*record linkage*") si riferisce all'insieme di metodologie e tecniche che consentono di identificare record riferiti ad una stessa unità statistica, ma contenuti in archivi diversi. Nel caso dell'indagine di copertura del 6° Censimento generale dell'agricoltura l'obiettivo del processo di abbinamento (o integrazione) è stato duplice:

- da un lato determinare il numero delle aziende agricole rilevate in entrambe le occasioni e, per differenza, il numero delle aziende agricole sfuggite ad una delle due rilevazioni;
- dall'altro lato effettuare l'operazione anche nei confronti delle unità contenute in tutte le fonti amministrative, utilizzate per predisporre la lista pre-censuaria⁷¹.

L'abbinamento tra i risultati definitivi del Censimento e i risultati dell'indagine di copertura è centrale per stimare i tassi di copertura del Censimento con il modello di Petersen (o *dual-system* o di tipo cattura-ricattura). Infatti, per stimare correttamente l'errore di sotto-copertura, la procedura di abbinamento che identifica le unità enumerate in entrambe le occasioni deve essere priva di errori di abbinamento, i quali, anche se molto contenuti, possono compromettere l'affidabilità delle stime del grado di copertura del Censimento.

⁷¹ Questa seconda finalità dell'abbinamento ha consentito di dare un contributo informativo alla determinazione delle regole di eleggibilità di unità giuridico-economiche contenute nelle fonti amministrative ai fini della realizzazione del registro statistico delle aziende agricole.



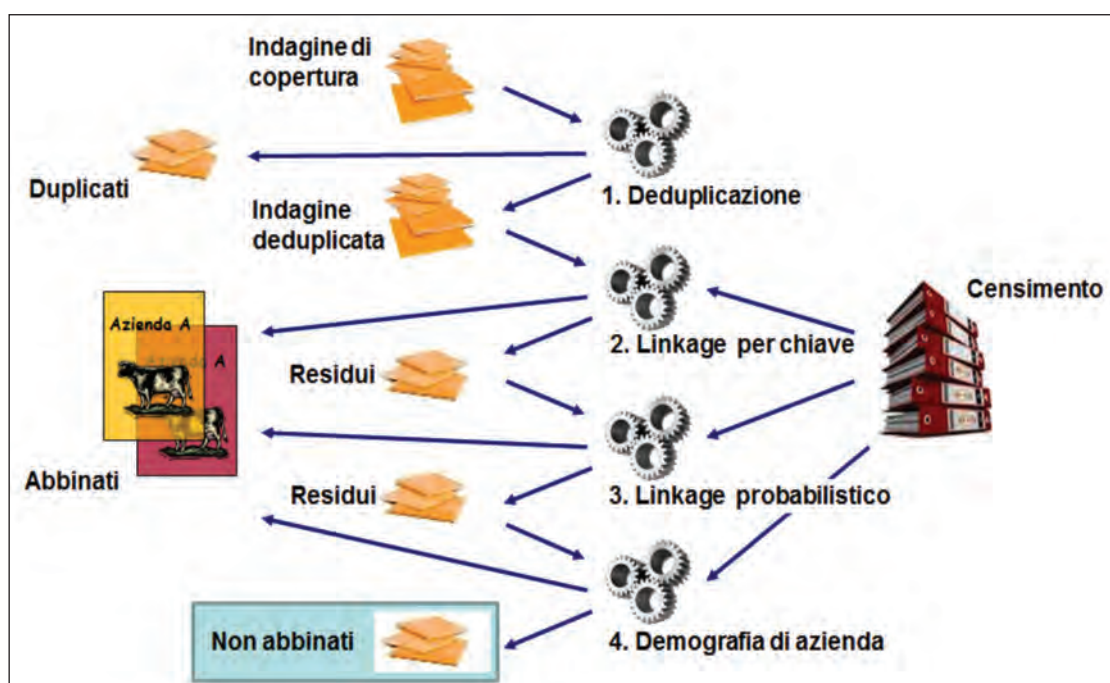
Pertanto, per assicurare la massima correttezza degli abbinamenti è stata adottata una procedura di *record linkage* strutturata in diverse fasi e iterazioni. Sono state adottate tecniche di *record linkage* sia deterministiche che probabilistiche e sono state considerate differenti regole di confronto nelle varie fasi della procedura. Le prime fasi sono state indirizzate all'identificazione degli abbinamenti più facili, per mezzo di procedure computazionali semplici e, per costruzione, prive di errore, lasciando gli abbinamenti più complessi alle fasi successive.

Un'accurata revisione manuale interattiva è stata applicata ai risultati più ambigui delle varie fasi per garantire la massima correttezza degli abbinamenti individuati anche attraverso le metodologie probabilistiche. Inoltre, le operazioni di *linkage* hanno previsto una fase finale di ricerca manuale, guidata da procedure automatiche, di legami multipli tra le unità e legami al netto di trasformazioni di natura giuridica o organizzativa che possono essere ricondotte a fenomeni di demografia di azienda agricola. Tale operazione si è resa necessaria per riuscire a descrivere la complicata natura delle unità da integrare, dovuta essenzialmente a tre cause: la difficoltà di applicare la definizione di azienda agricola per le unità di dimensioni ridotte; le frequenti trasformazioni di natura giuridica o organizzativa che l'azienda può subire nel tempo, mantenendo comunque delle regole di continuità nel territorio e nel tipo di produzione agraria; nonché l'organizzazione familiare e/o associativa di numerose aziende.

La complessa articolazione della strategia di *linkage* delineata è stata supportata e assistita dalla disponibilità e dall'utilizzo di strumenti informatici generalizzati, che implementano in maniera ottimale le sofisticate metodologie richieste.

Nella figura successiva è riportata una descrizione grafica della strategia complessiva.

Figura 2.7 - Strategia di linkage adottata - Workflow



Le procedure di abbinamento sono state realizzate in fasi successive che prevedono:

1. abbinamento esatto tramite codice identificativo univoco;
2. abbinamento probabilistico tramite tecniche di *record linkage*;
3. revisione manuale dei residui o dei casi più ambigui;
4. ricerca dei legami multipli e delle unità coinvolte in trasformazioni.

Il primo tipo di abbinamento è stato realizzato sfruttando il codice identificativo univoco (partita iva e/o codice fiscale del conduttore dell'azienda) presente in entrambe le liste da abbinare. Tale codice, purtroppo, non era privo di errori per tutte le unità delle due liste, quindi questa prima fase di abbinamento ha coinvolto essenzialmente i codici per cui si disponeva dell'informazione corretta.

L'abbinamento probabilistico, implementato in assenza o incompletezza del codice identificativo univoco, richiede l'applicazione di sofisticate tecniche di *record linkage*, basate sulla somiglianza degli attributi relativi alle unità provenienti dai due file. Tali tecniche vengono sempre più spesso utilizzate sia per integrare informazioni provenienti da fonti diverse sia per eliminare duplicazioni all'interno di uno stesso file. A tal fine, già da qualche anno è disponibile in Istat il software open-source RELAIS; nel caso in esame è stata utilizzata la versione beta 2.3 che mette a disposizione le metodologie più consolidate per il *record linkage* probabilistico ma anche *facilities* per l'applicazione di innovativi metodi di riduzione dello spazio di ricerca e l'introduzione di probabilità di abbinamento esogene. Per garantire la massima qualità del risultato dell'abbinamento, cioè per minimizzare l'errore di falso abbinamento e di mancato abbinamento, i casi più ambigui e le unità più rilevanti nella stima del tasso di copertura sono state revisionate manualmente, così come i record non abbinati dalle procedure automatiche (residui) e gli abbinamenti individuati dalle procedure di *linkage* probabilistico. Per questa attività è stato sviluppato un software *ad hoc*, al fine di consentire la revisione manuale dei risultati.

Si sottolinea infine che le attività di *linkage* sono state rivolte al riconoscimento non solo delle aziende agricole rilevate al Censimento, ma anche alle unità contenute nelle fonti amministrative da cui è stata tratta la lista pre-censuaria. Tale attività, seppure non strettamente necessaria per la valutazione del tasso di copertura e sotto-copertura del Censimento, ha permesso di interpretare le varie cause sia della copertura che della sotto-copertura legate alla "demografia" delle aziende, alla realtà familiare e ai legami multipli, alla mancata risposta o al rifiuto a rispondere, nonché alla scarsità di segnali nelle fonti amministrative. Queste analisi consentono di valutare l'entità di tali fenomeni ed eventualmente predirne il rischio, in ottica prospettica per la costruzione di un registro di aziende agricole.

Nell'ottica della creazione del registro, è infine importante sottolineare che, per ottimizzare i risultati di tutte le procedure di *record linkage*, è stato fondamentale aver rilevato nel questionario i codici fiscali, i nomi e cognomi e le date di nascita delle persone collegate in qualche modo all'intestatario-conduttore del terreno, in quanto coniugi, soci, proprietari, o affittuari. Queste informazioni sono state molto utili per le operazioni di *linkage* e per l'individuazione degli eventuali duplicati tra le liste.

In conclusione la strategia di abbinamento è stata applicata tra le aziende rilevate alla PES nei fogli di mappa campione e le unità contenute nei seguenti archivi:

- a) la lista delle aziende agricole attive e temporaneamente inattive (rispettivamente 1.620.884 e 34.070) rilevate al Censimento;
- b) la lista complessiva finale delle unità, pari a 2.137.030, che comprendono tutte quelle presenti nella lista pre-censuaria pari a 2.047.948 e le nuove unità individuate durante la fase di raccolta dei dati. Questa fonte contiene, oltre quelle che sono confluite nella lista del punto a) anche le aziende cessate, le aziende che hanno rifiutato di compilare il questionario o irreperibili, duplicazioni delle unità presenti nella lista di partenza e unità non eleggibili ad azienda agricola;



c) le unità agricole presenti nella lista BIFA (Base Informativa da Fonte Amministrativa) che complessivamente è di circa 5 milioni di record e implementata sulla base dello studio e dell'analisi di 17 fonti amministrative differenti, di settore o meno, compreso l'archivio completo delle aziende agricole di Agea (Agenzia per le Erogazioni in Agricoltura).

Al termine dell'abbinamento con tutte le fonti descritte è stato possibile non solo determinare il tasso di copertura del Censimento attraverso gli abbinamenti con la fonte a), ma è stato possibile dettagliare il tasso di sotto-copertura attraverso le informazioni provenienti dalle fonti b) e c). Il risultato di questa scomposizione è descritta nel Paragrafo 2.7.4.

2.7.2 Attività preliminari

Le attività preliminari alle operazioni di *linkage* sono state tre:

- riconciliazione degli schemi;
- standardizzazione di nomi e indirizzi;
- preparazione dei file di input per RELAIS.

Tali attività sono decisive per l'esito della fase di abbinamento in modo particolare quando attributi raccolti nel questionario in testo libero assumono il ruolo di variabili cardine per le procedure di *linkage*.

2.7.2.1 Riconciliazione degli schemi

Dall'analisi delle variabili disponibili sulle due fonti da confrontare (Censimento ed indagine di copertura) otteniamo la lista di attributi comuni che segue:

Prospetto 2.4 - Riconciliazione degli schemi

ATTRIBUTO	Tipo di dato	Note
Nome del conduttore	Testo libero	Nell'indagine di copertura nome e cognome vengono rilevati come due variabili distinte mentre nel Censimento viene rilevata una unica variabile
Tipo di conduzione	Codice	
Provincia di residenza del conduttore	Codice	
Comune di residenza del conduttore	Codice	
Indirizzo di residenza del conduttore	Testo libero	
CAP di residenza del conduttore	Codice	
Codice fiscale del conduttore	Codice	In caso di aziende può contenere la partita Iva
Provincia del centro aziendale	Codice	
Comune del centro aziendale	Codice	
Indirizzo del centro aziendale	Testo libero	
CAP del centro aziendale	Codice	
Superficie agricola totale	Numerico	
Superficie agricole utilizzata	Numerico	

Da notare che mentre gli attributi raccolti dall'indagine di copertura sono inseriti esclusivamente dal rilevatore, alcuni degli attributi del Censimento presentano due versioni

una denominata P (proveniente dalle liste di riferimento e prestampata sul modulo del Censimento) ed una denominata M (che è la versione modificata dal compilatore e che quindi si intende sostituisca quanto prestampato).

2.7.2.2 Standardizzazione delle variabili a testo libero

Il confronto di variabili a testo libero nasconde per sua natura una serie di difficoltà. Un confronto esatto è in generale troppo restrittivo perché anche minime differenze nel testo comportano la non corrispondenza e per contro un confronto approssimato, ma troppo lasco, rischia di associare valori diversi solo perché il testo presenta parti comuni magari poco significative.

In generale è sconsigliato l'utilizzo di tali variabili nelle procedure di *linkage*, a meno che non si sia confidenti di una alta qualità nella rilevazione delle stesse.

Se invece, come in questo caso, la qualità non è alta, ma il loro impiego è necessario perché di alto potere identificativo e perché i restanti attributi non sono sufficienti per la individuazione delle entità, è fortemente consigliata ogni possibile standardizzazione del valore degli attributi.

Nel nostro caso i tre attributi *“Nome del conduttore”*, *“Indirizzo del conduttore”* e *“Indirizzo del centro aziendale”* sono irrinunciabili in quanto sono gli unici con potere identificativo per le entità dove non sia valido il codice fiscale.

La qualità del *“Nome del conduttore”* non è elevata, perché può riportare sia un nome che una ragione sociale e alcune diciture possono essere abbreviate oppure omesse. Analogamente anche le variabili *indirizzo* sono di non elevata qualità, in quanto il formato previsto (dug - denominazione urbanistica generica, nome via e numero civico) non è sempre rispettato; va rilevato anche che la localizzazione dei centri agricoli è spesso di difficile determinazione e che accade di frequente la sovrascrittura dell' *“indirizzo del centro aziendale”* con *“l'indirizzo di residenza del conduttore”* o vice versa. Una prima procedura di standardizzazione degli indirizzi viene eseguita mediante il software di normalizzazione Egon, in possesso dell'Istat.

Successivamente si procede alla lavorazione di queste variabili eseguendo le operazioni di seguito elencate in sequenza:

1. conversione dei caratteri minuscoli in maiuscoli;
2. eliminazione delle diciture non significative (Prospetto 2.5);
3. eliminazione dei caratteri speciali;
4. eliminazione degli spazi.

Prospetto 2.5 - Standardizzazione delle variabili testuali

VARIABILE	Elenco diciture non significative		
Nome del conduttore	AGRICOLA	SRL	COOPERATIVA
	SOCIETA	F.LLI	EREDI
	S.N.C.	UTE	SAS
	S.A.S.	SOC.	FIGLI
	AZ.AGR.	SOC	SNC
	AZ.	SS	COOP.
	AGR.	S.S	CON
	S.R.L.	S.S.	SEDE

continua alla pagina successiva



Prospetto 2.5 - Standardizzazione delle variabili testuali

VARIABILE	Elenco diciture non significative		
Indirizzo di residenza del conduttore/ Indirizzo del centro aziendale	VIA	FRAZIONE	SNC
	CONTRADA	FRAZ	PARTICELLA
	C.DA	PROVINCIALE	PART.
	C/DA	VIALE	VICOLO
	C\DA	PIAZZA	COMUNALE
	CDA	P.ZZA	CASCINA
	LOCALITA	BORGO	C.NA
	LOCALIT	CORSO	CNA
	LOC	REGIONE	STATALE
	STRADA		

Le diciture non significative sono scelte tra i termini che si presentano più frequentemente tra i valori assunti dalle variabili nelle due fonti.

2.7.2.3 Preparazione dei file di input

È possibile fornire a RELAIS i *dataset* da processare con file di testo a tracciato variabile con separatore.

La generazione dei file di input è stata realizzata semplicemente con una interrogazione SQL al database archivio delle due fonti contenente i dati standardizzati. Produrre i data set in questo modo ha permesso di risolvere in origine alcune ambiguità sui dati. In particolare:

1. assegnando lo stesso nome alle variabili rappresentate nel tracciato record dei due file (il tracciato record è descritto nella prima riga), RELAIS individua automaticamente gli attributi comuni;
2. nella fonte Censimento a ogni variabile viene direttamente assegnato il valore M (modificato) se presente o il valore P (prestampato) in alternativa; in questo modo la procedura di *linkage* utilizza sempre il valore corrente.

2.7.3 I metodi di linkage

Le procedure di abbinamento, articolate in quattro fasi successive come illustrato nel Paragrafo 2.7.1, sono basate principalmente su due tecniche:

- abbinamento esatto tramite codice identificativo univoco;
- abbinamento probabilistico tramite modelli di *record linkage* più sofisticati.

In questo paragrafo vengono illustrate le tecniche adottate, i risultati e le difficoltà incontrate.

2.7.3.1 Linkage per chiave esatta

L'abbinamento tra le aziende agricole rilevate al Censimento e alla indagine di copertura è stato realizzato con una procedura in più passi, che si è avvalsa al meglio della presenza di variabili con alto potere discriminante e della qualità delle informazioni presenti nelle due fonti. In primo luogo, per le imprese che presentavano il codice fiscale del conduttore di azienda (o la Partita Iva) privo di errori, si è adottata una procedura di abbinamento deterministico esatto usando come chiave tale variabile.

Le aziende agricole presenti nel file dell'indagine di copertura ammontavano a 22.513 record. Il file del Censimento di riferimento ammonta a 1.620.884 aziende agricole.

Prima di procedere all'abbinamento deterministico esatto tra i record delle due rilevazioni è stato necessario eseguire procedure di de-duplicazione, già descritte nel Paragrafo 2.6.2, sull'archivio della copertura. Sono stati eliminati, in tal modo, 925 record per i quali si aveva una esatta duplicazione del codice fiscale (o Partita Iva del conduttore).

Una volta eliminati questi duplicati, si sono abbinate le aziende per chiave esatta: dei 21.588 record della copertura si sono abbinate 18.767 aziende agricole, ossia l'83,4 per cento del totale della copertura.

I record residui di tale attività, 2.821 unità, sono quindi passati alla fase successiva in cui si ricerca un abbinamento con metodi probabilistici alle aziende rilevate al Censimento.

2.7.3.2 Record linkage probabilistico

Nel modello probabilistico per l'individuazione degli abbinamenti, si ipotizza che la distribuzione del vettore dei confronti, y , i cui elementi k sono il risultato del confronto tra le variabili di *Match* comuni ai due insiemi e fortemente discriminanti per le unità, sia una mistura di due distribuzioni, una generata dalle coppie (a, b) che effettivamente rappresentano la stessa unità, distribuzione m , e una generata dalle coppie (a, b) che rappresentano unità diverse, distribuzione u . A partire dalla stima di tali distribuzioni, è possibile costruire il peso composto di abbinamento (Fellegi and Sunter, 1969), dato dal rapporto delle verosimiglianze:

$$r = \frac{m(y)}{u(y)} = \frac{\Pr(y|M)}{\Pr(y|U)}$$

dove M è l'insieme delle coppie che rappresentano degli abbinamenti e U è l'insieme delle coppie che non rappresentano degli abbinamenti, con $M \cup U = \Omega$ e $M \cap U = \emptyset$. In generale, la stima dei parametri delle distribuzioni viene generalmente ottenuta per mezzo dell'applicazione dell'algoritmo *Expectation-Maximization* (Jaro, 1985).

Sulla base del rapporto r , le coppie sono ordinate e sottoposte a un processo di classificazione negli insiemi M ed U in modo tale da minimizzare gli errori di classificazione e la dimensione dell'area per cui non viene presa nessuna decisione. Le coppie che hanno una probabilità di abbinamento superiore ad una certa soglia, T_m , sono classificate come *Match*; quelle al di sotto della soglia T_u come *non Match*; infine, per quelle con una probabilità compresa tra T_u e T_m non viene presa una decisione e sono mandate a revisione manuale.

Data la numerosità elevata delle unità che si vogliono abbinare, è necessario introdurre metodi di riduzione dello spazio di ricerca volti a ridurre la complessità del problema sia dal punto di vista statistico (stima dei parametri del modello troppo difficile se le coppie sono troppo poche rispetto alle non coppie nello spazio di ricerca creato) sia dal punto di vista informatico.

I 2.821 residui dell'indagine non abbinati per codice fiscale con le aziende del Censimento sono stati abbinati sulla base delle variabili residue, fortemente identificative, del conduttore e dell'azienda stessa. A causa dell'elevato numero delle possibili coppie si sono adottate in primo luogo tecniche per la riduzione dello spazio di ricerca e poi si è fatto ricorso al *record linkage* probabilistico con il modello di Fellegi -Sunter implementato nel software RELAIS, al fine di individuare tutti i reali abbinamenti.

Date le ipotesi del modello di Petersen (Wolter, 1986) si è adottata una strategia volta a minimizzare o ridurre al minimo la probabilità di falso abbinamento e contestualmente anche quella di mancato abbinamento.



Sono stati applicati in parallelo numerosi modelli, considerando link solo quelli che venivano confermati da modelli diversi (tecnica detta “*multiple passes*”).

Come metodi di riduzione dello spazio ricerca ci si è avvalsi del tradizionale bloccaggio, del *Sorted Neighborhood* e di una tecnica innovativa denominata *Simhash* (Borrelli et al, 2012).

Come variabili di bloccaggio (o di sorted) sono state usate, in modo alternativo o combinato:

- le variabili geografiche relative alla residenza del conduttore (provincia, comune, CAP);
- le variabili geografiche relative alla centro aziendale (provincia, comune, CAP);
- le variabili relative al nome o cognome standardizzato del conduttore.

Come variabili di abbinamento, scelte tra le k comuni ai due file, sono state usate, in modo alternativo:

- le variabili relative al conduttore dell'azienda (codice fiscale, nome standardizzato);
- le variabili geografiche relative alla residenza del conduttore (provincia, comune, CAP, indirizzo);
- le variabili geografiche relative alla residenza del centro aziendale (provincia, comune, CAP, indirizzo).

Sono state mandate a revisione manuale, per garantire la precisione dei risultati, le coppie per cui la probabilità di abbinamento era superiore 0,85.

Tra i 2.821 residui sono stati individuati 851 nuovi abbinamenti.

I modelli probabilistici sono stati usati anche per identificare abbinamenti tra le aziende di copertura e quelle presenti nella lista pre-censuaria.

Con gli stessi metodi probabilistici sono state individuate ulteriori 148 coppie, sfruttando i nominativi ausiliari che era possibile indicare sia nel questionario dell'indagine di copertura che del Censimento. Questi abbinamenti sono da attribuire a fenomeni di demografia di azienda agricola (trasformazioni di forma giuridica, scissione, fusione, cambi di conduttore o intestatario del terreno) o di legami multipli (più individui che ruotano intorno alla stessa azienda e/o più aziende condotte da uno stesso gruppo di individui). Come noto, si tratta di fenomeni entrambi molto rilevanti e frequenti nella realtà agricola italiana, soprattutto per le aziende di dimensioni contenute. Anche gli abbinamenti rintracciati per mezzo di queste informazioni su individui aggiuntivi sono stati verificati e confermati dalla revisione manuale.

Alla fine di questo passo sono rimasti 1.822 record non abbinati, i quali hanno ‘determinato’ la sotto-copertura del Censimento. Grazie all'utilizzo di tutte le fonti amministrative elencate nel Paragrafo 2.5.4 è stato possibile spiegare le ragioni della sotto-copertura e scomporre il tasso nelle sue componenti come illustrato nel paragrafo seguente.

2.7.4 Gli strumenti utilizzati per l'abbinamento

2.7.4.1 RELAIS

RE.L.A.IS (*Record linkage At Istat*) è un software che nasce nel 2006 in Istat per rispondere alla necessità di avere uno strumento software generalizzato per la fase di *record linkage*, ovvero per riconoscere se due record appartenenti a due diverse fonti di dati descrivono la stessa entità del mondo reale. RELAIS è un progetto open source, rilasciato con licenza EUPL. Tale scelta nasce dalla possibilità di riusare soluzioni già disponibili nella comunità scientifica. Da un punto di vista tecnologico, RELAIS è stato implementato usando tecnologie open source quali Java e R, come linguaggi di programmazione, e MySQL, come sistema di gestione delle basi di dati.

RELAIS si basa sull'idea che un progetto di *record linkage* non è un processo rigido, ma deve essere costruito dinamicamente sulla base dei vincoli specifici della singola applicazione. In particolare, un processo di RL è suddivisibile nelle seguenti fasi: (i) pre-processamento; (ii) creazione dello spazio di ricerca; (iii) scelta delle variabili di *Match*; (iv) scelta delle funzioni di confronto; (v) scelta del modello decisionale; (vi) applicazione di vincoli sul *linkage* (*linkage* 1:1). Per ognuna di queste fasi RELAIS mette a disposizione diverse tecniche che possono essere diversamente combinate, consentendo di ottenere specifici *workflow* per ogni applicazione. La creazione dei *workflow* è facilitata dalla presenza di una semplice interfaccia grafica e dall'accessibilità dei menù delle diverse tecniche, resi disponibili solo quando le fasi propedeutiche si sono concluse con successo. Per facilitare ulteriormente l'utente nella costruzione del *workflow* migliore, la scelta delle variabili di *Matching* e di bloccaggio è coadiuvata da due utilità di *profiling* dei dati di input.

Il risultato del processo di *linkage* è la classificazione di ogni coppia presente nello spazio di ricerca come *Match*, *Possible Match* o *NonMatch*; nel caso di *linkage* deterministico l'insieme dei *Possible Match* è sempre vuoto. Se è stato applicato un modello di *linkage* probabilistico la divisione nei tre gruppi è definita in base alla scelta da parte dell'utente della soglia T_u , sotto il cui valore le coppie sono classificate come *NonMatch*, e la soglia T_m , sopra il cui valore le coppie sono classificate *Match*, le coppie con valori di r ($r=m/u$) comprese tra T_u e T_m sono classificate come *Possible Match*.

2.7.4.2 Software per la revisione manuale

Tutte le possibili coppie dei vari modelli probabilistici scelti con probabilità di abbinamento inferiore alla soglia fissata per i *Match*, ma superiore a quella fissata per i *NonMatch* sono state sottoposte al lavoro di alcuni revisori al fine di identificare altri abbinamenti.

Il software per la revisione manuale consente ai revisori di visualizzare le informazioni relative ai record risultati *Possibile Match* durante la fase di *linkage* tra gli archivi generati dall'indagine di copertura del Censimento dell'agricoltura (copertura) e il Censimento dell'agricoltura (Censimento).

Gli esperti di dominio hanno selezionato un insieme di variabili comuni, ritenendole più significative e discriminanti per giudicare se due record corrispondono effettivamente alla stessa unità di rilevazione. In particolare le variabili scelte sono: (1) Nome e Cognome del conduttore dell'azienda; (2) Codice fiscale del conduttore dell'azienda; (3) Provincia di residenza del conduttore; (4) Comune di residenza del conduttore; (5) Indirizzo di residenza del conduttore; (6) Provincia di localizzazione del centro aziendale (CEA); (7) Comune di localizzazione del CEA; (8) Indirizzo di localizzazione del CEA; (9) Tipo di conduzione dell'azienda; (10) Superficie Agricola Utilizzata (SAU); (11) Superficie Totale (ST).

Le variabili sono suddivise per argomento in quattro schede:

1. Conduttore: contenente il nome e cognome del conduttore dell'azienda e il suo codice fiscale;
2. Residenza: contenente la provincia, il comune e l'indirizzo di residenza del conduttore;
3. Localizzazione CEA: contenente la provincia, il comune e l'indirizzo di localizzazione del CEA;
4. Informazioni Azienda: contenete il tipo di conduzione aziendale, la SAU e la SAT.

Per ogni scheda, il revisore deve decidere se le informazioni visualizzate sono concordanti, ovvero contengono (anche se espressa diversamente) la stessa informazione, oppure sono discordanti, ovvero contengono informazioni diverse.



In una ultima scheda (scheda di Decisione), il revisore stabilisce, in base alle decisioni precedentemente prese, se i dati presenti nella fonte Censimento e nella fonte copertura sono relativi alla medesima azienda.

Il software è stato realizzato in Java e, attraverso le librerie di connessione ad Oracle, accede direttamente alle basi di dati che archiviano le due fonti per presentare i dati.

2.7.5 Scomposizione della copertura/sotto-copertura

Il *record linkage* è stato effettuato con il Censimento e con gli archivi utilizzati, tra cui tutte le fonti amministrative che hanno concorso alla composizione della lista pre-censuaria. Tale approccio ha quindi consentito sia di stimare sia il tasso di copertura del Censimento, sia di ricercare l'abbinamento di tutte le aziende agricole rilevate alla PES ma non enumerate al Censimento con le unità presenti in tutti gli archivi amministrativi di partenza descritti nel Paragrafo 2.7.1.

Si possono descrivere alcuni casi per cui tali abbinamenti si sono verificati. Ad esempio, una generica azienda agricola, rilevata all'indagine di copertura, non si è abbinata con un'azienda rilevata al Censimento, ma con una presente nella lista pre-censuaria ed esitata come 'irreperibile' al Censimento. Poiché le informazioni rilevate alla PES hanno dimostrato che tale unità è eleggibile ad azienda agricola, questa ha contribuito alla determinazione del tasso di sotto-copertura, nella *componente* 'mancata risposta'. Un altro esempio è dato da una generica azienda rilevata alla PES, e non presente nella lista pre-censuaria, che si è abbinata con una unità agricola presente in uno degli archivi presi in considerazione. In questo caso, l'Istat in fase di preparazione della lista pre-censuaria ha identificato questa unità come 'non eleggibile', mentre le informazioni rilevate dalla PES hanno dimostrato che doveva essere inserita nella lista pre-censuaria ed enumerata dal Censimento. Tale unità ha contribuito alla determinazione del tasso di sotto-copertura, nella *componente* 'in lista'.

In questo modo è stato possibile scomporre la sotto-copertura del Censimento nelle seguenti componenti:

- '0 (NIL)' = azienda non presente in alcun archivio; determinata dalle unità rilevate all'indagine di copertura, ma senza alcun legame con tutte le unità del Censimento, della lista pre-censuaria, o quelle degli altri archivi;
- 'IL' = azienda presente in uno degli archivi presi in esame (esclusa la lista pre-censuaria);
- 'MR' = azienda presente in lista pre-censuaria, ma esitata al Censimento come non rispondente o non esistente (ossia non rilevata, per varie cause);
- 'NU' = azienda presente in lista pre-censuaria, rilevata, ma classificata al Censimento come azienda non eleggibile.

L'applicazione delle procedure di *record linkage* ha consentito la disaggregazione anche del tasso di copertura nelle seguenti componenti:

- '1 C' = azienda rilevata al Censimento;
- 'DE' = demografia di azienda agricola;
- 'LM' = azienda che presenta legami multipli con il Censimento.

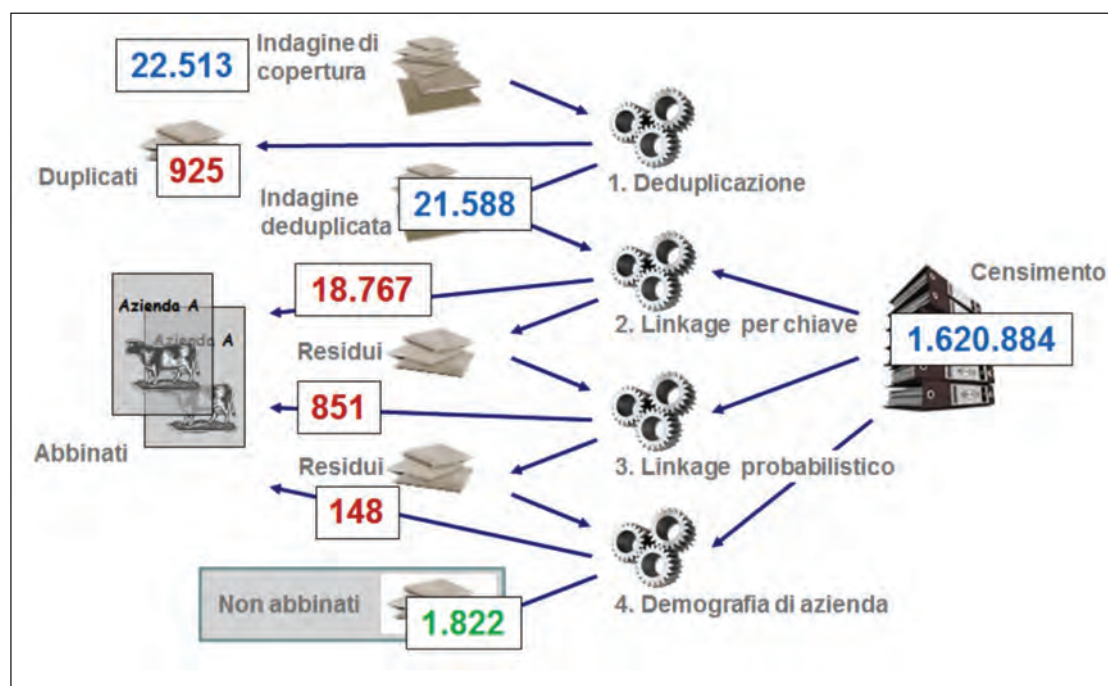
Con il termine "Demografia di azienda agricola" si deve intendere il risultato di un abbinamento tra un'azienda rilevata all'indagine di copertura e una rilevata al Censimento in cui è rispettata una continuità tra le unità agricole coinvolte; ossia è stata individuata la stessa azienda agricola che, tuttavia, nel tempo intercorso tra le due rilevazioni, ha subito un cambiamento nelle variabili anagrafiche, nella forma giuridica oppure nell'estensione della SAU o della superficie totale (ST).

Per “legami multipli con il Censimento” si intende un’unità rilevata alla PES che trova abbinamento con più aziende rilevate al Censimento o viceversa; in questo caso si è ricorsi all’utilizzo dell’archivio del Catasto dei terreni agricoli in cui si è arrivati all’abbinamento univoco tra terreno e azienda agricola, come nel caso di terreni divisi tra più persone della stessa famiglia, che sono complessivamente riconducibili ad un’unica azienda agricola.

2.7.6 Sommario dei risultati del Linkage

Come dettagliato in Figura 2.8 partendo dai 22.513 record rilevati dall’indagine di copertura le aziende non abbinate al termine delle procedure risultano 1.822.

Figura 2.8 - Risultati delle procedure di abbinamento



2.8 I risultati

Vengono ora presentati tutti i risultati dell’indagine di copertura, sia per domini di stima territoriali sia a livello di superficie agricola. Al fine di consentire una migliore lettura, si ricorda, ed è una novità di questa indagine, che il tasso di sotto-copertura⁷² viene scomposto in 4 componenti: azienda non presente in alcun archivio; azienda trovata in liste (esclusa la lista pre-censuaria); azienda esitata al Censimento come mancata risposta, rifiuto, irreperibile, eccetera (codificata come b2 o b3, rispettivamente “Azienda in lista non rilevata” e “Azienda in lista non esistente o doppia”), azienda presente in lista pre-censuaria (classificata non eleggibile al Censimento).

Si deve considerare, inoltre, che le componenti della sotto-copertura ci consentono

⁷² Per la spiegazione dettagliata delle componenti della copertura e della sotto-copertura si rimanda al Paragrafo 2.7.5.



di individuare anche la fase del Censimento in cui si è verificato l'errore. "Azienda non presente in alcuna lista" (NIL) denota inequivocabilmente un problema delle fonti amministrative che non conterrebbero l'universo completo delle aziende presenti in Italia o non sarebbero sufficientemente aggiornate. "Azienda trovata in liste (esclusa la pre-censuaria)" (IL) indica una difficoltà nella fase di individuazione, all'interno di tutte le fonti amministrative, delle unità eleggibili ad aziende agricole. "Azienda esitata al Censimento b2 o b3" (MR) denota un problema della fase di rilevazione, in cui alcune unità sono state o erroneamente classificate oppure non è stato possibile effettuare la rilevazione per rifiuto o irreperibilità del conduttore dell'azienda agricola. "Azienda in lista pre-censuaria ma classificata non eleggibile al Censimento" (NU) mostra un problema della fase di rilevazione e post rilevazione, in quanto delle aziende presenti in lista pre-censuaria sono state escluse dal campo di osservazione sulla base delle regole previste dal piano di controllo e correzione, effettuato al termine del Censimento.

Come si evince, ciascuno dei 4 dettagli di sotto-copertura rappresenta un errore collocabile cronologicamente nelle fasi dell'indagine censuaria. La *componente* NIL indica un problema nella fase di progettazione del Censimento, la *componente* IL mostra un problema nella fase di costruzione della lista pre-censuaria; la *componente* MR evidenzia un problema nella fase di rilevazione; la *componente* NU indica un problema nella fase di controllo e correzione del dato censuario.

Risulta una novità anche la scomposizione del tasso di copertura in tre componenti: azienda rilevata al Censimento; demografia di azienda agricola; azienda che presenta legami multipli con il Censimento. "Azienda rilevata al Censimento" indica l'abbinamento della stessa azienda, per la quale sono state individuate le medesime caratteristiche sia al Censimento sia alla PES. Per "Demografia di azienda agricola" si intende l'abbinamento della stessa azienda nella quale, tuttavia, sono stati riscontrati dei cambiamenti (probabilmente accaduti nel lasso di tempo intercorso tra le due indagini) che hanno modificato qualcuna delle variabili rilevate. La *componente* "Azienda che presenta legami multipli con il Censimento" denota un abbinamento non univoco, in cui, tuttavia, risulta chiara la riconducibilità alla stessa azienda agricola (per esempio, il caso dell'azienda divisa tra componenti dello stesso nucleo familiare).

Si ricorda, inoltre, che tutti i risultati sono espressi in percentuale, in quanto i tassi calcolati dal modello di Petersen (descritto nel Paragrafo 2.4.2) sono moltiplicati per cento.

A causa di una forte caduta campionaria non è stato possibile fornire stime attendibili per la Provincia autonoma di Bolzano.

2.8.1 I risultati a livello nazionale

Le aziende agricole stimate dall'indagine di copertura e realmente esistenti sul territorio nazionale nel periodo temporale di riferimento del Censimento sono 1.851.245. In tale stima sono incluse anche le aziende temporaneamente inattive e sono escluse quelle della Provincia autonoma di Bolzano.

I risultati a livello nazionale, mostrati nella Tavola 2.3 e scomposti per *componente*, spiegano che 3,9 punti percentuali di sotto-copertura sono relativi ad aziende agricole, rilevate alla PES ed esistenti sul territorio nazionale, che non sono state trovate in alcuna lista (o fonte amministrativa) utilizzata per la preparazione della lista pre-censuaria. Il 4,9 per cento di sotto-copertura è spiegato da aziende agricole che hanno trovato un abbinamento con una delle fonti amministrative, ma che non sono state incluse in lista pre-censuaria perché ritenute non eleggibili. Il 2,3 per cento di

sotto-copertura è dovuto a problemi legati all'errata classificazione di unità agricole (orto familiare, non esercita attività agricola, eccetera) o a mancata risposta (rifiuto, irreperibile, eccetera). Lo 0,6 per cento di sotto-copertura è caratterizzato da aziende presenti in lista pre-censuaria ed erroneamente classificate fuori dal campo di osservazione durante la fase di controllo e correzione dei dati censuari.

Tavola 2.3 - Tasso di sotto-copertura del 6° Censimento generale dell'agricoltura per componente - Italia

Componenti della Sotto-Copertura		Tasso Sotto-Copertura (%)
0 (NIL)	Azienda non presente in alcun archivio	3,9
IL	Azienda trovata in un archivio (esclusa la lista precensuaria)	4,9
MR	Azienda in lista precensuaria (non rispondente o non esistente al censimento)	2,3
NU	Azienda in lista precensuaria (non eleggibile al censimento)	0,6

La Tavola 2.4 presenta, a livello nazionale, la scomposizione del tasso di copertura nelle varie componenti. Come si vede, circa 81,4 punti percentuali di copertura sono spiegati da aziende rilevate alle due indagini e aventi, sostanzialmente, caratteristiche identiche. Il 5,2 per cento della copertura è classificabile come "Demografia di azienda agricola", ossia si tratta di unità rilevate alle due indagini che, pur essendo le stesse aziende agricole, presentano delle caratteristiche differenti (ad esempio denominazione diverse della stessa azienda alla quale sottendono legami familiari) probabilmente sopraggiunte nel range temporale di rilevazione. L'1,7 per cento di copertura si riferisce a legami multipli con il Censimento.

Tavola 2.4 - Tasso di copertura del 6° Censimento generale dell'agricoltura per componente - Italia

Componenti della Copertura		Tasso Copertura (%)
1 C	Azienda rilevata al censimento	81,4
DE	Demografia d'azienda agricola	5,2
LM	Azienda che presenta legami multipli con il censimento	1,7

2.8.2 I risultati a livello di ripartizione

La Tavola 2.5 presenta i risultati dei dettagli di sotto-copertura a livello delle 5 ripartizioni geografiche: Nord-ovest, Nord-est, Centro, Sud e Isole.

Nella colonna NIL, in cui sono presenti i valori di sotto-copertura dovuti ad aziende non presenti in alcuna fonte amministrativa, si rilevano risultati superiori al 4,5 per cento per Sud, Centro e Isole, mentre i valori sono di poco superiori all'1 per cento per le due ripartizioni del Nord. Nella colonna IL, ossia aziende presenti in fonti amministrative ma non incluse nella lista pre-censuaria, spicca il valore del Centro pari a circa 8,8 per cento; i valori delle altre ripartizioni oscillano tra il 3,2 per cento del Nord-est e il 5 per cento delle Isole. La quota di sotto-copertura dovuta a errata classificazione o mancata risposta al Censimento (colonna MR) risulta più rilevante nel Centro (3,9 per cento) mentre al Sud e nelle Isole la quota si attesta poco sopra il 2 per cento. L'errore dovuto all'errata classificazione di aziende agricole presenti in lista pre-censuaria (colonna NU) è assente nel Nord-ovest ed è prossimo allo zero nel Nord-est e nelle Isole.



Tavola 2.5 - Tasso di sotto-copertura del 6° Censimento generale dell'agricoltura per componente e ripartizione geografica (%)

RIPARTIZIONI GEOGRAFICHE	0 (NIL) Azienda non presente in alcun archivio	IL Azienda trovata in un archivio (esclusa la lista precensuaria)	MR Azienda in lista pre-censuaria (non rispondente o non esistente al censimento)	NU Azienda in lista precensuaria (non eleggibile al censimento)
Nord-ovest	1,4	3,9	1,6	-
Nord-est	1,3	3,2	1,0	0,1
Centro	5,3	8,8	3,9	0,9
Sud	4,5	4,4	2,4	1,1
Isole	6,0	5,0	2,1	0,1
Italia	3,9	4,9	2,3	0,6

La Tavola 2.6 presenta la scomposizione del tasso di copertura per ripartizioni geografiche. Si evidenzia la forte differenza tra i valori della colonna '1C', che si riferisce alle aziende con legami perfettamente univoci tra le due rilevazioni: mentre le ripartizioni del Nord hanno valori superiori al 90 per cento, al Centro, al Sud e nelle Isole si registrano valori compresi tra il 74 per cento e l'80 per cento. I risultati, inoltre, dimostrano che nelle Isole (e specialmente in Sardegna) sono più diffusi i fenomeni legati alla "Demografia di azienda agricola" e ai "Legami multipli".

Tavola 2.6 - Tasso di sotto-copertura del 6° Censimento generale dell'agricoltura per componente e ripartizione geografica (%)

RIPARTIZIONI GEOGRAFICHE	1 C Azienda rilevata al censimento	DE Demografia d'azienda agricola	LM Azienda che presenta legami multipli con il censimento
Nord-ovest	90,2	2,4	0,4
Nord-est	90,1	3,2	1,1
Centro	74,3	4,8	2,0
Sud	80,2	5,8	1,6
Isole	74,7	8,8	3,3
Italia	81,4	5,2	1,7

2.8.3 I risultati a livello regionale

La numerosità campionaria dei fogli di mappa catastali consente di calcolare i dettagli della copertura e della sotto-copertura a livello regionale, con un errore di campionamento decisamente contenuto. La PES è stata condotta in tutta Italia ma, a causa di un'eccezionale caduta campionaria, e quindi di una conseguente bassa significatività, non è possibile pubblicare i dati per la Provincia autonoma di Bolzano.

La Tavola 2.7 presenta i risultati della sotto-copertura dettagliati a livello regionale.

Nella colonna NIL (aziende non presenti in alcun archivio) si notano valori nulli in Valle d'Aosta ed Emilia-Romagna: in queste due regioni la PES non ha rilevato aziende non presenti nelle fonti amministrative. Risultano molto contenuti i tassi percentuali del Piemonte, del Friuli-Venezia Giulia e delle Marche. Tale dato, letto in ottica *Farm Register*, mostra la bontà delle fonti amministrative utilizzate e continuamente aggiornate in queste regioni.

La quota di sotto-copertura, dovuta ad aziende presenti in liste (IL), è quella che presenta maggiore variabilità; si passa dal valore nullo della Valle d'Aosta all'8,3 per cento

dell'Umbria, al 9,48 per cento del Lazio e al 10,54 per cento del Toscana. In queste ultime tre regioni (tutte del Centro) sembra necessario uno studio più approfondito relativamente ai criteri di eleggibilità delle unità agricole, presenti nelle fonti amministrative, ad aziende agricole. Anche per questo dettaglio di sotto-copertura l'Emilia-Romagna presenta un valore molto basso e prossimo all'unità.

La quota di sotto-copertura spiegata dalla mancata risposta (MR) presenta valori decisamente contenuti per tutte le regioni italiane (tranne che per il Lazio dove si registra circa il 6,5 per cento). Il lusinghiero risultato evidenzia l'elevata qualità complessiva della rilevazione censuaria.

La *componente* NU (aziende presenti in lista pre-censuaria e censite ma escluse dalle procedure censuarie di controllo e correzione) presenta una quota nulla di sotto-copertura per Piemonte, Valle d'Aosta, Lombardia, Friuli Venezia Giulia, Liguria, Emilia-Romagna e Provincia autonoma di Trento. Molto contenuti risultano tutti gli altri valori, tranne nel caso dell'Abruzzo che registra il 2,1 per cento.

Tavola 2.7 - Tasso di sotto-copertura del 6° Censimento generale dell'agricoltura per componente e regione (%)

REGIONI	0 (NIL) Azienda non presente in alcun archivio	IL Azienda trovata in un archivio (esclusa la lista precensuaria)	MR Azienda in lista pre-censuaria (non rispondente o non esistente al censimento)	NU Azienda in lista precensuaria (non eleggibile al censimento)
Piemonte	0,9	5,1	0,1	-
Valle d'Aosta/Vallée d'Aoste	-	-	0,4	-
Liguria	1,5	5,2	1,0	-
Lombardia	2,3	2,1	4,0	-
Bolzano/Bozen		-	-	-
Trento	2,7	2,5	1,4	-
Veneto	1,4	6,3	0,7	0,4
Friuli-Venezia Giulia	0,7	3,9	0,5	-
Emilia-Romagna	-	1,2	1,0	-
Toscana	4,5	10,5	2,9	1,4
Umbria	5,4	8,3	1,1	0,3
Marche	0,9	3,7	1,5	0,3
Lazio	7,5	9,5	6,5	1,0
Abruzzo	1,7	5,0	1,8	2,1
Molise	1,3	2,7	0,1	1,6
Campania	4,8	5,6	1,8	1,7
Puglia	2,9	4,6	2,3	0,4
Basilicata	5,8	4,0	2,7	1,4
Calabria	7,8	3,1	3,8	1,0
Sicilia	5,4	4,8	2,2	0,1
Sardegna	7,3	5,4	1,9	0,2
Italia	3,9	4,9	2,3	0,6

La Tavola 2.8 presenta i valori regionali relativi alla scomposizione del tasso di copertura. Come riscontrato nel caso delle ripartizioni geografiche, la modalità "1 C" presenta una elevata variabilità: si passa dal 65,4 per cento dell'Umbria al 97,1 per cento dell'Emilia-Romagna. Il Molise, tra le regioni del Sud, presenta il valore più elevato di copertura (93,3 per cento) per gli abbinamenti perfetti.



La quota di copertura dovuta alla “Demografia di azienda agricola” risulta sotto l’uno per cento per Emilia-Romagna e Molise e prossima all’unità per Valle d’Aosta e Liguria. Si segnalano ancora i valori più alti riscontrati in Umbria (11,5 per cento), Sardegna (9,1 per cento) e Sicilia (8,7 per cento).

La parte di copertura dovuta a legami multipli è sostanzialmente nulla in Emilia-Romagna e molto contenuta in Lombardia (0,6 per cento), Marche (0,8 per cento), Lazio (0,6 per cento), Abruzzo (0,9 per cento) e Molise (0,5 per cento). Risulta molto ingente la quota di legami multipli registrata in Sardegna (circa il 6 per cento).

Tavola 2.8 - Tasso di sotto-copertura del 6° Censimento generale dell’agricoltura per componente e regione (%)

REGIONI	1 C Azienda rilevata al censimento	DE Demografia d’azienda agricola	LM Azienda che presenta legami multipli con il censimento
Piemonte	92,1	1,7	-
Valle d’Aosta/Vallée d’Aoste	94,5	1,1	4,0
Liguria	89,9	1,1	1,2
Lombardia	87,0	4,0	0,6
Bolzano/Bozen	-	-	-
Trento	88,7	3,9	0,8
Veneto	83,8	5,3	2,1
Friuli-Venezia Giulia	88,9	3,2	2,9
Emilia-Romagna	97,1	0,6	0,1
Toscana	74,1	5,1	1,6
Umbria	65,4	11,5	8,0
Marche	88,2	4,5	0,8
Lazio	72,7	2,3	0,6
Abruzzo	84,4	4,1	0,9
Molise	93,3	0,5	0,5
Campania	80,5	4,0	1,7
Puglia	80,5	7,7	1,5
Basilicata	80,9	3,8	1,4
Calabria	75,4	6,7	2,2
Sicilia	76,8	8,7	2,1
Sardegna	70,2	9,1	6,0
Italia	81,4	5,2	1,7

2.8.4 I risultati per classe di Superficie agricola utilizzata

In questo paragrafo vengono presentati i risultati di sotto-copertura e copertura per aziende suddivise per classi di Superficie agricola utilizzata (SAU) rilevata.

Nella Tavola 2.9 sono presentati i risultati dei dettagli della sotto-copertura nazionale per classi di SAU. Nella classe di SAU pari a zero sono raccolte le aziende agricole esclusivamente zootecniche. Tutte le modalità di scomposizione della sotto-copertura (esclusa la prima riga relativa alle aziende solo zootecniche) presentano una distribuzione dei tassi percentuali decrescente, ossia le quote maggiori di errore sono attribuibili alle classi più basse. Il significato di tale fenomeno è la difficoltà di “catturare” le aziende agricole più piccole dove esiste certamente un problema definitorio: “è azienda agricola oppure no?“. In casi di aziende agricole al limite delle soglie fisiche di eleggibilità, è possibile che rilevatori diversi possano fornire interpretazioni differenti.

Nella modalità “0” (o NIL, aziende non presente in alcun archivio), la sotto-copertura, seppur con qualche lieve salto, passa da 7,3 punti percentuali della classe “fino a 0,99 ettari” alla completa copertura dell’ultima classe dove sono rappresentate le aziende con oltre 100 ettari. Tale fenomeno si presenta anche per le modalità IL e NU. La modalità che raccoglie le mancate risposte (MR) merita certamente un discorso a parte, poiché la distribuzione nelle classi di SAU non sembra rispettare una particolare regola. Ciò dimostra che il fenomeno dell’irreperibilità o del rifiuto è distribuito, più o meno uniformemente, su tutte le classi ad eccezione di quelle con SAU non inferiore a 20 ettari.

Tavola 2.9 - Tasso di sotto-copertura del 6° Censimento generale dell’agricoltura per componente e classe di Superficie agricola utilizzata (SAU) (%)

CLASSI DI SAU (in ettari)	0 (NIL)	IL	MR	NU
	Azienda non presente in alcun archivio	Azienda trovata in un archivio (esclusa la lista precensuaria)	Azienda in lista pre-censuaria (non rispondente o non esistente al censimento)	Azienda in lista precensuaria (non eleggibile al censimento)
0 - (0)	0,5	2,6	0,1	-
1 - (0,01 - 0,99)	7,3	10,2	2,5	1,3
2 - (1 - 1,99)	5,1	4,7	2,6	0,9
3 - (2 - 2,99)	4,2	2,8	2,6	0,3
4 - (3 - 4,99)	1,5	2,7	2,8	0,2
5 - (5 - 9,99)	0,7	0,8	2,0	0,3
6 - (10 - 19,99)	0,8	0,3	1,6	0,1
7 - (20 - 29,99)	0,2	-	2,6	-
8 - (30 - 49,99)	0,4	0,3	1,5	-
9 - (50 - 99,99)	0,5	1,0	1,4	-
10 - (100 ed oltre)	-	-	4,2	-
Italia	3,9	4,9	2,3	0,6

La Tavola 2.10 presenta la *componente* della copertura per classe di SAU delle aziende. Non sembrano esserci distribuzioni con particolare caratteristiche, se non per la modalità “1 C” in cui sono classificati gli abbinamenti univoci. Infatti, per tale modalità il tasso di copertura cresce, a partire dalla classe “0,01 - 0,99 ettari”, all’aumentare della classe di SAU: più l’azienda è grande e maggiore è la di rilevazione al censimento.



Tavola 2.10 - Tasso di copertura del 6° Censimento generale dell'agricoltura per componente e classe di Superficie agricola utilizzata (SAU) (%)

CLASSI DI SAU (in ettari)	1 C Azienda rilevata al censimento	DE Demografia d'azienda agricola	LM Azienda che presenta legami multipli con il censimento
0 - (0)	96,0	0,8	-
1 - (0,01 - 0,99)	71,0	5,6	2,1
2 - (1 - 1,99)	78,2	6,2	2,2
3 - (2 - 2,99)	82,7	5,2	2,2
4 - (3 - 4,99)	86,7	4,7	1,4
5 - (5 - 9,99)	88,7	6,4	1,1
6 - (10 - 19,99)	90,6	4,6	2,0
7 - (20 - 29,99)	92,9	3,8	0,6
8 - (30 - 49,99)	91,4	5,9	0,5
9 - (50 - 99,99)	91,2	5,6	0,2
10 - (100 ed oltre)	88,2	5,7	1,9
Italia	81,4	5,2	1,7

Nella Tavola 2.11 sono state unificate le classi di SAU dalla 2 alla 10, in modo che potesse essere chiaramente esplicitato il tasso di sotto-copertura delle piccole aziende agricole (fino a un ettaro) e quello di tutte le altre (oltre un ettaro).

La quota di sotto-copertura spiegata dalla *componente* NIL (azienda non presente in alcun archivio) mostra come il 7,3 per cento sia riferito alle aziende sotto l'ettaro e solo il 2,6 per cento alle altre. Nella *componente* IL (azienda presente in uno degli archivi presi in esame, esclusa la lista pre-censuaria) il fenomeno è maggiormente evidente: la sotto-copertura delle piccole aziende è pari ad oltre il 10 per cento, mentre quella delle "grandi" è pari al 2,5 per cento. Per la *componente* MR, come detto in precedenza, non sembra esserci una chiara distinzione tra le due classi di SAU, a dimostrazione che il fenomeno è decisamente equi-distribuito. Per la modalità NU si ha una sotto-copertura pari all'1,3 per cento per le aziende agricole entro l'ettaro di SAU e lo 0,4 per cento per le altre.

Tavola 2.11 - Tasso di sotto-copertura del 6° Censimento generale dell'agricoltura per componente e classe di Superficie agricola utilizzata (SAU) (%)

CLASSI DI SAU (in ettari)	0 (NIL) Azienda non presente in alcun archivio	IL Azienda trovata in un archivio (esclusa la lista precensuaria)	MR Azienda in lista pre-censuaria (non rispondente o non esistente al censimento)	NU Azienda in lista precensuaria (non eleggibile al censimento)
0 - (0)	0,5	2,6	0,1	-
1 - (0,01 - 0,99)	7,3	10,2	2,5	1,3
2 - (1,00 e oltre)	2,6	2,5	2,4	0,4
Italia	3,9	4,9	2,3	0,6

La Tavola 2.12 presenta la composizione della copertura per le 3 classi di SAU definite come sopra. Per la *componente* '1C' il tasso di copertura aumenta dalla prima alla seconda classe ossia cresce all'aumentare della dimensione. Per la *componente* DE - Demografia di azienda - non sembra essere presente un particolare trend tra le due classi, così come sembra poco rilevante la differenza nella *componente* LM.

Tavola 2.12 - Tasso di copertura del 6° Censimento generale dell'agricoltura per componente e classe di Superficie agricola utilizzata (SAU) (%)

CLASSI DI SAU (in ettari)	1 C Azienda rilevata al censimento	DE Demografia d'azienda agricola	LM Azienda che presenta legami multipli con il censimento
0 - (0)	96,0	0,8	-
1 - (0,01 - 0,99)	71,0	5,6	2,1
2 - (1,00 e oltre)	84,9	5,5	1,7
Italia	81,4	5,2	1,7

La Tavola 2.13 presenta i risultati nazionali delle componenti di sotto-copertura espressa in termini di Superficie Agricola Utilizzata. Come si vede dalla tavola, la quota di sotto-copertura di SAU dovuta ad aziende trovate alla PES e non presenti in alcuna fonte amministrativa è pari allo 0,9 per cento. La quota di sotto-copertura di SAU dovuta ad aziende rilevate all'indagine di copertura e negli archivi di confronto (esclusa la lista pre-censuaria) è pari all'1 per cento. La porzione di sotto-copertura di SAU imputabile a mancate risposte al Censimento è pari al 2,5 per cento mentre quella dovuta ad aziende rilevate alla PES ed escluse dalla lista censuaria dopo la fase di controllo e correzione è pari allo 0,1 per cento

Tavola 2.13 - Tasso di sotto-copertura della Superficie agricola utilizzata (SAU) per componente - Italia (%)

0 (NIL) Azienda non presente in alcun archivio	IL Azienda trovata in un archivio (esclusa la lista precensuaria)	MR Azienda in lista pre-censuaria (non rispondente o non esistente al censimento)	NU Azienda in lista precensuaria (non eleggibile al censimento)
0,9	1,0	2,5	0,1

2.8.5 La variabilità delle stime di copertura

A completamento dell'attività di stima dei tassi di copertura, è stata calcolata la misura relativa della variabilità di tali stime, espressa con il coefficiente di variazione. Com'è noto, il coefficiente di variazione (espresso in percentuale) è pari al rapporto tra l'errore standard della stima e la stima stessa, moltiplicato per cento.

Gli errori campionari, a livello Italia, di ripartizione e regionale, sono stati calcolati per dettaglio delle singole componenti della copertura e sotto-copertura. I dati sono riportati nell'Appendice 2.A - Variabilità delle stime di copertura' e mostrano valori molto contenuti assicurando l'assoluta significatività dei risultati e l'elevata qualità di tutto il piano di campionamento areale disegnato.



Appendice 2.A - Variabilità delle stime di copertura

Tavola 2.A.1 - Coefficiente di variazione (CV) del tasso di sotto-copertura per componente

Componenti della Sotto-Copertura		CV Tasso Sotto Copertura (%)
0 (NIL)	Azienda non presente in alcun archivio	10,8
IL	Azienda trovata in un archivio (esclusa la lista precensuaria)	6,2
MR	Azienda in lista precensuaria (non rispondente o non esistente al censimento)	6,8
NU	Azienda in lista precensuaria (non eleggibile al censimento)	13,0

Tavola 2.A.2 - Coefficiente di variazione (CV) del tasso di copertura per componente

Componenti della Copertura		CV Tasso Copertura (%)
1 C	Azienda rilevata al censimento	1,0
DE	Demografia d'azienda agricola	5,5
LM	Azienda che presenta legami multipli con il censimento	9,8

Tavola 2.A.3 - Coefficiente di variazione (CV) del tasso di sotto-copertura per componente e ripartizione geografica (%)

RIPARTIZIONI GEOGRAFICHE	0 (NIL) Azienda non presente in alcun archivio	IL Azienda trovata in un archivio (esclusa la lista precensuaria)	MR Azienda in lista pre-censuaria (non rispondente o non esistente al censimento)	NU Azienda in lista precensuaria (non eleggibile al censimento)
Nord-ovest	29,6	17,0	22,8	-
Nord-est	27,0	13,6	24,4	58,2
Centro	20,1	12,0	15,0	25,0
Sud	17,9	10,3	9,2	15,4
Isole	18,1	16,1	18,5	60,5
Italia	10,8	6,2	6,8	13,0

Tavola 2.A.4 - Coefficiente di variazione (CV) del tasso di copertura per componente e ripartizione geografica (%)

RIPARTIZIONI GEOGRAFICHE	1 C Azienda rilevata al censimento	DE Demografia d'azienda agricola	LM Azienda che presenta legami multipli con il censimento
Nord-ovest	1,2	21,4	46,4
Nord-est	1,0	16,0	24,9
Centro	2,9	10,2	33,9
Sud	1,8	7,8	12,6
Isole	3,4	13,1	16,2
Italia	1,0	5,5	9,8

Tavola 2.A.5 - Coefficiente di variazione (CV) del tasso di sotto-copertura per componente e regione (%)

REGIONI	0 (NIL) Azienda non presente in alcun archivio	IL Azienda trovata in un archivio (esclusa la lista precensuaria)	MR Azienda in lista pre-censuaria (non rispondente o non esistente al censimento)	NU Azienda in lista precensuaria (non eleggibile al censimento)
Piemonte	48,6	20,3	101,4	-
Valle d'Aosta/Vallée d'Aoste	-	-	107,0	-
Liguria	61,5	57,2	103,6	-
Lombardia	42,9	33,7	22,7	-
Bolzano/Bozen	-	-	-	-
Trento	38,2	30,7	46,5	-
Veneto	24,3	17,3	39,0	57,4
Friuli-Venezia Giulia	105,3	26,2	82,0	-
Emilia-Romagna	-	36,4	31,9	-
Toscana	17,8	16,3	22,7	29,2
Umbria	31,5	33,3	55,9	104,4
Marche	39,7	38,2	50,2	101,3
Lazio	31,4	20,0	21,1	42,6
Abruzzo	46,0	28,1	32,6	48,0
Molise	76,7	45,8	98,4	72,6
Campania	24,3	17,3	24,0	26,0
Puglia	22,2	21,1	15,2	32,5
Basilicata	68,4	35,4	25,2	41,9
Calabria	33,0	12,9	17,4	27,6
Sicilia	26,8	22,0	20,8	71,4
Sardegna	20,9	20,9	38,2	97,3
Italia	10,8	6,2	6,8	13,0



Tavola 2.A.6 - Coefficiente di variazione (CV) del tasso di copertura per componente e regione (%)

REGIONI	1 C Azienda rilevata al censimento	DE Demografia d'azienda agricola	LM Azienda che presenta legami multipli con il censimento
Piemonte	1,5	41,8	-
Valle d'Aosta/Vallée d'Aoste	4,3	66,4	99,1
Liguria	5,2	72,4	83,9
Lombardia	2,2	23,9	70,1
Bolzano/Bozen	-	-	-
Trento	1,9	26,7	67,8
Veneto	2,4	21,9	24,9
Friuli-Venezia Giulia	1,9	49,7	50,6
Emilia-Romagna	0,5	44,7	99,5
Toscana	3,0	16,8	32,9
Umbria	5,4	12,8	38,8
Marche	3,4	27,6	52,3
Lazio	5,9	25,3	46,8
Abruzzo	2,8	30,9	28,0
Molise	3,0	76,0	65,9
Campania	3,5	19,3	34,7
Puglia	2,5	10,6	18,5
Basilicata	6,7	16,1	45,3
Calabria	5,6	17,8	18,2
Sicilia	4,6	16,4	24,8
Sardegna	4,5	22,1	21,0
Italia	1,0	5,5	9,8

Tavola 2.A.7 - Coefficiente di variazione (CV) del tasso di sotto-copertura per componente e classe di Superficie agricola utilizzata (SAU) (%)

CLASSI DI SAU (in ettari)	0 (NIL)	IL	MR	NU
	Azienda non presente in alcun archivio	Azienda trovata in un archivio (esclusa la lista precensuaria)	Azienda in lista pre-censuaria (non rispondente o non esistente al censimento)	Azienda in lista precensuaria (non eleggibile al censimento)
0 - (0)	49,0	21,6	100,8	-
1 - (0,01 - 0,99)	10,7	6,9	11,8	15,6
2 - (1 -1,99)	14,7	10,1	12,2	23,0
3 - (2 -2,99)	18,6	17,9	17,4	47,1
4 - (3 - 4,99)	30,6	21,3	18,4	31,6
5 - (5 - 9,99)	29,4	30,8	17,9	40,2
6 - (10 - 19,99)	53,6	54,9	27,4	92,1
7 - (20 - 29,99)	96,5	-	33,7	-
8 - (30 - 49,99)	78,6	95,3	37,8	-
9 - (50 - 99,99)	71,3	61,8	45,2	-
10 - (100 ed oltre)	-	-	58,8	-
Italia	10,8	6,2	6,8	13,0

Tavola 2.A.8 - Coefficiente di variazione (CV) del tasso di copertura per componente e classe di Superficie agricola utilizzata (SAU) (%)

CLASSI DI SAU (in ettari)	1 C	DE	LM
	Azienda rilevata al censimento	Demografia d'azienda agricola	Azienda che presenta legami multipli con il censimento
0 - (0)	0,8	41,0	-
1 - (0,01 - 0,99)	1,9	8,4	12,2
2 - (1 -1,99)	1,7	8,3	15,4
3 - (2 -2,99)	1,7	13,7	22,9
4 - (3 - 4,99)	1,6	13,7	32,3
5 - (5 - 9,99)	1,2	12,2	26,2
6 - (10 - 19,99)	1,4	15,9	40,7
7 - (20 - 29,99)	1,5	29,4	48,4
8 - (30 - 49,99)	1,9	26,2	67,6
9 - (50 - 99,99)	2,1	30,0	95,1
10 - (100 ed oltre)	3,7	42,3	50,7
Italia	1,0	5,5	9,8



Tavola 2.A.9 - Coefficiente di variazione (CV) del tasso di sotto-copertura per componente e classe di Superficie agricola utilizzata (SAU) (%)

CLASSI DI SAU (in ettari)	0 (NIL)	IL	MR	NU
	Azienda non presente in alcun archivio	Azienda trovata in un archivio (esclusa la lista precensuaria)	Azienda in lista pre-censuaria (non rispondente o non esistente al censimento)	Azienda in lista precensuaria (non eleggibile al censimento)
0 - (0)	49,0	21,6	100,8	-
1 - (0,01 - 0,99)	10,7	6,9	11,8	15,6
2 - (1,00 e oltre)	14,2	8,7	8,2	18,7
Italia	10,8	6,2	6,8	13,0

Tavola 2.A.10 - Coefficiente di variazione (CV) del tasso di copertura per componente e classe di Superficie agricola utilizzata (SAU) (%)

CLASSI DI SAU (in ettari)	1 C	DE	LM
	Azienda rilevata al censimento	Demografia d'azienda agricola	Azienda che presenta legami multipli con il censimento
0 - (0)	0,8	41,0	-
1 - (0,01 - 0,99)	1,9	8,4	12,2
2 - (1,00 e oltre)	1,0	6,4	13,8
Italia	1,0	5,5	9,8

Appendice 2.B - Linearizzazione degli stimatori ${}_d\tilde{N}$, ${}_d\tilde{T}$, ${}_d^Y\tilde{T}$ e ${}_d^Y\tilde{S}$

Per quanto illustrato nel Paragrafo 2.4.2, le espressioni (4), (5), (6), (7) e (8) evidenziano che gli stimatori ${}_d\tilde{T}$, ${}_d\tilde{S}$, ${}_d\tilde{N}$, ${}_d^Y\tilde{T}$ e ${}_d^Y\tilde{S}$ sono funzioni non lineari degli stimatori lineari non distorti ${}_d\hat{X}_{+1}$ (comune a tutti), ${}_d\hat{X}_{11}$ (presente in ${}_d\tilde{T}$ e ${}_d\tilde{N}$), ${}_d\hat{X}_{21}$ (presente in ${}_d\tilde{S}$), ${}_d^Y\hat{X}_{11}$ e ${}_d^Y\hat{X}_{21}$ (presenti rispettivamente in ${}_d^Y\tilde{T}$ e ${}_d^Y\tilde{S}$) e del totale ${}_d\hat{X}_{+}$ (presente in ${}_d\tilde{N}$); si pone, quindi, il problema della determinazione dei corrispondenti valori attesi e varianze.

Tale problema è risolto ricorrendo al *metodo della linearizzazione* che consiste nell'approssimare lo stimatore non lineare con il suo *sviluppo in serie di Taylor*.

Al fine di illustrare il suddetto metodo, con riferimento al generico dominio d ($d = 1, \dots, D$), si indichi con ${}_d\tilde{\psi} = f({}_d\hat{\theta})$ uno stimatore del parametro ${}_d\psi$, in cui f è una funzione non lineare; ${}_d\hat{\theta} = ({}_d\hat{\theta}_1, \dots, {}_d\hat{\theta}_Q)$ e il generico ${}_d\hat{\theta}_q$ ($q = 1, \dots, Q$) è uno stimatore lineare non distorto del totale ${}_d\theta_q$ ($q = 1, \dots, Q$) della variabile ${}_d\Theta_q$ ($q = 1, \dots, Q$).

La condizione richiesta per l'applicabilità del metodo è che la funzione di più variabili f sia differenziabile almeno fino al secondo ordine in un intorno sufficientemente ampio del punto dello spazio euclideo q -dimensionale individuato dal vettore ${}_d\theta = ({}_d\theta_1, \dots, {}_d\theta_Q)$.

Lo sviluppo in serie di Taylor di ${}_d\tilde{\psi} = f({}_d\hat{\theta})$ intorno a ${}_d\theta$ rispetto alle variabili ${}_d\hat{\theta}_q$ conduce all'identità:

$${}_d\tilde{\psi} = f({}_d\theta) + \sum_{q=1}^Q g_q({}_d\theta)({}_d\hat{\theta}_q - {}_d\theta_q) + R_2 \quad (26)$$

dove:

$$g_q({}_d\theta) = \left[\frac{\partial f({}_d\hat{\theta})}{\partial {}_d\hat{\theta}_q} \right]_{{}_d\hat{\theta} = {}_d\theta}$$

è il valore della derivata parziale di ${}_d\tilde{\psi}$ rispetto a ${}_d\hat{\theta}_q$ calcolata nel punto ${}_d\theta$ e R_2 è il resto della formula di Taylor, funzione dei termini dello sviluppo di ordine superiore al primo. Quando la dimensione n del campione è sufficientemente elevata, R_2 è trascurabile rispetto ai termini⁷³ lineari.

Pertanto, essendo $f({}_d\theta) = {}_d\psi$, tenendo conto della (26), si può scrivere:

$${}_d\tilde{\psi} - {}_d\psi \cong \sum_{q=1}^Q g_q({}_d\theta)({}_d\hat{\theta}_q - {}_d\theta_q) \quad (27)$$

⁷³ È bene sottolineare che dovendo in tale contesto calcolare la varianza degli stimatori (4), (5), (6), (7) e (8), è sufficiente sviluppare la serie fino ai termini di ordine lineare. I termini di ordine quadratico sono necessari solitamente quando si vuole calcolare il valore atteso di una funzione non lineare.



Calcolando ora il valore atteso⁷⁴ di entrambi i membri della (27), si ha:

$$E({}_d\tilde{\psi}) - {}_d\psi \cong \sum_{q=1}^Q g_q({}_d\theta) [E({}_d\hat{\theta}_q) - {}_d\theta_q] = 0$$

dalla quale si desume che ${}_d\tilde{\psi}$ è uno stimatore approssimativamente corretto di ${}_d\psi$. Pertanto, elevando al quadrato entrambi i membri della (27) e calcolandone i valori attesi, si ottiene:

$$E({}_d\tilde{\psi} - {}_d\psi)^2 = MSE({}_d\tilde{\psi}) = V({}_d\tilde{\psi}) \cong V\left[\sum_{q=1}^Q g_q({}_d\theta) {}_d\hat{\theta}_q\right] \quad (28)$$

in cui $MSE({}_d\tilde{\psi})$ è il *mean squared error*, ossia l'errore quadratico medio, di ${}_d\tilde{\psi}$ che, data l'ipotesi di dimensione campionaria sufficientemente elevata, è anche uguale alla sua varianza.

Alla luce di quanto illustrato, contestualizziamo il suddetto metodo, applicandolo agli stimatori citati all'inizio di questa Appendice, partendo però - per semplicità espositiva - dallo stimatore ${}_d\tilde{N}$.

Dalla (6) si evince che ${}_d\tilde{N}$ è funzione non lineare di ${}_d\hat{X}_{+1}$ e ${}_d\hat{X}_{11}$; pertanto, sviluppando ${}_d\tilde{N} = {}_d\tilde{N}({}_d\hat{X}_{+1}, {}_d\hat{X}_{11})$ in serie di Taylor nell'intorno del punto di coordinate $({}_dX_{+1}, {}_dX_{11})$

e limitando lo sviluppo ai termini di ordine lineare, si ottiene:

$$\begin{aligned} {}_d\tilde{N} &\cong \left[{}_d\tilde{N}\right]_{\substack{{}_d\hat{X}_{+1} = {}_dX_{+1} \\ {}_d\hat{X}_{11} = {}_dX_{11}}} + \left[\frac{\partial {}_d\tilde{N}}{\partial {}_d\hat{X}_{+1}}\right]_{\substack{{}_d\hat{X}_{+1} = {}_dX_{+1} \\ {}_d\hat{X}_{11} = {}_dX_{11}}} ({}_d\hat{X}_{+1} - {}_dX_{+1}) + \left[\frac{\partial {}_d\tilde{N}}{\partial {}_d\hat{X}_{11}}\right]_{\substack{{}_d\hat{X}_{+1} = {}_dX_{+1} \\ {}_d\hat{X}_{11} = {}_dX_{11}}} ({}_d\hat{X}_{11} - {}_dX_{11}) = \\ &= {}_d\hat{N} + \left[\frac{\partial {}_d\tilde{N}}{\partial {}_d\hat{X}_{11}}\right]_{\substack{{}_d\hat{X}_{+1} = {}_dX_{+1} \\ {}_d\hat{X}_{11} = {}_dX_{11}}} ({}_d\hat{X}_{11} - {}_dX_{11}) + \left[-\frac{\partial {}_d\tilde{N}}{\partial {}_d\hat{X}_{11}^2}\right]_{\substack{{}_d\hat{X}_{+1} = {}_dX_{+1} \\ {}_d\hat{X}_{11} = {}_dX_{11}}} ({}_d\hat{X}_{11} - {}_dX_{11}) = \\ &= {}_d\hat{N} + \frac{\partial {}_d\tilde{N}}{\partial {}_d\hat{X}_{11}} \frac{\partial {}_d\tilde{N}}{\partial {}_d\hat{X}_{+1}} - \frac{\partial {}_d\tilde{N}}{\partial {}_d\hat{X}_{11}} \frac{\partial {}_d\tilde{N}}{\partial {}_d\hat{X}_{+1}} - \frac{\partial {}_d\tilde{N}}{\partial {}_d\hat{X}_{11}^2} \frac{\partial {}_d\tilde{N}}{\partial {}_d\hat{X}_{11}} + \frac{\partial {}_d\tilde{N}}{\partial {}_d\hat{X}_{11}} = \\ &= {}_d\hat{N} + \frac{\partial {}_d\tilde{N}}{\partial {}_d\hat{X}_{11}} \frac{\partial {}_d\tilde{N}}{\partial {}_d\hat{X}_{+1}} - \frac{\partial {}_d\tilde{N}}{\partial {}_d\hat{X}_{11}} \frac{\partial {}_d\tilde{N}}{\partial {}_d\hat{X}_{11}} \end{aligned}$$

Per ciò che riguarda ${}_d\tilde{\tau}$ e ${}_d\tilde{\xi}$, si illustrerà solo lo sviluppo del primo dei due, poiché la varianza del secondo stimatore non è necessario derivarla in quanto vale la relazione:

$$V_p\{{}_d\tilde{\xi}\} = V_p\{1 - {}_d\tilde{\tau}\} = V_p\{{}_d\tilde{\tau}\}.$$

74 Nelle formule che seguono, per non appesantire ulteriormente la notazione simbolica, è stato ommesso il pedice p nei simboli indicanti gli operatori di valore atteso, varianza ed errore quadratico medio anche se sono da intendersi rispetto al disegno campionario.

Sviluppando ${}_d\tilde{\tau} = {}_d\tilde{\tau}({}_d\hat{X}_{+1}, {}_d\hat{X}_{11})$ in serie di Taylor nell'intorno del punto di coordinate $({}_dX_{+1}, {}_dX_{11})$ e considerando solo i termini di ordine lineare, si ha:

$$\begin{aligned} {}_d\tilde{\tau} &\cong \left[{}_d\tilde{\tau} \right]_{\substack{{}_d\hat{X}_{+1} = {}_dX_{+1} \\ {}_d\hat{X}_{11} = {}_dX_{11}}} + \left[\frac{\partial {}_d\tilde{\tau}}{\partial {}_d\hat{X}_{+1}} \right]_{\substack{{}_d\hat{X}_{+1} = {}_dX_{+1} \\ {}_d\hat{X}_{11} = {}_dX_{11}}} ({}_d\hat{X}_{+1} - {}_dX_{+1}) + \left[\frac{\partial {}_d\tilde{\tau}}{\partial {}_d\hat{X}_{11}} \right]_{\substack{{}_d\hat{X}_{+1} = {}_dX_{+1} \\ {}_d\hat{X}_{11} = {}_dX_{11}}} ({}_d\hat{X}_{11} - {}_dX_{11}) = \\ &= {}_d\hat{\tau} + \left[-\frac{{}_d\hat{X}_{11}}{{}_d\hat{X}_{+1}^2} \right]_{\substack{{}_d\hat{X}_{+1} = {}_dX_{+1} \\ {}_d\hat{X}_{11} = {}_dX_{11}}} ({}_d\hat{X}_{+1} - {}_dX_{+1}) + \left[\frac{1}{{}_d\hat{X}_{+1}} \right]_{\substack{{}_d\hat{X}_{+1} = {}_dX_{+1} \\ {}_d\hat{X}_{11} = {}_dX_{11}}} ({}_d\hat{X}_{11} - {}_dX_{11}) = \\ &= {}_d\hat{\tau} - \frac{{}_dX_{11}}{{}_dX_{+1}^2} {}_d\hat{X}_{+1} + \frac{{}_dX_{11}}{{}_dX_{+1}} + \frac{1}{{}_dX_{+1}} {}_d\hat{X}_{11} - \frac{{}_dX_{11}}{{}_dX_{+1}} = \\ &= {}_d\hat{\tau} - \frac{{}_dX_{11}}{{}_dX_{+1}^2} {}_d\hat{X}_{+1} + \frac{{}_d\hat{X}_{11}}{{}_dX_{+1}} \end{aligned}$$

Passando ora agli stimatori dei tassi di copertura e sotto-copertura per *componente*, indicati rispettivamente con ${}^Y\tilde{S}$ e ${}^Y\tilde{S}$, considerando il fatto che la forma funzionale non lineare di essi è analoga a quella dello stimatore ${}_d\tilde{\tau}$, il loro sviluppo in serie di Taylor limitato ai termini di ordine lineare nell'intorno del punto di coordinate rispettivamente $({}_dX_{+1}, {}^YX_{11})$ e $({}_dX_{+1}, {}^YX_{21})$ si può ricavare facilmente sostituendo ${}^Y\hat{X}_{11}$ per il primo e ${}^Y\hat{X}_{21}$ per il secondo al posto di ${}_d\hat{X}_{11}$; pertanto, non verrà dimostrato.

Appendice 2.C - Varianze e stime delle varianze degli stimatori

${}_d\tilde{N}$, ${}_d\tilde{\tau}$, ${}^Y\tilde{\tau}$ e ${}^Y\tilde{S}$

Nell'Appendice 2.B si è visto come il problema della determinazione della varianza degli stimatori (4), (5), (6), (7) e (8) - che hanno una forma funzionale non lineare - possa essere risolto linearizzandoli attraverso lo sviluppo in serie di Taylor calcolato nell'intorno del punto di coordinate date dal valore atteso rispetto al disegno di campionamento degli stimatori di cui essi sono funzioni.

In questa Appendice descriveremo, quindi, come calcolare le varianze delle forme linearizzate degli stimatori ${}_d\tilde{\tau}$, ${}_d\tilde{S}$, ${}_d\tilde{N}$, ${}^Y\tilde{\tau}$ e ${}^Y\tilde{S}$.

Per una migliore comprensione degli sviluppi algebrici che seguono, è conveniente ripartire dalla (28), di cui riscriviamo le ultime due espressioni:

$$V({}_d\tilde{\psi}) \cong V \left[\sum_{q=1}^Q g_q({}_d\theta) {}_d\hat{\theta}_q \right].$$



Come emerge dall'espressione testé riportata, la varianza dello stimatore ${}_d\tilde{\psi}$ è approssimabile con la varianza della combinazione lineare degli stimatori ${}_d\hat{\theta}_1, \dots, {}_d\hat{\theta}_Q$ con coefficienti dati dalle derivate $g_q({}_d\theta)$ ($q = 1, \dots, Q$).

Si può, pertanto, applicare la nota formula della varianza di una combinazione lineare:

$$V({}_d\tilde{\psi}) \cong V\left[\sum_{q=1}^Q g_q({}_d\theta) {}_d\hat{\theta}_q\right] = \sum_{q=1}^Q [g_q({}_d\theta)]^2 V({}_d\hat{\theta}_q) + 2 \sum_{q=1}^{Q-1} \sum_{q' < q} g_q({}_d\theta) g_{q'}({}_d\theta) C({}_d\hat{\theta}_q, {}_d\hat{\theta}_{q'}). \quad (29)$$

La varianza di ${}_d\tilde{\psi}$ è quindi approssimata con una funzione lineare delle varianze e covarianze dei ${}_d\hat{\theta}_q$.

Ci preme sottolineare che, in generale, per costruire uno stimatore di $V({}_d\tilde{\psi})$ non basta sostituire nel secondo membro della (29) gli usuali stimatori delle varianze e covarianze dei ${}_d\hat{\theta}_q$, poiché le derivate parziali che vi figurano sono calcolate nel punto ${}_d\theta$, che è incognito. Tuttavia, nel contesto in esame⁷⁵, i ${}_d\hat{\theta}_q$ ($q = 1, \dots, Q$) sono rappresentati dagli stimatori ${}_d\hat{X}_{+1}, {}_d\hat{X}_{11}, {}_d\hat{X}_{21}, {}_d^Y\hat{X}_{11}$ e ${}_d^Y\hat{X}_{21}$ che sono stimatori non distorti rispettivamente di ${}_dX_{+1}, {}_dX_{11}, {}_dX_{21}, {}_d^YX_{11}$ e ${}_d^YX_{21}$. Quindi, i loro valori attesi sono proprio uguali ai corrispondenti totali; pertanto, possono essere sostituiti a questi ultimi.

Operando in tal modo, si ottiene lo stimatore della varianza

$$V({}_d\tilde{\psi}) \cong V\left[\sum_{q=1}^Q g_q({}_d\hat{\theta}) {}_d\hat{\theta}_q\right] = \sum_{q=1}^Q [g_q({}_d\hat{\theta})]^2 V({}_d\hat{\theta}_q) + 2 \sum_{q=1}^{Q-1} \sum_{q' < q} g_q({}_d\hat{\theta}) g_{q'}({}_d\hat{\theta}) C({}_d\hat{\theta}_q, {}_d\hat{\theta}_{q'})$$

in cui $g_q({}_d\hat{\theta})$ è il valore della derivata parziale di f nel punto ${}_d\hat{\theta}$.

Ciò premesso, di seguito sono descritti gli sviluppi che portano a determinare le espressioni (16), (21), (23) e (24) che compaiono nel Paragrafo 2.4.4.

A tal fine, conviene innanzitutto esplicitare le derivate $g_q({}_d\theta)$ ($q = 1, \dots, Q$) che compariranno come coefficienti nell'espressione della varianza (29) applicata ai su citati stimatori:

⁷⁵ In ogni caso, per campioni di dimensioni sufficientemente elevate, è ragionevole ritenere che le stime ${}_d\hat{\theta}_q$ ($q = 1, \dots, Q$) siano prossime ai rispettivi valori attesi ${}_d\theta_q$ ($q = 1, \dots, Q$) e, pertanto, che la sostituzione dei primi con questi ultimi sia lecita.

$$\begin{aligned}
\text{(a.1)} \quad & \left[\frac{\partial {}_d \tilde{N}}{\partial {}_d \hat{X}_{+1}} \right]_{\substack{{}_d \hat{X}_{+1} = {}_d X_{+1} \\ {}_d \hat{X}_1 = {}_d X_1}} = \left[\frac{{}_d X_{+1}}{{}_d \hat{X}_1} \right]_{\substack{{}_d \hat{X}_{+1} = {}_d X_{+1} \\ {}_d \hat{X}_1 = {}_d X_1}} = \frac{{}_d X_{+1}}{{}_d X_1} \\
\text{(a.2)} \quad & \left[\frac{\partial {}_d \tilde{N}}{\partial {}_d \hat{X}_1} \right]_{\substack{{}_d \hat{X}_{+1} = {}_d X_{+1} \\ {}_d \hat{X}_1 = {}_d X_1}} = \left[-\frac{{}_d X_{+1} + {}_d \hat{X}_{+1}}{{}_d \hat{X}_1^2} \right]_{\substack{{}_d \hat{X}_{+1} = {}_d X_{+1} \\ {}_d \hat{X}_1 = {}_d X_1}} = -\frac{{}_d X_{+1} + {}_d X_{+1}}{{}_d X_1^2} \\
\text{(b.1)} \quad & \left[\frac{\partial {}_d \tilde{T}}{\partial {}_d \hat{X}_{+1}} \right]_{\substack{{}_d \hat{X}_{+1} = {}_d X_{+1} \\ {}_d \hat{X}_1 = {}_d X_1}} = \left[-\frac{{}_d \hat{X}_1}{{}_d \hat{X}_{+1}^2} \right]_{\substack{{}_d \hat{X}_{+1} = {}_d X_{+1} \\ {}_d \hat{X}_1 = {}_d X_1}} = -\frac{{}_d X_1}{{}_d X_{+1}^2} \\
\text{(b.2)} \quad & \left[\frac{\partial {}_d \tilde{T}}{\partial {}_d \hat{X}_1} \right]_{\substack{{}_d \hat{X}_{+1} = {}_d X_{+1} \\ {}_d \hat{X}_1 = {}_d X_1}} = \left[\frac{1}{{}_d \hat{X}_{+1}} \right]_{\substack{{}_d \hat{X}_{+1} = {}_d X_{+1} \\ {}_d \hat{X}_1 = {}_d X_1}} = \frac{1}{{}_d X_{+1}} \\
\text{(c.1)} \quad & \left[\frac{\partial {}_d^Y \tilde{T}}{\partial {}_d^Y \hat{X}_{+1}} \right]_{\substack{{}_d^Y \hat{X}_{+1} = {}_d^Y X_{+1} \\ {}_d^Y \hat{X}_1 = {}_d^Y X_1}} = \left[-\frac{{}_d^Y \hat{X}_1}{{}_d^Y \hat{X}_{+1}^2} \right]_{\substack{{}_d^Y \hat{X}_{+1} = {}_d^Y X_{+1} \\ {}_d^Y \hat{X}_1 = {}_d^Y X_1}} = -\frac{{}_d^Y X_1}{{}_d^Y X_{+1}^2} \\
\text{(c.2)} \quad & \left[\frac{\partial {}_d^Y \tilde{T}}{\partial {}_d^Y \hat{X}_1} \right]_{\substack{{}_d^Y \hat{X}_{+1} = {}_d^Y X_{+1} \\ {}_d^Y \hat{X}_1 = {}_d^Y X_1}} = \left[\frac{1}{{}_d^Y \hat{X}_{+1}} \right]_{\substack{{}_d^Y \hat{X}_{+1} = {}_d^Y X_{+1} \\ {}_d^Y \hat{X}_1 = {}_d^Y X_1}} = \frac{1}{{}_d^Y X_{+1}} \\
\text{(d.1)} \quad & \left[\frac{\partial {}_d^Y \tilde{S}}{\partial {}_d^Y \hat{X}_{+1}} \right]_{\substack{{}_d^Y \hat{X}_{+1} = {}_d^Y X_{+1} \\ {}_d^Y \hat{X}_2 = {}_d^Y X_2}} = \left[-\frac{{}_d^Y \hat{X}_{21}}{{}_d^Y \hat{X}_{+1}^2} \right]_{\substack{{}_d^Y \hat{X}_{+1} = {}_d^Y X_{+1} \\ {}_d^Y \hat{X}_2 = {}_d^Y X_2}} = -\frac{{}_d^Y X_{21}}{{}_d^Y X_{+1}^2} \\
\text{(d.2)} \quad & \left[\frac{\partial {}_d^Y \tilde{S}}{\partial {}_d^Y \hat{X}_2} \right]_{\substack{{}_d^Y \hat{X}_{+1} = {}_d^Y X_{+1} \\ {}_d^Y \hat{X}_2 = {}_d^Y X_2}} = \left[\frac{1}{{}_d^Y \hat{X}_1} \right]_{\substack{{}_d^Y \hat{X}_{+1} = {}_d^Y X_{+1} \\ {}_d^Y \hat{X}_2 = {}_d^Y X_2}} = \frac{1}{{}_d^Y X_1}
\end{aligned}$$

Una volta determinate tali derivate, è agevole applicare la (29) agli stimatori ${}_d \tilde{N}$, ${}_d \tilde{T}$, ${}_d^Y \tilde{T}$ e ${}_d^Y \tilde{S}$; infatti, con semplici passaggi, si ha:

$$\begin{aligned}
\text{(a)} \quad V({}_d \tilde{N}) &= \left\{ \left[\frac{\partial {}_d \tilde{N}}{\partial {}_d \hat{X}_{+1}} \right]_{\substack{{}_d \hat{X}_{+1} = {}_d X_{+1} \\ {}_d \hat{X}_1 = {}_d X_1}} \right\}^2 V({}_d \hat{X}_{+1}) + \left\{ \left[\frac{\partial {}_d \tilde{N}}{\partial {}_d \hat{X}_1} \right]_{\substack{{}_d \hat{X}_{+1} = {}_d X_{+1} \\ {}_d \hat{X}_1 = {}_d X_1}} \right\}^2 V({}_d \hat{X}_1) + \\
&+ 2 \left\{ \left[\frac{\partial {}_d \tilde{N}}{\partial {}_d \hat{X}_{+1}} \right]_{\substack{{}_d \hat{X}_{+1} = {}_d X_{+1} \\ {}_d \hat{X}_1 = {}_d X_1}} \right\} \left\{ \left[\frac{\partial {}_d \tilde{N}}{\partial {}_d \hat{X}_1} \right]_{\substack{{}_d \hat{X}_{+1} = {}_d X_{+1} \\ {}_d \hat{X}_1 = {}_d X_1}} \right\} C({}_d \hat{X}_{+1}, {}_d \hat{X}_1) = \\
&= \frac{{}_d X_{+1}^2}{{}_d X_1^2} V({}_d \hat{X}_{+1}) + \frac{{}_d X_{+1}^2 + {}_d X_{+1}^2}{{}_d X_1^4} V({}_d \hat{X}_1) - 2 \frac{{}_d X_{+1}^2 + {}_d X_{+1}}{{}_d X_1^3} C({}_d \hat{X}_{+1}, {}_d \hat{X}_1) = \\
&= {}_d \hat{N}^2 \left[\frac{V({}_d \hat{X}_{+1})}{{}_d X_{+1}^2} + \frac{V({}_d \hat{X}_1)}{{}_d X_1^2} - 2 \frac{C({}_d \hat{X}_{+1}, {}_d \hat{X}_1)}{{}_d X_{+1} {}_d X_1} \right],
\end{aligned}$$

per ${}_d \tilde{N}$;



$$\begin{aligned}
(b) \quad V_{\frac{d}{d}\tilde{\mathcal{T}}} &\equiv \left\{ \left[\frac{\partial \frac{d}{d}\tilde{\mathcal{T}}}{\partial \frac{d}{d}\hat{X}_{+1}} \right]_{\substack{d\hat{X}_{+1}=dX_{+1} \\ d\hat{X}_{11}=dX_{11}}} \right\}^2 V_{\frac{d}{d}\hat{X}_{+1}} + \left\{ \left[\frac{\partial \frac{d}{d}\tilde{\mathcal{T}}}{\partial \frac{d}{d}\hat{X}_{11}} \right]_{\substack{d\hat{X}_{+1}=dX_{+1} \\ d\hat{X}_{11}=dX_{11}}} \right\}^2 V_{\frac{d}{d}\hat{X}_{11}} + \\
&+ 2 \left\{ \left[\frac{\partial \frac{d}{d}\tilde{\mathcal{T}}}{\partial \frac{d}{d}\hat{X}_{+1}} \right]_{\substack{d\hat{X}_{+1}=dX_{+1} \\ d\hat{X}_{11}=dX_{11}}} \right\} \left\{ \left[\frac{\partial \frac{d}{d}\tilde{\mathcal{T}}}{\partial \frac{d}{d}\hat{X}_{11}} \right]_{\substack{d\hat{X}_{+1}=dX_{+1} \\ d\hat{X}_{11}=dX_{11}}} \right\} c_{\left(\frac{d}{d}\hat{X}_{+1}, \frac{d}{d}\hat{X}_{11}\right)} = \\
&= \frac{d}{d} \frac{X_{+1}^2}{X_{+1}^4} V_{\frac{d}{d}\hat{X}_{+1}} + \frac{1}{d} \frac{1}{X_{+1}^2} V_{\frac{d}{d}\hat{X}_{11}} - 2 \frac{d}{d} \frac{X_{11}}{X_{+1}^3} c_{\left(\frac{d}{d}\hat{X}_{+1}, \frac{d}{d}\hat{X}_{11}\right)}
\end{aligned}$$

per $\frac{d}{d}\tilde{\mathcal{T}}$;

$$\begin{aligned}
(c) \quad V_{\frac{d}{d}^Y\tilde{\mathcal{T}}} &\equiv \left\{ \left[\frac{\partial \frac{d}{d}^Y\tilde{\mathcal{T}}}{\partial \frac{d}{d}\hat{X}_{+1}} \right]_{\substack{d\hat{X}_{+1}=dX_{+1} \\ d\hat{X}_{11}=dX_{11}}} \right\}^2 V_{\frac{d}{d}\hat{X}_{+1}} + \left\{ \left[\frac{\partial \frac{d}{d}^Y\tilde{\mathcal{T}}}{\partial \frac{d}{d}^Y\hat{X}_{11}} \right]_{\substack{d\hat{X}_{+1}=dX_{+1} \\ d\hat{X}_{11}=dX_{11}}} \right\}^2 V_{\frac{d}{d}^Y\hat{X}_{11}} + \\
&+ 2 \left\{ \left[\frac{\partial \frac{d}{d}^Y\tilde{\mathcal{T}}}{\partial \frac{d}{d}\hat{X}_{+1}} \right]_{\substack{d\hat{X}_{+1}=dX_{+1} \\ d\hat{X}_{11}=dX_{11}}} \right\} \left\{ \left[\frac{\partial \frac{d}{d}^Y\tilde{\mathcal{T}}}{\partial \frac{d}{d}^Y\hat{X}_{11}} \right]_{\substack{d\hat{X}_{+1}=dX_{+1} \\ d\hat{X}_{11}=dX_{11}}} \right\} c_{\left(\frac{d}{d}\hat{X}_{+1}, \frac{d}{d}^Y\hat{X}_{11}\right)} = \\
&= \frac{d}{d} \frac{X_{+1}^2}{X_{+1}^4} V_{\frac{d}{d}\hat{X}_{+1}} + \frac{1}{d} \frac{1}{X_{+1}^2} V_{\frac{d}{d}^Y\hat{X}_{11}} - 2 \frac{d}{d} \frac{X_{11}}{X_{+1}^3} c_{\left(\frac{d}{d}\hat{X}_{+1}, \frac{d}{d}^Y\hat{X}_{11}\right)}
\end{aligned}$$

per $\frac{d}{d}^Y\tilde{\mathcal{T}}$;

$$\begin{aligned}
(d) \quad V_{\frac{d}{d}^Y\tilde{\mathcal{S}}} &\equiv \left\{ \left[\frac{\partial \frac{d}{d}^Y\tilde{\mathcal{S}}}{\partial \frac{d}{d}\hat{X}_{+1}} \right]_{\substack{d\hat{X}_{+1}=dX_{+1} \\ d\hat{X}_{21}=dX_{21}}} \right\}^2 V_{\frac{d}{d}\hat{X}_{+1}} + \left\{ \left[\frac{\partial \frac{d}{d}^Y\tilde{\mathcal{S}}}{\partial \frac{d}{d}^Y\hat{X}_{21}} \right]_{\substack{d\hat{X}_{+1}=dX_{+1} \\ d\hat{X}_{21}=dX_{21}}} \right\}^2 V_{\frac{d}{d}^Y\hat{X}_{21}} + \\
&+ 2 \left\{ \left[\frac{\partial \frac{d}{d}^Y\tilde{\mathcal{S}}}{\partial \frac{d}{d}\hat{X}_{+1}} \right]_{\substack{d\hat{X}_{+1}=dX_{+1} \\ d\hat{X}_{21}=dX_{21}}} \right\} \left\{ \left[\frac{\partial \frac{d}{d}^Y\tilde{\mathcal{S}}}{\partial \frac{d}{d}^Y\hat{X}_{21}} \right]_{\substack{d\hat{X}_{+1}=dX_{+1} \\ d\hat{X}_{21}=dX_{21}}} \right\} c_{\left(\frac{d}{d}\hat{X}_{+1}, \frac{d}{d}^Y\hat{X}_{21}\right)} = \\
&= \frac{d}{d} \frac{X_{+1}^2}{X_{+1}^4} V_{\frac{d}{d}\hat{X}_{+1}} + \frac{1}{d} \frac{1}{X_{+1}^2} V_{\frac{d}{d}^Y\hat{X}_{21}} - 2 \frac{d}{d} \frac{X_{21}}{X_{+1}^3} c_{\left(\frac{d}{d}\hat{X}_{+1}, \frac{d}{d}^Y\hat{X}_{21}\right)}
\end{aligned}$$

per $\frac{d}{d}^Y\tilde{\mathcal{S}}$.

Per ottenere una stima delle varianze (a), (b), (c) e (d), basta semplicemente sostituire alle varianze e covarianze in esse coinvolte i rispettivi stimatori e al posto dei totali $\frac{d}{d}X_{+1}$, $\frac{d}{d}X_{11}$, $\frac{d}{d}^YX_{11}$ e $\frac{d}{d}^YX_{21}$ le corrispondenti stime campionarie.

Appendice 2.D - Numerosità di primo e secondo stadio teoriche e rispondenti per regione

La tavola che segue mostra gli elementi essenziali della struttura campionaria ottenuta attraverso il disegno di campionamento descritto nel Paragrafo 2.4.3.

Tavola 2.D.1 - Distribuzione dei comuni campione e dei fogli di mappa campione, dei comuni rispondenti e delle aziende intervistate per regione

REGIONI	Numero di comuni campione	Numero di fogli di mappa campione (a)	Numero di comuni rispondenti	Numero di aziende intervistate
Piemonte	42	127	42	1.013
Valle d'Aosta/Vallée d'Aoste	13	30	12	249
Liguria	16	73	12	213
Lombardia	38	172	38	822
Trento	24	1.641	24	454
Veneto	38	87	38	1.195
Friuli-Venezia Giulia	12	40	12	414
Emilia-Romagna	54	124	54	1.046
Toscana	50	124	43	1.007
Umbria	23	52	22	459
Marche	32	75	31	631
Lazio	29	82	29	1.337
Abruzzo	17	35	16	771
Molise	13	28	13	503
Campania	28	70	28	2.126
Puglia	34	87	34	3.540
Basilicata	16	35	16	1.361
Calabria	31	84	31	1.660
Sicilia	52	140	49	1.946
Sardegna	28	47	27	841
Italia	590	3.153	571	21.588

(a) Il dato relativo ai fogli di mappa della provincia di Trento si riferisce in realtà a particelle catastali, poiché in tale provincia è tuttora in vigore il Catasto austro-ungarico



Conclusioni

Le numerose innovazioni introdotte nel 6° Censimento generale dell'agricoltura hanno decisamente accresciuto l'importanza della misurazione della qualità del dato censuario. Verificare gli effetti di tali innovazioni, specialmente di quelle che hanno caratterizzato la preparazione della lista pre-censuaria, è considerato un obiettivo imprescindibile per l'Istat. L'utilizzo delle fonti amministrative al fine di creare una lista valida per la conduzione di un Censimento rappresenta una novità assoluta che non può e non deve rimanere un esercizio isolato nel panorama delle statistiche agricole. La prossima progettazione ed esecuzione di un registro annuale delle aziende agricole (*Farm Register*) deve necessariamente tenere in considerazione i risultati del Censimento e la sua valutazione di qualità.

In tale quadro si inseriscono le due rilevazioni post-censuarie, che mirano a valutare sia il grado di copertura sul territorio del Censimento sia il grado di attendibilità delle singole risposte fornite dai conduttori di azienda durante la rilevazione censuaria. Le due rilevazioni post-censuarie sono state eseguite per la seconda volta in Italia con le medesime caratteristiche⁷⁶. Infatti, precedentemente, la valutazione della qualità dell'informazione, fornita con i risultati del Censimento, era stata realizzata con tecniche di analisi dei dati, come documentato da appositi volumi a suo tempo pubblicati.

Le due indagini di controllo realizzate in occasione del 6° appuntamento censuario presentano, rispetto a quelle realizzate per il Censimento precedente, una dimensione campionaria maggiore che ha permesso non solo di fornire stime sulla precisione del Censimento a livello regionale, ma anche di disaggregare le informazioni sulle singole coltivazioni e allevamenti (indagine di misura), nonché sulle diverse componenti del tasso di copertura e sotto-copertura. Questa ricchezza di informazioni, che non ha precedenti anche a livello internazionale, consente non solo di interpretare nella maniera più completa possibile i dati censuari, ma permette anche di proiettare l'attenzione al futuro immediato, fatto di realizzazione periodica del *Farm Register* e delle numerose indagini campionarie del settore agricolo.

Brevemente si ricorda che: 1) *l'indagine di misura* è finalizzata a fornire valutazioni dell'errore di misurazione dovuto al processo di risposta e svolta mediante re-intervista telefonica su un campione di aziende rilevate al Censimento; 2) *l'indagine di copertura* è finalizzata a fornire una misurazione del grado di copertura del Censimento rispetto alla popolazione delle aziende agricole mediante un campione areale di circa 1.500 fogli di mappa del Catasto dei terreni.

La progettazione e tutte le fasi di realizzazione delle due indagini sono state caratterizzate da un'attenta cura dei processi fondamentali di preparazione ed esecuzione della rilevazione sul campo, in modo che fossero sempre mantenuti elevati standard qualitativi. Si vuole qui dare rilevanza in particolare alla fase di formazione dei rilevatori che, nel caso dell'indagine di misura, ha visto quattro appuntamenti formativi con tutti i rilevatori e un *de-briefing* a settimana per la soluzione dei problemi sorti durante la fase di rilevazione telefonica e che, nel caso dell'indagine di copertura è stata svolta dai ricercatori Istat, nei luoghi deputati dalle regioni o dagli Uffici territoriali dell'Istat, direttamente a tutti i rilevatori impiegati sul territorio. In entrambi i casi questo ha consentito di condurre le indagini di interesse con un livello qualitativo notevole.

I risultati dell'indagine di misura hanno certificato l'elevata qualità del dato censuario, mostrando valori distorsivi tendenzialmente negativi, e quindi interpretabili come una

⁷⁶ La prima esperienza di indagini post-censuarie così strutturate è stata effettuata in occasione del 5° Censimento generale dell'Agricoltura 2000.

contenuta sottostima del Censimento. Si prenda a titolo esemplificativo la variabile più importante, ossia la SAU: a livello Italia, la distorsione è pari a -1,3 per cento ossia il Censimento ha sottostimato tale variabile di 1,3 punti percentuali. Tutte le altre variabili rilevate all'indagine di misura presentano distorsioni decisamente basse, comprese variabili di rilevante importanza per il paese quali quelle relative alla superficie della vite e dell'olivo. A livello nazionale, nessuna distorsione calcolata supera i sette punti percentuali. Ovviamente, scendendo di livello territoriale, qualche distorsione presenta valori meno contenuti, ma si tratta di risultati la cui spiegazione è stata attribuita a problemi di rilevazione al livello locale.

L'interpretazione dei risultati dell'indagine di copertura è certamente fondamentale nell'ottica di costruzione del *Farm Register*. Il *record linkage* è stato appositamente impostato per abbinare le unità rilevate alla PES sia con il Censimento sia con tutti gli archivi di preparazione della lista pre-censuaria. Tale procedimento ha permesso la totale scomposizione del tasso di copertura e di sotto-copertura in differenti modalità. In particolare le modalità di scomposizione del tasso di sotto-copertura hanno permesso di quantificare:

- la quota dovuta a unità non presenti in alcun archivio di confronto utilizzato. Tale porzione di sotto-copertura è attribuibile a mancanze informative delle fonti amministrative utilizzate e presenti nel panorama nazionale;
- la quota dovuta ad aziende presenti negli archivi di confronto (escluso la lista pre-censuaria). Tale porzione è attribuibile alle regole adottate per definire l'eleggibilità di una unità ad azienda agricola;
- la quota dovuta ad aziende esitate al Censimento con i codici b2 (azienda in lista non rilevata) e b3 (azienda in lista non esistente); tale porzione comprende la mancata risposta ed errori di classificazione delle aziende da parte dei rilevatori;
- la quota dovuta ad aziende rilevate ed escluse dalla lista censuaria; tale porzione comprende errori dovuti alla fase di controllo e correzione dei dati censuari.

Come detto in precedenza, la scomposizione consente di attribuire quote dell'errore non campionario commesso dal Censimento alle varie fasi di sua preparazione e realizzazione. È necessario ricordare che nessun altro paese membro della Comunità europea pubblica risultati sull'errore di misura e di copertura del Censimento agricolo, mentre l'Istat, oramai, può vantare una consolidata tradizione.

I risultati dell'indagine di copertura sono decisamente contenuti, sia se confrontati con quelli della scorsa tornata censuaria (quando, però, non fu possibile una simile scomposizione), sia se confrontati con altre esperienze internazionali a livello mondiale. Ad esempio, lo *U.S. Bureau of Census* ha pubblicato i risultati dell'indagine di copertura del Censimento agricolo del 2007 mostrando che la quota di sotto-copertura dovuta a mancata risposta era pari al 14,7 per cento e la quota dovuta a errori di lista era pari al 16,2 per cento. Inoltre, in un documento di programmazione dell'ultimo Censimento, scriveva che gli Stati non avrebbero dovuto superare, per il Censimento del 2012, livelli di sotto-copertura pari al 13 per cento sia per la quota di mancata risposta sia per la quota di errore dovuto alle liste.

I risultati dell'indagine di copertura relativi alla sotto-copertura per classi di SAU mostrano che gli errori maggiori vengono commessi per le piccole aziende per le quali è molto sottile la differenza tra azienda e non azienda secondo la definizione statistica. I risultati mostrano, altresì, che il Censimento ha mancato solo poche medie e grandi aziende, per problemi legati alla mancata risposta o per difficoltà di verifica dell'effettivo esercizio di attività agricola. Dunque per queste dimensioni di azienda le fonti am-



ministrative si sono dimostrate largamente attendibili. Peraltro, i risultati della copertura delle unità di piccola e piccolissima dimensione dimostrano chiaramente che permane, nel nostro paese, un problema relativo alla definizione di azienda agricola. Infatti, porre esclusivamente delle soglie fisiche nella definizione a eleggibilità non sembra soddisfare la moltitudine di casi esistenti nel panorama nazionale. Sembra auspicabile, anche nell'ottica del *Farm Register*, l'apertura di una nuova discussione in merito alla definizione di azienda agricola nella quale, probabilmente, debba essere vagliata l'ipotesi di inserire anche delle soglie di valenza più strettamente economica.

In definitiva, i risultati delle due indagini di qualità costituiscono un passaggio fondamentale nell'ottica di costituzione di un registro delle aziende agricole sempre aggiornato, che rappresenta certamente una risorsa fondamentale sia per gli *stakeholder* del settore, al fine di produrre analisi sempre tempestive, sia per i *policy maker* al fine di determinare puntuali politiche territoriali di intervento, nonché per rendere disponibile un *frame* utilizzabile per le indagini campionarie del settore agricolo.

Riferimenti bibliografici

- J. Bethel (1989), Sample Allocation in Multivariate Surveys, *Survey Methodology*, 15, 47-57.
- F. Borrelli, L. Mancini, L. Marcone, L. Valentino, "Record linkage between Large Datasets: Evidence from the 15th Italian Population Census", in *Quaderni di statistica*, 05/2012; 14, pp. 149-152.
- N. Cibella, M. Fortini, D. Ichim, T. Tuoto, "Record linkage methods and techniques as proposed in RELAIS" in *Proceedings of Statistics Canada Symposium 2010, Social Statistics: The Interplay among Censuses, Surveys and Administrative Data*.
- G. Cicchitelli, A. Herzel, G.E. Montanari (1992), *Il campionamento statistico*, Il Mulino, Bologna.
- W. Cochran (1977), *Sampling Techniques*, J. Wiley, New York.
- C. De Vitiis, P.D. Falorsi, M. Ballin, G. Scepi (1998), Principi e metodi del software generalizzato per la definizione del disegno di campionamento nelle indagini sulle imprese condotte dall'Istat, *Statistica applicata*, 10(2), pp. 235-257.
- J.C. Deville, P. Lavallée (2006), Indirect Sampling: The Foundation of the Generalized Weight Share Method, *Survey Methodology*, vol.32, n.2, pp.165-176.
- J.C. Deville, C.E. Sarndal (1992), Calibration estimators in survey sampling, *Journal of the American Statistical Association*, 87, pp. 376-382.
- L. Fabbris (1991), Campioni di numerosità due o tre per strato selezionati con probabilità variabili: valutazione empirica di alcune stime di frequenze assolute", in *Atti della giornata di studio sul campionamento statistico. Annali di statistica, Serie IX*, Istat.
- I.P. Fellegi (1963), An Analysis of Response Variance, *Bulletin of the International Statistical Institute*, 40, pp. 758-759.
- I.P. Fellegi (1964), Response Variance and Its Components, *Journal of the American Statistical Association*, 59, pp. 1016-1041.
- I.P. Fellegi (1974), An Improved Method of Estimating the Correlated Response Variance, *Journal of the American Statistical Association*, 69, pp. 496-501.
- I.P. Fellegi, D. Holt (1976), A Systematic Approach to Automatic Edit and Imputation, *Journal of the American Statistical Association*, 71, 17-35.
- I.P. Fellegi, A.B. Sunter (1969), "A Theory for Record linkage", *Journal of the American Statistical Association*, 64, pp. 1183-1210.
- G. Forsman, I. Schreiner (1991), The Design and Analysis of Reinterview: an Overview, in: *Measurement Errors in Surveyce*, P. Biemer, R.M. Groves, L. Lyberg, N. Mathiowetz and S. Sudman (eds), New York, Wiley, 279-302.



- M. Latouche, J.M. Berthelot (1992), Use of Score Function to Prioritize and Limit Re-contacts in Editing Business Surveys, *Journal of Official Statistics*, vol. 8, No. 3, Part II.
- M.H. Hansen, W.N. Hurwitz, M.A. Bershad (1961), Measurement Errors in Censuses and Surveys, *Bulletin of the International Statistical Institute*, 38(2), pp. 359-374.
- M. Jaro (1985), "Advances in Record linkage Methodologies as Applied to Matching the 1985 Census of Tampa, Florida", *Journal of American Statistical Society*, 84(406), pp. 414-420.
- P. Lavallée (2007), *Indirect Sampling*, Springer, New York.
- P. Lavallée, L.P. Rivest (2012), Capture-Recapture Sampling and Indirect Sampling, *Journal of Official Statistics*, 28, n.1, pp.1-27.
- W.G. Madow (1949), On theory of systematic sampling, *Annals of Mathematical Statistics*, 20.
- M. Mazziotta, A. Bernardini, L. Soriani (2011), *Indagine di copertura del 6° Censimento generale dell'agricoltura - Manuale delle istruzioni*, Istat.
- M.N. Murthy (1967), *Sampling theory and methods*, Statistical Publishing Society, Calcutta.
- R.C. Petersen, J.C. Stevens, M. Gangli, E.G. Tangalos, J.L. Cummings, S.T. DeKosky (2001), Practice parameter: Early detection of Dementia: mild cognitive impairment (an evidence based review). Report of the quality standard subcommittee of the American Academy of Neurology. *Neurology* 56: 1133-1142.
- J.N.K. Rao (2003), *Small Area Estimation*, John Wiley & Sons, Hoboken, New Jersey.
- Relais, User's guide. Versione 2.2, autori vari, <http://www.istat.it/it/files/2011/03/Relais2.2UserGuide.pdf>.
- C.E. Särndal, B. Swensson, J. Wretman (1989), The weighted residual technique for estimating the variance of the general regression estimator of the finite population total, *Biometrika*, vol. 76, n. 3, pp. 527-537.
- A.C. Singh, C.A. Mohl (1996), Understanding calibration estimators in survey sampling, *Survey Methodology*, 22, pp. 107-115.
- K.M. Wolter (1985), *Introduction to Variance Estimation*, Springer-Verlag, New York.
- K.M. Wolter (1986), Some Coverage Error models for Census Data, *Journal of the American Statistical Association*, 81, 394, pp.338-346.
- R.S. Woodruff (1971), A Simple Method for Approximating the Variance of a Complicated Estimate, *Journal of the American Statistical Association*, vol.66, n. 334, pp. 411-414.
- D. Zardetto (2012), *ReGenesees: R Evolved Generalized Software for Sampling Estimates and Errors in surveys*, R package version 1.3, Istat, Italy, <http://www.istat.it/it/strumenti/metodi-e-software/software/regenesees>.

L'Istat misura la qualità del 6° Censimento generale dell'agricoltura attraverso la realizzazione di due indagini campionarie post-censuarie: l'indagine di misura e l'indagine di copertura. Nel presente fascicolo sono descritte tutte le fasi di tali rilevazioni, nonché i risultati che, per la prima volta, sono diffusi a livello regionale. L'interpretazione dei risultati è fondamentale non solo per la valutazione del Censimento, ma rappresenta anche un importante passaggio ai fini dello studio, della progettazione e della costruzione del Registro statistico delle aziende agricole (Farm Register). I modelli di rilevazione utilizzati nelle due indagini sono contenuti nel cd-rom collocato nel primo fascicolo dell'opera, *Il quadro normativo*.