

L'Indagine MET: logica, conduzione e metodologia¹

La logica dell'indagine strutturale MET²

Negli anni 2000 sono state realizzate diverse indagini di campo sulle imprese italiane con numerosità e caratteristiche tali da garantire una significatività di carattere generale per il sistema produttivo italiano.

L'indagine MET si è affermata come un prodotto unico per la sua estensione, per la sua rappresentatività (territoriale, settoriale e dimensionale) e per il rigore con cui viene condotta.

L'obiettivo è stato quello di fornire un quadro ampio e con sufficiente dettaglio territoriale, dimensionale e settoriale di alcuni aspetti significativi della vita delle imprese: si tratta, di gran lunga, della più vasta indagine del genere svolta in Italia con circa 25.000 imprese intervistate nel 2008 e poco meno di 23.000 nel 2009 e oltre 25.000 casi nel 2011.

Una numerosità così elevata si giustifica con la necessità di avere analisi rappresentative della struttura italiana, ma anche capaci di leggere e analizzare fenomeni relativamente rari e fondamentali quali quelli (per esempio Ricerca e Internazionalizzazione) legati al segmento più dinamico e innovativo del nostro sistema produttivo.

L'analisi è concentrata sulle caratteristiche strutturali delle imprese, sugli aspetti di competitività, sulle reti, sulle criticità emerse e sui fabbisogni esterni (anche di *policy*) che gli operatori manifestano.

Un aspetto determinante è rappresentato dalle esigenze di sostegno delle imprese (che, entro certi limiti, vengono considerati come parte della domanda di politiche pubbliche) in aree quali quelle della ricerca e dell'innovazione, dell'internazionalizzazione, della cooperazione tra imprese, degli investimenti e dei rapporti con il settore finanziario.

I risultati delle indagini precedenti sono stati oggetto di diverse pubblicazioni³, oltre che di numerose presentazioni in incontri scientifici dedicati ai temi trattati.

La cadenza temporale e l'orizzonte coperto attribuiscono, se possibile, ancora maggior rilievo all'indagine. Dopo la rilevazione del 2008 (conclusa 10 giorni prima del fallimento di Lehman Brothers) e quella dell'autunno del 2009 (prima fase della grande crisi), il 2011 si è caratterizzato come l'anno dell'approfondimento della fase fortemente negativa per l'industria italiana ed europea con peggioramenti di molti

¹ Autori: P. Ascani, R. Brancati, A. Maresca per il primo paragrafo; M. Centra e P.D. Falorsi per il secondo.

² L'intera impostazione dell'Indagine è stata curata da un Comitato Scientifico costituito da Giorgio Alleva, Giovanni Barbieri e Alberto Zuliani.

³ Rapporti MET, vari anni editi da Donzelli Editore e Fatti in cerca di idee di R. Brancati, Donzelli Ed. 2010.

parametri, ma anche con segnali apprezzabili di reazione da parte di fasce significative di operatori.

Nel 2011 l'indagine è stata riproposta (con una numerosità analoga alle precedenti pari a circa 25.000 casi) in modo da cogliere variazioni e aspetti di cambiamento della struttura produttiva con particolare riferimento alle problematiche indicate e a quelle strategiche per la crescita del sistema produttivo.

Tra gli aspetti di unicità del lavoro va segnalata la disponibilità dei database 2008, 2009 e 2011 con una "quota *panel*" che raggiunge i 16.000 casi per ciascuna coppia di anni e circa 8.000 per l'intero periodo. Ciò offre una possibilità di confronto tra le posizioni delle imprese rispetto alle diverse tematiche di interesse a seconda dei profili di mercato e di innovazione registrati nel recente passato e dal modo in cui la crisi ha avuto impatto sulla gestione aziendale.

La grande numerosità del campione consente approfondimenti dell'analisi per poter cogliere in modo rigoroso alcuni degli aspetti essenziali della struttura, sia con riferimento all'anno in esame, sia in termini diacronici potendo utilizzare un *panel* molto consistente di imprese presenti in tutte e tre le indagini considerate.

Con l'accresciuto ruolo dei mercati internazionali, per la gran parte degli operatori dinamici, un'analisi accurata delle forme e dell'evoluzione dei processi di internazionalizzazione (dalla presenza sui mercati esteri a forme più complesse) associata alle *performance* aziendali e inserita in un quadro più ampio sul ruolo del credito e sui vincoli indotti, si presenta come particolarmente utile per comprendere la struttura produttiva e la domanda di *policy* che essa esprime.

Un campione così numeroso, associato a dati di fonti diverse – si pensi ai dati di bilancio delle società di capitali, o i dati sull'occupazione – consente analisi accurate anche per valutare le condizioni di diversi gruppi e tipologie di imprese. Il confronto tra il campione di controllo e soggetti finanziati, per i quali occorrono le anagrafiche relative, consente indicazioni di rilievo sul profilo delle imprese interessate e primi riferimenti per un'analisi controfattuale dettagliata.

Il campionamento è stato tale, comunque, da garantire ampia significatività in ciascuna delle 20 regioni italiane.

L'origine dell'Indagine

Lo studio nasce come naturale evoluzione del Rapporto MET sviluppato nel corso dell'ultimo quinquennio. Si trattava, infatti, di associare a un'analisi dell'offerta di politiche pubbliche da parte delle amministrazioni, anche un approfondimento esteso sulla questione essenziale: l'articolazione e la differenziazione anche quantitativa della domanda da parte delle imprese.

Il Rapporto MET

Il Rapporto MET rappresenta l'unica ricostruzione analitica delle politiche pubbliche per le imprese. L'approccio seguito nel lavoro è di carattere pragmatico e, al di là dei problemi legati all'utilità stessa degli interventi, ha come scopo quello di fornire un'*accountability* accurata e aggregata di una politica di grande rilievo che,

in Italia come nelle principali nazioni a elevato sviluppo economico, assorbe importanti risorse.

Il lavoro, nei primi anni, verteva principalmente sull'offerta pubblica (di Governi regionali e nazionale) esaminando dimensioni dei flussi, orientamento e obiettivi perseguiti; nelle ultime due edizioni si è integrato lo studio con un'ampia indagine di campo al fine di ricostruire le caratteristiche strutturali delle imprese, alcune problematiche specifiche della gestione e soprattutto le loro esigenze.

Nelle edizioni 2006 e 2007 si è utilizzato un campione esteso (oltre 5.000 interviste realizzate), per dimensioni simile alle indagini Invind-Banca d'Italia e Mediocredito Centrale-Unicredit; la stessa indagine è stata associata ad analisi e interpretazioni della struttura produttiva: pur essendo di dimensioni largamente inferiori a quella utilizzata nel 2008, su alcune tematiche aggregate è possibile offrire un quadro diacronico.

Le esigenze attuali

Dalle reazioni avute da parte dei soggetti interessati è emerso l'interesse per un approfondimento ulteriore al fine di soddisfare esigenze di significatività territoriale (Regioni) con aspetti di dettaglio che facciano emergere caratteristiche e peculiarità dei soggetti più dinamici, di quelli più innovatori o anche di quelli che hanno bisogno di un sostegno temporaneo per uscire da situazioni di difficoltà.

È sorta, quindi, la necessità di realizzare un'indagine molto più estesa e si è ampliato il campione delle interviste arrivando a toccare circa 25.000 imprese⁴, realizzando una rilevazione unica, per dimensioni, nel panorama italiano (escludendo, ovviamente, la ricerca dell'ISTAT che presenta caratteristiche del tutto diverse e dunque non confrontabili con il nostro lavoro). I criteri di campionamento sono discussi nei prossimi paragrafi.

Obiettivi e Strategia

La questione fondamentale che è opportuno sottolineare, tuttavia, è l'obiettivo stesso dell'indagine e la strategia da seguire per estrarre informazioni complesse.

- L'obiettivo primario è quello di cogliere le esigenze delle imprese concentrandosi in modo particolare sui fabbisogni esterni necessari per la crescita e lo sviluppo. Per esempio, nel caso in cui un'impresa riconosca come indispensabile, per la competitività aziendale, una strategia innovativa. Il tipo di sostegno a questa attività può comunque assumere forme estremamente diverse e non tutte egualmente preferibili dalle imprese (formazione del personale interno, acquisizione di nuovo personale qualificato, consulenze, cooperazione con Centri di Ricerca e Università, acquisto di brevetti, interventi sul capitale di rischio, finanziamenti bancari, contributi). Capire caratteri e importanza di questi apporti esterni (privati e pubblici), delle difficoltà che il mercato incontra nell'offrirli e delle problematiche connesse è

⁴ Le interviste sono svolte secondo metodologia CATI e CAWI - tecniche miste web e telefoniche -; i dettagli tecnici completi sono, naturalmente, disponibili nel caso di interesse.

una questione rilevante e poco approfondita. A ciò si associa l'aspetto speculare rappresentato dai vincoli presenti (vincoli tecnici, finanziari, di mercato, *etc.*) che possono bloccare la strategia.

- Per ottenere un quadro informativo esauriente si devono raggiungere obiettivi intermedi dell'analisi riguardanti aspetti di struttura dell'impresa, caratteristiche di mercato, dinamismo e scelte di investimento, rapporti internazionali e competitività; questi dati, oltre a essere funzionali all'obiettivo principale, rappresentano, comunque, un patrimonio informativo utile per il suo dettaglio e per la sua tempestività (i dati sono resi disponibili con meno di 5 mesi di ritardo rispetto alla rilevazione).

Altri caratteri distintivi dell'indagine

Esistono anche diversi caratteri specifici e distintivi dell'indagine.

- L'indagine e il questionario sono dedicati a informazioni non altrimenti disponibili. Si è scelto di rilevare in maniera approfondita aspetti e problematiche specifiche e di non indugiare su domande che toccano profili e situazioni ricavabili in modo più approfondito da Banche Dati esterne. Per esempio nel caso di informazioni sul mercato del lavoro pare più opportuno lavorare sui dati INPS e di "incrociare l'informazione" con quelle ricavate dalla indagine MET per aspetti di mercato, di tecnologie e altro. Così pure i dati di bilancio e le informazioni di dettaglio sull'entità finanziaria degli aiuti sono meglio ricavabili da fonti esterne quali le Banche Dati sui bilanci delle imprese o le stesse Anagrafiche delle agevolazioni pubbliche. Diviene quindi essenziale mantenere codici adatti all'incrocio con banche dati sull'universo dei soggetti quali quelle citate e collegarle al campione da noi estratto.
- L'intera indagine viene svolta con il massimo rigore scientifico. Oltre alla serietà dei lavori svolti in passato, per le edizioni "estese" (2008, 2009 e 2011) ci si è avvalsi anche del lavoro di un comitato scientifico di grande valore⁵ e di una commissione tecnica dedicata a seguire le fasi del campionamento, dell'implementazione dell'indagine e della complessa fase di calibrazione e di stima del riporto all'universo per i diversi insiemi di riferimento.
- La ripetizione dell'indagine con cadenza biennale consente di utilizzare le distribuzioni conosciute dei fenomeni (attraverso i risultati dell'indagine precedente) per ampliare il campionamento negli strati per i quali è ragionevole attendersi una maggiore frequenza di fenomeni rari e di interesse (per esempio le attività di R&S).

Il Questionario

⁵ Al Comitato Scientifico hanno partecipato Giorgio Alleva, Giovanni Barbieri e Alberto Zuliani, oltre ai senior del progetto MET Carlo Andrea Bollino e Raffaele Brancati; mentre il Comitato Tecnico (nelle fasi di impostazione associato allo stesso Comitato Scientifico) è costituito da Marco Centra e Piero Falorsi.

Il questionario proposto è lungo e complesso, ma sulla base delle esperienze maturate negli anni passati si sono definite delle linee guida che hanno portato a risultati soddisfacenti; il modulo somministrato è stato arricchito dal punto di vista qualitativo e, al contempo, semplificato eliminando domande delicate che portavano a molte cadute nelle risposte, come per es. quelle sui dati di bilancio *et similia*).

Il questionario è così suddiviso:

1. *sezione informativa* (dati di base, informazioni su occupazione, struttura proprietaria, *etc.*);
2. *sezione mercati e struttura* (mercati di sbocco, vantaggi competitivi, debolezze, struttura dei costi, banche e finanza, *etc.*);
3. *sezione reti e filiere* (relazioni tra imprese, reti locali e “lunghe”, filiere, mercati di approvvigionamento e vendita, tipologia di beni prodotti, articolazione internazionale, subfornitura, *etc.*);
4. *sezione dinamismo* (programmi -passati e previsti- di investimento, crescita di occupazione e fatturato, obiettivi degli investimenti, scelte qualitative e indicazioni da altre sezioni);
5. *sezione Ricerca e Innovazione* (programmi di innovazione, ricerche, relazioni con soggetti esterni per la ricerca, aree di sviluppo, *etc.*);
6. *sezione Internazionale* (tipologia di presenza sui mercati esteri, alleanze, utilizzazione di strumenti nazionali e regionali di supporto, *etc.*);
7. *l'utilizzazione delle politiche per le imprese* (qualche elemento valutativo e informazioni per la stima dell'efficacia netta, aspettative e clima di fiducia delle imprese, *etc.*);
8. *la Finanza:* (modalità di finanziamento delle imprese, presenza di vincoli e di razionamento per diverse attività di sviluppo – investimenti e R&S – banca di riferimento e altre grandezze rilevanti);
9. *la domanda di intervento pubblico o di supporto esterno* (in questa sezione si stima la domanda complessiva di intervento pubblico da parte delle imprese sia nella forma di sostegno finanziario, che in quelle di servizi reali, di politiche per lo sviluppo locale, di sostegno alle attività di ricerca e innovazione, di politiche infrastrutturali con ricadute sull'attività delle aziende intervistate);
10. *il procurement* (è stata introdotta una sezione dedicata ad approfondire il *procurement* pubblico e i rapporti con la domanda pubblica);
11. *temi specifici* (si sono alternate analisi su Energia e ambiente, costi amministrativi nei rapporti con la P.A., rilevanza delle certificazioni, impatto della crisi, altro. Nell'indagine 2011 si sono approfonditi i temi legati all'utilizzo delle tecnologie informatiche nei rapporti con la P.A.);
12. *performance e dati di bilancio* (laddove disponibili – società di capitali – alle imprese del campione vengono allineati i dati di bilancio per le elaborazioni di rilievo).

Il Trattamento del campione

Le anagrafiche delle imprese campione sono state estratte con la tecnica del passo sistematico a partire dall'anagrafe ufficiale delle Camere di commercio gestita dalle Camere di Commercio Italiane. Tale banca dati tuttavia non descrive, o descrive soltanto per alcune di esse, il numero degli addetti delle imprese e i numeri di telefono delle imprese. Per ovviare a tale inconveniente il numero degli addetti è stato identificato *ex post*, rilevandolo direttamente in fase di rilevazione dati ed effettuazione delle interviste, mentre i numeri di telefono delle imprese estratte sono stati rilevati a partire dagli elenco degli abbonati al servizio telefonico nazionale (Pagine bianche)⁶.

Ai fini del trattamento del campione sono state predisposte due liste di anagrafiche: una lista di anagrafiche base (25.090 anagrafiche), e una lista di anagrafiche suppletiva, formata da un numero di unità campionarie di riserva quattro volte superiore rispetto a quello delle anagrafiche del campione base, per un totale di circa 100.000 anagrafiche, sommando insieme quelle della lista base e quelle della lista suppletiva di riserva.

All'interno di ciascun livello di stratificazione del campione base e del campione di riserva le anagrafiche delle imprese sono state gestite per mezzo di una procedura software che ha consentito il rispetto delle quote prefissate all'interno di ciascuno strato e all'interno di ciascuna cella. In caso di "caduta" di un'anagrafica, questa è stata sostituita soltanto con le anagrafiche delle liste di riserva del medesimo strato.

Il questionario è stato somministrato presso ogni singola azienda alla cosiddetta "persona di riferimento", tipicamente l'amministratore, un dirigente responsabile o l'imprenditore, a seconda delle caratteristiche e delle dimensioni dell'azienda. La persona di riferimento è stata identificata per mezzo di una prima telefonata all'azienda nel corso della quale è stata presentata l'iniziativa di ricerca e concordato un appuntamento telefonico per l'effettuazione dell'intervista. In totale quindi, presso ogni azienda, sono state effettuate almeno due telefonate.

Per lo svolgimento della rilevazione è stato implementato l'utilizzo dei moduli *Acs (Automatic Call Scheduling)*. L'utilizzo di tali moduli ha fatto sì che la lavorazione di ogni singola anagrafica campione potesse avere almeno sei esiti distinti: intervista concessa, intervista rifiutata, appuntamento telefonico, richiesta del questionario via fax o via mail, numero di telefono sempre occupato, sempre libero, sempre con una segreteria telefonica o con un fax, numero errato (al telefono risponde un utente diverso da quello desiderato). A seconda degli esiti dei contatti telefonici, questi le anagrafiche sono trattate in modo conseguente.

Intervista concessa - L'intervista è stata archiviata nel database della ricerca e avviata fase di editing per la verifica di coerenza logica dei risultati.

⁶ Per ogni singola impresa è stata rilevata la ragione sociale, l'indirizzo, il telefono e il Codice Ateco. Per le sole società di capitali è stato rilevato anche il Codice fiscale.

Intervista rifiutata - L'intervista è stata conservata nel database della ricerca (tabella rifiuti), implementando la procedura per la sostituzione dell'anagrafica con un'altra statisticamente valida estratta dagli elenchi suppletivi.

Appuntamento telefonico – In tutti i casi in cui la “persona di riferimento” non fosse disponibile al momento del primo contatto telefonico, si è provveduto a rilevare la data e l'ora per un appuntamento telefonico.

Richiesta del questionario via fax o via e-mail – In tutti i casi in cui le “persone di riferimento” in azienda hanno manifestato la richiesta di rispondere al questionario per iscritto si è provveduto a trasmettere loro il questionario via fax o via mail.

Il quadro delle attività

Sotto il profilo metodologico la raccolta dei dati si è configurata come un'indagine quantitativa realizzata con la tecnica delle interviste telefoniche, che sono state effettuate con il Sistema Cati (*Computer assisted telephone interview*), somministrando alle imprese un questionario strutturato, fatta eccezione per le grandi imprese che sono state contattate ricorrendo al Sistema Cawi (*Computer assisted web interview*).

L'effettuazione dell'indagine di campo è stata preceduta da un test pilota, svolto con l'obiettivo di verificare la coerenza e la somministrabilità del questionario di rilevazione e del Sistema Cati/Cawi utilizzato. La sperimentazione ha comportato lo svolgimento di alcune centinaia di interviste presso altrettante imprese “fuori campione”.

La tabella riportata di seguito riepiloga le attività svolte durante il periodo di somministrazione del questionario. L'attività di rilevazione si è conclusa al 30 novembre 2011 con un tasso di rifiuto del 15,2% e un tasso di successo del 33,7%.

Tabella 1. Riepilogo delle attività di indagine.

	Numero casi	%
Concesse	25.090	33,7%
Rifiuti	11.785	15,8%
Appuntamenti (solleciti email/fax inviati)	4.136	5,6%
Liberi/fax/segreterie	21.708	29,2%
Errati	11728	15,8%
Tot lavorate	74.447	100,0%

Per verificare se il tasso di risposta da parte delle imprese fosse in qualche modo correlato alla tipologia dell'azienda è stata svolta inoltre un'analisi degli esiti per singolo strato, nell'ambito dei quali si mette in evidenza che la tecnica di calibrazione ha comunque provveduto a eliminare le eventuali distorsioni della stima causati da tassi di risposta anomali.

Interviste alle grandi imprese

Tenuto conto dell'importanza delle grandi imprese nell'ambito del disegno del campione, tali aziende sono state rilevate non solo per mezzo dell'intervista telefonica Cati.

La rilevazione sul campione delle grandi imprese si è articolata nelle seguenti fasi operative:

- costruzione di un sito web ad accesso riservato con password (Sistema Cawi) con il questionario della rilevazione messo a disposizione dei soggetti da intervistare;
- effettuazione di un primo contatto telefonico per identificare presso ciascuna impresa la persona di riferimento alla quale effettuare l'intervista. A seconda dell'esito della telefonata sono stati comunicati alle imprese (direttamente al telefono, o via e-mail, o via fax) il login e la password per accedere al sito riservato e compilare il questionario messo a disposizione on line, oppure è stata effettuata l'intervista telefonicamente in tutti i casi in cui l'interlocutore avesse preferito procedere direttamente in questo senso, o in tutti in cui l'impresa non disponesse di una connessione Internet;
- effettuazione di un'azione di sollecito telefonico presso tutti coloro che entro un mese circa dalla data prevista di chiusura della rilevazione non avevano compilato il questionario via web, nonostante avessero precedentemente manifestato la propria volontà di procedere in tal senso. In tutti i casi in cui si è reso necessario le interviste mancanti sono state effettuate telefonicamente utilizzando il Sistema Cati.

Il Personale

Per lo svolgimento delle interviste è stato impiegato un gruppo di intervistatori (ricercatori assistenti) esperto nello svolgimento delle interviste con il sistema Cati, sotto la direzione e il controllo di due ricercatori coordinatori. L'addestramento del personale è stato sia di tipo teorico, sia di tipo pratico. Complessivamente la formazione di ogni intervistatore non è stata inferiore alle sei ore.

La Sperimentazione

L'attività di sperimentazione ha risposto alla necessità di verificare la tecnica dell'intervista telefonica e l'applicazione software del Cati per la gestione delle interviste, verificare la procedura per il rispetto dei criteri di sostituzione delle unità campione, verificare i parametri campionari e le misure di attendibilità delle stime (es. propensione alle mancate risposte da parte delle aziende campione, la tendenza alla concentrazione spuria di modalità; *etc.*).

La sperimentazione è stata effettuata mediante una indagine pilota che ha permesso di verificare sul campo che l'insieme delle soluzioni previste per lo svolgimento della ricerca fosse adeguato al conseguimento degli obiettivi previsti.

L'indagine pilota è stata realizzata con le stesse modalità dell'indagine vera e propria, e ha comportato l'effettuazione di 30 interviste presso altrettante aziende "fuori campione", di modo da non alterare il campione di indagine vero e proprio.

I controlli di Qualità

Aspetti di carattere generale

Il sistema dei controlli di qualità ha lo scopo di contenere il cosiddetto errore non-campionario con l'obiettivo ultimo di ottenere il massimo livello di affidabilità dell'indagine. Tale sistema è basato su tre sottosistemi principali, a loro volta articolati su una serie di azioni ben identificate. I tre sottosistemi sono il sottosistema delle azioni preventive, il sottosistema delle azioni da svolgere nel corso della rilevazione e il sottosistema delle azioni di valutazione.

Azioni preventive

L'obiettivo delle azioni preventive è quello di creare un clima favorevole allo svolgimento delle interviste facilitando il più possibile il compito degli intervistatori presso le aziende campione.

Nell'ambito della presente rilevazione sono state applicate tre azioni di tipo preventivo: la "doppia telefonata" alle imprese, l'istituzione di un sito web, ad accesso riservato con password, a disposizione delle imprese, e l'istituzione di un servizio di informazione.

Doppia telefonata

Questa tecnica ha lo scopo di presentare l'indagine all'impresa con una prima telefonata, identificare la "persona di riferimento" al suo interno, spiegarli i motivi che ci hanno indotto a cercare la sua collaborazione, informarla dei suoi diritti con riferimento alla privacy, e concordare l'effettuazione dell'intervista vera e propria (seconda telefonata).

Sito web

Il sito Web, ad accesso riservato con password, è stato destinato a tutte le aziende che hanno avuto o il bisogno di avere maggiori informazioni sull'indagine alla quale sono state chiamate a partecipare. Sul sito web è stata predisposta una copia del questionario di rilevazione, che nel caso le "persone di riferimento" in azienda, hanno potuto scaricare e visionare prima dell'effettuazione dell'intervista vera e propria.

Azioni nel corso della rilevazione

Lo scopo delle azioni in corso d'opera è stato quello di diminuire al massimo le possibilità di errore degli intervistatori in itinere. Tali azioni sono basate essenzialmente sui controlli automatici previsti dal Sistema Cati e dai report periodici dell'attività svolta compilati dal coordinatore della ricerca.

Controlli automatici

I controlli automatici consistono nella registrazione controllata dei dati inseriti e nell'identificazione automatica degli errori. Tali controlli riguardano in particolare le violazioni delle regole logico-formali (regole di compatibilità) relative ai limiti imposti dal dominio di variazione delle singole variabili.

Report periodici

I report periodici descrivono in corso d'opera lo stato di avanzamento della rilevazione dati, basandosi su indicatori quali il numero delle interviste effettuate

per ciascuno strato, il numero dei contatti telefonici effettuato per ottenere una intervista, il tasso di ricorso alle sostituzioni, il tasso dei rifiuti e dei numeri errati, la produttività per singolo intervistatore, e così via.

Azioni di valutazione

Lo scopo delle azioni di valutazione è quello di identificare gli errori effettuati nel corso della rilevazione, nell'ambito delle singole interviste, nonostante il sistema dei controlli automatici del Cati. L'obiettivo ultimo delle azioni di valutazione è quello di limitarne l'impatto sulla coerenza dell'indagine nel suo insieme.

Tali azioni sono basate sull'effettuazione di una indagine di controllo, che consiste nella ripetizione di alcune interviste presso un campione ridotto di aziende tra quelle intervistate per verificare il livello dello scostamento tra i risultati ottenuti per mezzo delle interviste originali e i risultati ottenuti per mezzo delle interviste di controllo realizzate ex novo. L'indagine di controllo ha comportato l'effettuazione di 150 interviste presso altrettante aziende, clusterizzate per intervistatore, con lo scopo di individuare eventuali scostamenti rilevanti tra le due dimensioni considerate (interviste originali e interviste di controllo).

Sono state effettuate, infine, delle verifiche sulla coerenza dei valori ricavati dall'indagine e riportati all'universo (in particolare sul numero degli addetti) rispetto alle informazioni disponibili per aggregazioni regionali e settoriali.

Interviste di controllo

Nei mesi di gennaio-febbraio 2012 sono state effettuate circa 5000 interviste di controllo, per verificare la stabilità delle risposte fornite dalle imprese intervistate.

La selezione delle interviste da effettuare per verifica è stata basata su un check delle risposte date durante la rilevazione base. Si sono verificati, infatti, i questionari riempiti e si sono analizzate le risposte che presentavano risultati apparentemente incoerenti tra diverse sezioni o che segnalavano comportamenti relativamente atipici.

L'esito soddisfacente ha consentito la validazione dei risultati raggiunti.

La metodologia di campionamento e di stima⁷

Popolazione di interesse

La popolazione di interesse cui si riferisce l'indagine MET è costituita dalle imprese appartenenti ai settori dell'industria in senso stretto e dei servizi alla produzione per tutte le classi dimensionali e con sede legale sull'intero territorio nazionale. Nello specifico l'universo considerato è composto da imprese che operano in 38 settori, definiti al terzo *digit* del Codice Ateco 2007⁸.

⁷ Autori: Marco Centra e Piero D. Falorsi.

⁸ I settori di attività delle imprese considerati nell'universo di interesse sono: il settore estrattivo (Codice Ateco B), l'industria alimentare, delle bevande e del tabacco (Codice Ateco C10, C11, C12), l'industria tessile e dell'abbigliamento (Codice Ateco C13, C14), l'industria del concio, cuoio e pelli

La dimensione campionaria dell'indagine, particolarmente elevata e pari a circa 25.000 unità, ha consentito l'effettuazione di ulteriori approfondimenti e sovracampionamenti regionali. Tali approfondimenti, di cui si è tenuto conto nella fase di disegno e di allocazione del campione, hanno riguardato sottopopolazioni di particolare interesse: le regioni in area Convergenza (come definito dal sistema dei fondi strutturali europei), le regioni a prevalente vocazione manifatturiera, il settore dell'industria in senso stretto, il comparto delle società cooperative, il segmento di imprese con spiccata propensione all'innovazione. Il disegno del campione ha, inoltre, considerato il fatto che in alcune regioni è stato richiesto un maggior livello di accuratezza delle stime prodotte, e un conseguente sovracampionamento, realizzato in virtù di un cofinanziamento regionale specifico. Per tenere conto di tale vincolo le regioni sono state divise in 3 gruppi, caratterizzati da diversi livelli di attendibilità delle stime.

Il disegno di campionamento ha tenuto conto delle sottopopolazioni di particolare interesse ai fini dell'indagine:

Regioni in area Convergenza. Tale popolazione è definita dalle imprese la cui sede legale è compresa nelle regioni in obiettivo convergenza, come definito dal sistema dei fondi strutturali europei, vale a dire: Campania, Puglia, Calabria e Sicilia.

Regioni a prevalente vocazione manifatturiera. Si tratta delle imprese che hanno la sede legale nelle regioni a elevata incidenza del settore manifatturiero, vale a dire compreso nella sezione D della classificazione ATECO 2002: Piemonte, Lombardia, Veneto Emilia Romagna e Marche.

Settore dell'industria in senso stretto. Comprende le imprese il cui settore di attività economica è compreso nella sezione D della classificazione ATECO 2002.

Comparto delle società cooperative. Comprende le società cooperative definite dalla forma giuridica.

Imprese con spiccata propensione all'innovazione. Tale segmento di imprese, particolarmente rilevante per le finalità dell'intero progetto, è composto dalle imprese con elevata propensione all'esportazione. Allo scopo di individuare tale segmento di imprese è stata delineata una strategia specifica di analisi, determinata dal particolare scenario circa il bagaglio informativo attualmente disponibile: i) non si dispone, indipendentemente dalla definizione adottata di impresa innovativa,

(Codice Ateco C15), l'industria del legno e dei prodotti in legno (Codice Ateco C16), l'industria della carta, stampa ed editoria (Codice Ateco C17, C18), la fabbricazione di coke, e le raffinerie di petrolio (Codice Ateco C19), l'industria chimica (Codice Ateco C20, C21), l'industria della gomma e delle materie plastiche (Codice Ateco C22), la lavorazione dei minerali non metalliferi (Codice Ateco C23), la metallurgia (Codice Ateco C24, C25), l'industria meccanica (Codice Ateco C28), l'industria elettronica (Codice Ateco C26, C27), la fabbricazione di mezzi di trasporto (Codice Ateco C29, C30), le altre industrie manifatturiere (Codice Ateco C31, C32), la produzione e distribuzione di energia elettrica, gas e acqua (Codice Ateco D), i settori dei trasporti, magazzinaggio e comunicazioni (Codice Ateco H), e infine quelli del noleggio, dell'informatica, della ricerca e dei servizi alle imprese (Codice Ateco J, M).

dell'informazione puntuale sulla popolazione; ii) le imprese innovative sono identificabili solo tramite informazioni rilevate dall'indagine 2009 e quindi disponibili unicamente sul campione delle relative imprese rispondenti. Ipotizzando che le variabili esplicative della caratteristica innovativa di un'impresa siano rimaste sostanzialmente invariate rispetto al 2009, l'indagine di tale anno è stata utilizzata per mettere a punto una procedura di analisi in grado di identificare gli strati dove più elevata è la frequenza attesa di imprese innovative: tramite un sovracampionamento di tali strati si è ottenuta una frequenza pianificata delle imprese innovative nei diversi domini di studio, definiti, come nell'indagine 2009, da aggregazione di strati elementari. Sul piano operativo è stata studiata la composizione delle imprese rispondenti all'indagine 2009 rispetto al possesso del carattere di innovatività e alle variabili di stratificazione. Circa gli aspetti computazionali è stata utilizzata una procedura classificatoria⁹ (*tree-based classification model*), in grado di identificare gli strati dove la frequenza attesa di imprese innovative è più elevata. Avendo individuato (nell'indagine del 2009) le caratteristiche degli strati in cui le imprese innovative presentavano una maggiore incidenza, si sono introdotti una serie ulteriore di vincoli nella procedura di allocazione del campione per ottenere un sovracampionamento negli strati suddetti, tale da garantire una frequenza attesa pianificata della quota di imprese innovative.

Disegno del campione

L'indagine è basata su un disegno di campionamento casuale stratificato, con selezione delle imprese negli strati con probabilità uguale e senza reimmissione. Gli strati sono definiti dall'aggregazione delle seguenti variabili: regione della sede legale dell'impresa, dimensione espressa in termini di classe di addetti settore di attività economica.

Il campione è stato allocato in modo da garantire un livello di accuratezza delle stime prodotte secondo alcuni domini di studio pianificati. Il disegno prevede in primo luogo la pianificazione dei singoli domini regionali; inoltre è stato assicurato un livello di significatività della stima, riferita a una generica proporzione di un carattere nella popolazione, per domini definiti congiuntamente dalla ripartizione geografica e del settore di attività, dalla ripartizione geografica e dalla classe dimensionale d'impresa. Sono stati infine considerate le sottopopolazioni di interesse definite in precedenza: tra queste ultime, le imprese con particolare propensione all'innovazione sono state definite in base ai risultati della scorsa indagine, identificando gli strati del campione con più alta incidenza di imprese innovative. Pianificando opportunamente la numerosità campionaria in tali strati è stato possibile controllare *ex-ante* il livello di accuratezza delle stime riferite alla sottopopolazione di imprese innovative.

⁹ *Tree-based classification model*.

L'indagine 2011 contiene inoltre una componente longitudinale, relativa a un sottocampione di imprese intervistate nella tornata 2009 e 2008. Tale elemento rappresenta, da un lato, un'opportunità preziosa data dalla possibilità di produrre stime di flusso, dall'altra impone un vincolo rilevante nel disegno del campione per l'indagine 2011, riducendo i margini per l'allocazione del campione, dal momento che una quota importante della distribuzione del campione negli strati risulta determinata dalle imprese presenti nella rilevazione 2009.

Nella tabella seguente sono riportate le partizioni delle variabili utilizzate nella stratificazione.

Tabella 2 Partizione del settore di attività economica.

Settore di attività	Sottosezione ATECO 2007
C10 – C11 – C12	Filiera dell'alimentare
C13 – C14 – C15	Filiera dell'abbigliamento
C16 – C31	Legno e mobili
C17 – C18	Carta, stampa ed editoria
C20 – C21 – C22	Gomma, plastica e chimica
C24 – C25	Metalli
C29 – C30	Fabbricazione di mezzi di trasporto
C28	Meccanica
C26 – C27	Macchine elettriche e apparecchiature elettroniche
B – D – C19 – C23 – C32	Altre industrie manifatturiere
H	Trasporti, poste e comunicazioni
J – M	Altri servizi alle imprese

Tabella 3. Partizione della classe dimensionale dell'impresa.

Dimensione impresa (addetti)
Micro 1-9
Piccole 10-49
Media 50-249
Grandi 250 e oltre

Al fine di definire un criterio di allocazione regionale che tenga conto della differente importanza delle regioni rispetto ai diversi obiettivi si è considerata la stima di una generica proporzione p , rispetto alla quale le regioni sono state divise in tre gruppi:

- ✓ **Gruppo A:** Regioni per le quali possono essere prodotte stime affidabili anche per caratteristiche *rare*, per le quali il valore della proporzione p è compreso tra il 2% e il 5%; in tali regioni le stime saranno affidabili anche per le caratteristiche *poco frequenti* (con p compreso tra il 6% e il 9%) e *frequenti* (con p superiore al 10%)

- ✓ **Gruppo B:** Regioni per le quali possono essere prodotte stime affidabili solo per le caratteristiche *poco frequenti o frequenti*;
- ✓ **Gruppo C:** Regioni per le quali possono essere prodotte stime affidabili solo per caratteristiche *frequenti*.

Per quanto riguarda il concetto di affidabilità, una stima è considerata affidabile se il suo corrispondente *coefficiente di variazione* % (espresso come rapporto % tra l'errore di campionamento assoluto e il valore della stima) è inferiore al 15%.

Tabella 4. Suddivisione delle regioni per affidabilità delle stime.

Gruppo A Affidabilità delle stime di caratteristiche rare	Gruppo B Affidabilità delle stime di caratteristiche poco frequenti	Gruppo C Affidabilità delle stime di caratteristiche frequenti
Lombardia	Emilia Romagna	Piemonte
Veneto	Umbria	Valle D Aosta
Toscana	Marche	Trentino A.A
Lazio	Puglia	Friuli V.G.
Campania	Calabria	Liguria
Sicilia	Sardegna	Abruzzo
		Molise
		Basilicata

Nel piano di vincoli che l'allocazione è chiamata a soddisfare è stato introdotto un dominio unico nazionale, allo scopo di controllare le stime aggregate a livello nazionale. Nell'algoritmo che ha prodotto il disegno di allocazione tale vincolo permette di controllare che lo scostamento dall'allocazione proporzionale non sia eccessivo, facendo lievitare l'*effetto del disegno*, con conseguente impatto negativo sull'affidabilità delle stime nazionali.

Il risultato della procedura di allocazione è riportato in tabella 5, dove è indicato, per ciascuno dei domini di stima presi in esame, il valore minimo della stima di una proporzione, p , che può essere considerato attendibile per il dominio stesso.

Si osservi che la procedura di allocazione utilizza separatamente le 4 classi dimensionali, in modo da controllare l'errore per le singole modalità. I domini definiti dalle distribuzioni congiunte rispetto a *regione e settore* e rispetto a *regione e dimensione* non permettono di controllare il livello di attendibilità delle stime secondo parametri di significatività statistica, specialmente per la distribuzione del settore all'interno di ciascuna regione.

Per tale motivo è stato introdotto un nuovo vincolo in cui la nidificazione di settore e dimensione all'interno della regione è limitata alle regioni con numerosità più elevata del campione.

Si è operato inoltre in modo da garantire una elevata attendibilità anche per caratteristiche poco frequenti nella popolazione, riferite alla distribuzione congiunta *regione x settore* per le regioni con maggiori numerosità, riportato nel dominio “*regione (numerosa) x settore*”.

Tabella 5. Livello minimo della stima per tipo di dominio.

Tipo dominio	Livello minimo di del valore <i>p</i> per cui la stima è affidabile
<i>Dimensione</i>	
Micro 1-9	0,003
Piccole 10-49	0,016
Media 50-249	0,019
Grandi 250 e oltre	0,001
<i>Regione</i>	
Piemonte	0,081
Valle D Aosta	0,121
Lombardia	0,025
Trentino A.A	0,068
Veneto	0,016
Friuli V.G.	0,088
Liguria	0,084
Emilia Romagna	0,031
Toscana	0,015
Umbria	0,056
Marche	0,037
Lazio	0,013
Abruzzo	0,088
Molise	0,083
Campania	0,019
Puglia	0,027
Basilicata	0,087
Calabria	0,031
Sicilia	0,024
Sardegna	0,060
<i>Settore</i>	
Tutte le modalità	0,117
<i>Regione x dimensione</i>	
Tutte le modalità	0,117
<i>Regione x settore</i>	
Tutte le modalità	0,656
<i>Regione (numerosa) x dimensione</i>	
Tutte le modalità	0,128
<i>Regione (numerosa) x settore</i>	
Tutte le modalità	0,150
<i>Nazionale</i>	
Tutte le modalità	0,003

Tabella 6. Allocazione del campione sezionale.

Variabili	Campione		Popolazione		Tasso di sondaggio
	Imprese	Valore %	Imprese	Valore %	
<i>Regione</i>					
Piemonte	1.678	6,7	113.647	7,4	1,5
Valle D'Aosta	331	1,3	3.532	0,2	9,4
Lombardia	2.417	9,6	308.887	20,1	0,8
Trentino-Alto Adige	794	3,2	25.624	1,7	3,1
Veneto	2.330	9,3	142.482	9,3	1,6
Friuli Venezia Giulia	644	2,6	30.412	2,0	2,1
Liguria	721	2,9	41.177	2,7	1,8
Emilia-Romagna	1.639	6,5	134.453	8,8	1,2
Toscana	2.500	10,0	118.497	7,7	2,1
Umbria	806	3,2	24.233	1,6	3,3
Marche	1.129	4,5	47.001	3,1	2,4
Lazio	2.404	9,6	146.086	9,5	1,6
Abruzzo	682	2,7	31.564	2,1	2,2
Molise	457	1,8	7.103	0,5	6,4
Campania	1.686	6,7	111.755	7,3	1,5
Puglia	1.275	5,1	80.709	5,3	1,6
Basilicata	639	2,5	12.050	0,8	5,3
Calabria	943	3,8	35.131	2,3	2,7
Sicilia	1.234	4,9	85.762	5,6	1,4
Sardegna	781	3,1	34.898	2,3	2,2
<i>Settore di attività</i>					
Filiera dell'alimentare	1.891	7,5	68.221	4,4	2,8
Filiera dell'abbigliamento	2.029	8,1	70.991	4,6	2,9
Legno e mobili	1.565	6,2	54.913	3,6	2,8
Carta, stampa ed editoria	1.370	5,5	28.837	1,9	4,8
Gomma, plastica e chimica	1.379	5,5	17.252	1,1	8,0
Metalli	2.292	9,1	94.242	6,1	2,4
Fabbricazione di mezzi di trasporto	722	2,9	7.520	0,5	9,6
Meccanica	1.799	7,2	40.442	2,6	4,4
Macchine elettriche e apparecchiature elettron.	1.595	6,4	46.608	3,0	3,4
Altre industrie manifatturiere	1.600	6,4	55.504	3,6	2,9
Trasporti, poste e comunicazioni	2.480	9,9	146.587	9,5	1,7
Altri servizi alle imprese	6.368	25,4	903.886	58,9	0,7
<i>Dimensione</i>					
1-2 addetti	6.787	27,1	1.148.462	74,8	0,6
3-5 addetti	5.559	22,2	193.289	12,6	2,9
6-10 addetti	3.081	12,3	81.699	5,3	3,8
10-49 addetti	6.195	24,7	95.363	6,2	6,5
50-249 addetti	2.647	10,6	13.822	0,9	19,2
250-W addetti	821	3,3	2.368	0,2	34,7
<i>Forma giuridica</i>					
Ditte individuali	4.070	16,2	1.011.632	65,9	0,4
Soc. di persone	5.881	23,4	238.016	15,5	2,5
Cooperative	3.128	12,5	20.851	1,4	15,0
Soc. di capitali	12.011	47,9	264.504	17,2	4,5
Totale	25.090	100,0	1.535.003	100,0	1,6

I domini regionali, pur trattati separatamente, soddisfano i parametri che ne definiscono l'attribuzione ai gruppi definiti in precedenza.

Il risultato dell'allocazione nei differenti domini di stima è riportato nella tabella 6.

La componente *panel* è stata allocata tenendo conto di tre elementi: i) la disponibilità di imprese in ciascuno strato già intervistate nell'indagine 2009; ii) l'attendibilità di stime longitudinali su domini di studio pianificati; iii) il livello di mancata risposta atteso sulla componente *panel*.

La capienza è stata ricavata dall'allocazione delle imprese rispondenti nel 2009, rispetto alla medesima stratificazione utilizzata per l'indagine 2011. L'allocazione della componente *panel* è stata pianificata secondo i domini definiti dalle partizioni marginali delle variabili di stratificazione, con la particolarità che la regione e la classe dimensionale sono state trattate in modo analogo a quanto fatto per l'allocazione del campione complessivo.

Ciò ha permesso, anche per le stime longitudinali, di trattare separatamente le 4 classi dimensionali e le regioni di particolare interesse.

Il tasso di risposta atteso è stato provvisoriamente fissato a un livello di 0.9 per tutti gli strati e la capienza per ciascuno strato è stata modificata di conseguenza.

Il risultato dell'allocazione nei differenti domini di stima per la componente *panel* è riportato nella tabella seguente.

Tabella 7. Allocazione del campione longitudinale.

Variabili	Campione		Popolazione		Tasso di sondaggio
	Imprese	Valore %	Imprese	Valore %	
<i>Regione</i>					
Piemonte	838	6,0	103.395	7,6	0,8
Valle D'Aosta	177	1,3	3.479	0,3	5,1
Lombardia	708	5,1	256.380	18,9	0,3
Trentino-Alto Adige	500	3,6	24.366	1,8	2,1
Veneto	1.406	10,1	134.872	9,9	1,0
Friuli Venezia Giulia	301	2,2	27.210	2,0	1,1
Liguria	317	2,3	37.426	2,8	0,8
Emilia-Romagna	635	4,6	109.262	8,0	0,6
Toscana	1.547	11,1	106.800	7,9	1,4
Umbria	419	3,0	23.260	1,7	1,8
Marche	564	4,1	42.942	3,2	1,3
Lazio	1.533	11,0	131.182	9,7	1,2
Abruzzo	490	3,5	30.837	2,3	1,6
Molise	327	2,4	6.980	0,5	4,7
Campania	1.086	7,8	103.621	7,6	1,0
Puglia	785	5,6	71.273	5,2	1,1
Basilicata	396	2,8	11.807	0,9	3,4
Calabria	759	5,5	33.047	2,4	2,3
Sicilia	756	5,4	67.911	5,0	1,1
Sardegna	365	2,6	32.860	2,4	1,1
<i>Settore di attività</i>					
Filiera dell'alimentare	1.012	7,3	55.518	4,1	1,8
Filiera dell'abbigliamento	1.147	8,2	54.977	4,0	2,1
Legno e mobili	982	7,1	47.480	3,5	2,1
Carta, stampa ed editoria	778	5,6	24.003	1,8	3,2
Gomma, plastica e chimica	755	5,4	16.161	1,2	4,7
Metalli	1.023	7,4	77.061	5,7	1,3
Fabbricazione di mezzi di trasporto	383	2,8	6.236	0,5	6,1
Meccanica	866	6,2	33.454	2,5	2,6
Macchine elettriche e apparecchiature elettron.	834	6,0	39.703	2,9	2,1
Altre industrie manifatturiere	927	6,7	45.922	3,4	2,0
Trasporti, poste e comunicazioni	1.246	9,0	123.261	9,1	1,0
Altri servizi alle imprese	3.956	28,4	835.137	61,5	0,5
<i>Dimensione</i>					
1-9 addetti	9.107	65,5	1.248.581	91,9	0,7
10-49 addetti	2.928	21,1	94.178	6,9	3,1
50-249 addetti	1.379	9,9	13.817	1,0	10,0
250-W addetti	495	3,6	2.335	0,2	21,2
<i>Forma giuridica</i>					
Ditte individuali	1.068	7,7	852.008	62,7	0,1
Soc. di persone	4.226	30,4	222.475	16,4	1,9
Cooperative	1.011	7,3	20.851	1,5	4,8
Soc. di capitali	7.604	54,7	263.578	19,4	2,9
Totale	13.909	100,0	1.358.912	100,0	1,0

Fase di stima

La fase di stima fa riferimento alla classe degli stimatori calibrati, che consentono, in presenza di determinate condizioni, di generare uno stimatore in grado di riprodurre una serie di totali noti nella popolazione di riferimento dell'indagine e, grazie all'impianto metodologico degli stimatori di regressione, permettono di recuperare parte della distorsione indotta sul campione dei rispondenti dalle tendenze non casuali delle mancate risposte totali.

La fase di stima ha previsto la costruzione di uno stimatore sezionale, riferito all'intera indagine condotta nel 2011, e uno longitudinale, riferito alla sola quota *panel* dell'indagine.

La procedura di costruzione dello stimatore calibrato sezionale, riferito all'intera indagine 2011, ha utilizzato una serie di informazioni ausiliarie, ricavate dalla popolazione di riferimento dell'indagine, introdotti come sistema di vincoli che il campione è chiamato a riprodurre. In tal modo le stime prodotte dal campione con l'applicazione dello stimatore calibrato risultano coerenti rispetto ai vincoli imposti.

La fonte informativa per la messa a punto del sistema di vincoli è l'Archivio Statistico delle Imprese Attive (Asia) diffuso dall'Istat, relativo all'anno 2009. I vincoli considerati nella procedura di costruzione dello stimatore comprendono i domini di stima pianificati a cui sono state aggiunte alcune partizioni più accurate delle variabili di stratificazione.

Il sistema di vincoli presenta la struttura riportata nella tabella seguente.

Tabella 8. Il sistema di vincoli.

Regione	x	Settore di attività
Regione	x	Dimensione
Regione	x	Forma giuridica
Settore di attività	x	Dimensione
Dimensione	x	Forma giuridica
Settore di attività	x	Forma giuridica

In base a tale sistema di vincoli lo stimatore consente al campione di riprodurre il totale delle imprese nella popolazione in ciascuna regione e, all'interno ciascuna regione, la composizione secondo il settore di attività economica, la classe dimensionale, definita in termini di addetti, e la forma giuridica. Inoltre per il livello nazionale lo stimatore riproduce la distribuzione congiunta delle imprese nella popolazione secondo il settore di attività e la classe dimensionale, il settore di attività e la forma giuridica, la classe dimensionale e la forma giuridica. Le partizioni delle variabili utilizzate, e le relative numerosità nel campione e nella popolazione, sono analoghe a quelle riportate nella precedente tabella 7.

La procedura di calibrazione ha utilizzato un algoritmo iterativo appositamente messo a punto.

E' stato inoltre costruito un ulteriore stimatore per la componente longitudinale del campione, in grado di riferire le stime prodotte dal campione alla popolazione compresente tra il 2009 e il 2011. Questa fase ha comportato un passo ulteriore, relativo alla definizione dei vincoli riferiti alla popolazione compresente. Non essendo disponibile la popolazione delle imprese in attività sia nel 2009 che nel 2011 si è proceduto stimando le distribuzioni di tale popolazione in modo endogeno dalle stime prodotte dall'indagine 2011, opportunamente ponderate con lo stimatore sezionale generato nella fase di stima precedente. La popolazione longitudinale è stata pertanto definita come la popolazione cui è riferita l'intero campione dell'indagine 2011, eliminando le imprese costituite nell'intervallo 2009-2011 e le imprese costituite prima del 2009 e cessate nello stesso periodo. Le nuove imprese nate tra il 2009 e il 2011 sono state identificate dall'anno di costituzione rilevato dall'indagine, mentre le imprese cessate non sono ovviamente comprese nell'indagine. Dalla popolazione così definita sono state ricavate le stime dei totali, utilizzati come vincoli nella procedura di calibrazione, ottenendo lo stimatore calibrato longitudinale, secondo il medesimo schema utilizzato per lo stimatore sezionale, ma con una partizione meno aggregata della classe dimensionale. Tale scelta è stata imposta dalla minore numerosità della componente longitudinale del campione, pari a poco meno di 14.000 imprese. Le distribuzioni delle variabili utilizzate nella procedura di calibrazione riferite al campione e alla popolazione longitudinale sono analoghe a quelle riportate nella precedente tabella 7.

Appendice: procedura per l'allocazione del campione nei domini pianificati

Il procedimento seguito per allocare il campione nei domini di studio pianificati utilizza una tecnica specifica per le indagini multiobiettivo, dove l'interesse si concentra oltre che sulla stima di un generico parametro riferito all'intera popolazione oggetto di studio, anche sul valore del medesimo parametro in specifiche sottopopolazioni, chiamate domini di studio. In genere, in assenza di ulteriori informazioni sulla popolazione di interesse, l'allocazione ottimale prevede un disegno proporzionale. Tuttavia l'allocazione proporzionale, pur massimizzando l'efficienza delle stime riferite all'intera popolazione, non assicura il controllo dell'accuratezza delle stime riferite ai singoli domini di studio. La tecnica utilizzata (Cochran, 1977) permette di predeterminare il livello di accuratezza, espresso in termini di varianza campionaria delle stime prodotte, nei domini di studio, controllando parallelamente l'attendibilità delle stime riferite all'intera popolazione.

Sia data una popolazione stratificata in H strati ($h=1,2,\dots,H$) di numerosità N_h ; siano definiti D domini, ottenuti tramite una determinata aggregazione di strati, di numerosità N_d , ($d=1,2,\dots,D$):

$$N_d = \sum_{h=1}^H N_h \cdot I_{h,d}$$

dove $I_{h,d}$ indica se lo strato h contribuisce alla definizione del dominio d :

$$I_{h,d} = \begin{cases} 1 & \text{se } h \in d \\ 0 & \text{se } h \notin d \end{cases}$$

La varianza campionaria della stima di una frazione P della popolazione nel dominio d è data da:

$$V_d = \sum_{h=1}^H \frac{S_h^2 \cdot (N_h - n_h)}{(N_h - 1) \cdot n_h} \cdot \frac{N_h^2}{N_d^2} \cdot I_{h,d}$$

approssimando la quantità $(N_h - 1) \cong N_h$ per popolazioni sufficientemente grandi, si ottiene:

$$V_d \cong \sum_{h=1}^H \frac{S_h^2 \cdot (N_h - n_h)}{N_h \cdot n_h} \cdot \frac{N_h^2}{N_d^2} \cdot I_{h,d}$$

La quantità V_d può essere scomposta in due addendi, uno dei quali dipende dalle quantità n_h , cioè dalla allocazione del campione negli strati, mentre l'altro è indipendente dall'allocazione del campione ed è funzione della partizione della popolazione nei domini:

$$V_d \cong \sum_{h=1}^H \frac{S_h^2 \cdot (N_h - n_h)}{N_h \cdot n_h} \cdot \frac{N_h^2}{N_d^2} \cdot I_{h,d} = \sum_{h=1}^H \frac{S_h^2 \cdot N_h^2}{N_d^2 \cdot n_h} \cdot I_{h,d} - \sum_{h=1}^H \frac{S_h^2 \cdot N_h}{N_d^2} \cdot I_{h,d}$$

Posto:

$$V_{d0} = - \sum_{h=1}^H \frac{S_h^2 \cdot N_h}{N_d^2} \cdot I_{h,d} ; \quad V_{d1} = \sum_{h=1}^H \frac{S_h^2 \cdot N_h^2}{N_d^2 \cdot n_h} \cdot I_{h,d} ; \quad V_{dh}^2 = \frac{S_h^2 \cdot N_h^2}{N_d^2} \cdot I_{h,d}$$

si ottiene:

$$V_d = V_{d0} + V_{d1} = V_{d0} + \sum_{h=1}^H \frac{V_{dh}^2}{n_h}$$

La scomposizione della varianza così ottenuta permette di controllare l'allocazione del campione negli strati h in modo da assicurare un livello predeterminato del valore della varianza V_d della stima riferita al dominio d .

Il problema si sostanzia quindi nel definire un'allocazione del campione, di numerosità non fissata, in grado di rispettare una serie di vincoli posti alla varianza delle stime per ciascun dominio di studio.

A tale scopo è utile il teorema di Kuhn-Tucker che risolve problemi di minimo vincolato nella programmazione non lineare.

Applicato al problema in questione il teorema di Kuhn-Tucker può essere formulato nel modo seguente (Bethel, 1989):

dato un vettore di vincoli $V^*, \{V_1^*, V_2^*, \dots, V_D^*\}$ (*upper bound*) delle varianze V_d esiste un vettore di moltiplicatori $\lambda_D, \{\lambda_1, \lambda_2, \dots, \lambda_D\}$ tali che l'allocazione

$$n_h = \sqrt{\sum_{d=1}^D \lambda_d \cdot V_{dh}^2} \quad \text{soddisfa i vincoli } V_d \leq V_d^*, \quad d = 1, 2, \dots, D.$$

La definizione dei moltiplicatori λ_d non è immediata e richiede un algoritmo iterativo dovuto a Chromy (1987) (illustrato in Centra e Falorsi, 2007) di seguito esposto.

Sia fissato un valore iniziale del vettore dei moltiplicatori λ_d , pari a 1 per ciascun dominio di studio:

$${}_0\lambda_d = 1 \quad \forall d$$

sia definita l'allocazione del campione al passo k in funzione dei moltiplicatori λ_d :

$${}_k n_h = \sqrt{\sum_{d=1}^D {}_k \lambda_d \cdot V_{dh}^2}$$

siano definiti i valori dei moltiplicatori λ_d al generico passo $k+1$ come funzione dei corrispondenti valori al passo precedente:

$${}_{k+1}\lambda_d = {}_k \lambda_d \cdot \left(\frac{\sum_{h=1}^H \frac{V_{dh}^2}{{}_k n_h}}{V_d^* - V_{d0}} \right)^2$$

Iterando la procedura l'intero vettore dei moltiplicatori converge verso un valore che soddisfa il vincolo imposto.

Il procedimento utilizzato ha prodotto l'allocazione del campione negli strati come definiti in precedenza. La procedura applicata impone una serie di vincoli espressi in termini di varianza massima ammessa su distribuzioni marginali rispetto alla distribuzione congiunta dei caratteri di stratificazione; i vincoli utilizzati nella procedura sono riportati nella tabella 5. La dimensione campionaria nei domini di studio risultante è riportata in tabella 6.

L'effetto del disegno, vale a dire la perdita in termini di accuratezza delle stime riferite all'intera popolazione dovuta al particolare piano di campionamento utilizzato, è pari a 2,52. Ciò significa che, a livello nazionale, la varianza della stima di una generica proporzione della popolazione prodotta dal disegno utilizzato è pari a circa due volte e mezzo la varianza del campione casuale semplice di pari numerosità. Il livello dell'effetto del disegno non è giudicato particolarmente alto e va messo in relazione con gli indubbi benefici dovuti alla possibilità di predeterminare i livelli di accuratezza delle stime nei domini di studio.

1. Riferimenti bibliografici

Brancati, R. (2010). *Fatti in cerca di idee. Il sistema italiano delle imprese e le politiche tra desideri e realtà*. Roma: Donzelli.

Brancati, R. (A cura di). (2012). *Crisi industriale e crisi fiscale. Rapporto MET 2012*. Roma: Donzelli.

Centra, M. e Falorsi, P. (2007). *Strategie di campionamento per il monitoraggio e la valutazione delle politiche, Temi&Strumenti*. ISFOL.

Cochran, W. G. (1977). *Sampling Techniques, 3rd ed.* New York: John Wiley and Sons.