

10. Appendici

10.1 Le indagini MET: campionamento, metodologia, utilizzi scientifici e istituzionali

La MET è un centro di ricerca indipendente nato nel 1992 da studiosi provenienti da diverse università italiane e da altre istituzioni pubbliche e private.

Dal 2000 MET ha sviluppato un proprio approfondimento dedicato alle politiche per le attività produttive che propone analisi approfondite sia dell'*offerta pubblica* (da parte delle amministrazioni comunitarie, nazionali e regionali con una rilevazione diretta degli interventi) che della *domanda derivante dai bisogni delle imprese* e sulla struttura industriale ricavata attraverso una vasta indagine che si svolge con cadenza biennale (vedi *infra*).

La MET si è caratterizzata, a livello nazionale, come fonte indipendente che predisponesse uno studio sugli interventi di sostegno alle imprese associato con la più vasta indagine diretta presso le imprese industriali e dei servizi alla produzione svolta in Italia e una tra le maggiori in Europa (a eccezione di alcune rilevazioni delle statistiche ufficiali).

Dal 2008 al 2020 si sono susseguite 7 *waves* di indagine con oltre 168.000 interviste per avere ampia rappresentatività regionale, settoriale e dimensionale (cfr. *infra*). Nel 2020 è stato effettuato un supplemento di indagine Covid a due mesi dalla conclusione dell'indagine biennale per analizzare il ruolo di alcune componenti strutturali nella crisi derivata dalla pandemia.

L'ampiezza del questionario consente di approfondire scelte strategiche, *performance* aziendali, struttura finanziaria e vincoli tanto da consentire di effettuare una analisi dei diversi profili associati a vincoli e criticità reali e finanziarie.

Prima di dettagliare scelte, metodologie e procedure delle indagini, che costituiscono parte essenziale del contributo alla qualità delle informazioni raccolte, è opportuno ricordare in estrema sintesi le caratteristiche che hanno guidato le attività di seguito illustrate e che hanno portato alla scelta di realizzare numerosità così elevate e inusuali.

In primo luogo la scelta di osservare l'intero universo delle imprese dell'industria e dei servizi alla produzione: la ragione è che per comprendere stasi e dinamismo

del sistema occorre analizzare anche le attività di minore dimensione che, pur non necessariamente crescendo nel breve periodo in termini di addetti o di fatturato, possono intraprendere sentieri di sviluppo di grande rilevanza e da analizzare.

In secondo luogo, la scelta di utilizzare un questionario molto completo con riferimento a tutte le principali aree aziendali, da quelle di struttura interna alle relazioni esterne, alla componente finanziaria e di qualità del capitale umano (addetti e manager). Ciò consente di incrociare e di porre in relazione fenomeni o attività diversi raramente confrontabili su basi omogenee.

In terzo luogo, le scelte di campionamento sono state particolarmente originali inserendo, accanto a domini di studio tradizionali (settore, dimensione, regione), anche sovra-campionamenti nelle aree in cui era più probabile rintracciare fenomeni rari e importanti come la R&S, o altri driver della competitività.

10.1.1 La logica dell'indagine strutturale MET

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: per poter raggiungere gli obiettivi è stato necessario arrivare a numerosità campionarie molto elevate, con una media, per ciascuna delle 7 wave, di oltre 24000 interviste. Le indagini comprendono inoltre una cospicua componente longitudinale, finalizzata alla stima della dinamica dei comportamenti delle imprese; la quota panel è disegnata in modo da produrre stime attendibili per specifiche disaggregazioni della popolazione di interesse dell'indagine.

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 ma 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 sono stati oggetto di diverse pubblicazioni⁸⁷, 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 parametri, mentre il 2013 segnala ancora livelli di attività calanti a fronte di una accentuata variabilità ed eterogeneità dei fenomeni.

Nel 2019, infine, l'indagine è stata chiusa nel gennaio 2020 a due mesi dall'emergere della crisi da Covid-19.

La elevata numerosità campionaria 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 la componente *panel* del campione di imprese, particolarmente consistente e presente in tutte 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.

⁸⁷ Rapporti MET, vari anni editi da Donzelli Editore e Fatti in cerca di idee di R. Brancati, Donzelli Ed. 2010.

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. Così 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.

L'intera indagine viene svolta con il massimo rigore scientifico. Ci si è avvalsi anche del lavoro di un comitato scientifico di grande valore e di una commissione tecnica dedicata a seguire le fasi del campionamento, dell'implementazione dell'indagine e della complessa fase di stima e della costruzione 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), ottimizzando in tal modo l'efficienza delle stime prodotte

Il Questionario

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:

- *sezione informativa* (dati di base, informazioni su occupazione, struttura proprietaria, gruppi, *etc.*);
- *sezione mercati e struttura* (mercati di sbocco, vantaggi competitivi, debolezze, struttura dei costi, banche e finanza, *etc.*);
- *competenze e criticità* (livello formativo del personale, formazione ed esperienze del management, provenienza, genere, fabbisogni e criticità, soluzioni, *etc.*);
- *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.*);
- *sezione dinamismo* (programmi – passati e previsti – di investimento, crescita di occupazione e fatturato, obiettivi degli investimenti, scelte qualitative e indicazioni da altre sezioni);

- *sezione Ricerca e Innovazione* (programmi di innovazione, ricerche, relazioni con soggetti esterni per la ricerca, aree di sviluppo, *etc.*);
- *sezione Internazionale* (tipologia di presenza sui mercati esteri, alleanze, utilizzazione di strumenti nazionali e regionali di supporto, *etc.*);
- *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.*);
- *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);
- *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);
- *la sostenibilità sociale ed ambientale nelle azioni delle imprese* (è stata introdotta una sezione con domande specifiche);
- *temi specifici* (si sono alternate analisi su Energia e ambiente, costi amministrativi nei rapporti con la P.A., rilevanza delle certificazioni, impatto della crisi, procurement pubblico, i temi legati all'utilizzo delle tecnologie informatiche);
- *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).

La possibilità di “incrociare” così tanti fenomeni diversi rappresenta un unicum dell'indagine.

Il Trattamento delle liste campionarie nella fase di campo

Le anagrafiche delle imprese del 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 elenchi disponibili.

Ai fini del trattamento delle liste campionarie sono state predisposte due liste di anagrafiche: una lista di anagrafiche base, 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 un primo contatto nel corso del quale è stata presentata l'iniziativa di ricerca.

Circa il 75% delle interviste è stato effettuato in modalità web, predisponendo link appropriati sulla piattaforma dedicata, mentre la quota rimanente è stata realizzata in modalità CATI, ovvero telefonica.

Per lo svolgimento della rilevazione è stato implementato l'utilizzo e 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.

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.

Il quadro delle attività

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”.

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.

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.

10.1.2 La metodologia: il disegno campionario e la fase di stima⁸⁸

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 nell'intero territorio nazionale. Nello specifico l'universo considerato è composto da imprese che operano in 38 settori, definiti al terzo *digit* del Codice Ateco⁸⁹.

⁸⁸ Di Marco Centra (INAPP) e Piero Demetrio Falorsi (già Istat).

⁸⁹ La classificazione Ateco utilizzata nell'ambito del presente documento si riferisce a quella in vigore fino al 31 dicembre 2007. In sede di estrazione dell'unità statistiche ed in sede di rilevazione, si è tenuto conto tuttavia non solo di tale vecchia classificazione, ma anche della nuova classificazione Ateco delle imprese. I settori di attività delle imprese considerati nell'universo di interesse sono: il settore estrattivo (Codice Ateco C), l'industria alimentare, delle bevande e del tabacco (Codice Ateco DA), l'industria tessile e dell'abbigliamento (Codice Ateco DB), l'industria del concio, cuoio e pelli (Codice Ateco DC),

La dimensione campionaria dell'indagine, particolarmente elevata e pari a circa 25.000 unità, ha consentito l'effettuazione di ulteriori approfondimenti e sovra-campionamenti 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 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 tenuto conto del fatto che in alcune regioni è stato richiesto un maggior livello di accuratezza delle stime prodotte rispetto ad altre, e un conseguente sovra-campionamento, 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 a prevalente vocazione manifatturiera. Si tratta delle imprese che hanno la sede legale nelle regioni ad elevata incidenza del settore manifatturiero, vale a dire compreso nella sezione D della classificazione ATECO 2007: 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 2007.

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 del particolare scenario circa il bagaglio informativo attualmente disponibile: i) non si dispone, indipendentemente dalla definizione adottata di impresa innovativa, dell'informazione puntuale sulla popolazione; ii) le imprese innovative sono identificabili solo tramite informazioni rilevate dall'indagine precedente e quindi disponibili unicamente sul campione delle relative imprese rispondenti. Ipotizzando che le variabili esplicative

l'industria del legno e dei prodotti in legno (Codice Ateco DD), l'industria della carta, stampa ed editoria (Codice Ateco DE), la fabbricazione di coke, e le raffinerie di petrolio (Codice Ateco DF), l'industria chimica (Codice Ateco DG), l'industria della gomma e delle materie plastiche (Codice Ateco DH), la lavorazione dei metalli non metalliferi (Codice Ateco DI), la metallurgia (Codice Ateco DJ), l'industria meccanica (Codice Ateco DK), l'industria elettronica (Codice Ateco DL), la fabbricazione di mezzi di trasporto (Codice Ateco DM), le altre industrie manifatturiere (Codice Ateco DN), la produzione e distribuzione di energia elettrica, gas e acqua (Codice Ateco E), i settori dei trasporti, magazzinaggio e comunicazioni (Codice Ateco I – con l'esclusione delle agenzie di viaggio e degli operatori turistici, codice 63.3), e infine quelli del noleggio, dell'informatica, della ricerca e dei servizi alle imprese (Codice Ateco K – con l'esclusione delle attività immobiliari, codice 70).

della caratteristica innovativa di un'impresa siano rimaste sostanzialmente invariate rispetto all'anno di riferimento precedente, 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 da aggregazione di strati elementari. Sul piano operativo è stata studiata la composizione delle imprese rispondenti all'indagine precedente rispetto al possesso del carattere di innovatività e alle variabili di stratificazione. Circa gli aspetti computazionali è stata utilizzata una procedura classificatoria *tree-based classification model*, in grado di identificare gli strati dove la frequenza attesa di imprese innovative è più elevata. Avendo individuato (nell'indagine dell'anno precedente) 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 e un corrispondente livello di attendibilità delle stime prodotte.

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, 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 delle stime, riferite 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, come sopra indicato.

L'indagine di ciascun anno contiene inoltre una componente longitudinale, relativa a un sottocampione di imprese intervistate nelle tornate precedenti. 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, 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 precedente.

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

Partizione del settore di attività economica

Settore di attività	Sottosezione ateco 2007
Filiera dell'alimentare	DA
Filiera dell'abbigliamento	DB- DC
Legno e mobili	DD -D36.1
Carta, stampa ed editoria	DE
Gomma, plastica e chimica	DH - DG - DF
Metalli	DJ
Fabbricazione di mezzi di trasporto	DM
Meccanica	DK
Macchine elettriche e apparecchiature elettroniche	DL
Altre industrie manifatturiere	E -C DI - DN (escluso 36.1)
Trasporti, poste e comunicazioni	I - 60 - 61 - 62 - 63 - 64
Altri servizi alle imprese	K - 71 - 72 - 73 - 74

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:

1. **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%);

2. **Gruppo B:** Regioni per le quali possono essere prodotte stime affidabili solo per le caratteristiche *poco frequenti o frequenti*;
3. **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%.

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 3, 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; in tal modo è stato possibile tener conto della marcata eterogeneità dimensionale del tessuto produttivo italiano, ottimizzando al contempo la numerosità del campione in ciascuna delle classi dimensionali. 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 $\times$ settore* per le regioni con maggiori numerosità, riportato nel dominio "*regione (numerosa) $\times$ settore*".

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

La componente *panel* è stata allocata tenendo conto di tre elementi: i) la disponibilità di imprese in ciascuno strato già intervistate nell'indagine precedente; 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 periodo passato, rispetto alla medesima stratificazione utilizzata per l'indagine in corso. L'allocazione della componente *panel* è stata pianificata secondo i domini definiti dalle partizioni marginali delle variabili di stratificazione, con la particola-

rità 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.

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, e uno longitudinale, riferito alla sola quota panel dell'indagine.

La procedura di costruzione dello stimatore calibrato sezionale 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 tramite il portale i.stat. 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.

Regione	x	Settore di attività		
Regione	x	Dimensione		
Regione	x	Forma giuridica		
Regione	x	Classe di fatturato		
Ripartizione geografica	x	Settore di attività	x	Dimensione

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 dimensio-

nale, definita in termini di addetti, e la forma giuridica. Inoltre lo stimatore riproduce la distribuzione congiunta delle imprese nella popolazione secondo la ripartizione geografica, il settore di attività e la classe dimensionale. Le partizioni delle variabili utilizzate, e le relative numerosità nel campione e nella popolazione, sono analoghe a quelle riportate nella precedente Tabella 4.

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

è 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 2011 e il 2013. 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 nell'anno di riferimento dell'indagine che in quello dell'indagine precedente si è proceduto stimando le distribuzioni di tale popolazione in modo endogeno dalle stime prodotte dall'indagine, opportunamente ponderate con lo stimatore sezionale generato nella fase di stima precedente. La popolazione longitudinale è stata pertanto definita come la popolazione cui è riferito l'intero campione dell'indagine, eliminando le imprese costituite nell'intervallo tra indagine precedente e corrente e le imprese costituite prima della rilevazione passata e cessate nello stesso periodo. Le nuove imprese nate 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.

10.1.3 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 multi-obiettivo, 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, secondo il modello dell'allocazione ottima di Neyman. 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 ($b=1,2,\dots,H$) di numerosità N_b ; 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 b 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:

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

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

La definizione dei moltiplicatori λ_d non è immediata e richiede un algoritmo iterativo dovuto a Chromy (1987) (illustrato in Centra 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.

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.

10.1.4 Utilizzi istituzionali dei dati delle Indagini MET e riconoscimenti

Di seguito si riportano alcuni dei lavori istituzionali che hanno direttamente utilizzato e pubblicato dati delle Indagini MET.

- ICE-ISTAT (2013), Rapporto ICE 2013, utilizzazione dati ed elaborazioni da Indagine MET. 2011.
- ISTAT, Istituto nazionale di statistica (2018). Rapporto sulla competitività dei settori produttivi. Utilizzo dati tratti dall' Indagine MET (2015) per un esame del profilo strategico delle imprese industriali e per uno studio del ruolo delle attività di innovazione sulla performance economica e occupazionale delle imprese.
- ISTAT, Istituto nazionale di statistica (2018). Rapporto annuale. La situazione del paese. Utilizzo dati tratti dall' Indagine MET (2018).
- ISTAT, Istituto nazionale di statistica (2020). Collaborazione strutturata con MET su metodologia per la produzione di “Nuovi indicatori sulla struttura e i comportamenti delle imprese industriali” che utilizzi dati Indagini MET.
- MISE, Ministero dello Sviluppo Economico (2018). Utilizzo indagine MISE-MET su Imprese e tecnologia 4.0.
- MISE Ministero dello Sviluppo Economico MIUR Ministero dell'università e della Ricerca, vari anni utilizzo dati delle Indagini MET per predisposizione Piani e strategie di valutazione.

Il Presidente della Repubblica italiana Giorgio Napolitano, ha concesso 2 Medaglie come riconoscimento sia per l'Indagine MET 2009 che per quella del 2011.

Nel 2020 la MET è stata riconosciuta come ente di ricerca qualificato dall'ISTAT Istituto Centrale di Statistica.