



Metodi e fonti: note metodologiche

3 Luglio 2017

Per informazioni: statistiche@bancaditalia.it
www.bancaditalia.it/statistiche/index.html

Indagine sulle imprese industriali e dei servizi

Aspetti generali

La Banca d'Italia conduce indagini campionarie sulle imprese dal 1972¹. Fino al 1998 l'indagine ha riguardato le sole imprese della trasformazione industriale con 50 addetti e oltre. Dal 1999 il campo di osservazione si è ampliato, estendendosi al complesso delle imprese manifatturiere e ai settori energetico-estrattivo e, dal 2001, anche alle imprese con 20-49 addetti. Dal 2002 la popolazione di riferimento è stata allargata alle imprese dei servizi privati non finanziari² con 20 addetti e oltre.

Dal 2006 la rilevazione si è estesa anche al settore delle costruzioni con 20 addetti e oltre. L'indagine sul 2013 ha esteso il campione delle costruzioni alle imprese con 10-19 addetti, al fine di rendere l'indagine maggiormente rappresentativa degli andamenti economici delle imprese di minore dimensione, particolarmente importanti nel settore, caratterizzato da dimensione media d'impresa (in termini di addetti) molto inferiore a quella dei settori dell'industria e dei servizi privati non finanziari.

In seguito alla revisione metodologica del 1987, lo schema di estrazione adotta un disegno probabilistico stratificato non proporzionale, che è rimasto nella sostanza immutato, sebbene la numerosità complessiva sia aumentata nel tempo. L'indagine utilizza lo stimatore di Horvitz-Thompson, con opportune riponderazioni per il trattamento degli *outlier*. Nel corso degli ultimi anni, soprattutto per le variabili con più elevata presenza di dati anomali, le stime sono state rese robuste mediante tecniche di *winsorizzazione*.

Fino al 2003 erano utilizzati per la classificazione dimensionale gli addetti a fine anno; dal 2004, sia per omogeneità con le rilevazioni esterne sia per evitare effetti stagionali sul numero di occupati, le imprese sono classificate in base al numero di addetti medi annui.

Nel seguito sono descritti nel dettaglio i vari aspetti metodologici che caratterizzano l'indagine³.

1 L'anno è quello di riferimento dell'indagine, che viene condotta nei primi mesi dell'anno successivo.

2 Sono esclusi dall'indagine: le attività d'intermediazione finanziaria (su cui la Banca d'Italia dispone di una grande mole di dati, raccolti per le finalità istituzionali di vigilanza e politica monetaria) e assicurativa; la Pubblica amministrazione e i settori scolastico e sanitario e gli altri servizi pubblici, sociali e personali.

3 Per ulteriori dettagli sulle indagini precedenti si rinvia alla prima pubblicazione periodica riguardante l'indagine (Banca d'Italia Indagine sulle imprese industriali e dei servizi. Anno di riferimento 2003 in Supplementi al Bollettino Statistico – Indagini campionarie n. 55, 20 ottobre 2005).

La composizione dell'universo e del campione

L'universo di riferimento (tav. 11) è costituito dalle imprese con sede amministrativa in Italia, con almeno 20 addetti, appartenenti ai settori di attività economica dell'industria in senso stretto e dei servizi privati non finanziari; con almeno 10 addetti per quello delle costruzioni. Dall'indagine sul 2010 si utilizza la classificazione delle attività economiche Ateco 2007 (tav. 1)⁴.

Negli anni più recenti il campione è composto da circa 5.000 imprese (circa 3.000 imprese dell'industria in senso stretto, circa 1.000 imprese dei servizi privati non finanziari, circa 600 delle costruzioni). Le tavole I1 e I2 dell'Appendice statistica riportano la numerosità della popolazione e la percentuale di imprese sondate dall'indagine. La frazione sondata è più elevata per le imprese con almeno 50 addetti e per quelle con sede nel Sud e nelle Isole.

Tav. 1

Definizione dei settori di attività economica

	Sezione Ateco 2007	Divisione Ateco 2007	Settore di attività economica	Aggregazioni di settori utilizzate nelle tavole
Industria in senso stretto	C	10-12	Alimentari, bevande e tabacco	Altre manifatturiere
		13-15	Tessili, abbigliamento, pelli e calzature	Tessili, abbigliamento, pelli e calzature
		19-22	Chimica, gomma e plastica	Chimica, gomma e plastica
		23	Minerali non metalliferi	Altre manifatturiere
		24-30; 33	Metalmeccanica	Metalmeccanica
		16-18; 31-32	Altra industria manifatturiera (legno, fabbricazione pasta carta, altre industrie manifatturiere)	Altre manifatturiere
	B	05-09	Estrazioni di minerali da cave e miniere	Energetiche ed estrattive
D	35	Fornitura di energia elettrica	Energetiche ed estrattive	
E	36-39	Fornitura di acqua	Energetiche ed estrattive	
Costruzioni	F	41-43	Costruzioni	Costruzioni
Servizi privati non finanziari	G	45-47	Commercio all'ingrosso e al dettaglio, riparazioni	Commercio, alberghi e ristorazione
	I	55-56	Attività dei servizi di alloggio e di ristorazione	
	H	49-53	Trasporti, magazzinaggio	Trasporti, magazzinaggio e comunicazioni
	J	58-63	Servizi di informazione e comunicazione	
L, M, N ^(a)	68-75; 77-82	Altri servizi a imprese e famiglie	Altri servizi a imprese e famiglie	

(a) Comprende: L = attività immobiliari; M = attività professionali, scientifiche e tecniche; N = noleggio, agenzie di viaggio, servizi di supporto alle imprese.

4 Si veda Istat (2009), Metodi e Norme, n. 40.

Il disegno campionario

Lo schema di campionamento dell'indagine è di tipo stratificato. Gli strati sono costituiti dalle combinazioni di settore di attività economica (sulla base della classificazione riportata alla tav. 1), classe dimensionale (in termini di addetti)⁵ e sede amministrativa dell'impresa⁶.

Per la determinazione della numerosità campionaria si opera in due fasi: nella prima s'individua la numerosità a livello di classe dimensionale, utilizzando il metodo noto come *optimum allocation to strata*⁷, che consente di minimizzare la varianza delle medie e delle variazioni campionarie delle principali variabili rilevate (occupazione, fatturato e investimenti). Nella seconda fase la numerosità di ciascuna classe dimensionale è ripartita tra le regioni e i settori di attività economica in proporzione al numero d'impresе della popolazione d'interesse appartenenti allo strato⁸.

Le imprese da rilevare sono selezionate sulla base degli archivi INPS, Infocamere e altre liste reperite autonomamente dalle Filiali, allo scopo di ridurre al minimo il rischio di copertura incompleta della popolazione. Le imprese rilevate nell'edizione precedente sono sempre ricontattate, se fanno ancora parte della popolazione d'interesse⁹, mentre quelle non più disposte a collaborare sono sostituite con altre omogenee per tipo di attività economica e classe dimensionale.

La fase di rilevazione, il questionario e i comportamenti di risposta

La rilevazione viene svolta annualmente tra febbraio e maggio, con riferimento all'attività economica di fine anno precedente; le interviste sono condotte dalle Filiali della Banca.

Dall'edizione relativa al 2010, il caricamento dei dati è interamente realizzato tramite un'applicazione accessibile via *web*.

L'indagine raccoglie ogni anno informazioni su aspetti anagrafici e strutturali, occupazione, investimenti, fatturato, risultato d'esercizio, capacità produttiva, indebitamento¹⁰. Le sezioni monografiche sono dedicate a specifici temi di interesse¹¹.

Per contenere l'onere di risposta, i quesiti monografici possono essere occasionalmente posti solo a un sottoinsieme causale del campione¹².

5 Le classi di addetti, riferite fino al 2003 al numero di addetti a fine anno e dal 2004 al numero medio di addetti nell'anno, sono: 20–49 (a partire dal 2013, per il campione delle imprese di costruzioni è stata aggiunta la classe 10–19 addetti), 50–99, 100–199, 200–499, 500–999, 1.000–4.999 addetti, 5.000 addetti e oltre. Le unità con 5.000 addetti e oltre formano uno strato autorappresentativo.

6 Piemonte e Valle d'Aosta sono considerate un'unica regione.

7 Cfr., ad esempio, Cochran W. G., *Sampling Techniques*, New York, John Wiley & Sons, 1977.

8 La numerosità campionaria regionale viene poi ripartita tra le Filiali, che curano direttamente la rilevazione. In alcune aree del Paese si procede, se necessario, a un sovracampionamento per le autonome esigenze di studio delle realtà economiche locali.

9 Tipiche cause di uscita dalla popolazione d'interesse sono la variazione di attività economica e la diminuzione del numero di addetti al di sotto della soglia d'inclusione: cfr. par. 4.

10 Per le imprese delle costruzioni si rileva, in luogo d'investimenti e fatturato, il valore della produzione (sia quella totale sia quella realizzata in opere pubbliche).

11 Per maggiori dettagli cfr. i questionari al link: <https://www.bancaditalia.it/statistiche/tematiche/indagini-famiglie-impresе/impresе-industriali/distribuzione-microdati/questionario-indagine-impresе/index.html>.

12 Per la ripartizione dei quesiti nei questionari destinati a un gruppo d'impresе A o a un gruppo d'impresе B, si vedano i questionari.

I contenuti e la struttura dei questionari vengono definiti nel periodo immediatamente precedente alla rilevazione. Le nuove domande vengono sottoposte a test attraverso una rilevazione pilota affidata alle Filiali, al fine di verificare la comprensibilità del testo e la disponibilità delle informazioni presso le imprese.

Per occupazione, investimenti e fatturato, le informazioni vengono richieste con riferimento a tre periodi: l'anno appena concluso (dato di preconsuntivo), l'anno precedente (dato di consuntivo) e quello successivo (dato di previsione).

La tavola Tav. I3 riporta i tassi di risposta all'indagine.

Il controllo di qualità dei dati e l'imputazione dei dati mancanti

I dati raccolti sono sottoposti a un processo di controllo di qualità. In particolare si verifica la compatibilità dei valori con quelli ammessi dalla domanda, la coerenza intertemporale dei dati *panel* e quella di alcuni specifici importi e la presenza di valori anomali. La validazione dei dati avviene in modo gerarchico: ogni livello di controllo si applica ai dati ritenuti validi al livello precedente.

Un primo controllo del questionario viene effettuato dal personale della Banca d'Italia che cura la rilevazione, che, sulla base della competenza tecnica e della conoscenza dell'economia locale, effettua una verifica qualitativa dei dati raccolti¹³.

La procedura informatica di caricamento impedisce l'acquisizione di dati esterni all'intervallo di definizione della variabile o incompatibili con i vincoli di coerenza interni al questionario. I dati accettati dalla procedura d'immissione possono risultare però anomali rispetto a soglie statisticamente plausibili su base storica o provenienti da informazioni esterne: in questo caso essi vengono evidenziati e sottoposti all'attenzione dei rilevatori, i quali sono tenuti a verificarli e, se del caso, a ricontattare l'impresa per avere chiarimenti. L'eventuale conferma viene registrata in un apposito campo. Solo a quel punto il dato scrutinato viene considerato valido ai fini dell'indagine.

Nella successiva fase dei controlli vengono usate tecniche di *editing* fondate su assunzioni di tipo statistico che consentono di identificare le osservazioni anomale rispetto alle distribuzioni *cross-section* (su scala naturale o logaritmica).

Un'ulteriore procedura di controllo, detta di *selective editing*, è in grado di fornire una lista di priorità per le verifiche da effettuare a fronte di osservazioni anomale valutate sulla base dell'importanza di ognuna per la stima finale. I valori della singola variabile vengono confrontati con quello previsto da un semplice modello di regressione; a partire da tale statistica viene costruito, sulla base di un'approssimazione di Taylor, un punteggio (*score*) per ciascuna impresa, che dipende dall'effetto che il singolo valore avrebbe sulla stima finale condizionatamente al modello. Più lo *score* è elevato (cioè maggiore è l'impatto del valore osservato sulla stima finale secondo il modello) più urgente è la verifica del valore di quella variabile per quell'impresa¹⁴. Il procedimento viene integrato per un insieme di più variabili, compilando una graduatoria rispetto a una funzione che sintetizza gli *score* riferiti alle singole variabili. Questa tecnica consente di elevare la qualità delle stime riducendo al contempo il *respondent burden* nella fase finale delle elaborazioni, poiché limita il riesame

13 In questa fase i rilevatori, utilizzando una scheda che riporta i dati delle principali variabili forniti dalle imprese che hanno partecipato all'indagine precedente, effettuano anche una prima verifica di coerenza.

14 Cfr. Battipaglia P., Selective editing to increase efficiency in survey data processing. An application to the Bank of Italy's Business Survey on Industrial Firms, in: Irving Fisher Committee Bulletin 13, December 2002, pp. 149–154.

e l'eventuale ricontatto solo a quelle imprese che hanno un impatto giudicato significativo sulle stime finali.

Le imprese interessate da eventi straordinari quali ad esempio fusioni o scorpori sono incluse nel campione solo se in grado di fornire informazioni omogenee per il triennio rilevato nella stessa indagine (informazione a consuntivo, preconsuntivo e previsione). A tal fine, i dati possono essere ricostruiti considerando l'evento straordinario come già avvenuto all'inizio dell'anno di consuntivo, oppure come se l'evento non si fosse affatto verificato. Se una tale ricostruzione non è possibile, l'impresa viene esclusa dalle elaborazioni. Sebbene tale pratica possa costituire una fonte di distorsione delle stime dei livelli (quali gli investimenti totali), essa conferisce maggiore stabilità alle stime delle variazioni e dei valori medi per addetto, che sono il principale obiettivo dell'indagine.

Le imprese che partecipano all'indagine possono talvolta incontrare difficoltà nel rispondere ad alcuni quesiti. Se la mancata risposta riguarda le variabili principali, come ad esempio la spesa programmata per investimenti, il fatturato o l'occupazione previsti, si procede all'imputazione dei dati mancanti.

L'imputazione si avvale in generale di stimatori per rapporto, utilizzando come denominatore il numero di addetti medi dell'impresa nell'anno di riferimento che consente di cogliere l'effetto di scala¹⁵. In alcuni casi nella ricostruzione ci si avvale anche dei dati storici dell'impresa, sotto forma di effetti individuali. In questo modo si stima un livello per addetto ottenuto come combinazione tra una media generale di tipo *cross-section* e una media calcolata solo sui valori storici dell'impresa. I livelli ai tempi t e $t+1$ sono ricostruiti in sequenza, calcolando variazioni medie in opportune celle contenenti imprese omogenee¹⁶.

I dati elementari imputati sono identificati da appositi *flag*¹⁷. La percentuale di dati imputati è comunque molto contenuta per le variabili fondamentali a livello di consuntivo e preconsuntivo. Le domande relative alle previsioni, in particolare quelle inerenti agli investimenti, tendono ad avere un tasso di mancata risposta relativamente più elevato. Può risultare elevato, in funzione della difficoltà dei quesiti, anche il tasso di mancata risposta delle sezioni monografiche.

La ponderazione

L'indagine adotta una procedura di ponderazione stratificata che si compone di due fasi.

In una prima fase si usano come strati le combinazioni di settore di attività economica e classe dimensionale. A ogni impresa si attribuisce un peso iniziale, dato dal rapporto tra numero d'impresе contenute nella cella di strato e numero d'impresе del campione. Sia h la generica cella di strato e al suo interno N_h il numero d'impresе della

15 Per l'analisi delle condizioni sotto cui lo stimatore per rapporto è preferibile rispetto alla media. Cfr., ad esempio, Cicchitelli F. et al., Il campionamento statistico, Bologna, Il Mulino, 1994.

16 Nella fase d'imputazione si utilizzano ulteriori accorgimenti. Ad esempio se una cella nella quale andrebbe calcolata una media contiene un numero esiguo d'impresе, essa viene unita a celle contigue secondo la classe dimensionale, o l'area geografica. Nelle celle vengono inoltre calcolate medie robuste, che limitano l'influenza degli outlier nel processo di ricostruzione. In alcuni casi si adottano soluzioni particolari che sfruttano dati rilevati nello stesso questionario, ad esempio in presenza di vincoli aritmetici tra le variabili o di legami meno stretti, ma tali da consentire un'affidabile ricostruzione del dato mancante. Ad esempio, se un'impresa non dichiara il numero di assunzioni effettuate nell'anno, esso è calcolato aggiungendo il numero di cessazioni alla differenza tra l'occupazione a fine e inizio anno; analogamente si procede se manca il numero di cessazioni.

17 Cfr. <https://www.bancaditalia.it/statistiche/tematiche/indagini-famiglie-impresе/impresе-industriali/distribuzione-microdati/index.html>: Descrizione delle variabili dell'Indagine sulle impresе industriali e dei servizi.

popolazione d'interesse e n_h quello del campione¹⁸. Il peso della prima fase per ogni impresa dello strato h è pertanto:

$$(1) \quad w_h^{(0)} = \frac{N_h}{n_h}$$

Per costruzione, la somma dei pesi di ogni cella restituisce pertanto la numerosità della popolazione d'interesse in essa contenuta.

In una seconda fase si effettua una stratificazione a posteriori, utilizzando la tecnica del *raking*¹⁹, per tenere conto anche dell'area geografica k in cui è localizzata la sede amministrativa dell'impresa. Il peso iniziale è modificato con un fattore di aggiustamento f_k , consentendo di ottenere i pesi finali:

$$(2) \quad w_{hk}^{(1)} = w_h^{(0)} f_k$$

la cui somma coincide, nelle celle di post-strato, con il numero d'impresе della popolazione d'interesse ivi contenute.

Il sistema di pesi finale non tiene dunque conto di tutte le interazioni tra settore, classe dimensionale e area geografica, sia perché alcune combinazioni possono presentare poche unità campionarie, sia perché un sistema di pesi siffatto potrebbe condurre a stime eccessivamente variabili per alcuni domini di studio. La stratificazione a posteriori si limita perciò a ricostruire le combinazioni tra:

- a) Nord Ovest, Nord Est, Centro, Sud e Isole;
- b) imprese con 20-49 addetti e con 50 addetti e oltre (per le imprese di costruzioni dal 2013 si considera anche la classe 10-19 addetti);
- c) imprese manifatturiere, energetiche ed estrattive, delle costruzioni e dei servizi (queste ultime distinte per sezione; cfr. par. 2).

In ogni indagine, i pesi vengono ricalcolati sulla base della distribuzione della popolazione d'interesse alla data più recente disponibile²⁰.

Le stime campionarie

Per una generica variabile X , la stima dell'aggregato (\mathcal{J}) è effettuata con uno stimatore corretto del totale di tipo Horvitz-Thompson²¹, dato da:

$$(3) \quad \hat{\mathcal{J}} = \sum_{h=1}^L \hat{\mathcal{J}}_h = \sum_{h=1}^L w_h \left(\sum_{i=1}^{n_h} x_{i,h} \right) = \sum_{h=1}^L N_h \bar{x}_h ,$$

dove L è il numero totale di strati e \bar{x}_h è lo stimatore media campionaria per lo strato h -esimo.

18 Il simbolo n_h indica la numerosità campionaria effettiva. In questo modo si effettua implicitamente una correzione dei pesi per tenere conto delle mancate risposte totali (unit non-response).

19 La tecnica denominata iterative proportional fitting (o raking) consente di allineare in modo simultaneo i pesi campionari alla distribuzione di alcune caratteristiche note da fonti esterne. Cfr., per esempio, Verma V., *Advanced sampling methods: manual for statistical trainers*, Tokyo, Statistical Institute for Asia and the Pacific, 2000, pp. 6.13-6.21.

20 La popolazione d'impresе utilizzata è stata ottenuta da Istat, *Archivio statistico delle imprese attive*, 2013. Gli aggiornamenti sono normalmente rilasciati dall'Istat con riferimento all'universo di circa due anni prima. Le stime sono riviste periodicamente per tenere conto dei cambiamenti nella popolazione di riferimento dell'indagine.

21 Cfr., ad esempio, Cicchitelli F. et al., *Il campionamento statistico*, Bologna, Il Mulino, 1994.

La stima dei tassi di variazione, ad esempio del fatturato o degli investimenti, viene effettuata usando come stimatore il rapporto tra le sommatorie dei livelli per le singole imprese, ponderate con i coefficienti di riporto all'universo, con una formula del tipo:

$$(4) \quad \hat{r}_{t,t-1} = \frac{\hat{f}^t}{\hat{f}^{t-1}},$$

dove \hat{f}^t e \hat{f}^{t-1} rappresentano rispettivamente lo stimatore dell'aggregato al tempo t e $t-1$.

I livelli dell'aggregato che entrano a numeratore e denominatore della formula vengono rilevati presso l'impresa nell'anno di riferimento, anche quando essa aveva partecipato all'edizione precedente dell'indagine²².

Per la stima delle variazioni e del livello degli investimenti per addetto dall'indagine sul 1999 vengono impiegate tecniche di stima robusta²³. La distribuzione empirica di questa variabile risulta infatti ricca di *outlier*, anche in conseguenza della natura del fenomeno oggetto di studio: la misurazione dei livelli, il cui rapporto dà origine al tasso di variazione, è resa problematica dalla compresenza di fattori quali il carattere tipicamente pluriennale della spesa, fattori d'incertezza, errori di classificazione e altri errori di misura.

Utilizzando la tecnica nota come “winsorizzazione del secondo tipo” le variazioni e i livelli delle variabili per addetto, localizzate al di sopra o al di sotto di soglie (*cut-off*) determinate sulla base della distribuzione empirica sono ricondotte alla soglia stessa, in proporzione alla frazione sondata. Il valore *winsorizzato*, y_i^{wins} , è determinato dalla formula seguente:

$$(5) \quad y_i^{wins} = \begin{cases} f * y_i + (1 - f) * J & \text{se } y_i < J \\ f * y_i + (1 - f) * K & \text{se } y_i > K \\ y_i & \text{altrimenti} \end{cases}$$

dove f è la frazione sondata e J e K le soglie (*cut-off*) rispettivamente inferiore e superiore della distribuzione di y_i .²⁴

La tecnica è stata adottata anche per il fatturato a partire dall'indagine sul 2005 e dal 2006 per le imprese delle costruzioni, limitatamente alla produzione totale e a quella in opere pubbliche.

In alcuni casi l'indagine rileva direttamente dei tassi di variazione dei fenomeni economici (è questo il caso, ad esempio, della capacità produttiva tecnica) oppure delle

22 La stima della dinamica dei fenomeni basata sui dati di una singola indagine si è dimostrata assai più stabile rispetto a quella ricavabile dal confronto dei valori rilevati in indagini contigue, che in alcuni casi scontano sia modifiche strutturali delle imprese di cui è difficile tenere conto, sia problemi di classificazione e di misura. All'interno dello stesso questionario, questi aspetti sono tenuti maggiormente sotto controllo, rendendo più accurata la stima delle variazioni. Tale metodologia non consente tuttavia di tenere conto pienamente dell'ingresso o dell'uscita d'impresе dalla popolazione d'interesse.

23 Sulle tecniche di stima robusta in generale cfr., ad esempio: Andrews D. F. et al., *Robust estimates of location*, Princeton, Princeton University Press, 1972 o anche: Hoaglin D. C. et al. (eds.) *Understanding robust and exploratory data*, New York, John Wiley & Sons, 1983. Un riferimento teorico classico è Huber P. J., *Robust statistics*, New York, John Wiley & Sons, 1981.

24 In letteratura, la modifica dei valori basata sulla (5) viene chiamata “Type 2 Winsorization”; quella che non tiene conto della frazione sondata viene chiamata “Type 1 Winsorization”: in quest'ultimo caso, i valori che si trovano oltre la soglia vengono posti pari al cut-off. Per una descrizione dettagliata della procedura di winsorizzazione utilizzata per la stima delle variazioni degli investimenti cfr. Battipaglia P., *Robust Estimates of Investments from the Bank of Italy's Business Survey*, Statistics Research Report, London, London School of Economics, 2000.

percentuali (ad esempio delle ore di straordinario). In questo caso, la stima per l'intera popolazione viene calcolata come media dei tassi di variazione individuali, ponderata con il prodotto del coefficiente di riporto all'universo per il relativo importo del fenomeno (o, se non disponibile, per una sua *proxy*).

I deflatori per i livelli d'investimento e fatturato sono calcolati come medie robuste settoriali dei deflatori individuali rilevati direttamente presso le imprese, usando come fattore di ponderazione il prodotto tra coefficiente di riporto all'universo e l'importo del fatturato²⁵.

Gli errori standard

Il calcolo analitico della varianza di stimatori non lineari ricavati da un disegno di campionamento stratificato non proporzionale, con pesi aggiustati per tenere conto di variabili di post-stratificazione, risulta piuttosto complesso²⁶ e giustifica il ricorso a tecniche di ricampionamento²⁷.

In particolare si utilizza il metodo *jackknife*, che risulta particolarmente adatto a tener conto della struttura imposta ai dati dalla natura del disegno campionario, garantendo al contempo adeguate proprietà asintotiche²⁸.

Se T_n è il valore dello stimatore riferito a un campione di numerosità n e $T_{n-1,i}$ è il valore dello stesso stimatore calcolato sul campione in cui si è eliminata l'unità i -esima (*leave-one-out method*), si calcolano dapprima gli "pseudovalori" $\tilde{T}_{n,i}$ definiti come:

$$(6) \quad \tilde{T}_{n,i} = nT_n - (n-1)T_{n-1,i} \quad 1 \leq i \leq n$$

Lo stimatore *jackknife* della varianza di T_n è²⁹:

$$(7) \quad \hat{V}_{JACK}(T_n) = \frac{1}{n(n-1)} \sum_{i=1}^n \left(\tilde{T}_{n,i} - \frac{1}{n} \sum_{j=1}^n \tilde{T}_{n,j} \right)^2$$

Per l'industria in senso stretto e i servizi, gli errori standard più elevati si riscontrano per la stima della variazione degli investimenti, a causa della loro intrinseca variabilità (tav. I4). Più ridotti risultano gli errori standard relativi alle variazioni di fatturato e occupazione. Nelle analisi per dominio, ad esempio per classe dimensionale dell'impresa o per area geografica, la precisione si riduce rispetto alle stime per il totale del campione, circostanza di cui è necessario tenere conto nell'analisi dei risultati.

Gli errori standard relativi alle previsioni delle principali variabili rilevate nell'indagine sono calcolati utilizzando la tecnica dell'imputazione multipla, che consente di tenere conto che una quota non trascurabile dei dati è stata imputata a causa delle mancate risposte parziali (come descritto nel paragrafo 5). La tecnica consiste nella replicazione indipendente

25 Per le imprese delle costruzioni si ottengono i valori complessivi della produzione in costruzioni a prezzi costanti usando il deflatore Istat più recente per le costruzioni, mentre per il valore della produzione di opere pubbliche si utilizza il deflatore Istat più aggiornato per la voce "altre costruzioni", che esclude le abitazioni.

26 Cfr., ad esempio, il capitolo 7 di Särndal C.-E., Swensson B., Wretman J., *Model Assisted Survey Sampling*, New York, Springer-Verlag, 1992.

27 Un testo di riferimento classico è Wolter K. M., *Introduction to variance estimation*, New York, Springer-Verlag, 1985.

28 L'adeguatezza del metodo è anche supportata dal confronto di proprietà teoriche ed empiriche con altri metodi di replicazione. Cfr., ad esempio, Shao J., Tu D., *The Jackknife and Bootstrap*, New York, Springer, 1995, che contiene un'approfondita analisi comparativa dei due metodi di stima della varianza.

29 Si veda Tukey J. W. (1958). Bias and confidence in not quite large samples. *Annals of Mathematical Statistics*, 29, 614.

di un certo numero di *dataset* contenenti le osservazioni complete (nel caso specifico sono stati estratti 25 campioni di *bootstrap* dal campione originario). Dopo aver riallineato i pesi di disegno alle variabili di post-stratificazione, è stato replicato il processo d'imputazione della variabile di previsione. Se si indica con \hat{t} lo stimatore d'interesse e con m il numero campioni replicati, la varianza di \hat{t} si stima con la seguente espressione³⁰:

$$(8) \quad \hat{v}(\hat{t}) = \frac{1}{m} \sum_{j=1}^m \hat{v}_{p(s)}(\hat{t}_j^*) + \left(1 + \frac{1}{m}\right) \sum_{j=1}^m \frac{(\hat{t}_j^* - \hat{t}_{mi})^2}{m-1}, \text{ dove } \hat{t}_{mi} = \frac{1}{m} \sum_{j=1}^m \hat{t}_j^*$$

Il termine $\hat{v}_{p(s)}(\hat{t}_j^*)$ indica la varianza stimata sul j -esimo campione replicato per lo stimatore \hat{t} , usando l'informazione del piano di campionamento $p(s)$. La prima sommatoria rappresenta la varianza media rispetto alle replicazioni (*within imputation variance*), mentre la seconda (detta *between imputation variance*) si interpreta come la variabilità generata dal processo d'imputazione. La tavola I5 presenta i risultati, dai quali risulta la maggiore variabilità dei dati di previsione rispetto a quelli di consuntivo.

La tavola I6 presenta gli errori standard per i principali fenomeni rilevati per le imprese delle costruzioni, sia a consuntivo sia a previsione: anche in questo caso si riscontra la maggiore incertezza di questi ultimi.

30 Cfr., ad esempio, il capitolo 4 di Lehtonen H., Pahkinen E., *Practical Methods for Design and Analysis of Complex Surveys*, New York, Wiley, 2004.