



BANCA D'ITALIA  
EUROSISTEMA

# Questioni di Economia e Finanza

(Occasional Papers)

Cosa rivelano le informazioni esterne sulla distorsione delle stime campionarie? il caso dell'indagine *Invind*

di Leandro D'Aurizio e Giuseppina Papadia

Aprile 2016

Numero

329





BANCA D'ITALIA  
EUROSISTEMA

# Questioni di Economia e Finanza

(Occasional papers)

Cosa rivelano le informazioni esterne sulla distorsione delle stime campionarie? il caso dell'indagine *Invind*

di Leandro D'Aurizio e Giuseppina Papadia

*La serie Questioni di economia e finanza ha la finalità di presentare studi e documentazione su aspetti rilevanti per i compiti istituzionali della Banca d'Italia e dell'Eurosistema. Le Questioni di economia e finanza si affiancano ai Temi di discussione volti a fornire contributi originali per la ricerca economica.*

*La serie comprende lavori realizzati all'interno della Banca, talvolta in collaborazione con l'Eurosistema o con altre Istituzioni. I lavori pubblicati riflettono esclusivamente le opinioni degli autori, senza impegnare la responsabilità delle Istituzioni di appartenenza.*

*La serie è disponibile online sul sito [www.bancaditalia.it](http://www.bancaditalia.it).*

ISSN 1972-6627 (stampa)

ISSN 1972-6643 (online)

*Stampa a cura della Divisione Editoria e stampa della Banca d'Italia*

# COSA RIVELANO LE INFORMAZIONI ESTERNE SULLA DISTORSIONE DELLE STIME CAMPIONARIE? IL CASO DELL'INDAGINE *INVIND*

di Leandro D'Aurizio\* e Giuseppina Papadia\*\*

## Sommario

Il lavoro analizza due distorsioni delle stime di *Invind*, indagine annuale della Banca d'Italia condotta su un panel di imprese: 1) quella causata dalle unità che entrano e escono dal campione, differenti dalle unità con partecipazione più stabile (*panel attrition*); 2) quella derivata dall'aggiornamento tardivo delle informazioni sulla popolazione di riferimento, necessarie per calcolare i pesi del campione. Dal raffronto di indicatori economici (disponibili su un archivio integrato) delle imprese uscite e/o entrate con quelli delle imprese permanenti, emerge che la *panel attrition* ha impatto trascurabile sulle stime aggregate ufficiali, determinate dalle imprese più grandi, caratterizzate da partecipazione continuativa. La partecipazione discontinua delle imprese piccole, i risultati economici peggiori delle uscite dal campione e l'età media relativamente elevata delle imprese rilevate richiedono una valutazione attenta delle stime non influenzate dalla scala dimensionale. Infine, per gli anni meno recenti il lavoro analizza come variano le stime utilizzando informazioni aggiornate sulla popolazione, verificando che l'aggiornamento ritardato distorce significativamente le stime aggregate solo se la dimensione della popolazione è fortemente instabile, con effetti trascurabili per le stime non influenzate dalla scala dimensionale. La parte finale contiene alcune proposte di evoluzione dell'indagine, in considerazione dei risultati emersi dal lavoro.

**Classificazione JEL:** C82, C83.

**Parole chiave:** indagini campionarie sulle imprese, campioni *panel*, *attrition*, informazioni esterne, integrazione di fonti informative.

## Indice

1. Introduzione.....	5
2. La <i>panel attrition</i> : aspetti generali e proposte recenti della letteratura .....	6
3. La <i>panel attrition</i> nell'indagine <i>Invind</i> .....	7
4. L'archivio integrato per la misura della <i>panel attrition</i> su <i>Invind</i> .....	9
5. Risultati economici delle imprese e <i>panel attrition</i> nell'indagine <i>Invind</i> .....	11
6. I ritardi di aggiornamento della popolazione di riferimento.....	15
7. Conclusioni .....	17
Bibliografia.....	19
Tavole e grafici.....	20

---

\* IVASS, Servizio Studi e Gestione Dati.

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



## 1. Introduzione

Il lavoro ha l'obiettivo di verificare la presenza di distorsione nelle stime di un'indagine campionaria tramite l'utilizzo di informazioni esterne. In particolare, ci si focalizza sull'indagine *Invind* (Banca d'Italia, 2014), condotta annualmente dalla Banca d'Italia su un campione di imprese di circa 4.000 unità, rappresentativo della popolazione delle imprese industriali e dei servizi non finanziari con 20 addetti e oltre. Il campione è di tipo *panel*, cioè formato da imprese che si cerca di mantenere sempre nel campione nelle varie edizioni dell'indagine. Le informazioni sulla popolazione di riferimento sono di fonte ufficiale (Istat, 2014) e sono utilizzate per costruire il sistema di ponderazione che rende i risultati dell'indagine rappresentativi dell'intera popolazione.

Vengono considerate due possibili fonti di distorsione:

- 1) quella derivante dal fatto che alcune imprese che hanno partecipato a una edizione dell'indagine non partecipano alla successiva, in base a fattori né predeterminati nel disegno campionario (quindi non tenuti sotto controllo) né casuali e dal fatto che le unità individuate in sostituzione di quelle uscenti sono in qualche misura diverse da queste (*panel attrition*; per esempio, cfr. Martin *et al.*, 2001);
- 2) quella causata dal ritardo di aggiornamento dei dati sulla popolazione di imprese utilizzata per calcolare i pesi di riporto all'universo. Questi ultimi rendono le stime campionarie rappresentative della popolazione stessa.

La prima fonte di distorsione è analizzabile tramite un archivio esterno contenente i dati di bilancio delle imprese della popolazione di riferimento. Per ogni impresa del *panel*, esso consente di disporre di dati relativi sia agli anni in cui l'impresa ha partecipato sia a quelli in cui l'impresa è stata assente. In questo modo è possibile capire se le imprese *panel* che entrano o escono da una data edizione dall'indagine abbiano caratteristiche differenti da quelle che partecipano in modo continuativo.

La seconda fonte di distorsione è misurabile per tutti gli anni in cui si hanno informazioni aggiornate sulla popolazione (al momento della preparazione di questo studio si tratta degli anni di indagine fino al 2012). In questo modo si ha una misura delle possibili distorsioni delle stime normalmente prodotte, che utilizzano informazioni sulla popolazione o non aggiornate o non del tutto coerenti con le caratteristiche dell'indagine.

Per lo studio di entrambe le fonti di distorsione, si è scelto di utilizzare come archivio esterno la base dati *Cerved*, che copre l'universo delle società di capitali italiane e raccoglie:

- bilanci d'esercizio e altre informazioni integrative;
- bilanci riclassificati secondo gli schemi della Centrale;
- bilanci consolidati dei principali gruppi, anche nella versione riclassificata<sup>1</sup>.

I principali risultati del lavoro sono i seguenti. L'impatto della *panel attrition* è trascurabile per le stime aggregate, dipendenti in misura rilevante dai valori relativi alle imprese di maggiori dimensioni, che

---

<sup>1</sup> La scelta è ricaduta su *Cerved* per la ricchezza del contenuto informativo e per la disponibilità dei dati. La potenziale alternativa era rappresentata dal Registro Imprese, realizzato e gestito da infoCamere per le Camere di Commercio e definito dall'art. 60 del Codice dell'Amministrazione Digitale (CAD) un archivio di dati di interesse nazionale, in quanto fonte ufficiale, attendibile e garantita dalla legge di informazioni sulle imprese italiane e sui loro soci e amministratori. Non è stato possibile adottare questa soluzione per la mancanza dei dati di bilancio nella versione del Registro disponibile in Banca d'Italia.

partecipano all'indagine in modo continuativo. Le stime diffuse e commentate nelle pubblicazioni ufficiali sono di questo tipo. Le imprese più piccole tendono invece a partecipare in modo più discontinuo, così come quelle fondate in anni relativamente recenti rispetto all'entrata nel campione, mentre le aziende che escono in un dato anno dall'indagine tendono ad avere risultati economici meno brillanti delle altre: questi fattori vanno considerati nella valutazione delle stime che non risentono della scala dimensionale delle imprese. Per quel che riguarda il ritardo di aggiornamento delle informazioni sulla popolazione, esso rende necessaria una misura di cautela nel valutare le stime aggregate (se sono ponderate per essere rappresentative della popolazione stessa). Esse sono infatti distorte nel caso di un forte *trend* nella popolazione non registrato in modo tempestivo. Questo effetto è assente per le stime semplici, che non tengono conto della scala dimensionale delle imprese.

Il lavoro è organizzato come segue. Il secondo paragrafo descrive in generale i problemi della *panel attrition*. Il terzo paragrafo propone un metodo di misurazione dei suoi effetti, con riferimento all'indagine *Invind*. Il quarto descrive l'archivio integrato utilizzato per lo studio. Il quinto dà conto degli effetti della *panel attrition* su *Invind*, tramite una stima *logit* del processo di entrate e uscite del campione, prosegue con una analisi grafica e si chiude con il confronto tra alcune stime campionarie e le evidenze dell'archivio integrato. Il sesto paragrafo esamina il problema del ritardo di aggiornamento dei dati ufficiali sulla popolazione di riferimento. Il settimo e ultimo è dedicato alle conclusioni, che propongono possibili soluzioni ai problemi individuati e futuri argomenti di ricerca.

## **2. La *panel attrition*: aspetti generali e proposte recenti della letteratura**

Le indagini *panel* rilevano le stesse unità in occasioni ripetute d'indagine. Questa scelta rende possibile lo studio delle transizioni (Fabbris, 1989) e presenta il vantaggio operativo di non dover selezionare un nuovo campione nelle edizioni dell'indagine successive alla prima.

Nella pratica, avviene una erosione (*attrition*) del campione originario quando nelle indagini successive alla prima il *panel* non è più composto dalle stesse unità. Le uscite possono riguardare sia unità che rifiutano di partecipare alle indagini successive sia quelle che nel corso del tempo non appartengono più alla popolazione di riferimento (entrambe possono essere considerate componenti dell'*attrition* nelle varie definizioni presenti in letteratura). Le uscite dal *panel* possono essere compensate sia da unità che avevano temporaneamente cessato di partecipare all'indagine e rientrano nel campione, sia da nuove unità, non facenti parte del campione originario o che solo recentemente sono entrate a far parte della popolazione.

Se l'*attrition* comporta una diminuzione della quota di unità della popolazione incluse nel campione (frazione sondata), si ha un aumento naturale della variabilità delle stime. Se il processo generatore dell'*attrition* è casuale, le stime campionarie non sono affette da distorsione (Little e Rubin, 2002). Si genera invece distorsione, spesso difficilmente misurabile, se l'*attrition* dipende dalle variabili oggetto di studio dell'indagine. Un'altra componente di distorsione si ha se i tassi di entrata e uscita dalla popolazione di riferimento (indicati anche come nascite e morti), stimati dal *panel*, sono distorti rispetto ai valori effettivi della popolazione. In tal caso, la distribuzione per età stimata dal *panel* differisce da quella "vera" della popolazione, creando una fonte di distorsione anche per le stime delle variabili correlate con l'età stessa.

Per ovviare a questi inconvenienti, si adottano in genere alcune soluzioni operative che prevedono di sostituire le unità uscite dal campione con altre il più possibile simili rispetto alle caratteristiche osservabili utilizzate per disegnare l'indagine, nonché degli schemi di rotazione periodica di una parte del *panel*, che permettano di tener conto della evoluzione della popolazione rispetto alla composizione iniziale (Trivellato, 1999).



In questo modo si mantengono la numerosità campionaria originaria e la variabilità delle stime a livelli prossimi a quelli programmati in fase di disegno e al contempo si tiene conto della dinamica della popolazione. Se tuttavia l'*attrition* non è casuale, l'utilizzo delle sostituzioni non è sufficiente per eliminare la distorsione, perché le unità uscite e quelle entrate possono differire per caratteristiche non controllabili, osservate o meno, potenzialmente correlate con le variabili di interesse dell'indagine.

La recente letteratura statistica ha esaminato alcune soluzioni per tener conto della *panel attrition*. Quando mancano informazioni esterne, l'utilizzo di tecniche eterogenee di correzione dei dati mancanti generati dall'*attrition* può fornire indicazioni utili. Al riguardo Black *et al.* (2007) utilizzano una pluralità di metodi di imputazione, dai più semplici (imputazione tramite medie) ai più sofisticati (metodi Bayesiani di ricampionamento), per ricostruire i valori mancanti in un campione di dati relativi al traffico automobilistico del Regno Unito e suggeriscono di osservare la variabilità dei risultati per giudicare la tipologia della distorsione da dato mancante e il relativo impatto sulle stime. Deng *et al.* (2013) sottolineano la difficoltà di seguire i trend effettivi delle variabili di interesse in una indagine *panel* con *attrition* e la necessità di disporre di fonti esterne per compensarne gli effetti negativi. Utilizzando le edizioni più recenti dell'indagine statunitense *Survey on Income and Program Participation*, gli autori propongono come fonte esterna l'utilizzo di campioni supplementari (*refreshment samples*), formati da nuovi rispondenti selezionati casualmente. A questi è somministrato lo stesso questionario del campione principale e le differenze tra le risposte fornite dai due campioni sono utilizzate per correggere la distorsione. Le informazioni esterne sono tipicamente utilizzate per costruire sistemi di correzione dei pesi originari di espansione alla popolazione di riferimento, al fine di compensare gli effetti della *panel attrition*. Afonso (2015) applica la tecnica di correzione dei pesi per compensare la presenza di dati mancanti causati dall'*attrition* in un *panel* di microdati bancari derivati da Bankscope, per studiare l'evoluzione recente del margine di interesse netto dei 15 maggiori paesi dell'area dell'euro. I pesi sono adottati nell'ambito di stimatori basati sul metodo generalizzato dei momenti (GMM). La correzione genera stime consistenti, in alcuni casi risolvendo il problema di segni inattesi dei coefficienti dei modelli proposti, non compatibili con gli assiomi della teoria economica.

### 3. La *panel attrition* nell'indagine *Invind*

Il processo di *attrition* preso in esame in questo lavoro è quello osservato in due indagini consecutive di *Invind*, indicate con  $I_{t-1}$  e  $I_t$ : alcune imprese partecipano a  $I_{t-1}$  ma non a  $I_t$  e sono quindi sostituite da imprese non presenti in  $I_{t-1}$ . Nell'indagine *Invind* si cerca di mantenere la rappresentatività del *panel* tramite regole di sostituzione che, in ogni edizione, permettono di rimpiazzare le imprese uscite con altre, aventi sede principale nella stessa regione, e attività economica e dimensioni il più possibile simili<sup>2</sup>. Si considera l'evoluzione del *panel* dall'edizione dell'indagine relativa al 2002, primo anno in cui la popolazione di riferimento (imprese dell'industria in senso stretto e dei servizi finanziari con 20 addetti e oltre) e numerosità campionaria (circa 4.000 imprese) sono diventate quelle attuali (Banca d'Italia, 2005)<sup>3</sup>.

Nel periodo 2002-2013, il processo di uscita e entrata nel campione riguarda una quota considerevole di imprese (tav. 1): in media, un quinto delle imprese esce dal campione ed è sostituito da nuove imprese; di conseguenza solo l'80% delle imprese di una edizione è presente anche nella successiva (nel periodo considerato la numerosità campionaria è aumentata del 6,5%). Tra le imprese uscite dal campione, i tre quarti restano attive e continuano a far parte della popolazione di riferimento dell'indagine, mantenendo un numero

---

<sup>2</sup> Si rinuncia a studiare le caratteristiche delle stime longitudinali effettuabili su un'indagine *panel* e la loro eventuale distorsione, causata da un processo di *attrition* più complesso di quello in studio (ad esempio, derivato da imprese che, osservate in ripetute occasioni di indagine, hanno una partecipazione discontinua).

<sup>3</sup> L'indagine relativa ad un anno è condotta nei primi 4-5 mesi dell'anno successivo.

di addetti non inferiore a 20; il 14,2%, senza elevate oscillazioni annuali, esce perché non più attiva; il restante 10% diviene non più rilevabile, perché il numero di addetti scende al di sotto della soglia dei 20 addetti e, in misura più trascurabile, perché interessata da trasformazioni societarie (incorpori, fusioni, ecc.). La dimensione media delle imprese che permangono tra due edizioni successive, pari a 366 addetti, è notevolmente superiore a quella delle imprese entrate, che a sua volta risulta maggiore di quella delle uscite. Anche il numero di anni di attività delle imprese permanenti (in media 34 anni) tende a essere superiore a quello delle imprese entrate o uscite<sup>4</sup>. Le ultime due colonne della tavola mostrano che un terzo circa delle imprese entrate in un'edizione in realtà già apparteneva al *panel*, poiché presente in indagini antecedenti.

Il processo di entrata e uscita presenta una bassa variabilità temporale. In particolare, alcuni eventi esogeni importanti quali la crisi economica iniziata nel 2009 e la modifica della rete di rilevazione non sembrano aver avuto un impatto significativo sul fenomeno<sup>5</sup>.

A livello settoriale, il processo di *attrition* riguarda in misura superiore le imprese dei servizi privati non finanziari (tav. 2). Queste ultime sono state inserite nel *panel* più tardi (nel 2002) e pertanto risultano meno fidelizzate e meno propense a garantire una collaborazione continuativa all'indagine<sup>6</sup>, anche perché caratterizzate da minore stabilità della struttura societaria rispetto alle imprese industriali<sup>7</sup>.

La presenza di questo elevato livello di *panel attrition* suggerisce di verificare se le imprese entrate e uscite abbiano caratteristiche peculiari rispetto a quelle permanenti, specialmente per gli indicatori correnti dei risultati economici delle imprese (ROA, ROE, fatturato, utile). Più specificamente, se le imprese uscite presentassero risultati sistematicamente peggiori di quelli delle imprese entrate, il processo di auto-selezione potrebbe creare nel corso del tempo un campione con indicatori economici migliori di quelli della popolazione di riferimento.

La disponibilità di informazioni sui risultati economici di tutte le imprese del *panel* permette di attuare sia un'analisi **prospettica** (che, da una data occasione d'indagine, osserva le transizioni all'occasione successiva) sia una **retrospettiva** (che, da una data occasione d'indagine, osserva le transizioni occorse rispetto all'occasione precedente).

Formalmente, la visione prospettica considera l'indagine  $I_t$  condotta al tempo  $t$  e la successiva indagine  $I_{t+1}$  condotta al tempo  $t+1$  e misura al tempo  $t$  i risultati economici:

- delle imprese che partecipano a  $I_t$  ma non a  $I_{t+1}$  (uscite in  $t+1$ );
- delle imprese che non partecipano a  $I_t$ , ma partecipano a  $I_{t+1}$  (entrate in  $t+1$ );
- delle imprese che partecipano a entrambe le edizioni (permanent).

In modo speculare, la visione retrospettiva considera l'indagine  $I_t$  condotta al tempo  $t$  e la precedente indagine  $I_{t-1}$  condotta al tempo  $t-1$  e misura al tempo  $t$  i risultati economici:

---

<sup>4</sup> Si tenga presente che la correlazione tra dimensione d'impresa e numero di anni di attività è estremamente debole, pari a 0,03 per l'intero periodo esaminato. Il risultato deriva anche dalla debole propensione delle imprese italiane alla crescita dimensionale, ampiamente documentato in letteratura (Trento, 2012).

<sup>5</sup> Nel periodo tra settembre 2008 e novembre 2009 sono state chiuse 35 filiali della Banca d'Italia, in precedenza impegnate nella rilevazione delle imprese con sede principale nel loro territorio, successivamente rilevate dalle filiali rimaste operative nella regione.

<sup>6</sup> Se si considerano a livello annuale i dati sintetizzati in tavola 2, non riportati per brevità, si riscontra che la partecipazione all'indagine delle imprese dei servizi è migliorata nel corso del periodo esaminato, ma si mantiene ancora a livelli inferiori a quelli delle imprese industriali.

<sup>7</sup> Si veda al riguardo Banca d'Italia (2015), p. 58.

- delle imprese che hanno partecipato a  $I_{t-1}$ , ma non a  $I_t$  (uscite in  $t$ );
- delle imprese che non hanno partecipato a  $I_{t-1}$ , ma partecipano a  $I_t$  (entrate in  $t$ );
- delle imprese che partecipano a entrambe le edizioni (permanenti).

Per le imprese che escono dal campione, la visione prospettica misura i loro risultati nell'anno precedente l'uscita dal campione. In questo modo si può stimare l'eventuale legame di causalità tra l'uscita dal campione e i risultati economici dell'impresa uscita. Analoghe considerazioni valgono per le imprese che entrano nel campione, per le quali la visione prospettica misura i risultati economici dell'anno precedente l'entrata nel campione.

La visione prospettica e quella retrospettiva utilizzate assieme, consentono di misurare i risultati economici delle imprese uscite dal campione sia nell'anno precedente l'uscita sia in quello successivo. In questo modo è possibile verificare se eventuali differenze tra questi risultati e quelli delle imprese che permangono nell'indagine subiscano variazioni tra i due anni. La stessa verifica è possibile per le imprese che entrano nell'indagine in una data edizione, ma non avevano partecipato alla precedente.

Le grandezze aziendali utilizzate nel lavoro sono o variabili monetarie (ad esempio utile e fatturato) o indici di bilancio adimensionali (come il ROA o il ROE). Per le prime, rileva misurare la loro variazione media tra i due tempi  $t-1$  e  $t$ , mentre per le variabili adimensionali interessa misurare una media al tempo  $t$  e seguirne l'evoluzione nel tempo.

Con riferimento alla generica impresa  $i$  del campione di  $n_t$  unità rilevate al tempo  $t$ , si indica con  $y_{it}$  la generica variabile monetaria e con  $x_{it}$  la generica variabile adimensionale;  $z_{it}$  è usata per rappresentare una qualsiasi variabile correlata con la dimensione dell'impresa (ad esempio fatturato o numero di addetti), utilizzabile per ponderare le variabili adimensionali in modo che la loro media tenga conto della scala dell'impresa e infine  $w_{it}$  simboleggia il peso di espansione alla popolazione di riferimento.

La tavola 3 mostra gli indici di sintesi dei *risultati economici* delle imprese del campione che saranno utilizzati nel seguito. Si distingue tra indici aggregati, riportati nelle pubblicazioni ufficiali nella versione ponderata (Banca d'Italia, 2014) e indici semplici, che non tengono conto della differente dimensione delle imprese.

#### **4. L'archivio integrato per la misura della *panel attrition* su *Invind***

##### ***Caratteristiche generali dell'archivio integrato***

Il metodo descritto nel paragrafo precedente è applicabile se per ogni impresa del *panel* di *Invind* si hanno le variabili riguardanti i risultati economici per tutto l'arco di tempo considerato 2002-2013, anche se l'impresa ha partecipato a un numero ridotto di edizioni dell'indagine (il caso limite è quello di un'impresa presente in una sola edizione).

Fatta eccezione per un numero trascurabile di ditte individuali, le imprese del campione di *Invind* presentano la forma giuridica di società. L'archivio *Cerved*, che contiene i dati di bilancio di tutte le società italiane dal 1993, è stato quindi integrato all'archivio dei dati di *Invind* (fig. 1). Per costruire l'archivio integrato è stata usata come chiave di accoppiamento la Partita Iva, che identifica univocamente le imprese in *Cerved* ed è inoltre disponibile per le imprese del *panel* di *Invind*. L'archivio integrato è ulteriormente

arricchito con le informazioni estratte dagli archivi *Infocamere* per recuperare i dati relativi agli addetti delle imprese, non sempre presenti su *Cerved*<sup>8</sup>.

Il tasso d'integrazione delle informazioni *Invind-Cerved* è piuttosto elevato: per oltre il 90% delle imprese in un dato anno d'indagine è possibile utilizzare la relativa partita Iva per recuperare i dati di bilancio *Cerved* relativi allo stesso anno (tav. 4)<sup>9</sup>. Gli addetti sono recuperati in una percentuale di casi inferiore (circa 80%), perché non figurano sempre tra le informazioni di bilancio<sup>10</sup>.

La qualità del *matching* tra le due basi dati si misura anche in termini di distanza media tra le informazioni presenti in entrambi le fonti, quali ad esempio occupazione media e fatturato. Quasi l'80% degli scostamenti dei valori del fatturato è inferiore al 5% (fig. 2)<sup>11</sup>, mentre un 10% circa degli scostamenti è compreso tra il 5% e il 10%. Gli scostamenti di ordine superiore riguardano quindi non più di un decimo delle imprese del campione. L'accostamento è ancora migliore per gli occupati.

Un ulteriore indicatore di qualità del *matching* è il coefficiente di correlazione tra i dati rilevati sul campione *Invind* e quelli presenti su *Cerved* sia per gli addetti sia per il fatturato; la radice quadrata del coefficiente rappresenta un indice di affidabilità del dato di fonte campionaria (Biancotti *et al.*, 2004), che vale circa 0,98 per entrambe le variabili<sup>12</sup>. Queste evidenze convalidano la scelta degli archivi *Cerved*.

Poiché nell'archivio *Cerved* l'informazione sul numero di addetti è disponibile in una minoranza dei casi, è stato necessario imputarlo dove risultava mancante. Solo questa ricostruzione rende possibile due passi importanti delle analisi presentate di seguito:

- stimare la probabilità di entrata di una impresa in una edizione dell'indagine, se non presente in quella precedente, al fine di comprendere quali fattori siano associati a tale comportamento (le imprese eleggibili ad entrare sono infatti solo quelle con almeno 20 addetti);
- verificare se un'impresa uscita da un'edizione dell'indagine continui a far parte della popolazione di riferimento (se il numero di addetti dell'impresa resta superiore a 19 addetti).

La ricostruzione degli addetti è stata effettuata tramite un modello *OLS*, che stima per ogni anno una relazione lineare tra numero medio di addetti dell'anno e alcune variabili economiche di *performance*, controllando gli effetti di settore e area geografica. Il modello adottato è stato selezionato tra un gruppo di modelli di struttura simile in base ai seguenti criteri:

- utilizzare covariate disponibili nell'archivio senza elevate percentuali di valori mancanti;

---

<sup>8</sup> L'informazione sugli addetti è stata estratta dagli archivi *Infocamere* per una percentuale esigua di imprese *Cerved* (compresa tra il 3,5% e l'8%).

<sup>9</sup> Il dato non disponibile corrisponde a casi di imprese *Invind* con partita Iva mancante, oppure presente ma priva di una partita Iva corrispondente negli archivi *Cerved*. Una causa della mancata corrispondenza è il ritardo di aggiornamento degli archivi *Cerved* (i dati dell'anno  $t$  sono disponibili solo parzialmente nell'anno  $t+1$  e solo nell'anno  $t+2$  sono disponibili quasi completamente). Inoltre, è da tener presente che una piccola parte delle imprese *Invind* con una corrispondenza su *Cerved* è priva dei dati sugli indicatori economici (in percentuale inferiore allo 0,5%).

<sup>10</sup> Per il complesso delle imprese presenti in *Cerved*, il numero di addetti è presente soltanto nel 12% dei casi.

<sup>11</sup> Gli indicatori sono calcolati per le sole imprese che nell'anno d'indagine non segnalano modifiche strutturali dovute a eventi straordinari quali scorpori, fusioni, etc., che rendono di per sé le due fonti non confrontabili. Questi fenomeni riguardano in media tra il 3 e il 5% delle imprese di ogni edizione annuale e la loro frequenza cresce con la dimensione d'impresa (fig. 3).

<sup>12</sup> Questa elevata affidabilità non si ha per tutte le variabili. Ad esempio, per l'anno di costituzione dell'impresa l'indice assume valore pari a 0,82, ancora soddisfacente.

- soddisfare elementari considerazioni di logica economica, pur restando nell'ambito dei modelli associativi senza pretese di causalità;
- presentare buone caratteristiche di adattamento ai dati (in termini di *R-square*);
- ricostruire in modo soddisfacente i totali di imprese e addetti, disponibili da fonte statistica ufficiale<sup>13</sup>, secondo i più importanti fattori di classificazione della popolazione di riferimento.

Il modello scelto è stato stimato all'interno di strati formati da combinazioni di attività economica e area geografica della sede principale dell'impresa<sup>14</sup> e presenta la seguente struttura:

$$Addetti_{t,j,i} = \beta_{0t,j} + \beta_{1t,j}fatt_{t,j,i} + \beta_{2t,j}fatt_{t,j,i}^2 + \beta_{3t,j}roa_{t,j,i} + \beta_{4t,j}roe_{t,j,i} + \varepsilon_{t,j,i}$$

Dove *i* denota l'impresa *i*-esima, *j* lo strato *j*-esimo, *t* l'anno e *fatt* il fatturato<sup>15</sup>.

Il grado di somiglianza tra la distribuzione di imprese e addetti della popolazione di imprese con 20 addetti e oltre così ricostruita e quella della fonte statistica ufficiale è soddisfacente, specialmente in termini di frequenze relative (si veda tav. 5 per il confronto relativo al 2012. I confronti per gli anni precedenti presentano livelli di accostamento molto simili e non sono riportati per brevità)<sup>16</sup>.

## 5. Risultati economici delle imprese e *panel attrition* nell'indagine *Invind*

### *La propensione a entrare ed uscire dal campione*

Per studiare la propensione di entrata o uscita dal campione in una data edizione dell'indagine, si è adoperato un modello *logit* stimato su un *pooling* di osservazioni, relative agli anni 2003-2013 per le entrate e al 2002-2012 per le uscite. Gli effetti temporali sono considerati inserendo *dummy* di anno tra le covariate, mentre gli errori standard associati ai coefficienti del modello tengono opportunamente conto del fatto che la stessa impresa è ripetuta più volte nel *set* di dati usato per la stima.

Per la stima della probabilità di entrare nel campione nell'anno di indagine *t*, le imprese eleggibili a entrare sono quelle della popolazione con 20 addetti e oltre presenti negli archivi *Cerved* dell'anno *t-1* che per quell'anno non hanno partecipato all'indagine. Tra le covariate figurano le seguenti variabili di risultato economico d'impresa: ROA, ROE, variazione dell'utile e del fatturato (totali e per addetto), riferite all'anno precedente l'entrata nel campione per attenuare possibili problemi di endogeneità. Il modello comprende, oltre a *dummy* di anno, anche *dummy* di settore, area geografica della sede principale dell'impresa e classe dimensionale. Le variabili economiche sono inserite nel modello una alla volta (colonne 1-6, tav. 6) e poi tutte insieme (colonna 7). L'ultima colonna 8 è relativa alla stima che non comprende tra le covariate le variazioni dell'utile e del fatturato, che risentono maggiormente delle operazioni straordinarie di variazione del perimetro aziendale, non tracciabili negli archivi *Cerved*. Quest'ultima stima è quindi maggiormente affidabile, in quanto le covariate non sono strettamente correlate alla dimensione d'impresa e risentono meno di tali operazioni. I coefficienti associati alle variabili economiche, se significativi, hanno tutti segno

<sup>13</sup> Si tratta dell'Archivio Statistico delle Imprese Attive, realizzato dall'Istat (Istat, 2014), per il quale sono disponibili soltanto informazioni aggregate.

<sup>14</sup> Si veda l'appendice per il dettaglio delle aree geografiche e dei settori utilizzati.

<sup>15</sup> Il costo del lavoro, ottimale in questo processo di ricostruzione, non è utilizzabile per l'elevata percentuale di valori mancanti (in media 29%). Il fatturato rappresenta comunque una *proxy* soddisfacente della variabile, dato che la correlazione tra i due indicatori, calcolata per le imprese ove sono entrambi disponibili, è molto elevata negli strati dove si stima il modello di ricostruzione (in media 0,85).

<sup>16</sup> Si tenga presente che si analizzano tutte le edizioni di *Invind* fino a quella del 2013, per le quali si considera come popolazione di riferimento quella di fonte Istat relativa al 2012, la più aggiornata disponibile al momento della preparazione del lavoro (Istat, 2014).

positivo, ad eccezione dell'utile, che presenta segno negativo e modesta significatività solo nella colonna (7). In generale, migliori risultati economici dell'impresa aumentano la sua propensione a partecipare all'indagine. Le due variabili con maggiore effetto sono, nell'ordine, il ROA e il ROE. Risultano anche significativi i coefficienti associati alle *dummy* dimensionali, che segnalano la minore propensione a entrare delle imprese con 20-199 addetti. A livello geografico, l'ingresso delle imprese del Nord è meno frequente rispetto a quelle del Centro Sud.

La stima della probabilità di uscire dal campione è condotta in modo analogo (tav. 7), considerando come imprese eleggibili per l'uscita tutte quelle partecipanti a un'edizione dell'indagine di un dato anno, al quale sono riferite le variabili economiche usate come covariate (si escludono le imprese che l'anno successivo escono dalla popolazione di interesse dell'indagine). Troviamo un risultato speculare rispetto a quello verificato per la propensione a entrare: peggiori risultati economici tendono ad accrescere la propensione ad uscire dal campione nella successiva edizione. Anche per le uscite, le variabili maggiormente rilevanti sono il ROA e il ROE. Le imprese con meno di 200 addetti tendono maggiormente a uscire dal campione, specialmente quelle con 20-49 addetti. A livello di ripartizione geografica, la difficoltà di mantenere le imprese nel campione riguarda soprattutto il Nord Est.

La maggiore difficoltà a mantenere e, al contempo, a far entrare nel *panel* le imprese del Nord Est non si traduce in una diminuzione della dimensione campionaria per l'area, perché la popolazione di imprese contattabili è molto numerosa e compensa la minore frequenza di imprese contattate.

### ***Risultati economici delle imprese del campione e processo di entrata e uscita***

L'archivio integrato descritto nel paragrafo 4 rende possibile calcolare per ogni anno di indagine gli indicatori presentati in tavola 3, riguardo alle imprese entrate, uscite e permanenti e confrontarli con gli indicatori relativi al campione totale. Il confronto può essere effettuato sia nell'approccio prospettico sia in quello retrospettivo.

Vengono presentati dapprima gli indici semplici relativi alle variazioni del fatturato e agli andamenti di ROA e ROE (fig. 4). Le imprese entrate presentano tendenzialmente valori superiori a quelli delle imprese permanenti nell'anno precedente il loro ingresso, in particolare per la variazione del fatturato. Le *performance* di queste imprese tendono tuttavia ad assomigliare a quelle del resto del campione nell'anno d'entrata.

Questa evidenza descrittiva è confermata anche da un'analisi condizionale che modella la propensione a entrare nel campione utilizzando lo stesso *set* di regressori dell'analisi presentata in tavola 6 (tav. 8). La differenza è che in questa seconda stima i regressori relativi alle *performance* economiche delle imprese sono riferiti all'anno della loro entrata nel campione, anziché all'anno precedente. Si verifica che i coefficienti di tavola 8 relativi agli indicatori economici tendono a essere inferiori o con minore significatività di quelli di tavola 6.

Per le imprese uscite, i relativi indici sono sistematicamente inferiori a quelli degli altri gruppi di imprese, sia nell'anno precedente l'uscita dal campione sia in quello di uscita (fig. 4). Nell'anno precedente l'uscita, le imprese uscite che resteranno ancora attive presentano risultati migliori rispetto alle imprese uscite che chiuderanno. Le imprese uscite tendono a peggiorare i loro risultati rispetto all'anno precedente. Per il ROE e il ROA, il peggioramento è meno forte per le imprese uscite che continuano ad appartenere alla popolazione di interesse dell'indagine.

Per le variazioni dell'utile (non mostrate per brevità), gli andamenti delle imprese entrate sono simili a quelli delle imprese permanenti sia l'anno precedente l'entrata nel campione sia in quello successivo. Le

variazioni dell'utile delle imprese uscite sono inferiori rispetto a quelle delle imprese permanenti solo nell'anno di uscita dal campione.

Gli andamenti degli indici aggregati per fatturato, ROA e ROE (fig. 5), nonché quelli dell'utile (anche in questo caso non mostrati per brevità) non evidenziano risultati sistematicamente diversi per i gruppi di imprese considerati e quindi risentono in misura limitata del processo di entrate e uscite delle imprese dal campione nel corso del tempo.

Le considerazioni sviluppate finora restano valide anche per gli indici calcolati nella versione ponderata.

Questa analisi descrittiva fa emergere due evidenze:

1. le imprese uscite hanno in media risultati economici meno brillanti delle imprese permanenti e quindi anche delle imprese entrate, che tendono a essere simili a quelle permanenti nell'anno di entrata nel campione;
2. l'effetto tende a non essere presente se si tiene conto della dimensione delle imprese nel calcolo degli indicatori.

Questi risultati trovano riscontro anche in un'analisi econometrica condotta tramite un semplice modello lineare che, anno per anno e su tutti gli anni nel loro complesso, regredisce separatamente alcune variabili di *performance* rispetto a una *dummy* che individua le imprese entrate nel campione e a una serie di caratteristiche strutturali. La stima è ristretta alle imprese entrate o uscite nell'anno (di queste ultime, solo quelle che continuano ad avere almeno 20 addetti, restando nella popolazione di riferimento), senza considerare quelle permanenti<sup>17</sup>. La regressione semplice è replicata utilizzando il fatturato come variabile di scala, per tenere conto della diversa dimensione delle imprese.

Le stime effettuate sul *pooling* dei dati mostrano sinteticamente le differenze tra imprese entrate e uscite ed evidenziano che tale scarto diventa inferiore e non significativo quando le imprese più grandi hanno maggiore influenza sulle stime (tav. 9)<sup>18</sup>.

La regressione semplice per i singoli anni evidenzia le migliori *performance* delle imprese entrate presenti in gran parte degli anni per il ROA e la dinamica del fatturato. Per il ROE, i valori del coefficiente significativi a livello annuale sono concentrati in alcuni anni, mentre sono assenti per la dinamica dell'utile. L'utilizzo del fattore di scala dimensionale tende ad attenuare la significatività dello scarto tra imprese entrate e uscite per il ROA e la dinamica del fatturato (il segno del coefficiente si inverte in alcuni anni per il secondo indicatore). Per il ROE, l'effetto di scala nei singoli anni non ha un *pattern* riconoscibile. Infine, per la dinamica dell'utile l'effetto dimensionale rende significative alcune differenze tra imprese entrate e uscite, ma i confronti sono in taluni casi a favore di queste ultime.

### ***Risultati economici del campione Invind e della popolazione di riferimento a confronto***

Un quesito importante cui rispondere è se la *panel attrition* influenzi negativamente la capacità del campione di rappresentare la popolazione di riferimento.

---

<sup>17</sup> Sono state escluse dall'analisi le imprese con modifiche della struttura aziendale nei vari periodi utilizzati, per avere dati confrontabili a livello di singola impresa tra due anni consecutivi.

<sup>18</sup> I risultati sono presentati sia per il complesso del campione, sia per le sole imprese manifatturiere, per la particolare attenzione che l'indagine ha dedicato al settore nel corso degli anni.

Si utilizzano gli archivi *Cerved* per calcolare statistiche relative alla popolazione, formata dalle imprese con 20 addetti e oltre<sup>19</sup>. Questa scelta consente di confrontare le stime *Invind* con i corrispondenti valori relativi alla popolazione per un ampio insieme di indicatori economici registrati nei bilanci aziendali. La disponibilità dei microdati per la popolazione consente il confronto sia per gli indici aggregati sia per quelli semplici.

Allo scopo, si calcolano gli indici della tavola 3 sul campione *Invind* nella versione ponderata, per renderli rappresentativi dell'intera popolazione, e si confrontano con i corrispondenti indici semplici computati sulla popolazione di riferimento ricostruita sugli archivi *Cerved* (fig. 6). Le variabili messe a confronto sono: fatturato, ROE, ROA e utile. Il fatturato è rilevato nell'indagine ed è anche presente negli archivi *Cerved*, mentre ROE, ROA e utile sono presenti solo in questi ultimi. Per queste tre ultime variabili, la comparazione non valuta quindi l'informazione raccolta in *Invind*, ma mira a verificare se le analisi campionarie che utilizzano congiuntamente variabili rilevate in *Invind* e variabili di bilancio di fonte *Cerved* risentano di problemi di auto-selezione del campione.

Si descrive il confronto separatamente per le singole variabili.

1) **Variazione del fatturato.** Gli indici aggregati e, soprattutto, quelli semplici presentano andamenti qualitativamente simili, ma quelli calcolati su *Invind* sono inferiori a quelli calcolati su *Cerved*. Parte di questa differenza sistematica potrebbe spiegarsi con la presenza nei due archivi delle imprese con modifiche del perimetro aziendale, che vengono corrette in *Invind*, ma non in *Cerved*. Un'analisi delle imprese del campione *Invind* con modifiche del perimetro aziendale, rappresentativa dell'intera popolazione, mostra che sono prevalenti le modifiche che comportano un aumento delle dimensioni dell'impresa<sup>20</sup>. Pertanto la dinamica del fatturato stimata su *Cerved*, al lordo delle modifiche del perimetro aziendale, potrebbe risentire della sovrastima dovuta a questa componente.

2) **Andamenti del ROA.** Gli indici del ROA semplici e aggregati calcolati su *Invind* seguono l'andamento dei corrispondenti indici sulle imprese *Cerved* e sono lievemente al di sopra di questi ultimi, ad eccezione degli ultimi due anni 2012-2013.

3) **Andamenti del ROE.** Le serie del ROE di *Invind* sono sempre al di sopra di quelle *Cerved* e presentano andamenti tendenziali diversi. Dal 2010, le serie dei due archivi tendono ad avvicinarsi e nell'ultimo anno disponibile (2013) l'indice aggregato del ROE calcolato su *Cerved* è lievemente superiore a quello stimato da *Invind*.

4) **Variazione dell'utile.** Le serie *Invind* e *Cerved* hanno andamenti tendenziali e valori molto simili (ad eccezione del 2010, quando l'indice aggregato delle imprese *Cerved* diventa molto più elevato).

In sintesi, questi andamenti mostrano in alcuni casi differenze persistenti, maggiormente accentuati per il ROE e il ROA, ma non evidenziano nel campione *Invind* la presenza di distorsione uniforme rispetto alla popolazione di riferimento.

---

<sup>19</sup> La fonte statistica ufficiale più confrontabile (Istat, 2015) è di natura censuaria solo per le imprese con almeno 100 addetti. Essa è affetta da mancata risposta parziale e totale, che richiede complesse integrazioni con archivi amministrativi. Le stime diffuse consentirebbero esclusivamente confronti tra indici aggregati e solo per alcune delle variabili utilizzate in questo lavoro.

<sup>20</sup> Le evidenze di questa analisi, non mostrate per brevità, sono disponibili su richiesta.



## ***Età delle imprese del campione *Invind* e della popolazione di riferimento a confronto***

Si può confrontare la distribuzione del numero di anni di attività delle imprese della popolazione, ricostruita tramite l'archivio integrato, con quella stimata da *Invind* (tav. 10)<sup>21</sup>. Il confronto consente di verificare se l'evoluzione del campione abbia nel tempo condotto a caratteristiche di età delle imprese partecipanti all'indagine diverse da quelle della popolazione di riferimento. Il confronto rivela che media e mediana stimate tramite l'indagine sono superiori a quelle della popolazione (rispettivamente 28 e 25 anni, contro 16 e 13), evidenziando un processo di auto-selezione nel campione di imprese di più antica costituzione.

### **6. I ritardi di aggiornamento della popolazione di riferimento**

Le informazioni in forma aggregata sulla popolazione di riferimento sono adoperate per costruire il sistema di pesi di espansione dei dati elementari, necessario per ottenere stime campionarie rappresentative della popolazione stessa. Per l'indagine *Invind*, tali informazioni sono fornite dall'Istat con un ritardo di aggiornamento di 2-3 anni rispetto alla chiusura dell'indagine e la prima produzione delle stime. Negli anni successivi, i pesi di espansione sono aggiornati man mano che l'Istat rende disponibili gli aggiornamenti.

L'edizione dell'indagine *Invind* relativa all'anno  $t$ , condotta nei primi 4-5 mesi dell'anno  $t+1$  rileva il livello delle principali variabili di interesse (occupazione, fatturato e investimenti) per gli anni  $t-1$ ,  $t$  e  $t+1$  (quest'ultimo dato rappresenta una previsione formulata in corso d'anno), in modo da stimare nella stessa indagine le variazioni a consuntivo intercorse tra  $t$  e  $t-1$  e quelle previste tra  $t+1$  e  $t$ . Di conseguenza, ogni edizione dell'indagine dovrebbe utilizzare tre distinte popolazioni di riferimento. Questo approccio non è al momento utilizzato, perché richiederebbe ripetuti aggiornamenti del sistema di ponderazione nel corso degli anni. Si utilizza soltanto l'informazione sulla popolazione relativa all'anno  $t$ , aggiornata appena possibile. L'effetto di questo aggiornamento sulle stime è limitato.

Questa scelta non comporta distorsioni se la numerosità della popolazione di riferimento si mantiene stabile, come è avvenuto fino al 2006 (fig. 7). L'evoluzione successiva mostra un aumento del numero complessivo di imprese nel 2007-2008, seguito da una lieve ma costante diminuzione nel corso del 2009-2012, probabile effetto della crisi economica iniziata in Italia alla fine del 2008. Si nota inoltre una progressiva diminuzione del numero di imprese industriali, parzialmente compensata da un aumento del numero di imprese dei servizi. Questo fenomeno è principalmente imputabile alle trasformazioni strutturali dell'economia italiana, che hanno comportato un calo progressivo della quota di PIL prodotta dal settore industriale e, in misura minore, alle revisioni dei sistemi di classificazione delle attività economiche. Queste ultime tendono ad allocare nel settore dei servizi alcune attività di tipo accessorio (manutenzione, assistenza al cliente, etc.) originariamente classificate come industriali (Istat, 2010).

Si è ritenuto utile verificare come variano gli indicatori ponderati presentati in tavola 3 se si usano due distinte popolazioni di riferimento relative al tempo  $t-1$  e al tempo  $t$  per stimare la dinamica a consuntivo, anziché un'unica popolazione relativa al tempo  $t$ <sup>22</sup>. In questa analisi gli indicatori ponderati sono stati calcolati per le variabili di maggior interesse dell'indagine *Invind*: occupazione, fatturato e investimenti.

L'analisi è stata condotta dapprima per gli indici aggregati ponderati. Per questi il sistema di ponderazione unico sovrastima il tasso di variazione in presenza di un *trend* di decrescita della popolazione di riferimento e lo sottostima se la popolazione di riferimento è in crescita. La distorsione è stata marcata per

---

<sup>21</sup> Per rendere confrontabili i dati, si utilizza la data di fondazione di fonte *Cerved* sia per la popolazione sia per il campione, anche se quest'ultima è rilevata nell'indagine.

<sup>22</sup> Per brevità non si estende l'esercizio alle previsioni di variazione tra gli anni  $t$  e  $t+1$ .

gli anni 2007-2009 (in particolare per gli investimenti nel 2008), periodo in cui la dinamica della popolazione mostra variazioni più accentuate. Negli anni successivi, la sovrastima delle variazioni indotta dal *trend* di diminuzione del numero di imprese è stata contenuta (fig. 8, metà superiore).

L'entità della distorsione (in termini di rapporto tra le due variazioni) risulta maggiore per la dinamica dell'occupazione. Il risultato era atteso, dato che la qualità di questa stima dipende direttamente dalla corretta classificazione dimensionale della popolazione di imprese.

Infine, anche per gli indici semplici si è verificato l'effetto dell'utilizzo di due popolazioni di riferimento per singola edizione dell'indagine. Questi indici sono calcolati utilizzando l'espressione:

$$\frac{1}{\sum_{i=1}^{n_t} w_{i,t-1} \cap t} \sum_{i=1}^{n_t} \frac{y_{i,t}}{y_{i,t-1}} w_{i,t-1} \cap t \quad (1)$$

dove  $w_{i,t-1} \cap t$  indica il peso riferito alla popolazione di imprese esistenti sia al tempo  $t-1$  sia al tempo  $t$ <sup>23</sup>. Per gli investimenti, si utilizza la mediana ponderata di  $\frac{y_{i,t}}{y_{i,t-1}}$  anziché la sua media. La scelta si giustifica perché la variabile presenta variabilità ed asimmetria estreme, che la differenziano da fatturato e occupazione<sup>24</sup>. Per tutte queste stime, l'utilizzo della doppia popolazione di riferimento per ogni singola edizione comporta variazioni molto limitate (fig. 8, metà inferiore).

Per occupazione e fatturato, la variazione trascurabile dell'indice semplice ponderato si spiega esaminando l'espressione della differenza tra l'indice semplice originario e quello ricalcolato. Indicando con  $n$  la numerosità campionaria, essa si scrive come:

$$\sum_{i=1}^{n_t} \frac{y_{i,t}}{y_{i,t-1}} \left( \frac{w_{i,t}}{\sum_{i=1}^{n_t} w_{i,t}} - \frac{w_{i,t-1} \cap t}{\sum_{i=1}^{n_t} w_{i,t-1} \cap t} \right) \quad (2)$$

Si tratta di una somma di  $n$  addendi formati dal prodotto tra due quantità,  $\frac{y_{i,t}}{y_{i,t-1}}$  sempre positiva e quelle tra parentesi senza un segno predefinito.

Per l'indice aggregato, la differenza tra l'espressione originaria e quella ricalcolata con il doppio sistema di ponderazione si scrive invece come:

$$\sum_{i=1}^{n_t} \left( \frac{y_{i,t}}{y_{i,t-1}} \left( w_{i,t} y_{i,t-1} \left( \frac{1}{\sum_{i=1}^{n_t} w_{i,t} y_{i,t-1}} - \frac{1}{\sum_{i=1}^{n_t} w_{i,t-1} y_{i,t-1}} \right) \right) \right) \quad (3)$$

Come la (2), l'espressione (3) è una somma di  $n$  addendi, ognuno dato dal prodotto tra due quantità. La prima è la stessa della (2), la seconda in parentesi ha segno costante, positivo o negativo<sup>25</sup>, per tutti gli addendi. Ne consegue quindi che il valore assoluto della espressione (3) sarà superiore a quello della (2), per tutti i casi in cui le due grandezze  $\left| \frac{y_{i,t}}{y_{i,t-1}} \right|$  e  $\left| \frac{w_{i,t}}{w_{i,t-1}} \right|$  hanno intervalli di variazione limitati, situazione che caratterizza le situazioni reali con variabili diverse dagli investimenti.

<sup>23</sup> Questi pesi sono ottenuti ipotizzando che in ogni strato la popolazione presente al tempo  $t-1$  e anche al tempo  $t$  sia quella minima presente nei due anni considerati. Si tratta di una approssimazione per eccesso, valida quanto più le entrate e le uscite dalla popolazione sono trascurabili rispetto ai livelli della popolazione nei due anni considerati.

<sup>24</sup> Queste caratteristiche si spiegano con il fatto che le imprese tendono a non investire in tutti gli anni e di conseguenza la variazione per impresa degli investimenti è caratterizzata da forte variabilità, con molti valori concentrati nelle code della distribuzione. La asimmetria della distribuzione è accentuata dal fatto che la variazione percentuale ha limite inferiore pari a -100, ma non è limitata superiormente. I casi con valore di  $y_{i,t-1}=0$  sono stati posti pari a 0,01 per considerare anch'essi nell'analisi.

<sup>25</sup> Ricordiamo che tutte le grandezze considerate sono maggiori di zero.

## 7. Conclusioni

Questo lavoro studia la distorsione delle stime dell'indagine *Invind*, condotta dalla Banca d'Italia su un campione *panel* di imprese, come emerge dal confronto con due fonti esterne:

- un archivio integrato di dati di bilancio che fornisce informazioni economiche e strutturali per ogni impresa che ha partecipato all'indagine per tutto l'arco temporale considerato (2012-2013);
- dati aggregati sulla distribuzione della popolazione di riferimento dell'indagine, disponibili per il periodo 2012-2012.

L'archivio integrato è stato usato per verificare la presenza di eventuali effetti distorsivi della *panel attrition*, definita come processo di uscita di imprese da un'edizione d'indagine e entrata di nuove imprese nell'edizione successiva.

Per quanto riguarda gli effetti della *panel attrition*, dal lavoro emerge che gli stimatori maggiormente dipendenti dalle imprese di grandi dimensioni, correntemente diffusi nelle pubblicazioni ufficiali (Banca d'Italia, 2014), ne risentono in misura limitata. Al contrario, le analisi micro-econometriche che tendono a utilizzare indicatori rappresentativi del comportamento medio d'impresa, senza tenere conto della diversa dimensione delle differenti unità, necessitano di maggiore cautela. Le principali distorsioni derivano dal fatto che le imprese che escono dal campione presentano risultati inferiori alla media del campione, sia nell'anno di uscita sia in quello successivo e dalla sovra-rappresentazione delle imprese con maggior numero di anni di attività. La distorsione generata dalle imprese inserite nell'indagine per sostituire le uscite è più lieve, in quanto esse tendono ad avere risultati economici lievemente migliori della media del campione solo nell'anno precedente l'entrata.

Saranno utili futuri approfondimenti sui sistemi di correzione dei pesi di espansione dell'indagine (Sarndal e Lundstrom, 2005), atti a correggere le possibili distorsioni dei modelli micro-econometrici, maggiormente sensibili ai problemi della *panel attrition*<sup>26</sup>.

Dal punto di vista operativo della conduzione dell'indagine, emerge l'esigenza di un controllo più accurato del processo di entrate e uscite dal campione, dal momento che esso tende a interessare imprese con risultati economici particolarmente negativi per le uscite. In generale rafforzerebbe la rappresentatività del campione l'integrazione delle regole che governano il processo di sostituzione delle imprese uscite in modo da considerare l'evoluzione della popolazione per composizione settoriale e dimensionale, a livello regionale.

Il confronto delle stime, semplici e aggregate, calcolate sul campione e su una ricostruzione della popolazione di riferimento, mostra differenze che richiedono ulteriori approfondimenti<sup>27</sup>, ma non la presenza di distorsioni uniformi per tutte le variabili esaminate.

Si è però osservato che le imprese di *Invind* sono caratterizzate da un numero di anni di attività superiore a quello della popolazione di riferimento. Si ha quindi una tendenza all'autoselezione nel campione di imprese più anziane della media. Controllando per altre caratteristiche d'impresa, la presenza sul mercato da un numero più o meno lungo di anni è associata alle *performance* economiche, ma il legame è di intensità

---

<sup>26</sup> Per un esame critico dei casi in cui l'utilizzo dei pesi nei modelli è raccomandabile si vedano Faiella (2010) e Solonet *al.* (2013).

<sup>27</sup> Si tenga presente che alcune differenze sono dovute al fatto che la popolazione di riferimento, non disponibile direttamente, è stata ricostruita (Paragrafo 3).

variabile per gli indicatori considerati e risulta non unidirezionale<sup>28</sup>. La potenziale distorsione di questo effetto è controllabile includendo informazioni sulla distribuzione delle imprese per classe di età nella fase di post-stratificazione dei pesi, quando essi sono aggiornati negli archivi storici dell'indagine.

Per quel che riguarda gli effetti del ritardato aggiornamento delle informazioni sulla popolazione di riferimento, occorre tener presente che *Invind* rileva nella stessa indagine i valori relativi a tre anni consecutivi  $t-1$ ,  $t$  e  $t+1$  per le variabili di maggior interesse (il valore più recente è una previsione rilevata nel corso dell'anno). Il ritardo giustifica la soluzione attualmente adottata di utilizzare per ogni indagine informazioni relative alla sola popolazione dell'anno  $t$ . La scelta può tuttavia creare distorsione nelle stime aggregate ponderate: positiva se la dimensione della popolazione decresce, negativa se aumenta; le stime semplici non sono invece affette da questo problema. Se la dimensione della popolazione ha quindi una forte variazione non registrata tempestivamente, occorre cautela nella valutazione delle stime aggregate ponderate delle indagini più recenti, diffuse nelle pubblicazioni ufficiali. Nei prossimi anni il problema potrebbe diventare meno rilevante, per il progressivo miglioramento della tempestività della diffusione delle statistiche ufficiali. Nel frattempo, potrebbe essere utile inserire negli archivi storici dell'indagine il doppio sistema di ponderazione per correggere la distorsione delle edizioni meno recenti dell'indagine, nonché investigare le tecniche e le fonti di aggiustamento della distribuzione della popolazione ufficiale, utili a compensare questo tipo di distorsione.

---

<sup>28</sup> Questa evidenza si ricava da semplici regressioni degli indicatori economici analizzati nella tavola 9 sullo stesso arco temporale. Le regressioni sono state stimate su tutto il campione. Il numero di anni di attività dell'impresa non ha una associazione significativa con l'andamento del ROA, mentre si accompagna a migliori risultati in termini di ROE e utile. Il legame tra durata dell'attività e andamento del fatturato, seppure significativo, è di modesta entità e di segno negativo. I risultati di questa analisi non sono mostrati per brevità e sono disponibili su richiesta.

## Bibliografia

- Afonso L.(2015), *Correcting for Attrition in Panel Data Using Inverse Probability Weighting: An application To the EU15 Bank System*, Lisbon School of Economics and Management, Working Paper.
- Banca d'Italia (2005), Supplementi al Bollettino Statistico, Indagini campionarie, Indagine sulle imprese industriali e dei servizi, Anno di riferimento 2003, Anno XV, 20 Ottobre 2005.
- Biancotti C., D'Alessio G., Neri A. (2004), *Errori di misura nell'indagine sui bilanci delle famiglie italiane*, Temi di Discussione Banca d'Italia n. 520.
- Banca d'Italia (2014), Supplementi al Bollettino Statistico, Indagini campionarie, Indagine sulle imprese industriali e dei servizi, Anno di riferimento 2013, Nuova serie, Anno XXIV, 24 Luglio 2014.
- Banca d'Italia (2015), Supplementi al Bollettino Statistico, Indagini campionarie, Indagine sulle imprese industriali e dei servizi, Anno di riferimento 2014, Nuova serie, Anno XXV, 1° Luglio 2015.
- Black C., Broadstock D.C., Collins A., Hunt L. (2007), *A practical guide to developments in data imputation methods*, [www.tecmagazine.com](http://www.tecmagazine.com).
- Brick J.M. (2013), *Unit Nonresponse and Weighting Adjustments: A Critical Review*, Journal of Official Statistics, Vol. 29, No. 3, 2013, pp. 329–353.
- Deng Y., Hillygus D. S., Reiter J.P., Si Y., Zhen S. (2013), *Handling Attrition in Longitudinal Studies: The Case for Refreshment Samples*, Statistical Science, Vol. 28, No. 2, pp. 238–256.
- Deville J.-C., Sarndal C.-E., Sautory O. (1993), *Generalized Raking Procedures in Survey Sampling*, Journal of the American Statistical Association, Vol. 88, No. 423 (Sep., 1993), pp. 1013-1020.
- Fabbris L. (1989), *L'indagine campionaria*, Nuova Italia Scientifica.
- Faiella I.(2010), *The use of survey weights in regression analysis*, Bank of Italy's Working Paper n. 739.
- Istat (2010), <http://www.istat.it/it/strumenti/definizioni-e-classificazioni/ateco-2007>.
- Istat (2014), Archivio Asia <http://www.istat.it/it/archivio/archivio+asia>.
- Istat (2015), *Struttura e competitività delle imprese*, <http://www.istat.it/it/archivio/14020>.
- Little R., Rubin D. (2002), *Statistical Analysis with Missing Data*, John Wiley and Sons Inc., USA.
- Martin E., Abreu D., Winters F., *Money and Motive: Effects of Incentives on Panel Attrition in the Survey of Income and Program Participation*, Journal of Official Statistics, Vol. 17, No. 2, 2001, pp. 267–284.
- Sarndal C.-E., Lundstrom S. (2005), *Estimation in Surveys with Nonresponse*, Wiley.
- Solon G., Haider S.J., Wooldridge J. (2013), *What are we weighting for?*, NBER Working Paper n. 18859.
- Trento S. (2012), *Il capitalismo italiano*, Il Mulino – Bologna.
- Trivellato U. (1999), *Issues in the Design and Analysis of Panel Studies: a Cursory Review*, Quality & Quantity, Vol. 33, pp. 339-352-

## **Tavole e Grafici**

Tavola 1

Indagine *Invind*: imprese permanenti, uscite e entrate nelle varie edizioni dell'indagine (2002-2013)

Anno	Totale imprese del campione	Imprese permanenti nell'indagine successiva		Imprese uscite dall'indagine successiva						Imprese assenti nell'indagine precedente e entrate in quella corrente				
		%	num. imp.	%	num. imp.	di cui (% su imprese uscite):					%	num. imp.	di cui: presenti in indagini precedenti	
						Ancora attive con 20 addetti e oltre	Ancora attive con meno di 20 addetti	Non più rilevabili per incorpori, fusioni, etc.	Non più attive	Totale			%	num. imp.
2002	3.969	78,2	3.105	21,8	864	69,9	7,1	2,9	20,1	100,0	---	---	---	---
2003	4.135	78,4	3.240	21,6	895	74,2	9,3	2,0	14,5	100,0	24,9	1030	2,6	109
2004	4.226	80,0	3.382	20,0	844	74,1	7,2	2,1	16,6	100,0	23,3	986	3,4	142
2005	4.386	79,8	3.498	20,2	888	77,5	8,2	2,0	12,3	100,0	22,9	1004	4,4	195
2006	4.252	80,3	3.415	19,7	837	75,5	9,0	1,6	14,0	100,0	17,7	754	4,7	199
2007	4.063	79,2	3.217	20,8	846	74,6	10,3	1,7	13,5	100,0	15,9	648	5,3	214
2008	3.952	80,3	3.175	19,7	777	78,2	8,1	2,3	11,3	100,0	18,6	735	6,8	270
2009	3.921	80,6	3.162	19,4	759	78,0	7,9	2,0	12,1	100,0	19,0	746	7,4	291
2010	3.937	82,5	3.248	17,5	689	78,4	8,9	0,4	12,3	100,0	19,7	775	8,0	313
2011	4.120	82,4	3.396	17,6	724	77,6	7,3	1,7	13,4	100,0	21,2	872	8,8	361
2012	4.213	82,1	3.460	17,9	753	74,1	8,1	2,0	15,8	100,0	19,4	817	8,6	362
2013	4.215	---	---	---	---	---	---	---	---	---	17,9	755	11,8	498
<b>Media (%)</b>		<b>80,3</b>		<b>19,7</b>		<b>75,6</b>	<b>8,3</b>	<b>1,9</b>	<b>14,2</b>	<b>100,0</b>	<b>20,0</b>		<b>6,5</b>	
<b>Dimensione media (num. addetti)</b>		<b>366</b>				<b>227</b>					<b>252</b>			
<b>Numero medio anni di attività</b>		<b>34</b>				<b>31</b>					<b>29</b>			

Tavola 2

**Indagine *Invind*: tassi di permanenza, uscita e entrata medi per le edizioni dell'indagine (2002-2013)**  
(%)

	Media industria in senso stretto	Media servizi privati non finanziari
Imprese permanenti nell'indagine successiva	81,1	78,3
Imprese uscite dall'indagine successiva	18,9	21,7
Imprese assenti nell'indagine precedente e entrate in quella corrente	19,0	23,0

Tavola 3

**Indici di sintesi dei risultati economici delle imprese**

<b>Indici aggregati</b>		
	<b>Variabile monetaria<sup>(a)</sup></b>	<b>Variabile adimensionale<sup>(b)</sup></b>
Indice non ponderato	$\frac{\sum_{i=1}^{n_t} y_{it}}{\sum_{i=1}^{n_t} y_{it-1}}$	$\frac{1}{\sum_{i=1}^{n_t} z_{it}} \sum_{i=1}^{n_t} x_{it} z_{it}$
Indice ponderato	$\frac{\sum_{i=1}^{n_t} y_{it} w_{it}}{\sum_{i=1}^{n_t} y_{it-1} w_{it}}$	$\frac{1}{\sum_{i=1}^{n_t} z_{it} w_{it}} \sum_{i=1}^{n_t} x_{it} z_{it} w_{it}$
<b>Indici semplici</b>		
	<b>Variabile monetaria<sup>(a)</sup></b>	<b>Variabile adimensionale<sup>(b)</sup></b>
Indice non ponderato	$\frac{1}{n_t} \sum_{i=1}^{n_t} \frac{y_{it}}{y_{it-1}}$	$\frac{1}{n_t} \sum_{i=1}^{n_t} x_{it}$
Indice ponderato	$\frac{1}{\sum_{i=1}^{n_t} w_{it}} \sum_{i=1}^{n_t} \frac{y_{it}}{y_{it-1}} w_{it}$	$\frac{1}{\sum_{i=1}^{n_t} w_{it}} \sum_{i=1}^{n_t} x_{it} w_{it}$

(a) Es.: fatturato e utile. – (b) Es.: ROA e ROE.



Tavola 4

Indagine *Invind*: percentuale di imprese dei campioni annuali con dati economici relativi all'anno disponibili dagli archivi *Cerved* (2002-2013)

	Imprese dell'industria in senso stretto	Imprese dei servizi privati non finanziari	Totale
2002	87,8	89,6	88,2
2003	87,6	89,1	88,0
2004	88,6	91,8	89,4
2005	92,4	93,0	92,5
2006	93,3	92,6	93,1
2007	93,1	93,4	93,2
2008	92,1	94,2	92,7
2009	93,2	93,5	93,2
2010	93,4	94,8	93,8
2011	93,8	95,1	94,2
2012	94,3	94,6	94,4
2013	90,3	88,9	89,9
<b>Totale</b>	<b>91,6</b>	<b>92,6</b>	<b>91,9</b>

Tavola 5

Distribuzione delle imprese e degli addetti delle imprese *Cerved* con occupazione con 20 addetti e oltre ricostruita per i casi mancanti (2012)

	Archivio Statistico Imprese Attive (ASIA)		Cerved	
	Imprese (%)	Addetti (%)	Imprese (%)	Addetti (%)
Classe dim. le				
20-49 addetti	67,7	22,0	67,0	21,7
50-199 addetti	26,1	25,5	26,3	25,9
200-499 addetti	4,2	13,9	4,6	15,1
500 addetti e oltre	1,9	38,6	2,1	37,3
Settore				
Manifattura	46,8	39,5	44,8	41,2
Energetico-estrattivo	2,8	4,2	3,5	4,6
Servizi	50,4	56,3	51,7	54,3
Area geografica <sup>(1)</sup>				
Nord Ovest	35,2	40,6	33,1	39,0
Nord Est	28,0	24,4	26,5	24,0
Centro	19,0	22,4	23,4	24,5
Sud e Isole	17,8	12,5	17,0	12,6
	Totale imprese	Totale addetti	Totale imprese	Totale addetti
	64.571	5.926.256	57.701	5.321.570

(1) Sede amministrativa dell'impresa.

Indagine *Invid*: probabilità di entrare nel campione l'anno successivo (2002-2012)<sup>(1)</sup>

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Indicatori economici</b>								
roa	0,01430 ***						0,00675 ***	0,00656 ***
roe		0,00002 ***					0,00002 ***	0,00002 ***
variazione utile per addetto (%)			0,00293 **				0,00760 **	0,00107
variazione utile (%)				0,00318 **			-0,00841 *	
variazione fatturato per addetto (%)					0,00417		-0,01500	-0,02400
variazione fatturato (%)						-0,01810	-0,02570	
<b>Classe dimensionale (baseline: 500 e oltre)</b>								
20-49 addetti	-1,56690 ***	-1,55390 ***	-1,40620 ***	-1,51010 ***	-1,40870 ***	-1,51670 ***	-1,40160 ***	-1,39930 ***
50-199 addetti	-0,42320 ***	-0,41660 ***	-0,41650 ***	-0,40920 ***	-0,42060 ***	-0,41350 ***	-0,40650 ***	-0,40620 ***
200-499 addetti	0,55840 ***	0,54940 ***	0,55330 ***	0,59020 ***	0,55950 ***	0,59460 ***	0,54130 ***	0,54000 ***
<b>Area geografica (baseline: Sud e isole)<sup>(2)</sup></b>								
Nord ovest	-0,66690 ***	-0,66570 ***	-0,69190 ***	-0,65750 ***	-0,69150 ***	-0,65590 ***	-0,70470 ***	-0,70410 ***
Nord est	-0,20530 ***	-0,20420 ***	-0,23530 ***	-0,21530 ***	-0,23360 ***	-0,21020 ***	-0,25200 ***	-0,25180 ***
Centro	0,04110	0,03780	0,08120	0,07900	0,07900 ***	0,07360 **	0,08390 ***	0,08460 ***
<b>Attività economica (baseline: Alberghi e ristorazione)</b>								
Industrie alimentari, bevande e tabacco	0,99350 ***	0,96780 ***	0,86760 ***	0,90750 ***	0,86940 ***	0,91130 ***	0,84660 ***	0,84580 ***
Trasporti e comunicazioni	-0,78300 ***	-0,80100 ***	-0,62370 ***	-0,65320 ***	-0,62300 ***	-0,66470 ***	-0,59270 ***	-0,59100 ***
Altri serv. priv. non fin.	-1,19170 ***	-1,09850 ***	-0,93930 ***	-1,08330 ***	-0,94480 ***	-1,09390 ***	-0,94360 ***	-0,94450 ***
Industrie tessili, dell'abbigliamento, pelli cuoio e calzature	0,39410 ***	0,38290 ***	0,25330 ***	0,32250 ***	0,25690 ***	0,32840 ***	0,27800 ***	0,27830 ***
Fabbricazione di coke, industria chimica, gomma e plastica	0,67720 ***	0,65880 ***	0,56390 ***	0,59320 ***	0,56780 ***	0,60140 ***	0,53920 ***	0,53950 ***
Industria della lavorazione dei minerali non metalliferi	0,61070 ***	0,57130 ***	0,48630 ***	0,54550 ***	0,48730 ***	0,55150 ***	0,48100 ***	0,48120 ***
Industria metalmeccanica	0,27770 ***	0,26160 ***	0,18660 ***	0,23290 ***	0,19090 ***	0,23700 ***	0,16830 ***	0,16760 ***
Altre industrie manifatturiere	0,40130 ***	0,39400 ***	0,33610 ***	0,37080 ***	0,33620 ***	0,37290 ***	0,32790 ***	0,32800 ***
Energetico-estrattivo	-0,25220 ***	-0,22470 ***	-0,20390 ***	-0,17310 **	-0,21300 ***	-0,17640 **	-0,20730 ***	-0,20740 ***
Commercio ingrosso e dettaglio	-0,21010 ***	-0,24430 ***	-0,20120 ***	-0,28220 ***	-0,19280 ***	-0,27630 ***	-0,20210 ***	-0,20280 ***
numero osservazioni	1.022.729	1.018.066	590.433	701.037	597.639	727.512	590.045	590.045

(1) \*\*\*: valore del p-value inferiore a 0,01, \*\*: valore del p-value compreso tra 0,01 e 0,05, \*=valore del p-value compreso tra 0,05 e 0,1. Valori dei regressori inferiori al 1° o superiori al 99° percentile posti pari al percentile di riferimento. Errori standard calcolati considerando la stessa impresa ripetuta nel tempo come *cluster*. Tutte le stime includono dummy di anno. - (2) Riferita alla sede amministrativa dell'impresa.

Tavola 7

Indagine *Invid*: probabilità di uscire dal campione nella successiva edizione (2002-2012)<sup>(1)</sup>

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
roa	-0,02440 ***						-0,00807 ***	-0,00900 ***
roe		-0,00390 ***					-0,00286 ***	-0,00290 ***
variazione utile per addetto (%)			-0,00501 ***				-0,01670	-0,00166
variazione utile (%)				-0,00504 ***			0,01550	
variazione fatturato per addetto (%)					-0,24800 ***		0,19670	-0,12230
variazione fatturato (%)						-0,44670 ***	-0,39640 **	
<b>Classe dimensionale (baseline: 500 e oltre)</b>								
20-49 addetti	0,29870 ***	0,30230 ***	0,26120 ***	0,30000 ***	0,30480 ***	0,30030 ***	0,25240 ***	0,25620 ***
50-199 addetti	0,04560	0,05190	0,06260	0,04790	0,06760	0,06820 **	0,06000	0,05900
200-499 addetti	-0,09260 **	-0,09270 **	-0,09270 **	-0,10640 ***	-0,11020 ***	-0,10730 ***	-0,08720 **	-0,08860 **
<b>Area geografica (baseline: Sud e isole)<sup>(2)</sup></b>								
Nord ovest	0,03000	0,03100	0,02140	0,02360	0,02510	0,02220	0,02560	0,02870
Nord est	0,30690 ***	0,30420 ***	0,30290 ***	0,29370 ***	0,29740 ***	0,29770 ***	0,31870 ***	0,31970 ***
Centro	-0,20110	-0,20710	-0,21130	-0,21490	-0,20510	-0,20440 ***	-0,20060	-0,20150
<b>Attività economica (baseline: Alberghi e ristorazione)</b>								
Industrie alimentari, bevande e tabacco	-0,25430 ***	-0,24220 ***	-0,24270 ***	-0,22910 ***	-0,26000 ***	-0,25660 ***	-0,22990 ***	-0,23160 ***
Trasporti e comunicazioni	0,03490	0,04710	0,03510	0,03950	0,06400	0,06780	0,04160	0,03820
Altri serv. priv. non fin.	0,25430 ***	0,20030 ***	0,16590 **	0,17550 ***	0,14820 **	0,15630 ***	0,19370 ***	0,18970 ***
Industrie tessili, dell'abbigliamento, pelli cuoio e calzature	0,05230	0,05280	0,08810	0,08520	0,06300	0,05180	0,04570	0,05250
Fabbricazione di coke, industria chimica, gomma e plastica	0,02930	0,03560	0,02270	0,01130	0,02950	0,03150	0,04900	0,04820
Industria della lavorazione dei minerali non metalliferi	-0,08870	-0,08910	-0,05690	-0,06550	-0,08260	-0,08950	-0,08140	-0,07660
Industria metalmeccanica	-0,11320 ***	-0,12620 ***	-0,12900 ***	-0,13340 ***	-0,13590 ***	-0,13320 ***	-0,11680 ***	-0,11770 ***
Altre industrie manifatturiere	-0,08560 *	-0,08120	-0,06610	-0,06290	-0,07090	-0,07730	-0,08600	-0,08220
Energetico-estrattivo	-0,13500	-0,11850	-0,13660	-0,14770	-0,13950	-0,13640	-0,11770	-0,11940
Commercio ingrosso e dettaglio	0,05530	0,06390	0,04770	0,05670	0,07100	0,07380 *	0,06290	0,05950
numero osservazioni	38.708	37.732	34.828	36.027	41.275	41.279	34.247	34.247

(1) \*\*\*: valore del p-value inferiore a 0,01, \*\*: valore del p-value compreso tra 0,01 e 0,05, \* = valore del p-value compreso tra 0,05 e 0,1. Valori dei regressori inferiori al 1° o superiori al 99° percentile posti pari al percentile di riferimento. Errori standard calcolati considerando la stessa impresa ripetuta nel tempo come *cluster*. Tutte le stime includono *dummy* di anno. - (2) Riferita alla sede amministrativa dell'impresa.

Tavola 8

**Indagine *Invid*: legame associativo tra propensione a entrare nel campione e caratteristiche d'impresa (2003-2013)<sup>(1)</sup>**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Indicatori economici</b>								
roa	0,00864 ***						0,00541 ***	0,00621 ***
roe		0,00002 ***					0,00001 ***	0,00001 ***
variazione utile per addetto (%)			0,00041				-0,00200	-0,00062
variazione utile (%)				0,00109			0,00175	
variazione fatturato per addetto (%)					-0,25550		-0,35850	-0,28630
variazione fatturato (%)						-0,10580	0,19240	
<b>Classe dimensionale (baseline: 500 e oltre)</b>								
20-49 addetti	-1,56640 ***	-1,55670 ***	-1,40420 ***	-1,55840 ***	-1,43440 ***	-1,56610 ***	-1,43620 ***	-1,43080 ***
50-199 addetti	-0,42020 ***	-0,41120 ***	-0,44380 ***	-0,41940 ***	-0,44820 ***	-0,42190 ***	-0,43620 ***	-0,43680 ***
200-499 addetti	0,57770 ***	0,56560 ***	0,54220 ***	0,57280 ***	0,57250 ***	0,58040 ***	0,56590 ***	0,56150 ***
<b>Area geografica (baseline: Sud e isole)<sup>(2)</sup></b>								
Nord ovest	-0,69530 ***	-0,69980 ***	-0,71770 ***	-0,69000 ***	-0,72660 ***	-0,69110 ***	-0,73690 ***	-0,73830 ***
Nord est	-0,22850 ***	-0,23780 ***	-0,26720 ***	-0,22340 ***	-0,27260 ***	-0,22090 ***	-0,28970 ***	-0,29030 ***
Centro	0,05040	0,05590	0,06790	0,05330	0,08880 ***	0,04970 *	0,09540 ***	0,09210 ***
<b>Attività economica (baseline: Alberghi e ristorazione)</b>								
Industrie alimentari, bevande e tabacco	0,97080 ***	0,93820 ***	0,89940 ***	0,96100 ***	0,90040 ***	0,97420 ***	0,87240 ***	0,87480 ***
Trasporti e comunicazioni	-0,74720 ***	-0,72190 ***	-0,69160 ***	-0,73600 ***	-0,62740 ***	-0,75000 ***	-0,59130 ***	-0,59670 ***
Altri serv. priv. non fin.	-1,07470 ***	-1,02510 ***	-0,94680 ***	-1,03450 ***	-0,96680 ***	-1,04530 ***	-0,96400 ***	-0,95880 ***
Industrie tessili, dell'abbigliamento, pelli cuoio e calzature	0,38470 ***	0,38420 ***	0,30360 ***	0,36870 ***	0,28640 ***	0,36700 ***	0,31340 ***	0,30960 ***
Fabbricazione di coke, industria chimica, gomma e plastica	0,65080 ***	0,62750 ***	0,57630 ***	0,65090 ***	0,57050 ***	0,66080 ***	0,54040 ***	0,54120 ***
Industria della lavorazione dei minerali non metalliferi	0,56780 ***	0,53290 ***	0,50130 ***	0,54990 ***	0,48790 ***	0,55130 ***	0,48610 ***	0,48210 ***
Industria metalmeccanica	0,27660 ***	0,26680 ***	0,25650 ***	0,28110 ***	0,25310 ***	0,28580 ***	0,23160 ***	0,23490 ***
Altre industrie manifatturiere	0,38060 ***	0,36100 ***	0,35230 ***	0,36660 ***	0,34530 ***	0,36830 ***	0,33590 ***	0,33340 ***
Energetico-estrattivo	-0,25520 ***	-0,24540 ***	-0,26990 ***	-0,23730 ***	-0,26190 ***	-0,23620 ***	-0,26680 ***	-0,26270 ***
Commercio ingrosso e dettaglio	-0,22910 ***	-0,24150 ***	-0,18330 ***	-0,24690 ***	-0,18640 ***	-0,23780 ***	-0,19900 ***	-0,19670 ***
<b>numero osservazioni</b>	888.508	868.249	634.578	862.571	642.407	893.524	634.154	634.154

(1) \*\*\*: valore del p-value inferiore a 0,01, \*\*: valore del p-value compreso tra 0,01 e 0,05, \* = valore del p-value compreso tra 0,05 e 0,1. Valori dei regressori inferiori al 1° o superiori al 99° percentile posti pari al percentile di riferimento. Errori standard calcolati considerando la stessa impresa ripetuta nel tempo come *cluster*. Tutte le stime includono *dummy* di anno. - (2) Riferita alla sede amministrativa dell'impresa.

**Indagine *Invind*: legame tra entrata nel campione e principali indicatori economici  
(2003-2013)<sup>(1)(2)(3)(4)</sup>**

Tutto il campione								
	Regressione semplice				Regressione con fattore di scala			
	Variazione del fatturato	ROA	ROE	Variazione dell'utile	Variazione del fatturato	ROA	ROE	Variazione dell'utile
2003	3,769 ***	2,335 ***	1,367	27,519	-2,321 **	1,351 ***	4,614 *	398,934 ***
2004	1,613	0,600	1,484	-24,311	4,869 ***	-0,474	-4,170	56,106
2005	6,071 ***	1,409 ***	5,585 **	69,410	-1,271	0,335	0,859	-80,208 **
2006	4,577 ***	2,221 ***	3,281	60,170	2,156 **	0,876 **	0,439	-14,813
2007	3,421 ***	2,195 ***	4,297 *	-8,553	1,456	0,345	5,496 ***	-26,960
2008	4,868 ***	3,520 ***	4,022	34,575	1,844 **	1,871 ***	8,080 ***	-468,128 ***
2009	4,948 ***	2,740 ***	25,342 ***	136,912	-4,899 ***	1,739 ***	18,495 ***	10,774
2010	5,892 ***	2,497 ***	5,639 **	72,544	1,400	0,505	1,175	75,920
2011	3,760 ***	2,066 ***	6,782 ***	150,704	-2,483 **	-0,318	-3,682 *	-224,208 ***
2012	6,306 ***	4,262 ***	7,397 ***	59,313	8,442 ***	4,488 ***	8,905 ***	145,227 **
2013	7,092 ***	3,180 ***	4,428	17,229	-0,702	-0,335	11,316 ***	-3,032
Stima pooled <sup>(5)</sup>	4,889 ***	2,488 ***	6,547 ***	59,609 **	0,537	0,417	3,880	27,622

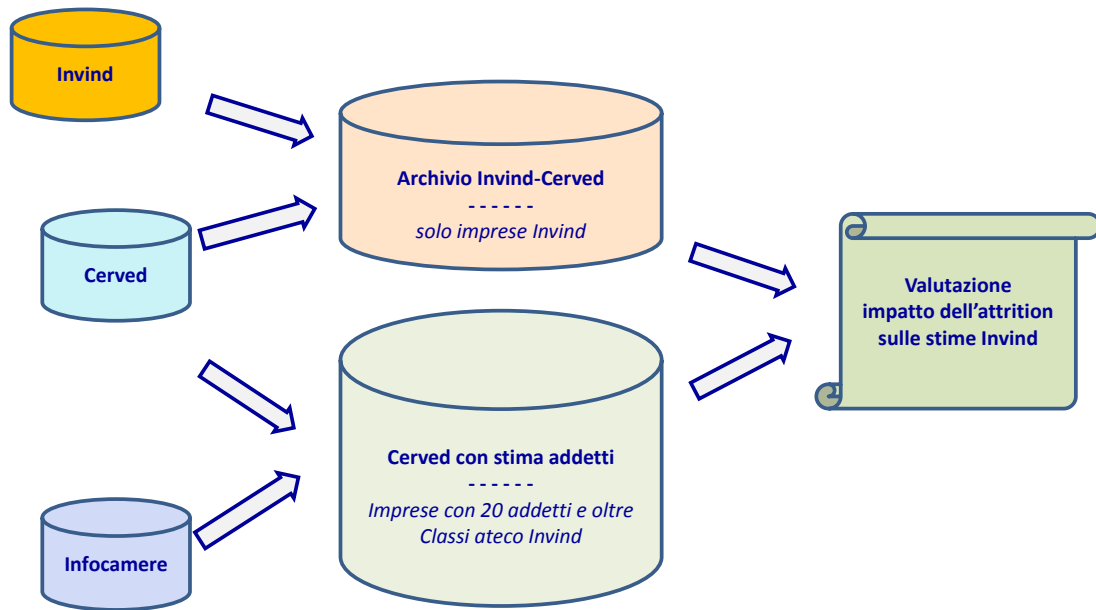
Solo imprese manifatturiere								
	Regressione semplice				Regressione con fattore di scala			
	Variazione del fatturato	ROA	ROE	Variazione dell'utile	Variazione del fatturato	ROA	ROE	Variazione dell'utile
2003	4,135 **	2,278 ***	1,166	39,913	-1,776	1,275 **	4,196	754,836 ***
2004	2,430	1,143 **	0,038	-24,815	1,419	-1,181 **	-15,500 ***	161,757 **
2005	7,863 ***	1,715 ***	7,329 **	55,563	0,446	-0,867	2,869	-72,443
2006	6,351 ***	2,753 ***	7,595 ***	43,851	3,239 **	2,541 ***	10,728 ***	-7,906
2007	3,088 *	2,321 ***	4,222	27,301	2,243 **	1,437 ***	6,085 **	-46,000
2008	6,686 ***	4,242 ***	5,166 *	184,032	4,802 ***	1,487 **	9,868 ***	202,264 **
2009	6,316 ***	3,452 ***	31,040 ***	58,151	-1,262	2,578 ***	27,898 ***	-57,976
2010	5,372 **	2,933 ***	7,168 **	104,516	-3,489 **	0,861	0,639	84,850
2011	5,197 **	2,509 ***	8,673 ***	174,638	4,351 ***	-2,656 ***	-10,299 ***	-342,436 ***
2012	5,567 ***	4,502 ***	10,985 ***	39,762	0,139	3,834 ***	7,300 **	192,330 **
2013	8,577 ***	3,580 ***	4,504	30,048	4,007 ***	-0,543	-2,939	-56,372
Stima pooled <sup>(5)</sup>	5,713 ***	2,887 ***	8,011 ***	65,240 **	0,315	0,723 *	2,750	82,872

(1) La tavola riporta coefficienti e significatività della *dummy* relativa all'entrata nel campione rispetto all'uscita. - (2) Stime OLS separate (dipendenti: variabili indicate in colonna, espresse come indici percentuali), stimate solo sulle imprese entrate e uscite nell'anno (queste ultime, solo se mantengono un numero di addetti non inferiore a 20). - (3) \*\*\*: valore del p-value inferiore a 0,01, \*\*: valore del p-value compreso tra 0,01 e 0,05, \*: valore del p-value compreso tra 0,05 e 0,1. Errori standard stimati con la correzione di White per l'eteroschedasticità - (4) Covariate aggiuntive rispetto alla *dummy* di entrata: settori di attività economica, area geografica della sede principale, classe dimensionale. - (5) Errori standard calcolati considerando la stessa impresa ripetuta nel tempo come *cluster*. Covariata aggiuntiva rispetto alle stime annuali: *dummy* di anno.

**Anni di attività delle imprese per la popolazione di riferimento e stimati dall'indagine *Invind***

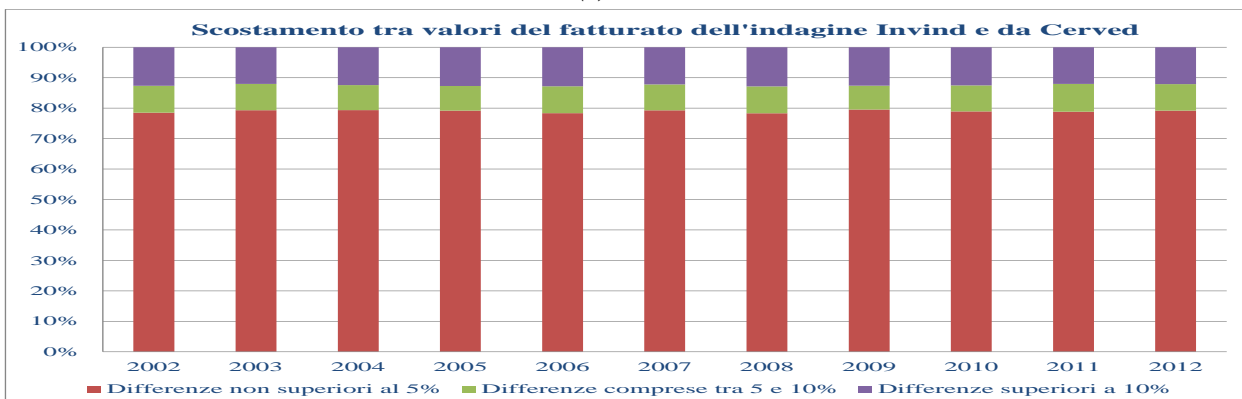
	Popolazione ricostruita		<i>Invind</i>	
	Media	Mediana	Media	Mediana
2002	13	9	25	22
2003	13	10	26	23
2004	13	9	26	23
2005	15	11	26	24
2006	20	17	27	24
2007	20	18	28	25
2008	20	18	28	25
2009	20	18	29	26
2010	21	18	30	27
2011	20	18	30	27
2012	21	19	30	28
2013	21	19	31	29
<b>Media generale</b>	<b>16</b>	<b>13</b>	<b>28</b>	<b>25</b>

**Figura 1 – Schema di costruzione dell'archivio integrato**



**Figura 2 – Scostamento tra dati rilevati in *Invind* e dati degli archivi *Cerved***

(a) *Fatturato*



(b) *Occupati medi*

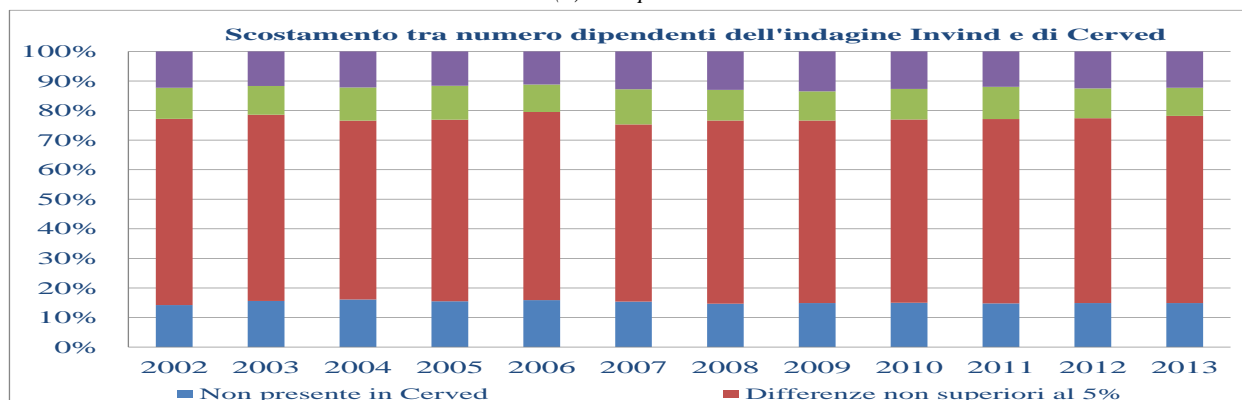
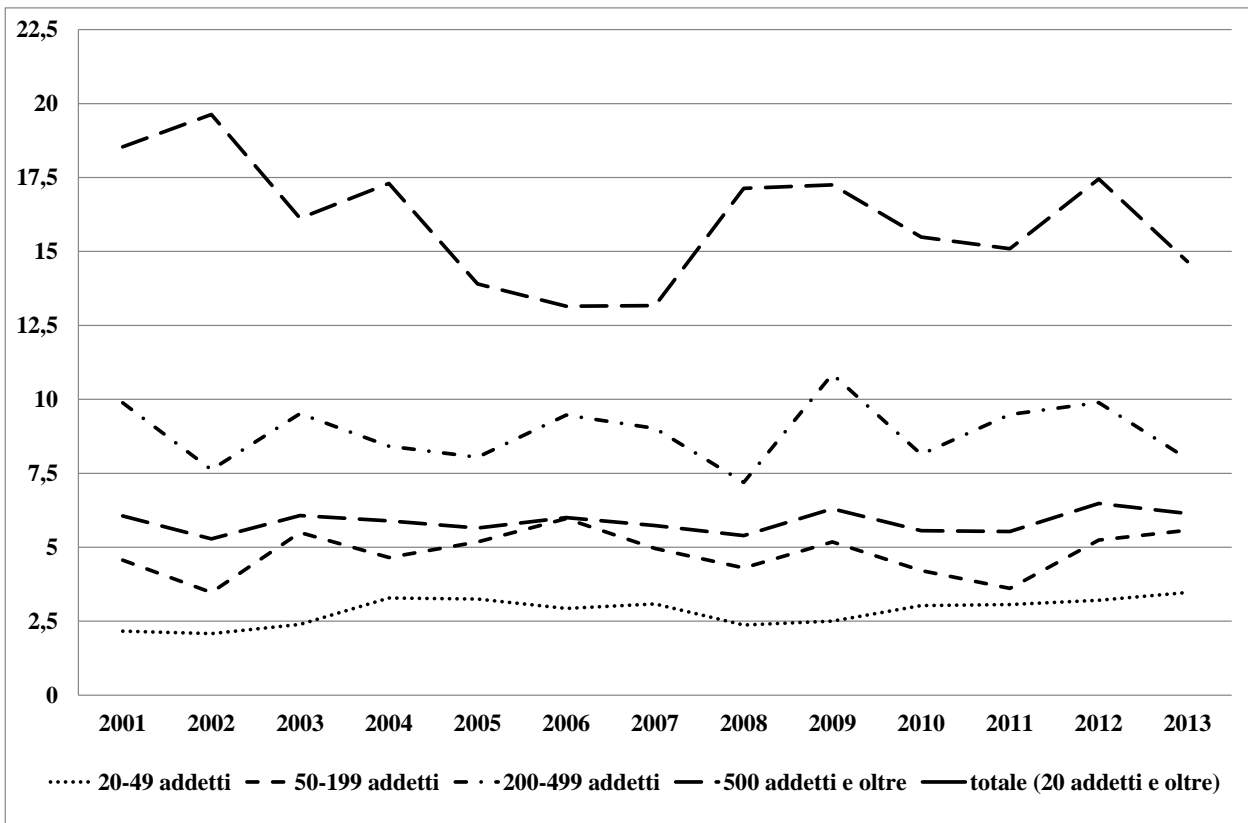


Figura 3 –Indagine *Invind*: quota di imprese con trasformazioni societarie

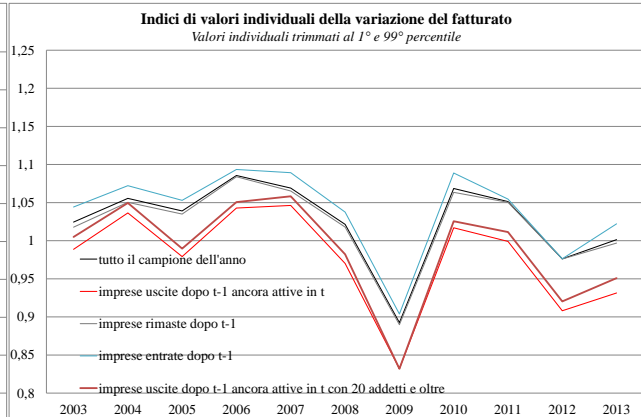
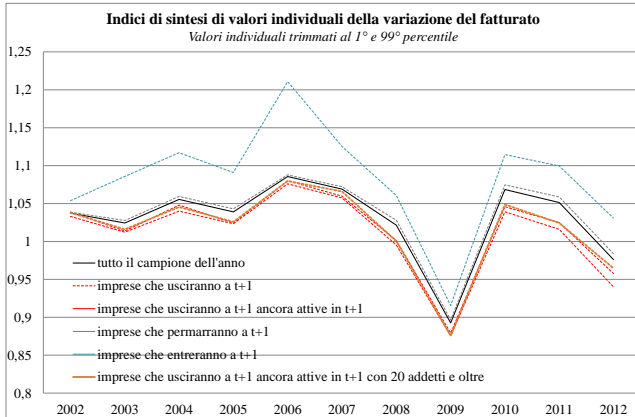


**Figura 4 – Indici semplici di variazione del fatturato e degli andamenti di ROA e ROE**

**Variazione annuale del fatturato**

**(a) Visione prospettica**

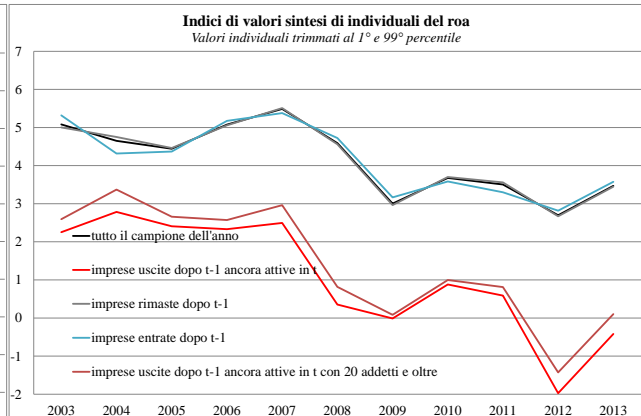
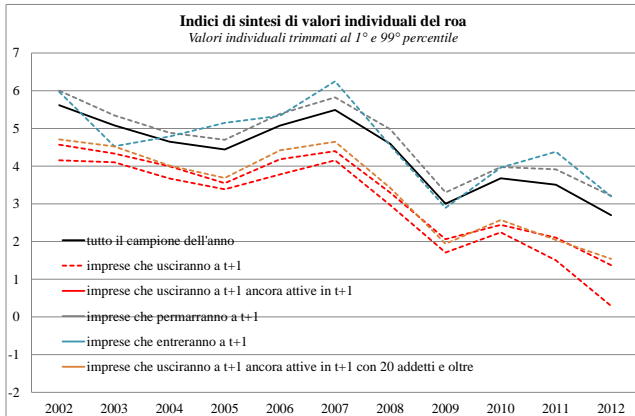
**(b) Visione retrospettiva**



**ROA**

**(a) Visione prospettica**

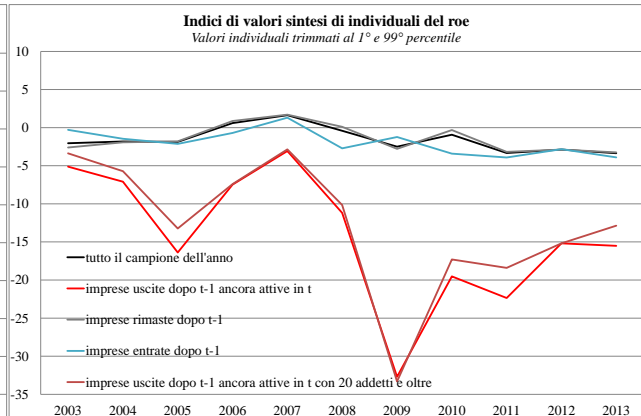
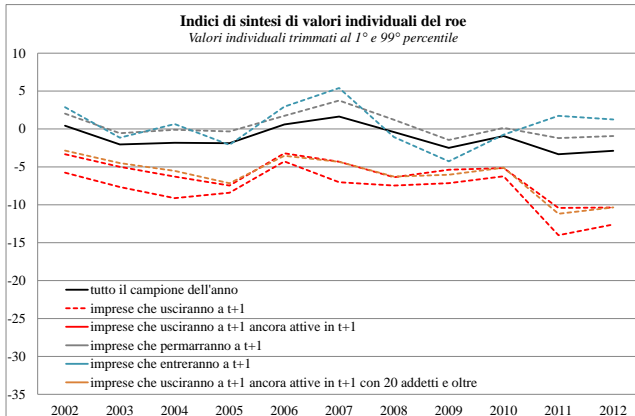
**(b) Visione retrospettiva**



**ROE**

**(a) Visione prospettica**

**(b) Visione retrospettiva**



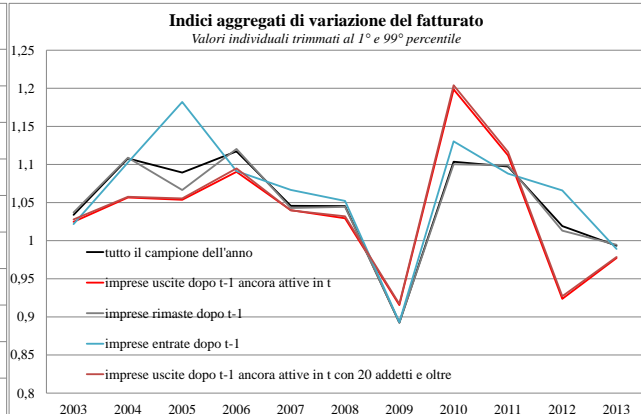
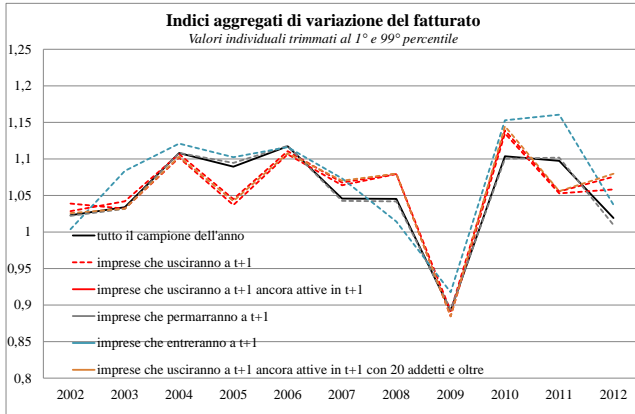


**Figura 5 – Indici aggregati di variazione del fatturato e degli andamenti di ROA e ROE**

**Variazione annuale del fatturato**

**(a) Visione prospettica**

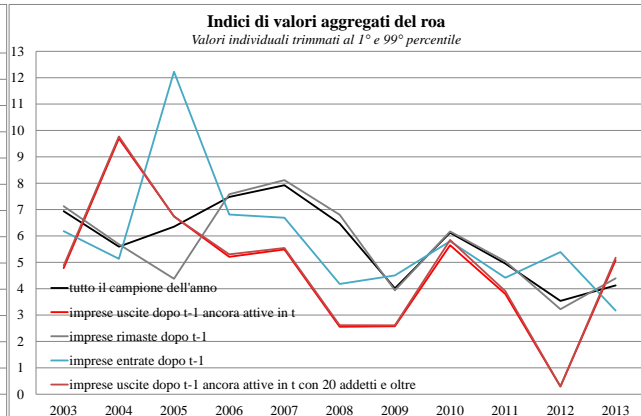
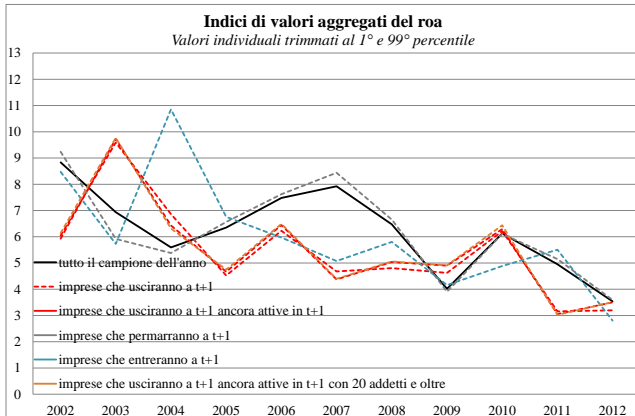
**(b) Visione retrospettiva**



**ROA**

**(a) Visione prospettica**

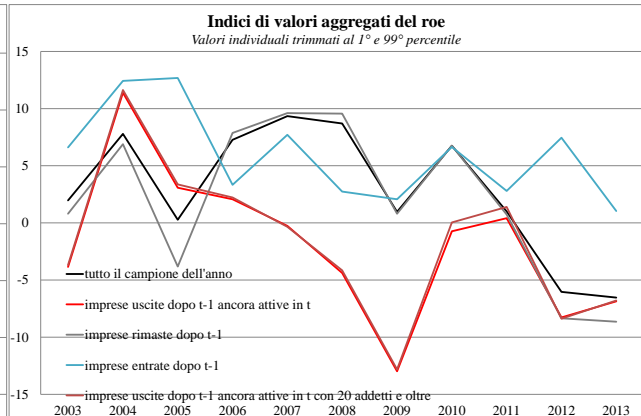
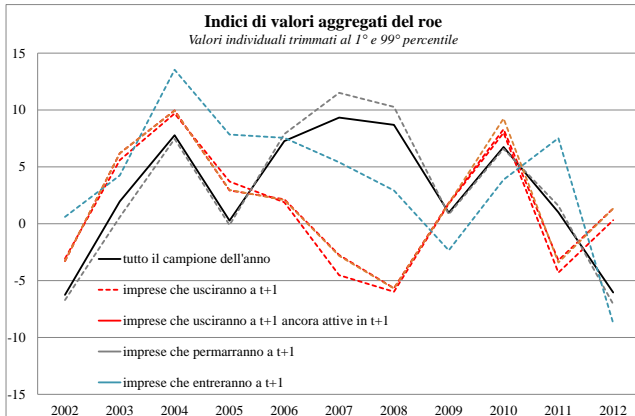
**(b) Visione retrospettiva**



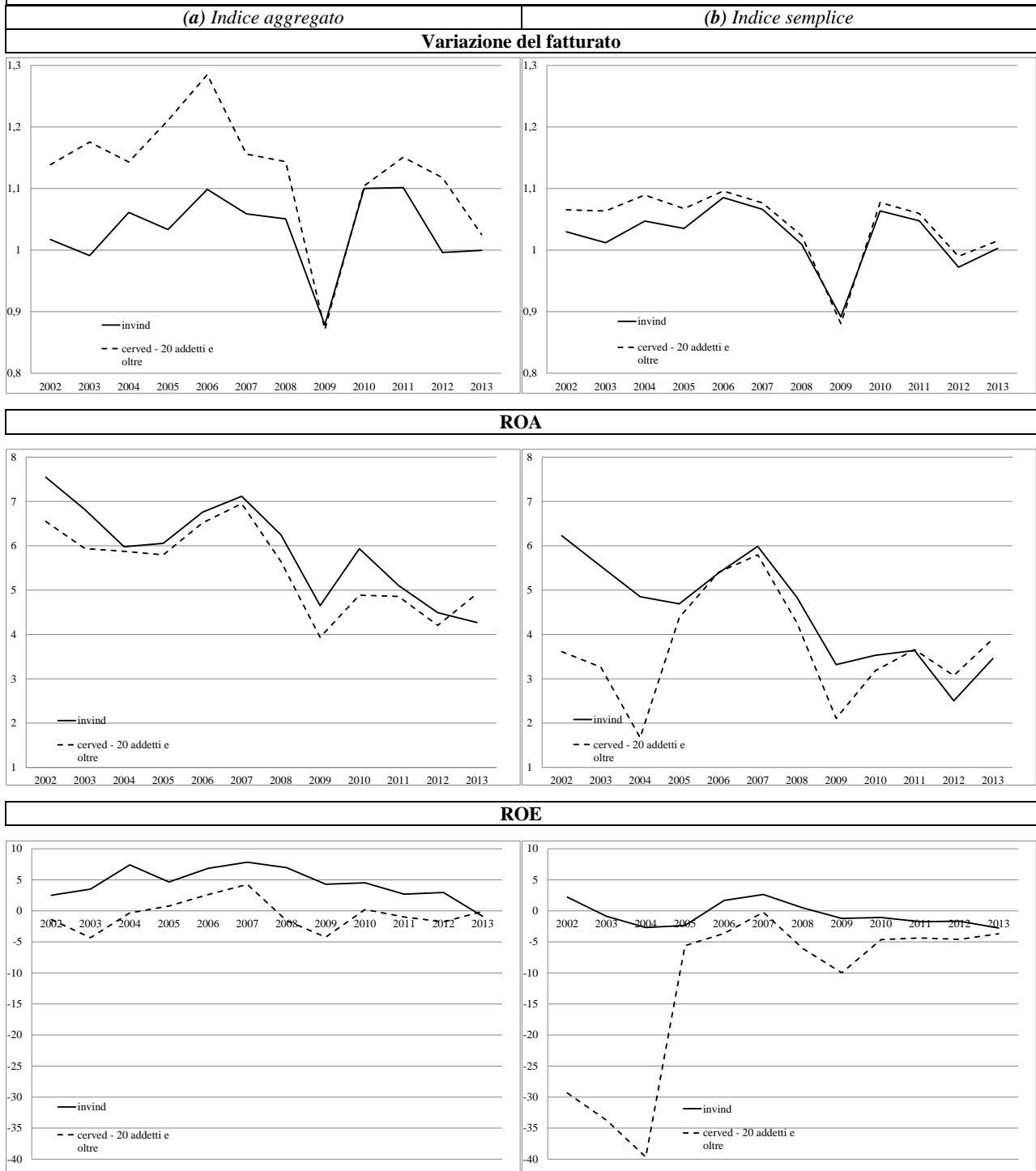
**ROE**

**(a) Visione prospettica**

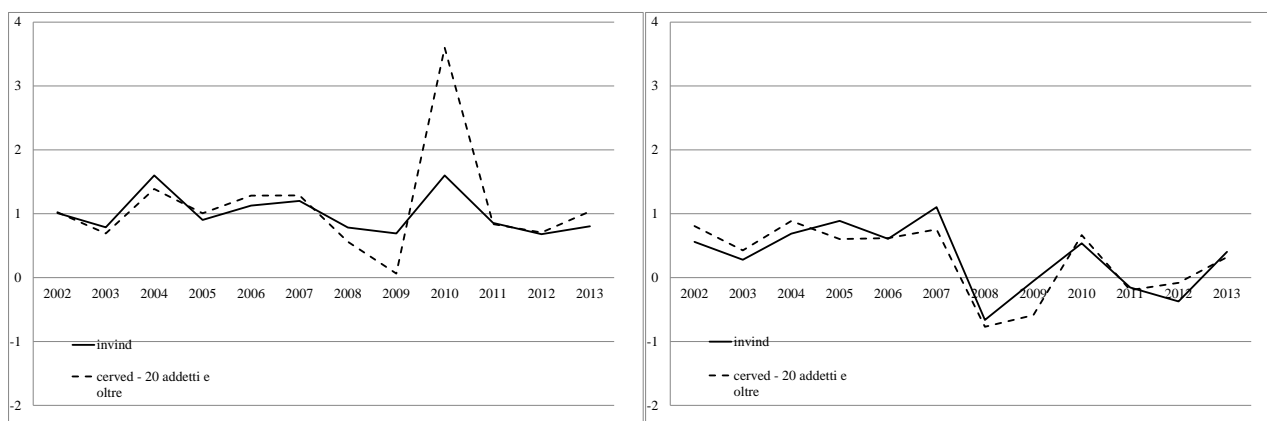
**(b) Visione retrospettiva**



**Figura 6 – Indici ponderati stimati in *Invind* e indici semplici stimati su *Cerved* a confronto**

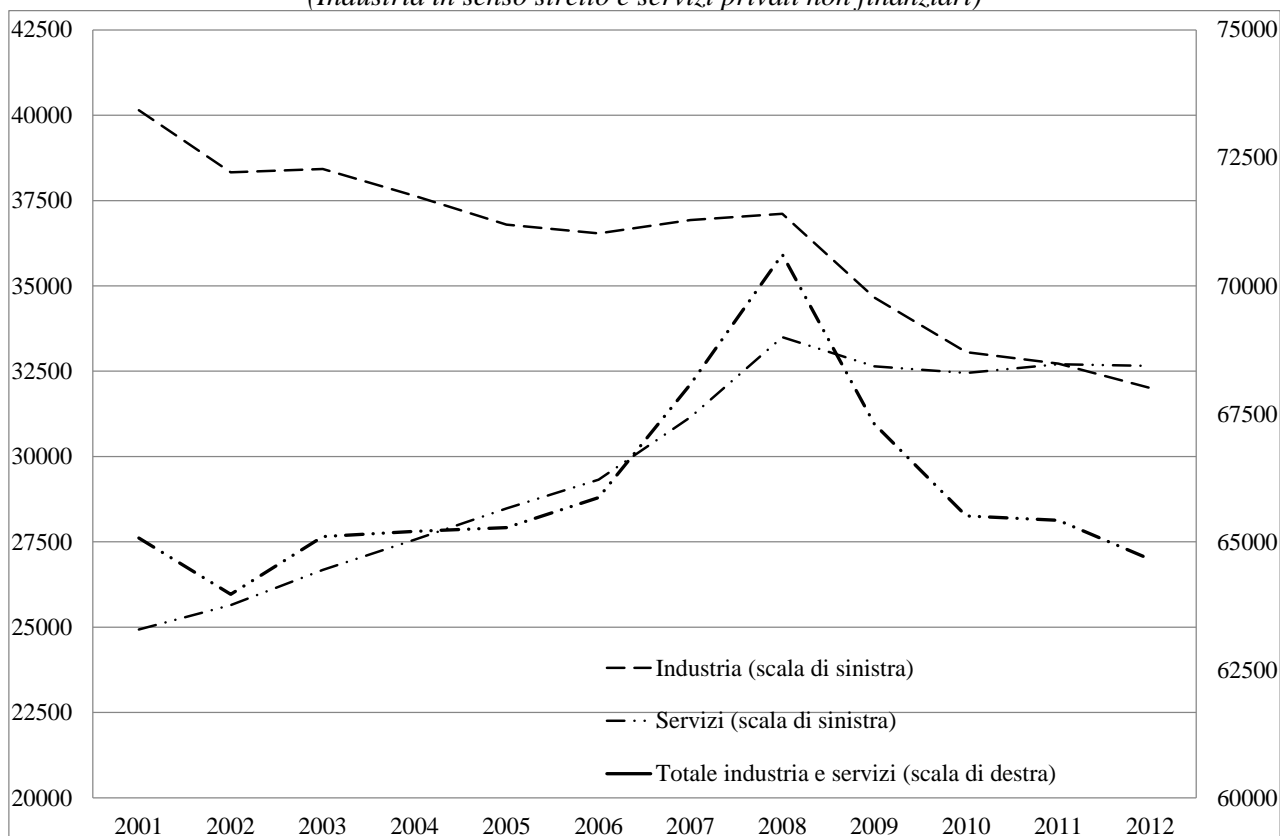


### Variazione dell'utile



(1) Indici ponderati per *Invid*. Fatturato a prezzi costanti 2013.

**Figura 7 – Imprese italiane con 20 addetti e oltre, 2001-2012<sup>(1)</sup>**  
*(Industria in senso stretto e servizi privati non finanziari)*

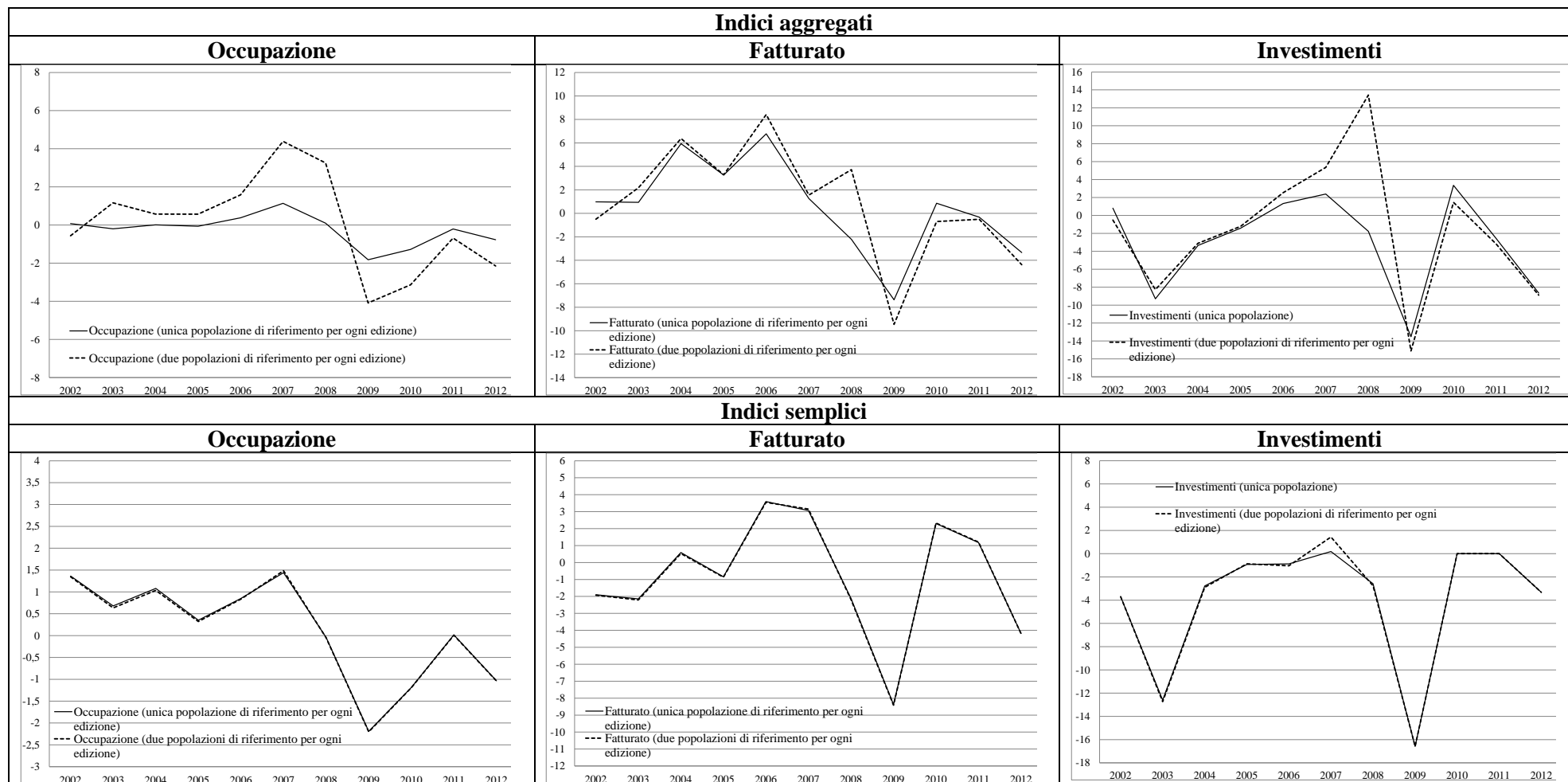


(1) Fonte: Istat, Archivio Statistico delle Imprese Attive.

**Figura 8 – Indici aggregati e semplici per la variazione di occupazione, fatturato e investimenti**

*(Invind, 2002-2012)<sup>(1)</sup>*

*(variazioni percentuali, valori individuali trimmati al 1° e 99° percentile)*



(1) Indici ponderati espressi come variazioni percentuali. Fatturato a prezzi costanti 2013.

## Appendice

### Raggruppamenti di attività economica e aree geografiche della sede principale dell'impresa utilizzati per la ricostruzione degli addetti dell'archivio integrato

<b>Raggruppamenti di attività economica</b>
Estrazione di minerali energetici
Estrazione di minerali non energetici
Industrie alimentari, delle bevande e del tabacco
Industrie tessili e dell'abbigliamento
Industrie conciarie, fabbricazione di prodotti in cuoio pelle e similari
Industria del legno e dei prodotti in legno
Fabbricazione della pasta-carta, della carta e del cartone, dei prodotti di carta; stampa ed editoria
Fabbricazione di coke, raffinerie di petrolio, trattamento dei combustibili nucleari
Fabbricazione di prodotti chimici e di fibre sintetiche e artificiali
Fabbricazione di articoli in gomma e materie plastiche
Fabbricazione di prodotti della lavorazione di minerali non metalliferi
Metallurgia, fabbricazione di prodotti in metallo
Fabbricazione di macchine ed apparecchi meccanici
Fabbricazione di macchine elettriche ed apparecchiature elettriche ed ottiche
Fabbricazione di mezzi di trasporto
Altre industrie manifatturiere
Produzione e distribuzione di energia elettrica, gas e acqua
Commercio all'ingrosso e al dettaglio; riparazione di autoveicoli, motocicli e di beni personali e per la casa
Alberghi e ristoranti
Trasporti, magazzinaggio e comunicazioni
Attività immobiliari, noleggio, informatica, ricerca, altre attività professionali e imprenditoriali
<b>Localizzazione della sede principale dell'impresa</b>
Nord ovest
Nord est
Centro
Sud e isole