

Conti economici delle imprese – Anno 2004

Indice

1 - Aspetti metodologici della rilevazione sulle imprese con 100 addetti ed oltre	Pag. 1
2 - Aspetti metodologici della rilevazione sulle imprese con 1-99 addetti	" 23
3 - Calcolo e presentazione sintetica degli errori campionari	" 44
Avvertenze	" 63
Bibliografia	" 64
Glossario	" 66

1 - Aspetti metodologici della rilevazione sulle imprese con 100 addetti ed oltre ^(*)

1.1 - Caratteristiche della rilevazione

L'indagine Sci (Rilevazione sul Sistema dei conti delle imprese) si rivolge a tutte le imprese italiane con almeno 100 addetti che operano nei settori industriali e dei servizi, con l'esclusione di alcune divisioni dell'intermediazione monetaria e finanziaria, delle assicurazioni e dei servizi domestici. Al di sotto della soglia dei 100 addetti viene estratto un campione di piccole e medie imprese alle quali viene inviato un questionario ridotto rispetto al questionario Sci.

La lista di partenza dell'indagine Sci viene estratta dall'archivio statistico delle imprese attive dell'Istat (Asia). I dati vengono raccolti sia tramite un questionario inviato a mezzo posta, sia tramite questionario elettronico. Successivamente, sulla base del monitoraggio dei rispondenti, si effettuano solleciti finalizzati alla riduzione del numero di mancate risposte. I dati raccolti si riferiscono sia all'impresa, classificata secondo l'attività economica prevalente, sia alle unità funzionali (unità di produzione omogenea) dell'impresa stessa, onde poter fornire dati omogenei per settore di attività economica.

Il questionario è stato predisposto per la rilevazione dei dati economico-finanziari e patrimoniali delle imprese. Dal 1996 le voci di bilancio vengono richieste secondo quanto stabilito dalla IV Direttiva Cee sui bilanci. Alcune voci del questionario sono state esposte e strutturate allo scopo di ottenere uno schema di bilancio riclassificato secondo modalità atte a permettere una corretta analisi economica e dei principali aspetti della gestione aziendale.

L'indagine ha anche lo scopo di accertare le voci dei ricavi e dei costi delle imprese per la determinazione del valore aggiunto e di altri aggregati da utilizzare per le valutazioni dei conti economici nazionali e della tavola intersettoriale dell'economia italiana.

1.2 - Unità di osservazione

La principale unità di rilevazione è costituita dall'impresa, alla quale si affianca l'unità funzionale¹. Questa costituisce un'unità di rilevazione aggiuntiva solo per le imprese con dimensione pari o superiore a 200 addetti e attive in più categorie di attività economica. A queste imprese viene inviato insieme al questionario generale anche il Mod. Sci-Uf, finalizzato alla rilevazione dei dati relativi alle varie categorie di attività economica esercitate dall'impresa in aggiunta all'attività principale in cui si richiedono i dati per divisione di azienda o unità funzionale (Uf).

Questa articolazione dei modelli di rilevazione consente, sia pure entro certi limiti, di disaggregare alcuni dei risultati principali per attività economiche omogenee. Infatti i dati vengono elaborati e presentati sia per aggregazioni di imprese sia per aggregazioni di Uf. Quest'ultima unità è derivata dall'impresa, separandone i principali aggregati economici sulla base delle diverse linee di produzione, identificate dalle categorie della classificazione Ateco2002 (codici a 4 cifre). I dati per impresa, allocati nel settore di produzione prevalente, misurano quindi la dimensione economica e la *performance* delle imprese, ma non consentono di misurare con precisione la dimensione dei diversi settori di attività economica. Questi possono essere definiti in modo più preciso attraverso i dati delle unità funzionali, seppure per un insieme ridotto di indicatori.

(*) A cura di Alessandro Zeli, responsabile della rilevazione e coordinatore delle diverse fasi del processo di produzione, validazione e diffusione dei risultati. Roberto Nardecchia ha curato la stesura dei paragrafi 1.5.5 e 1.7. Patrizia Perini e Leopoldo Nascia hanno coordinato la fase di correzione e revisione dei modelli. Rossana Albertini, Andreina Cifelli, Diego Distefano, Anna Marcone, Manuela Silvi e Paola Urilli hanno curato la fase di acquisizione e revisione dei modelli. Adele Borin e Roberto Nardecchia hanno curato la fase di integrazione delle mancate risposte. La predisposizione delle tavole è stata curata da Luciano Fanfoni e Tiziana Di Francescantonio.

¹ Si veda la definizione nel Glossario.

Si deve poi tener conto che alle imprese plurilocalizzate vengono richiesti alcuni dati a livello territoriale (dipendenti, spese per il personale e investimenti) in maniera da consentire la disaggregazione regionale dei correlati aggregati.

Il valore aggiunto regionale, che difficilmente le imprese con stabilimenti in più regioni potrebbero determinare partendo dai dati della loro contabilità generale, viene stimato nel corso del processo di elaborazione in base alla ripartizione del valore aggiunto, di ciascuna impresa, proporzionalmente alla distribuzione regionale delle spese di personale da essa stessa indicata.

1.3 - Campo di osservazione

Le rilevazioni relative all'anno 2004 sono state, pertanto, condotte seguendo la versione italiana della classificazione Nace (Ateco 2002)². La rilevazione sul sistema dei conti delle imprese interessa le imprese con 100 addetti ed oltre, esercitanti un'attività industriale o terziaria nei seguenti settori: estrattivo (sezione C della classificazione delle attività economiche Ateco 2002); manifatturiero (sezione D); produzione e distribuzione di energia elettrica, gas e acqua (sezione E); costruzioni (sezione F); commercio e riparazione (sezione G); alberghiero e della ristorazione (sezione H); trasporti e telecomunicazioni (sezione I); immobiliare, di noleggio, informatica, professionale ed imprenditoriale (sezione K); istruzione (sezione M); sanità e servizi sociali (sezione N) ed altri servizi pubblici, sociali e personali (sezione O, con l'esclusione della divisione 91).

Vengono, inoltre, rilevati alcuni comparti della sezione J (Intermediari monetari e finanziari), tuttavia la diffusione dei risultati dell'indagine per questi ed altri comparti della sezione J sarà oggetto di una diversa pubblicazione.

1.4 - Definizione degli aggregati e stima del valore aggiunto

Con l'indagine Sci vengono rilevate le principali variabili economiche occorrenti al calcolo del valore aggiunto e di altri aggregati che, pur non essendo direttamente legati alla sua determinazione, assumono un'importanza fondamentale per l'analisi economica dei settori industriali, commerciali e dei servizi.

Pertanto, ai fini di una interpretazione corretta dei risultati dell'indagine, si rinvia al Glossario per le definizioni delle variabili e al prospetto 1.1 per la schematizzazione delle definizioni di fatturato, valore aggiunto e valore della produzione.

Prospetto 1.1 - Schema riassuntivo delle definizioni di fatturato, valore aggiunto e valore della produzione

FATTURATO	VALORE AGGIUNTO	VALORE DELLA PRODUZIONE
Ricavi da:	Fatturato	Fatturato
Vendite di prodotti fabbricati dall'impresa	+/- variazioni delle scorte di beni e servizi	+/- variazione delle scorte di prodotti finiti e in lavorazione
Vendite di merci acquistate in nome proprio e rivendute senza trasformazione	+ produzione capitalizzata	+/- variazione delle scorte di beni da rivendere senza trasformazione
Lavorazioni in conto terzi su materie prime e semilavorati i terzi	+ Altri redditi operativi (con esclusione dei sussidi)	- acquisti di beni e servizi da rivendere senza trasformazione
Attività di intermediazione	- Acquisti	+ produzione capitalizzata
Introiti lordi del traffico (per le imprese di trasporto)	- Imposte sui prodotti e sulla produzione	+ altri redditi operativi (con esclusione dei sussidi)
Prestazioni di servizi a terzi		

Coerentemente alla definizione dettata dal regolamento sulle statistiche strutturali sulle imprese, il valore aggiunto è calcolato sottraendo alla somma algebrica dei ricavi per la vendita di beni e servizi, della variazione delle rimanenze di prodotti finiti ed in corso di lavorazione e dei semilavorati, della variazione dei lavori in corso di ordinazione, dell'incremento di immobilizzazioni per lavori interni e di altri ricavi e proventi non

² Istat. *Classificazione delle attività economiche Ateco 2002 (derivata dalla Nace rev 1.1)*. Roma: Istat, 2003. (Metodi e norme, n. 18).

finanziari e non straordinari, la somma algebrica dei costi per l'acquisto di materie prime e sussidiarie, di prodotti energetici e di merci da rivendere senza trasformazione, dei costi per l'acquisto di servizi (lavorazioni affidate a terzi, trasporto, intermediazione, pubblicità, studi e ricerche, consulenze, informatica, premi di assicurazione contro i danni, canoni per licenze d'uso di *royalties* e brevetti, smaltimento rifiuti, bancari ed altri), dei costi per il godimento di beni di terzi (fitti passivi su fabbricati strumentali, quote di leasing pagate nell'esercizio, canoni di locazione per beni strumentali diversi dagli immobili), dei costi di formazione del personale, degli altri oneri di gestione non finanziari e non straordinari, delle imposte indirette sui prodotti e sulla produzione e della variazione delle rimanenze di materie prime, sussidiarie, di consumo e di merci da rivendere senza trasformazione.

1.5 - Preparazione, invio, raccolta e revisione dei questionari

1.5.1 - Preparazione dei questionari

La fase di preparazione dei questionari consiste nelle seguenti operazioni: controllo delle voci di bilancio inserite, definizione dei dettagli richiesti rispetto al modello base (IV Direttiva), eventuali aggiunte o cancellazioni di voci richieste, aggiornamento delle definizioni e spiegazione delle stesse contenute nel libretto delle istruzioni (Guida alla compilazione), allestimento dei questionari di rilevazione elencati di seguito:

- a) questionario Sci;
- b) questionario Sci-Uf (per la rilevazione delle variabili economiche necessarie alla ricostruzione del valore aggiunto, degli investimenti e dei valori regionali per le divisioni di imprese o principali attività secondarie);

La rilevazione è effettuata tramite invio (postale ed elettronico) alle imprese di un questionario riportante le varie sezioni del bilancio, come da indicazioni contenute nella IV Direttiva della Comunità Europea in materia di bilanci e recepite nel nostro ordinamento con la legge n. 69 del 26 marzo del 1990 e con il decreto legislativo n. 127 del 9 aprile 1991.

Nel 2002 è stata effettuata una revisione del questionario al fine di diminuire l'onere statistico per le imprese. In particolare si è operata una riduzione delle informazioni richieste ricavabili da fonti amministrative come la sezione relativa allo stato patrimoniale. Questa sezione è stata quella che ha subito i tagli maggiori con circa 70 voci eliminate; le informazioni che vengono ancora richieste nell'ambito dello stato patrimoniale sono quelle relative alle immobilizzazioni e ad una loro maggior disaggregazione rispetto a quanto richiesto dagli usuali schemi di bilancio. Oltre allo stato patrimoniale sono state eliminate o ridotte le sezioni C, D ed E del conto economico.

Sono stati aggiunti quesiti riguardanti i lavoratori 'atipici' che devono essere classificati come interinali, collaboratori coordinati e continuativi o altre tipologie. Per gli 'atipici' vengono richieste anche informazioni su ore lavorate, sesso e spesa sostenuta. Sono state introdotte domande sulla scomposizione secondo il sesso di alcune voci riguardanti le retribuzioni e il costo del lavoro complessivo.

Il questionario si compone di 8 sezioni. La prima contiene tutte le voci relative al conto economico, la seconda registra alcune voci dello stato patrimoniale, mentre la terza e la quarta si riferiscono agli aspetti occupazionali; in particolare, la terza sezione richiede informazioni circa l'occupazione totale e gli addetti suddivisi per qualifica professionale e sesso, la quarta sezione riguarda i costi sostenuti per il personale. La quinta sezione registra l'acquisizione dei capitali fissi effettuata nell'esercizio, mentre la sesta riguarda un insieme di voci che non è possibile inserire nello schema di bilancio precedentemente descritto. Nella settima sezione si richiedono alcuni dati disaggregati per regione.

La sezione 8 è utilizzata a partire dall'anno di riferimento 2002 per richiedere informazioni con l'obiettivo di comprendere il comportamento complessivo delle imprese al di là dei meri risultati economici (indagine multiscopo). In particolare sono state aggiunte 3 sottosezioni riguardanti: la delocalizzazione ed esternalizzazione delle attività produttive, il comportamento 'sociale' delle imprese, le relazioni internazionali delle imprese (legami *outwards*). Oltre all'introduzione di queste nuove sottosezioni nel questionario è proseguita la rilevazione della spesa per la protezione dell'ambiente iniziata nell'anno di riferimento 2001 (tramite un questionario *ad hoc*: Sci.Amb). I dati rilevati nella sezione 8 sono oggetto di pubblicazioni specifiche e non sono, pertanto, inseriti nella presente pubblicazione.

1.5.2 - Preparazione della lista di partenza

Poiché la rilevazione è censuaria e la lista di partenza comprende tutte le imprese che, secondo l'archivio Asia disponibile (relativo all'anno t-1 rispetto all'anno di riferimento) al momento del lancio della rilevazione, sono comprese nell'universo di riferimento.

Al momento dell'estrazione della lista di partenza vengono imposte le seguenti condizioni:

1. l'impresa deve impiegare almeno 100 addetti in media annua;
2. l'impresa deve svolgere una delle attività contenute nel campo di osservazione dell'indagine;
3. l'impresa deve avere un periodo di attività nell'anno t-1 non inferiore ai 6 mesi;
4. l'impresa deve avere una ragione sociale coerente con le definizioni accettate a livello nazionale e comunitario (non deve essere cioè un ente pubblico, un'istituzione privata o non profit).

Una volta predisposta la lista di partenza, alle imprese con più unità funzionali (imprese multifunzionali) viene associato un codice identificativo caratterizzante. La lista delle imprese multifunzionali è predisposta sulla base delle variazioni rispetto all'anno precedente comunicate dalle imprese multifunzionali. Per tale operazione si fa riferimento ai seguenti criteri metodologici:

- individuazione dell'insieme delle imprese con più di 200 addetti e con unità locali come registrato nel file Asia unità locali a sua volta creato a partire dai dati raccolti dal Censimento intermedio dell'industria e dei servizi (Ciis);
- all'interno dell'insieme di cui sopra, individuazione del sottoinsieme di imprese aventi unità locali con codice Ateco differente dall'impresa madre a livello della classe di attività economica (4^a cifra Ateco);
- creazione di un'unità funzionale per ognuna delle Ateco (differenti a livello di 4 cifre, classe);
- verifica della presenza di imprese multifunzionali monolocalizzate tramite un controllo con l'indagine Prodcum;
- aggiornamento annuale della lista successivamente a comunicazioni delle imprese;
- verifica che la singola Uf rappresenti almeno il 5 per cento del fatturato dell'impresa madre (altrimenti viene considerata non rilevante);
- aggiornamento della lista tramite reiterazione del processo descritto ad ogni censimento o aggiornamento dell'archivio Asia unità locali.

Una volta creata la lista di partenza, anche per le imprese multifunzionali, vengono scorporate le liste relative alle imprese localizzate nelle provincia autonome di Trento e Bolzano. Tali liste sono inviate ai competenti istituti statistici provinciali che provvedono all'invio, alla raccolta e alla correzione dei questionari e alla transazione dei risultati all'Istat.

1.5.3 - Invio raccolta, registrazione del pervenuto e pre-check

Le bozze dei questionari in formato Pdf vengono inviate alla tipografia assieme ai file di personalizzazione contenenti gli indirizzi e i dati anagrafici. La tipografia provvede alla stampa dei questionari, alla loro personalizzazione e alla spedizione postale.

I questionari compilati che ritornano all'Istat vengono registrati manualmente nel sistema di gestione del pervenuto e dei solleciti con codici di risposta opportuni. Essi vengono smistati, seguendo le attività economiche di competenza, ai vari revisori che provvedono alla codifica di alcuni valori (per esempio dei codici regionali) e alla verifica di alcune incoerenze (ad esempio l'unità di misura, la presenza delle principali voci di bilancio: fatturato, valore della produzione, costi). Conclusa questa operazione di *pre-check* i questionari vengono inviati in registrazione esterna non controllata.

Ai fini dell'assistenza alla compilazione del questionario la struttura operativa dell'Istat preposta alla rilevazione effettua un servizio telefonico di *help-desk*.

I questionari pervenuti vengono classificati e preregistrati. In particolare, vengono evidenziate situazioni particolari come i casi di imprese che hanno cessato l'attività, di imprese in corso di liquidazione, di imprese

fuori campo di osservazione, di imprese trasferite, di imprese regolarmente rispondenti. La fase di preregistrazione degli arrivi è funzionale alla gestione di un archivio con i riferimenti delle imprese rispondenti da utilizzarsi nella gestione delle fasi di sollecito

Le imprese vengono sollecitate più volte a fornire le informazioni richieste. Si effettuano, infatti, due solleciti, di cui il secondo con rinvio in allegato del modello di rilevazione.

Una volta completata la fase di sollecito vengono ulteriormente contattate telefonicamente le imprese più grandi al fine di disporre di un quadro completo almeno per le imprese maggiori e per quelle con più unità funzionali.

1.5.4 - Preparazione del questionario elettronico

Dall'anno di riferimento 2003 nella rilevazione, in parallelo con l'invio del questionario cartaceo, è stato introdotto l'utilizzo del questionario elettronico. L'opportunità di compilare la versione elettronica del questionario in sostituzione di quella cartacea è stata offerta a tutte le imprese.

L'istituto, con una comunicazione a mezzo della posta ordinaria, ha inviato a tutte le imprese della rilevazione la presentazione della versione elettronica e le istruzioni per scaricare il questionario personalizzato ed effettuare il suo invio elettronico.

L'introduzione del questionario elettronico ha consentito il miglioramento della qualità dei questionari e ha implicato una modifica delle procedure di acquisizione e di registrazione dei questionari come illustrato nei paragrafi successivi. Per l'indagine 2004 il questionario elettronico è stato utilizzato per la seconda volta e il gradimento delle imprese è stato particolarmente elevato, infatti, al termine della rilevazione, 2.732 imprese hanno mostrato una preferenza verso il suo utilizzo rappresentando una percentuale pari al 59,7 per cento del totale rispondenti.

1.5.4.1 Questionario elettronico: fasi di allestimento, spedizione, raccolta

1) Preparazione questionari e istruzioni

La preparazione del questionario elettronico utilizza la procedura di preparazione della versione cartacea a cui si rinvia (Paragrafo 1.5.1)

2) Allestimento dei questionari in formato Excel:

L'allestimento prevede la preparazione, da parte dell'unità operativa Ssi/B, del questionario in formato Excel. Successivamente il questionario Excel viene predisposto per il *data capturing* automatico dei dati inseriti dalle imprese e vengono inseriti diversi controlli automatici formali e di congruenza nei dati.

L'unità Ssi/B predispose il piano delle specifiche di registrazione e il piano di controllo preliminare automatico dei dati pervenuti.

Il formato elettronico prevede la versione excel dei seguenti questionari e delle istruzioni:

- a. questionario Sci;
- b. questionario Sci.Amb (quesiti sulla spesa per la protezione dell'ambiente);
- c. questionario Sci.Uf (per la rilevazione della variabili economiche necessarie alla ricostruzione del valore aggiunto, degli investimenti e dei valori regionali per le divisioni di imprese o principali attività secondarie);
- d. libretto per le istruzioni.

3) Preparazione della lista di partenza.

Per realizzare la lista di partenza dei questionari elettronici è stata utilizzata la lista dei questionari cartacei a cui si rinvia (Paragrafo 1.5.2).

4) Web e spedizione.

Con la lista di spedizione, viene creato un elenco di *password* per consentire l'accesso, esclusivamente al rispondente, alla sezione del sito in cui scaricare e caricare i questionari da parte delle imprese.

Infatti la modalità di compilazione telematica prevede l'invio di una lettera informativa alle imprese con una *password* provvisoria e con le istruzioni di accesso al sito *Indata* dell'Istat dove le imprese possono scaricare il questionario in formato Excel.

Il formato Excel adottato contiene alcuni vincoli e calcoli reimpostati finalizzati ad una compilazione controllata da parte dell'impresa al fine di eliminare all'origine errori dovuti alla mancata quadratura dei dati o errori dovuti all'unità di misura.

Dopo la compilazione il questionario deve essere caricato sul sito *Indata* dell'Istat, per essere, in seguito, inviato al *server* del servizio Ssi.

Assieme alla procedura di invio viene effettuata anche la predisposizione dei programmi per la gestione del *data-capturing* e del *pre-check*.

5) Solleciti.

Le imprese vengono sollecitate con le stesse procedure del questionario cartaceo poiché non sono previste procedure di sollecito elettronico tramite l'invio di e-mail.

6) Registrazione del pervenuto e *pre-check*.

I questionari acquisiti dal sito *Indata*, dopo essere stati inviati al server del servizio Ssi, vengono controllati da un programma di *pre-check* che crea un record per ogni questionario e assegna un codice di qualità secondo criteri predefiniti. Il codice di qualità ricalca la codifica delle varie tipologie di risposta predisposta a livello centralizzato per tutto il servizio Ssi (tale codifica permette, ad esempio, di individuare i codici doppi, i questionari incompleti, le ditte cessate, fallite, ecc.).

Con cadenza settimanale l'unità Ssi/B crea un file con i codici identificativi delle imprese e i codici di risposta che immesso nel programma di gestione dei solleciti consente il monitoraggio delle risposte (curva dei ritorni) e la gestione dei solleciti.

Se il questionario supera tutti i controlli del programma di *pre-check* i dati relativi al questionario Sci, con esclusione della sezione anagrafica, della sezione multiscopo e della sezione ambiente vengono caricati nel database Oracle dell'indagine da dove viene effettuata (tramite il programma di *check*) la revisione.

Dopo il *pre-check* la procedura crea altri 5 file:

- un file con i dati relativi alla parte anagrafica per l'aggiornamento dell'archivio Asia;
- un file con i dati dell'indagine sull'ambiente;
- tre file con i dati della parte multiscopo che vengono inviati alle rispettive indagini di riferimento.

Se un questionario non viene classificato automaticamente dal programma di *pre-check*, questo rimane sul server di acquisizione e viene analizzato con un controllo manuale.

7) Al termine della revisione tutti i questionari dell'indagine Sci, compresi quelli relativi alle unità funzionali, vengono caricati nella base dati di lavoro (che include anche i record relativi ai questionari cartacei revisionati).

1.5.5 - Risultati della raccolta per l'anno di riferimento 2004

Per la rilevazione relativa all'anno 2004 sono state interpellate 9.923 imprese alle quali sono stati inviati direttamente per posta i modelli sopra descritti. Sono giunti all'Istat circa 5.360 questionari (con un tasso di risposta pari al 54,1 per cento). Circa 4.560 sono stati classificati come corretti dopo accurati controlli strutturali e dinamici volti ad accertare la qualità dei singoli dati. Circa 2.730 questionari validi (pari al 60 per cento del totale) sono pervenuti in formato elettronico.

Per risolvere le mancate risposte si è proceduto ad una integrazione delle variabili d'interesse attribuendo a ciascuna impresa non rispondente i valori *pro capite* (calcolati sugli addetti) rilevati in una impresa delle stesse dimensioni, operante nella stessa classe di attività economica e nella stessa regione (Paragrafo 1.6.3). Le percentuali di copertura dell'indagine in termini di imprese sono riportate, per ciascuna sezione, sottosezione, divisione e gruppi di attività economica nella tavola A1. In essa vengono presentati, con la stessa

disaggregazione, il peso degli addetti e del valore aggiunto rilevati rispetto a quelli stimati per l'universo delle imprese. I dati sono presentati con una disaggregazione che arriva fino al livello di gruppo (3 cifre Ateco2002).

Dalla tavola 2.1 si evince che le percentuali di copertura si collocano, in media, intorno al 46 per cento; in ogni caso, in termini di addetti e valore aggiunto aziendale, le percentuali di copertura raggiungono valori maggiori. Le percentuali di copertura della rilevazione (in termini di addetti) sono in relazione con la maggiore o minore concentrazione del settore; questa relazione è verificata in misura più ampia nei settori industriali, dove le maggiori coperture si registrano nel settore dell'estrazione dei minerali energetici dove la copertura è del 99 per cento in termini di addetti, nelle telecomunicazioni (97,8 per cento), nella raffinazione (84,7 per cento) e nel settore della costruzione di altri mezzi trasporto (72,6 per cento). La percentuale di copertura diminuisce fortemente in altri settori, come ad esempio l'industria della produzione di apparecchi radiotelevisivi (50,3 per cento) o della produzione di prodotti in metallo, che registrano un valore pari al 51,4 per cento. Nel settore dei trasporti si registrano percentuali di copertura oltre l'80 per cento: 79,0 per cento nei trasporti terrestri e 75,3 per cento nei trasporti aerei. Da notare il basso tasso di risposta nel settore della attività immobiliari (27,3 per cento).

1.5.6 - Procedure di revisione dei questionari delle imprese e delle unità funzionali

I questionari compilati in maniera valida, ossia riportanti almeno i dati relativi a spese e ricavi, vengono sottoposti ad una fase di correzione preliminare, volta a determinare l'esistenza delle caratteristiche fondamentali del modello, quali la presenza dell'attività economica svolta dall'impresa, la localizzazione ecc.

I *record* preparati dalla registrazione esterna vengono caricati nel database *relazionale* dell'indagine grazie al quale i revisori possono (tramite il programma di *check*) accedere ai dati e procedere alla revisione dei singoli questionari.

Nella prima fase della revisione i questionari vengono sottoposti ad un programma di controllo e quadratura automatica con il quale si individuano le correzioni da apportare ai modelli. Il programma, tramite una successione di forzature, riesce a quadrare e validare una parte dei questionari. La revisione dei questionari per i quali si richiede una conoscenza più approfondita dei fondamenti analitici del bilancio viene demandata alla verifica puntuale di revisori esperti specializzati in uno o più settori di attività economica.

Ove si verificano mancate risposte parziali, si ricorre all'imputazione analitica e statistica del dato mancante: in base all'insieme delle restanti risposte valide e tramite le relazioni di bilancio esistenti tra le variabili del questionario si ricostruiscono, per ciascuna impresa, le voci non compilate. L'imputazione analitica è effettuata in maniera sia automatica sia manuale.

Completata questa parte della revisione, vengono corretti e validati tutti i questionari che presentano valori che comportano problemi nella fase di aggregazione a livello di strato di pubblicazione, interpellando, se del caso, nuovamente le imprese. Questa fase di correzione si avvale di ulteriori controlli automatici per l'individuazione di valori fuori scala e per una verifica di congruenza dinamica con i dati d'impresa che si riferiscono agli anni precedenti. La correzione dei questionari relativi alle Uf viene effettuata al termine della revisione di tutti i questionari delle imprese multifunzionali. La correzione delle Uf necessita di particolare attenzione poiché occorre quadrare i dati all'interno di ogni singolo questionario Uf e rendere coerente la somma di tutti i questionari Uf con il totale di impresa. Anche per la correzione delle Uf i revisori si avvalgono dell'aiuto di un programma di controllo e correzione. Una volta validati i questionari, si passa alla integrazione delle mancate risposte totali, alla cui descrizione è dedicato il paragrafo 1.6.

Un ulteriore controllo viene effettuato in sede di costruzione delle tavole di pubblicazione, per quanto concerne la tutela del segreto statistico e della riservatezza dei dati individuali.

1.6 - Procedure di integrazione delle mancate risposte totali

1.6.1 - Le fasi dell'integrazione

La fase di integrazione delle mancate risposte totali viene realizzata in diversi passi. Viene innanzitutto individuato l'insieme delle imprese che devono essere integrate. Questa operazione è effettuata a partire

dall'archivio Asia che fornisce informazione sulle imprese nate, cessate e sulle trasformazioni avvenute nel periodo di riferimento.

Successivamente viene lanciata la procedura di integrazione tramite donatore per la quale le imprese rispondenti forniscono i dati alle imprese non rispondenti a meno di un fattore di ponderazione che tiene conto delle diverse dimensioni delle imprese.

A partire dall'anno di riferimento 2000 sono stato messi a disposizione dell'indagine Sci i bilanci civilistici depositati dalle imprese di capitale presso le camere di commercio. Questa base dati rappresenta una fonte preziosa ai fini della ricostruzione dell'informazione non rilevata a causa delle mancate risposte.

Una volta integrate con il metodo del donatore tutte le imprese non rispondenti, si passa ad una successiva fase di integrazione che utilizza il database dei bilanci civilistici. In primo luogo si accoppiano le imprese integrate tramite donatore con il file contenente i dati di bilancio.

Per le imprese accoppiate vengono, quindi, sostituite le voci stimate con il metodo del donatore con quelle reali dichiarate dalle imprese nei bilanci civilistici. Questo avviene per le principali voci di bilancio: fatturato, costi, spese per il personale, eccetera.

I dettagli di tali variabili, che sono richiesti solo nel questionario Sci ma non sono presenti nei bilanci, vengono stimati di nuovo tenendo come vincolo il valore totale della voce proveniente dai bilanci. Gli addendi sono riponderati utilizzando pesi calcolati come rapporti tra i valori stimati con il metodo del donatore ed il nuovo totale proveniente dai bilanci. Alla fine del processo gli addendi vengono quadrati col totale.

Le imprese che non si accoppiano con i record contenuti nel file dei bilanci rimangono integrate col metodo del donatore. Si tratta, generalmente, di imprese non di capitale il cui numero si aggira intorno al 13 per cento del totale delle imprese dell'universo Sci (Paragrafo 1.7).

Qui di seguito si illustrano nel dettaglio i singoli passi della procedura di integrazione delle mancate risposte totali.

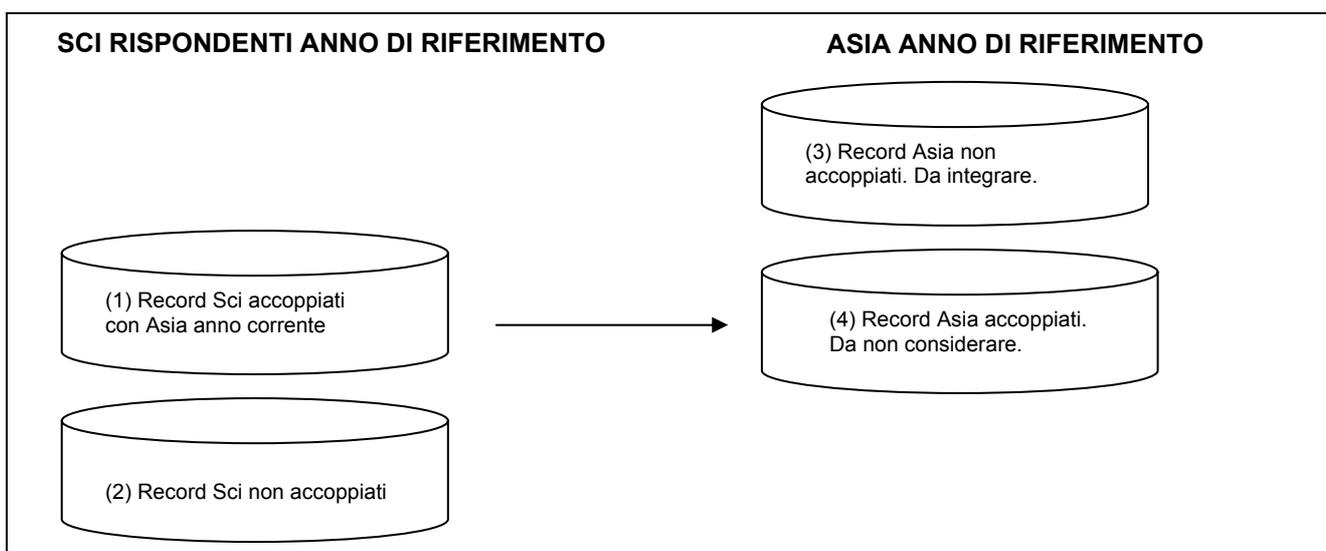
1.6.2 - Allestimento della lista di integrazione delle mancate risposte totali

Al momento della chiusura dell'indagine tutti i record dell'impresa sono corretti e quadrati con i record relativi alle Uf. In questo momento e una volta acquisito l'archivio Asia relativo all'anno di riferimento dell'indagine si inizia la fase di integrazione delle mancate risposte totali.

Il primo passo consiste nell'eliminare dalla lista di partenza i record delle imprese non rispondenti (o, se si preferisce, vengono mantenuti solamente i record delle imprese rispondenti).

A questo punto si accoppiano i record dei rispondenti con i record del nuovo archivio Asia e si ottiene per differenza una prima approssimazione del file dei non rispondenti da integrare (Figura 1.1).

Figura 1.1 – Procedure di individuazione delle imprese da integrare



A questo punto occorre risolvere tre problemi:

1. individuare i motivi per i quali alcuni record Sci non si sono accoppiati ai record dell'archivio Asia corrente;
2. individuare le possibili differenze tra le variabili comuni ai record accoppiati;
3. ricostruire, tramite l'archivio degli eventi di fusione e scorporo in possesso dell'archivio Asia le modificazioni intercorse tra l'anno t e l'anno t-1 per le imprese interessate all'indagine Sci.

Lo strumento utilizzato per risolvere tali problematiche è il citato archivio degli eventi. Per il primo problema, infatti, si tratta di imprese che molto probabilmente hanno subito eventi di trasformazione tali da rendere necessaria una modifica del codice Asia (modifica del codice fiscale, scissione in più imprese, eccetera) tramite l'archivio degli eventi che contiene il codice di partenza dell'anno t-1, gli eventi che le imprese hanno subito e i codici di arrivo all'anno t è possibile ricostruire le trasformazioni e recuperare imprese nel file di Asia dei record non accoppiati e di conseguenza inserirli nel file dei record accoppiati.

Anche per le imprese Sci accoppiate è necessario verificare che i dati corrispondano a quelli registrati nell'archivio Asia ed in particolare il numero degli addetti. Infatti tra un anno ed un altro un'impresa rispondente può essere stata coinvolta in operazioni di scissione ed aver dichiarato, ad esempio, a Sci tutti gli addetti dell'impresa di partenza presentando un questionario consolidato coerente con i dati relativi all'occupazione, mentre Asia, tramite le sue fonti, registra tutte le imprese che hanno subito trasformazioni nel periodo di riferimento con i relativi addetti.

Una volta operati tutti i controlli descritti sopra, si prenderà in considerazione il file (3) (Figura 1.2) che conterrà tutte le imprese da integrare (con i vari metodi descritti successivamente) al fine di giungere ad una stima dell'universo Sci relativo all'anno corrente.

1.6.3 - La fase di integrazione delle mancate risposte totali tramite donatore

In tutte le rilevazioni statistiche, campionarie o totali, si riscontrano mancate risposte. La parte di informazione che non viene fornita dalle imprese non rispondenti produce, ovviamente, una distorsione nei valori finali che devono essere forniti come risultato dell'indagine. Per ovviare a questo tipo di problemi si possono applicare diverse metodologie.

Per l'indagine Sci nei casi di mancata risposta totale si ricorre, in una prima fase, alla stima delle variabili economiche delle imprese non rispondenti attraverso metodologie di imputazione. È stata utilizzata una strategia di imputazione tramite donatore, che consiste nell'individuare, per ogni impresa non rispondente (ricevente), un profilo economico ad essa applicabile, desumibile, all'interno di insiemi (strati) di imprese caratterizzati da analoga dimensione aziendale, regione di appartenenza e settore di attività economica, da imprese che assumono il ruolo di "donatore"³.

Si ha, pertanto, per ogni strato l, la seguente situazione:

$$1) \quad Y_l = Y_l^R + Y_l^{NR}$$

dove Y_l rappresenta l'ammontare totale della variabile Y nello strato l da stimare; $Y_l^R = \sum y_{ij}^R$ (con $i=1, \dots, m$) è l'ammontare, conosciuto, della variabile Y per le m imprese rispondenti nello strato l; $Y_l^{NR} = \sum y_{ij}^{NR}$ (con $j=m+1, \dots, n$) rappresenta l'ammontare, incognito, della variabile Y da attribuire alle n-m imprese non rispondenti.

La parte di informazione incognita viene stimata tramite la replicazione dei dati di imprese estratte dall'insieme costituito dalle m imprese rispondenti nello stesso strato l dell'impresa mancante.

Quindi:

$$2) \quad \text{stim } y_{j,l}^{NR} = \alpha y_{i,l}$$

dove:

³ Cfr. Istat. Relazione finale del gruppo di lavoro sulla *Definizione delle metodologie applicabili alle rilevazioni ed indagini sulle statistiche economiche circa il problema delle mancate risposte parziali e totali*. Roma: Istat, 1994.

$$3) \quad \alpha = x_j / x_i$$

con X si rappresenta una variabile ausiliaria atta a ponderare il valore della variabile stimata Y (nel caso della presente rilevazione X rappresenta il numero degli addetti). La ponderazione viene utilizzata sia al fine di stimare più precisamente il valore mancante sia per ricreare la variabilità che altrimenti andrebbe persa a causa dell'appiattimento del dato dell'impresa mancante sul dato del donatore.

A questo fine si pone un ulteriore vincolo al processo di estrazione e scelta del donatore, rappresentato da:

$$4) \quad \text{Max}(Z) = n$$

Dove Z è il numero di imprese mancanti alle quali una singola impresa rispondente può "donare" i propri dati. In genere si è posto $n = 4$. Una volta effettuato il processo di integrazione si sommano i dati delle imprese integrate e delle imprese rispondenti in maniera da ottenere la stima di Y_i .

1.6.4 - Integrazione tramite fonte amministrativa (bilanci civilistici depositati dalle imprese presso le camere di commercio)⁴

La seconda fase di integrazione ha come punto di partenza la base dei microdati corretti e già integrati con le usuali procedure di integrazione tramite donatore.

Si sono, quindi, sostituiti i dati stimati delle imprese non-rispondenti con i valori provenienti dai bilanci civilistici. Dal momento che nella rilevazione Sci le variabili richieste presentano un livello di dettaglio superiore a quello riportato nel conto economico della fonte bilanci, le sottovoci non incluse nella fonte bilanci sono state ricostruite ripartendo i totali noti sulla base delle proporzioni ottenute con l'attuale procedura di integrazione di Sci. Per le variabili non incluse nell'insieme di voci contenute nei bilanci civilistici sono stati lasciati i valori ottenuti attraverso l'usuale procedura di integrazione. Tutte le variabili stimate sono state successivamente quadrate con una procedura gerarchica.

Tramite la fonte bilanci sono disponibili 22 variabili relative al conto economico, 4 variabili relative allo stato patrimoniale e 6 relative al costo del lavoro.

Si tratta di 7 variabili di primo livello (capoconti individuati tramite lettera più il totale del costo del lavoro) e di 25 variabili di secondo livello (capoconti individuati da lettera e cifra più alcuni aggregati relativi al costo del lavoro).

In particolare sono disponibili dal file dei bilanci civilistici le seguenti variabili:

Conto economico

- valore della produzione (A), e le 5 componenti (A1-A5);
- costi della produzione (B), e le 9 componenti (B6-B14);
- proventi e oneri finanziari (C), e le 3 componenti (C15-C17);
- imposte sugli utili lordi (T);
- utili netti (U).

Stato Patrimoniale

Attivo

- immobilizzazioni (B), e le 3 componenti (B1-B3).

Per lo Stato Patrimoniale, oltre alle variabili elencate sopra, sono disponibili altre 16 voci che rappresentano variabili di terzo livello e che vengono utilizzate (se presenti) nell'integrazione delle mancate risposte totali.

Costo del lavoro

- totale del costo del lavoro;

⁴ R. Monducci, P.D. Falorsi, A. Pallara, A. Russo. "Prime esperienze sull'utilizzo integrato di fonti statistiche e amministrative per la produzione di statistiche strutturali sui risultati economici delle imprese". In *Temi di ricerca ed esperienze sull'utilizzo a fini statistici di dati di fonte amministrativa*. Roma: Franco Angeli, 2003. (Economia, n. 365-239).

- salari e stipendi;
- oneri sociali;
- quiescenza;
- altri costi;
- trattamento di fine rapporto.

Queste variabili sono utilizzate per integrare le circa 120 variabili della rilevazione Sci inerenti il conto economico, lo stato patrimoniale e il costo del lavoro.

Per quanto riguarda il conto economico e il costo del lavoro vi sono in totale 80 variabili da integrare; di queste 28 corrispondono dal punto di vista della definizione a quelle contenute nella fonte bilanci; le rimanenti 52 sono variabili di terzo livello che la rilevazione Sci richiede e che non sono incluse nello schema di conto economico base adottato dalla fonte bilanci. La procedura di integrazione prevede, innanzitutto, l'integrazione diretta delle 28 variabili di primo e secondo livello; queste variabili vengono sottoposte a quadratura, con i rispettivi totali; per la stima delle variabili di terzo livello si utilizzano come pesi i valori ricavati dalla procedura di integrazione tramite donatore. In particolare, le variabili di secondo livello integrate con la fonte bilanci vengono distribuite sulle variabili di terzo livello secondo le medesime proporzioni stimate con la procedura di integrazione tramite donatore.

Per quanto riguarda lo stato patrimoniale, le 38 variabili richieste sono in parte integrate con le voci di bilancio. Vengono invece stimati con la medesima procedura utilizzata per il conto economico alcuni di cui non presenti sul file bilanci, come: la disaggregazione per singole voci delle immobilizzazioni immateriali e materiali e i relativi valori al lordo del fondo di ammortamento.

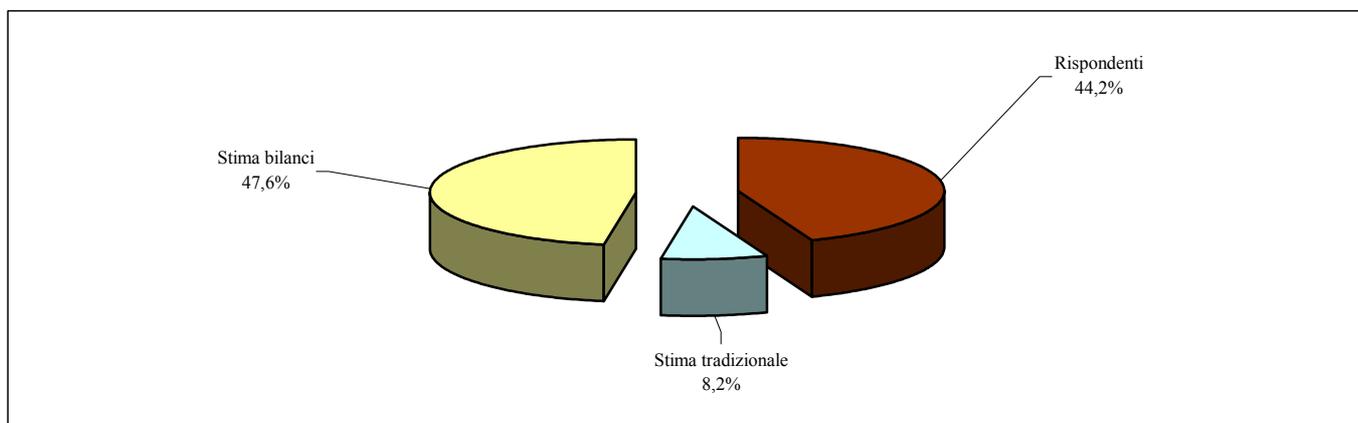
I risultati ottenuti, riportati nel paragrafo 1.7, mostrano che le differenze di risultato fra le due procedure di integrazione sono nel complesso di notevole entità. Inoltre, all'aumentare del livello di dettaglio dei dati prodotti (settori e dimensioni aziendali), emergono differenze più consistenti, specialmente nei settori dove si registra un più esiguo numero di imprese.

1.7 - Risultati della metodologia di integrazione tramite fonte amministrativa

Il totale delle imprese appartenenti al campo di osservazione dell'indagine Sci contenute nell'Archivio Asia per l'anno di riferimento 2004 è pari ad un totale di 10.313.

L'analisi dei dati stimati e integrati nell'ambito della conduzione dell'indagine Sci, sempre relativamente all'anno 2004 e distinti per fonte, mostra che la quota delle imprese rispondenti e valide è del 44,2 per cento. Si è utilizzata, invece, la base dei dati dei bilanci civilistici per ottenere una stima dei dati economici per una quota pari al 47,6 per cento del totale delle imprese, mentre la quota delle imprese integrate con il metodo tradizionale (tramite donatore) è del 8,2 per cento. (Figura 1.2)

Figura 1.2 – Imprese per modalità di risposta e di integrazione - Anno 2004 (valori percentuali)



La raccolta delle informazioni provenienti dalle imprese rispondenti e dai bilanci civilistici depositati dalle società di capitali presso le camere di commercio copre, pertanto, circa il 92 per cento del campo di osservazione dell'indagine.

Per quanto riguarda le imprese non rispondenti, risulta che il dato di bilancio ha consentito di reperire informazioni per l' 85 per cento delle 5.750 imprese da integrare nella rilevazione. L'ampia copertura è certamente favorita dalla natura giuridica delle imprese con 100 addetti ed oltre che sono costituite per lo più in società di capitali. Per analizzare l'efficienza dei processi di raccolta e correzione attuati nell'ambito dell'indagine Sci si è operato un confronto fra i dati rilevati nell'indagine (imprese rispondenti) ed i dati presenti nell'archivio dei bilanci.

Il confronto è stato condotto con riferimento a due variabili importanti e ampiamente confrontabili dal punto di vista definitivo (fatturato e costo del lavoro) analizzando la distribuzione delle differenze percentuali per le 4.086 imprese che si sono potute accoppiare in entrambe le basi dei dati. I risultati evidenziano un ottimo accostamento tra le due fonti per le variabili considerate come risulta dai prospetti seguenti.

Prospetto 1.1 - Analisi della distribuzione delle differenze percentuali del fatturato tra la fonte bilanci e la fonte Sci (rilevato) per impresa - Anno 2004

Numero imprese	4.086
Quinto percentile	-0,5
Venticinquesimo percentile	0,0
Mediana	0,0
Settantacinquesimo percentile	0,0
Novantacinquesimo percentile	0,08

Il prospetto 1.1 presenta la distribuzione delle differenze percentuali relative al fatturato; per il 90 per cento delle imprese tale distribuzione varia tra il -0,5 per cento (quinto percentile) e lo 0,08 per cento (novantacinquesimo percentile); tali variazioni, calcolate, a livello micro per ogni impresa, come differenza tra il dato presente sul bilancio amministrativo ed il dato rilevato dall'indagine.

Prospetto 1.2 - Analisi della distribuzione delle differenze percentuali del costo del lavoro tra la fonte bilanci e la fonte Sci (rilevato) per impresa - Anno 2004

Numero imprese	4.086
Quinto percentile	-0.4%
Venticinquesimo percentile	0,0%
Mediana	0,0%
Settantacinquesimo percentile	0,0%
Novantacinquesimo percentile	0,4%

Il prospetto 1.2 mostra le differenze percentuali calcolate relativamente al costo del lavoro; anche per questa variabile, si può osservare che il 90 per cento della distribuzione delle differenze tra le due fonti è contenuta in un intervallo piuttosto piccolo. I limiti sono, infatti, dati dal -0,4 per cento (quinto percentile) e dallo 0,4 per cento (novantacinquesimo percentile).

Prospetto 1.3 - Analisi della distribuzione delle differenze percentuali del fatturato tra la fonte bilanci e la fonte Sci (rilevato) per impresa e classi di addetti - Anno 2004

	Classi di addetti	
	100-249	250 ed oltre
Numero imprese	2488	1598
Quinto percentile	-0,6	-0,2
Venticinquesimo percentile	0,0	0,0
Mediana	0,0	0,0
Settantacinquesimo percentile	0,0	0,0
Novantacinquesimo percentile	0,06	0,02

La corrispondenza dei dati provenienti dai bilanci e dalla rilevazione è evidenziata anche dal fatto che la mediana, il primo ed il terzo quartile hanno un valore pari a zero per le due variabili in esame.

Inoltre il confronto tra i dati rilevati nell'indagine e i dati presenti nell'archivio dei bilanci civilistici è stato condotto anche per classi di addetti. L'esame dei risultati ottenuti mostra che la distribuzione delle differenze percentuali per impresa non è influenzata dalla dimensione aziendale come risulta dai prospetti seguenti:

Prospetto 1.4 - Analisi della distribuzione delle differenze percentuali del costo del lavoro tra la fonte bilanci e la fonte Sci (rilevato) per impresa e classi di addetti - Anno 2004

	Classi di addetti	
	100-249	250 ed oltre
Numero imprese	2488	1598
Quinto percentile	-0,3	-0,5
Venticinquesimo percentile	0,0	0,0
Mediana	0,0	0,0
Settantacinquesimo percentile	0,0	0,0
Novantacinquesimo percentile	0,6	0,2

In conclusione si può affermare che le procedure di correzione e revisione dei dati per le imprese rispondenti sono efficaci e producono risultati assolutamente coerenti al *benchmark* di riferimento rappresentato dai bilanci civilistici. Peraltro il guadagno in termini di qualità del dato dovuto all'utilizzo dell'integrazione tramite fonte amministrativa delle non rispondenti è notevole, allineando anche questa parte di informazioni al *benchmark*.

Tavola A1 – Grado di copertura della rilevazione SCI per sezione, sottosezione, divisione, gruppi e classi di attività economica – Anno 2004

Codice di attività	ATTIVITÀ ECONOMICA	Numero imprese		Percentuali di copertura		
		Esistenti	Rispondenti	Numero imprese	Numero addetti	Valore aggiunto aziendale
C	ESTRAZIONE DI MINERALI	14	9	64,3	95,2	99,3
CA	ESTRAZIONE DI MINERALI ENERGETICI	4	3	75,0	99,0	99,8
10	ESTRAZIONE DI CARBON FOSSILE E LIGNITE; ESTRAZIONE DI TORBA	1	1	100,0	100,0	100,0
10.1	Estraz.ed agglomerazione di carbon fossile	1	1	100,0	100,0	100,0
11	ESTRAZ. DI PETROL. GREGGIO, GAS NATURALE; SERVIZI CONNESSI ALL'ESTRAZ. DI PETROL. E GAS	3	2	66,7	98,9	99,8
11.1	Estraz. di petrolio greggio e di gas naturale	1	1	100,0	100,0	100,0
11.2	Attività dei servizi connessi all'estrazione di petrolio, gas, esclusa la prospezione	2	1	50,0	58,6	59,4
CB	ESTRAZIONE DI MINERALI NON ENERGETICI	10	6	60,0	69,2	78,2
14	ALTRE INDUSTRIE ESTRATTIVE	10	6	60,0	69,2	78,2
14.1	Estraz. di pietra	2	1	50,0	66,3	73,7
14.2	Estraz. di ghiaia, sabbia e argilla	2	0	0,0	0,0	0,0
14.3	Estrazione di minerali per le industrie chimiche e la fabbricazione di concimi	1	0	0,0	0,0	0,0
14.4	Produzione di sale	1	1	100,0	100,0	100,0
14.5	Estraz. di altri minerali e prodotti di cava n.a.c.	4	4	100,0	100,0	100,0
D	ATTIVITÀ MANIFATTURIERE	4.909	2.348	47,8	61,7	65,0
DA	INDUSTRIE ALIMENTARI, DELLE BEVANDE E DEL TABACCO	407	195	47,9	63,9	67,5
15	INDUSTRIE ALIMENTARI E DELLE BEVANDE	406	194	47,8	63,4	65,5
15.1	Produtz., lavorazione e conservazione di carne e di prodotti a base di carne	74	35	47,3	58,0	61,8
15.2	Lavoraz. e conservazione di pesci e di prodotti a base di pesce	6	2	33,3	35,0	11,2
15.3	Lavorazione e conservazione di frutta e ortaggi	59	20	33,9	54,4	50,8
15.4	Fabbr. di oli e grassi vegetali e animali	7	4	57,1	71,5	49,5
15.5	Industria lattiero-casearia	70	36	51,4	61,7	67,4
15.6	Lavorazione delle granaglie e di prodotti amidacei	7	3	42,9	46,3	48,8
15.7	Fabbr.di prodotti per l'alimentazione degli animali	16	8	50,0	62,7	79,3
15.8	Fabbr. di altri prodotti alimentari	119	61	51,3	74,2	78,6
15.9	Industria delle bevande	48	25	52,1	57,2	49,0
16	INDUSTRIA DEL TABACCO	1	1	100,0	100,0	100,0
16.0	Industria del tabacco	1	1	100,0	100,0	100,0
DB	INDUSTRIE TESSILI E DELL'ABBIGLIAMENTO	517	238	46,0	57,5	57,6
17	INDUSTRIE TESSILI	334	167	50,0	60,0	58,6

Tavola A1 segue – Grado di copertura della rilevazione SCI per sezione, sottosezione, divisione, gruppi e classi di attività economica – Anno 2004

Codice di attività	ATTIVITÀ ECONOMICA	Numero imprese		Percentuali di copertura		
		Esistenti	Rispondenti	Numero imprese	Numero addetti	Valore aggiunto aziendale
17.1	Preparazione e filatura di fibre tessili	69	38	55,1	65,3	63,9
17.2	Tessitura di materie tessili	103	50	48,5	65,9	62,6
17.3	Finissaggio dei tessili	56	28	50,0	63,9	62,4
17.4	Confezioni di articoli in tessuto, esclusi gli articoli di vestiario	14	6	42,9	38,5	46,2
17.5	Altre industrie tessili	41	19	46,3	54,3	54,6
17.6	Fabbr. di tessuti a maglia	13	5	38,5	47,1	56,5
17.7	Fabbr. di articoli in maglieria	38	21	55,3	45,1	45,0
18	CONFEZIONI DI ARTICOLI DI VESTIARIO; PREPARAZIONE E TINTURA DI PELLICCE	183	71	38,8	53,0	56,2
18.1	Confezione di vestiario in pelle	1	1	100,0	100,0	100,0
18.2	Confezione di altri articoli di vestiario ed accessori	180	69	38,3	52,8	56,1
18.3	Preparazione e tintura di pellicce; confezione di articoli in pelliccia	2	1	50,0	66,7	60,2
DC	INDUSTRIE CONCIARIE, FABBRICAZIONE DI PRODOTTI IN CUOIO, PELLE E SIMILI	132	48	36,4	47,9	56,9
19	PREPARAZ. E CONCIA DEL CUOIO; FABBRICAZ. DI ARTICOLI DA VIAGGIO, BORSE, SELLERIA, CALZATURE	132	48	36,4	47,9	56,9
19.1	Preparazione e concia del cuoio	23	8	34,8	41,9	42,3
19.2	Fabbr. di articoli da viaggio, borse, artic. da correggiaio e selleria	16	7	43,8	72,8	78,0
19.3	Fabbricazione di calzature	93	33	35,5	43,8	56,0
DD	INDUSTRIA DEL LEGNO E DEI PRODOTTI IN LEGNO	83	39	47,0	53,6	55,7
20	INDUSTRIA DEL LEGNO E DEI PRODOTTI IN LEGNO E SUGHERO, ARTICOLI DI PAGLIA	83	39	47,0	53,6	55,7
20.1	Taglio, piallatura e trattamento del legno	9	5	55,6	65,1	67,0
20.2	Fabbr. di fogli da impiallacciatura; fabbricazione di compensati, pannelli, ecc.	24	9	37,5	52,3	54,1
20.3	Fabbricazione di elementi di carpenteria in legno e falegnameria per l'edilizia	31	17	54,8	64,3	72,2
20.4	Fabbricazione di imballaggi in legno	2	1	50,0	47,9	31,9
20.5	Fabbr. di altri prod. in legno; fabbric. di art. in sughero, paglia, ecc.	17	7	41,2	31,6	30,2
DE	FABBRICAZIONE PASTA-CARTA, CARTA E PRODOTTI DI CARTA, STAMPA ED EDITORIA	260	126	48,5	62,0	59,7
21	FABBRICAZIONE DELLA PASTA-CARTA, DELLA CARTA E DEI PRODOTTI DI CARTA	117	63	53,8	70,6	74,8
21.1	Fabbr. della pasta-carta, della carta e del cartone	21	12	57,1	78,9	81,6
	Fabbricazione di articoli di carta e di cartone	96	51	53,1	66,7	71,8
22	EDITORIA, STAMPA E RIPRODUZIONE DI SUPPORTI REGISTRATI	143	63	44,1	54,3	50,2
22.1	Editoria	62	28	45,2	57,0	47,7
22.2	Stampa e attività dei servizi connessi alla stampa	79	34	43,0	51,8	56,2
22.3	Riproduzione di supporti registrati	2	1	50,0	35,5	12,9

Tavola A1 segue – Grado di copertura della rilevazione SCI per sezione, sottosezione, divisione, gruppi e classi di attività economica – Anno 2004

Codice di attività	ATTIVITÀ ECONOMICA	Numero imprese		Percentuali di copertura		
		Esistenti	Rispondenti	Numero imprese	Numero addetti	Valore aggiunto aziendale
DF	FABBR.DI COKE, RAFF.DI PETROLIO, TRATTAMENTO DEI COMBUSTIBILI NUCLEARI	27	18	66,7	84,7	92,3
23	FABBR.DI COKE, RAFF.DI PETROLIO, TRATTAMENTO DEI COMBUSTIBILI NUCLEARI	27	18	66,7	84,7	92,3
23.1	Fabbricazione di prodotti di cokeria	1	1	100,0	100,0	100,0
23.2	Fabbricazione di prodotti petroliferi raffinati	26	17	65,4	84,4	92,1
DG	FABBRICAZIONE DI PRODOTTI CHIMICI E DI FIBRE SINTETICHE E ARTIFICIALI	361	205	56,8	66,7	69,4
24	FABBRICAZIONE DI PRODOTTI CHIMICI E DI FIBRE SINTETICHE E ARTIFICIALI	361	205	56,8	66,7	69,4
24.1	Fabbricazione di prodotti chimici di base	88	53	60,2	75,9	82,3
24.2	Fabbricazione di pesticidi e di altri prodotti chimici per l'agricoltura	6	4	66,7	71,6	72,0
24.3	Fabbricazione di pitture, vernici e smalti, inchiostri da stampa e mastici	38	20	52,6	54,7	57,1
24.4	Fabbricazione di prodotti farmaceutici, chimici, botanici per usi medicinali	130	76	58,5	66,9	67,6
24.5	Fabbricazione di saponi e detergenti, di prodotti per la pulizia e la lucidatura	46	27	58,7	69,9	76,0
24.6	Fabbricazione di altri prodotti chimici	41	20	48,8	52,2	48,3
24.7	Fabbricazione di fibre sintetiche e artificiali	12	5	41,7	41,8	54,0
DH	FABBRICAZIONE DI ARTICOLI IN GOMMA E MATERIE	289	126	43,6	54,6	55,3
25	FABBRICAZIONE DI ARTICOLI IN GOMMA E MATERIE PLASTICHE	289	126	43,6	54,6	55,3
25.1	Fabbricazione di articoli in gomma	54	24	44,4	68,9	71,6
25.2	Fabbricazione di articoli in materie plastiche	235	102	43,4	46,5	46,7
DI	FABBRICAZIONE DI PRODOTTI DELLA LAVORAZIONE MINERALI NON METALLIFERI	280	156	55,7	69,5	74,2
26	FABBRICAZIONE DI PRODOTTI DELLA LAVORAZIONE NON METALLIFERI	280	156	55,7	69,5	74,2
26.1	Fabbricazione di vetro e di prodotti in vetro Fabbr.di prod.ceramici non refrattari, non destinati all'edilizia e di prod.ceramici refrattari	47	26	55,3	78,1	81,5
26.2	Fabbr.di piastrelle e lastre in ceramica per pavimenti e rivestimenti	28	13	46,4	54,1	55,2
26.3	Fabbr.di mattoni, tegole ed altri prodotti per l'edilizia in terracotta	77	53	68,8	79,0	82,4
26.4	Produzione di cemento, calce, gesso	17	10	58,8	57,9	66,4
26.5	Fabbr. di prodotti in calcestruzzo, cemento o gesso	18	11	61,1	84,0	84,0
26.6	Taglio, modellatura e finitura della pietra	70	33	47,1	50,1	51,8
26.7	Fabbr. di altri prodotti in minerali non metalliferi	11	3	27,3	23,5	25,4
26.8	Fabbr. di altri prodotti in minerali non metalliferi	12	7	58,3	59,4	65,3
DJ	PRODUZIONE DI METALLO E FABBRICAZIONE DI PRODOTTI IN METALLO	735	343	46,7	58,5	61,8
27	PRODUZIONE DI METALLI E LORO LEGHE	225	122	54,2	66,4	68,0

Tavola A1 segue – Grado di copertura della rilevazione SCI per sezione, sottosezione, divisione, gruppi e classi di attività economica – Anno 2004

Codice di attività	ATTIVITÀ ECONOMICA	Numero imprese		Percentuali di copertura		
		Esistenti	Rispondenti	Numero imprese	Numero addetti	Valore aggiunto aziendale
27.1	Produzione di ferro, di acciaio e di ferroleghie (CECA)	43	27	62,8	79,1	76,4
27.2	Fabbricazione di tubi	34	21	61,8	56,2	60,0
27.3	Altre attività di prima trasformazione del ferro e acciaio, ferr. non CECA	21	15	71,4	63,3	59,9
27.4	Produzione di metalli di base preziosi e non ferrosi	55	27	49,1	54,6	64,1
27.5	Fusione di metalli	72	32	44,4	56,4	49,6
28	FABBR. E LAVOR. DEI PRODOTTI IN METALLO, ESCLUSE MACCHINE E IMPIANTI	510	221	43,3	51,4	53,8
28.1	Fabbr. di elementi da costruzione in metallo	83	26	31,3	35,3	37,1
28.2	Fabbr. di cisterne, serbatoi, conten. metall.; Fabbr. di radiat. e caldaie	43	20	46,5	56,1	55,7
28.3	Fabbr. di gener. di vapore, escl. le caldaie per risc. Centrale ad acqua	5	3	60,0	45,8	50,3
28.4	Fucinatura, imbutitura, stampaggio, profilatura dei metalli; metallurgia delle polveri	88	40	45,5	50,0	55,8
28.5	Trattamento e rivestimento dei metalli, lavorazione di meccanica generale c/terzi	103	34	33,0	41,8	41,7
28.6	Fabbricazione di articoli di coltelleria, utensili ed oggetti diversi in metallo	47	25	53,2	62,0	64,8
28.7	Fabbr. di altri prodotti metallici	141	73	51,8	59,5	60,8
DK	FABBR. DI MACCHINE ED APP. MECCANICI, COMPRESI INSTALL., MONTAG., RIPARAZ. E MANUTENZ.	798	396	49,6	61,4	63,7
29	FABBR. DI MACCHINE ED APP. MECCANICI, COMPRESI INSTALL. MONT., RIPARAZ. E MANUTENZ.	798	396	49,6	61,4	63,7
29.1	Fabbr. di macchine e apparecchi per la produzione e l'utilizzo dell'energia meccanica	186	97	52,2	66,4	68,3
29.2	Fabbr. di altre macchine di impiego generale	252	119	47,2	56,6	59,3
29.3	Fabbr. di macchine per l'agricoltura e la silvicoltura	29	13	44,8	64,4	64,2
29.4	Fabbr. di macchine utensili (compresi inst., manut., riparaz.)	82	42	51,2	61,0	62,5
29.5	Fabbr. di altre macchine per impieghi speciali	167	84	50,3	62,8	63,2
29.6	Fabbr. di armi, sistemi d'arma e munizioni	11	8	72,7	76,5	79,5
29.7	Fabbr. di apparecchi per uso domestico n.a.c	71	33	46,5	57,5	60,7
DL	FABBR. DI MACCHINE ELETTRICHE E DI APPARECCHI ELETTRICHE ED OTTICHE	459	198	43,1	57,2	62,6
30	FABBR. DI MACCHINE PER UFFICIO, DI ELABORATORI E SISTEMI INFORMATICI	11	5	45,5	64,0	(a)
30.0	Fabbr. di macchine per ufficio, di elaboratori e sistemi informatici	11	5	45,5	64,0	(a)
31	FABBRICAZIONE DI MACCHINE ED APPARECCHI ELETTRICI N.A.C.	223	103	46,2	60,5	64,3
31.1	Fabbr. di motori, generatori e trasformatori elettrici	45	22	48,9	52,1	50,4
31.2	Fabbr. di app. per la distrib. e il controllo dell'elettricità	46	28	60,9	80,6	84,7
31.3	Fabbr. di fili e cavi isolati	24	12	50,0	41,2	47,4
31.4	Fabbr. di accumulatori, pile e batterie di pile	5	1	20,0	26,1	17,6
31.5	Fabbr. di app. di illuminazione e di lampade elettriche	17	10	58,8	72,8	77,0
31.6	Fabbr. di altri apparecchi elettrici n.a.c.	86	30	34,9	54,3	53,7

Tavola A1 segue – Grado di copertura della rilevazione SCI per sezione, sottosezione, divisione, gruppi e classi di attività economica – Anno 2004

Codice di attività	ATTIVITÀ ECONOMICA	Numero imprese		Percentuali di copertura		
		Esistenti	Rispondenti	Numero imprese	Numero addetti	Valore aggiunto aziendale
32	FABBRICAZIONE DI APPARECCHI RADIOTELEVISIVI E DI APPARECCHIATURE PER LE COMUNICAZIONI	107	35	32,7	50,3	58,6
32.1	Fabbr.di tubi e valvole elettron.e di altri comp. elettronici	33	13	39,4	69,8	81,9
32.2	Fabbr.di app. trasmettenti per la radio diffusione e la televisione	58	16	27,6	37,6	40,6
32.3	Fabbr.di app.riceventi per la radiodiffusione e la televisione, di app.per registrazione e riproduzione del suono o dell'immagine e di prod.connessi	16	6	37,5	29,5	34,6
33	FABBR.DI APPAR. MEDICALI, DI APPARECCHI DI PRECISIONE, DI STRUMENTI OTTICI E OROLOGI	118	55	46,6	59,3	60,1
33.1	Fabbr. di apparecchi medicali e chirurgici e di apparecchi ortopedici	38	15	39,5	50,6	40,9
33.2	Fabbr. di strum. e appar.di misur.,contr.,prova, navigazione	43	22	51,2	65,4	67,1
33.3	Fabbr. di apparecchi per il controllo dei processi industriali	13	4	30,8	59,3	63,6
33.4	Fabbr. di strumenti ottici e di attrezzature fotografiche	23	13	56,5	57,3	64,5
33.5	Fabbricazione di orologi	1	1	100,0	100,0	100,0
DM	FABBRICAZIONE DI MEZZI DI TRASPORTO	309	150	48,5	68,8	70,4
34	FABBRICAZIONE DI AUTOVEICOLI, RIMORCHI E SEMIRIMORCHI	212	101	47,6	67,1	66,8
34.1	Fabbricazione di autoveicoli	24	13	54,2	81,5	81,0
34.2	Fabbr. di carrozzerie per autoveicoli;fabbricazione di rimorchi e semirimorchi	30	11	36,7	50,3	39,5
34.3	Fabbr. di parti ed accessori per autoveicoli per loro motori	158	77	48,7	55,0	59,4
35	FABBRICAZIONE DI ALTRI MEZZI DI TRASPORTO	97	49	50,5	72,6	75,9
35.1	Industria cantierista: costruzioni navali e riparazione di navi e imbarcazioni	28	13	46,4	77,6	73,6
35.2	Costruz.di locomotive, anche da manovra, e di materiale rotabile ferro-tranviario	18	7	38,9	46,6	62,5
35.3	Costruzione di aeromobili e di veicoli spaziali	22	16	72,7	91,0	94,2
35.4	Fabbr. di motocicli e biciclette	29	13	44,8	46,1	40,0
DN	ALTRE INDUSTRIE MANIFATTURIERE	252	110	43,7	54,4	53,4
36	FABBRICAZIONE DI MOBILI, ALTRE INDUSTRIE MANIFATTURIERE	247	108	43,7	54,3	53,3
36.1	Fabbr. di mobili	185	82	44,3	56,5	57,1
36.2	Gioielleria e oreficeria	16	11	68,8	66,5	51,1
36.3	Fabbr. di strumenti musicali	3	0	0,0	0,0	0,0
36.4	Fabbr. di articoli sportivi	6	1	16,7	23,1	21,4
36.5	Fabbr. di giochi e giocattoli	8	0	0,0	0,0	0,0
36.6	Altre industrie manifatturiere n.a.c.	29	14	48,3	57,7	54,4
37	RECUPERO E PREPARAZIONE PER IL RICICLAGGIO	5	2	40,0	57,1	62,3
37.1	Recupero e preparazione per il riciclaggio di cascami e rottami metallici	1	1	100,0	100,0	100,0
37.2	Recupero e preparazione per il riciclaggio di cascami e rottami non metallici	4	1	25,0	48,8	55,0

Tavola A1 segue – Grado di copertura della rilevazione SCI per sezione, sottosezione, divisione, gruppi e classi di attività economica – Anno 2004

Codice di attività	ATTIVITÀ ECONOMICA	Numero imprese		Percentuali di copertura		
		Esistenti	Rispondenti	Numero imprese	Numero addetti	Valore aggiunto aziendale
E	PRODUZIONE E DISTRIBUZIONE DI ENERGIA ELETTRICA, DI GAS, DI VAPORE E ACQUA CALDA	127	86	67,7	50,6	55,8
40	PRODUZIONE E DISTRIBUZIONE DI ENERGIA ELETTRICA, DI GAS, DI VAPORE E ACQUA CALDA	86	64	74,4	52,6	57,9
40.1	Produzione e distribuzione di energia elettrica	35	27	77,1	45,5	52,0
40.2	Produz.di gas; distribuzione di combustibili gassosi mediante condotta	49	37	75,5	81,8	93,0
40.3	Produzione e distribuzione di vapore ed acqua calda	2	0	0,0	0,0	0,0
41	RACCOLTA, DEPURAZIONE E DISTRIBUZIONE D'ACQUA	41	22	53,7	38,6	28,0
41.0	Raccolta, depurazione e distribuzione d'acqua	41	22	53,7	38,6	28,0
F	COSTRUZIONI	417	178	42,7	54,0	59,0
45	COSTRUZIONI	417	178	42,7	54,0	59,0
45.1	Preparazione del cantiere edile	10	5	50,0	67,4	79,3
45.2	Costruzione completa o parziale di edifici; genio civile	304	142	46,7	59,6	63,8
45.3	Installazione dei servizi in un fabbricato	90	28	31,1	31,9	32,7
45.4	Lavori di completamento degli edifici	12	3	25,0	23,4	28,1
45.5	Noleggio macchine e attrezzature per la costruzione o la demolizione, con manovratore	1	0	0,0	0,0	0,0
G	COMMERCIO ALL'INGROSSO E AL DETTAGLIO; RIPARAZIONE DI AUTOVEICOLI, MOTOCICLI ECC.	1.105	463	41,9	47,5	58,2
50	COMMERCIO, MANUTENZIONE E RIPARAZIONE DI AUTOVEICOLI E MOTOCICLI; VENDITA AL DETT. CARB.	112	52	46,4	54,6	82,0
50.1	Commercio di autoveicoli	85	38	44,7	54,3	83,7
50.2	Manutenzione e riparazione di autoveicoli	4	3	75,0	85,8	86,7
50.3	Commercio di parti e accessori di autoveicoli	15	9	60,0	56,9	74,9
50.4	Commercio, manutenzione e riparazione di motocicli, accessori e pezzi di ricambio	1	1	100,0	100,0	100,0
50.5	Vendita al dettaglio di carburanti per autotrazione	7	1	14,3	17,6	23,2
51	COMMERCIO ALL'INGROSSO E INTERMEDIARI DEL COMMERCIO AUTOVEICOLI E MOTOCICLI ESCLUSI	481	228	47,4	53,8	62,1
51.1	Intermediari del commercio	5	1	20,0	12,3	7,7
51.2	Comm. all'ingrosso di materie prime agricole e di animali vivi	6	3	50,0	54,3	61,0
51.3	Comm. all'ingrosso di prodotti alimentari, bevande e tabacco	103	42	40,8	41,8	57,5
51.4	Comm. all'ingrosso di altri beni di consumo finale	212	106	50,0	58,7	64,1
51.5	Comm. all'ingrosso di prodotti intermedi non agricoli, rottami, cascami	81	38	46,9	59,2	60,1
51.8	Comm. all'ingrosso di macchinari e attrezzature	66	34	51,5	65,2	70,9
51.9	Comm. all'ingrosso di altri prodotti	8	4	50,0	10,6	6,0
52	COMM. AL DETTAGLIO, ESCL. AUTOVEICOLI E MOTOCICLI; RIPARAZ. DI BENI PERSON. E PER CASA	512	183	35,7	44,5	49,8
52.1	Comm. al dettaglio in esercizi non specializzati	277	106	38,3	46,8	55,1
52.2	Comm. al dettaglio di prodotti alimentari, bevande e tabacco in esercizi specializzati	8	3	37,5	46,8	42,4

Tavola A1 segue – Grado di copertura della rilevazione SCI per sezione, sottosezione, divisione, gruppi e classi di attività economica – Anno 2004

Codice di attività	ATTIVITÀ ECONOMICA	Numero imprese		Percentuali di copertura		
		Esistenti	Rispondenti	Numero imprese	Numero addetti	Valore aggiunto aziendale
52.3	Comm. al dettaglio di prodotti farmaceutici, cosmetici, profumeria	28	13	46,4	36,5	37,7
52.4	Comm. al dettaglio di altri prodotti (esclusi quelli di seconda mano)	176	51	29,0	37,5	39,7
52.5	Comm. al dettaglio di articoli di seconda mano	1	0	0,0	0,0	0,0
52.6	Comm. al dettaglio al di fuori dei negozi	21	10	47,6	44,9	42,0
52.7	Riparazione di beni di consumo personali e per la casa	1	0	0,0	0,0	0,0
H	ALBERGHI E RISTORANTI	274	77	28,1	49,0	47,8
55	ALBERGHI E RISTORANTI	274	77	28,1	49,0	47,8
55.1	Alberghi	129	44	34,1	45,5	52,1
55.2	Campeggi ed altri alloggi per brevi soggiorni	15	2	13,3	23,3	28,2
55.3	Ristoranti	29	5	17,2	38,6	33,2
55.4	Bar	8	1	12,5	2,2	1,4
55.5	Mense e fornitura di pasti preparati	93	25	26,9	60,8	59,9
I	TRASPORTI, MAGAZZINAGGIO E COMUNICAZIONI	895	429	47,9	80,8	90,3
60	TRASPORTI TERRESTRI; TRASPORTI MEDIANTE CONDOTTE	307	169	55,0	79,0	85,0
60.1	Trasporti ferroviari	16	11	68,8	92,2	92,9
60.2	Altri trasporti terrestri	287	154	53,7	70,8	73,6
60.3	Trasporti mediante condotte	4	4	100,0	100,0	100,0
61	TRASPORTI MARITTIMI E PER VIE D'ACQUA	46	22	47,8	52,4	59,5
61.1	Trasporti marittimi e costieri	45	22	48,9	54,6	63,7
61.2	Trasporti per vie d'acqua interne (compresi i trasporti lagunari)	1	0	0,0	0,0	0,0
62	TRASPORTI AEREI	19	7	36,8	75,3	66,9
62.1	Trasporti aerei di linea	16	6	37,5	78,4	71,3
62.2	Trasporti aerei non di linea	3	1	33,3	14,1	16,7
63	ATTIVITÀ DI SUPPORTO ED AUSILIARIE DEI TRASPORTI; ATTIVITÀ DELLE AGENZIE DI VIAGGIO	484	208	43,0	61,5	80,1
63.1	Movimentazione merci e magazzinaggio	327	124	37,9	37,8	50,8
63.2	Altre attività connesse ai trasporti	93	49	52,7	78,8	90,1
63.3	Attività delle agenzie di viaggio e degli operatori turistici, att. di assist. turist.	28	13	46,4	53,0	45,7
63.4	Attività delle altre agenzie di trasporto	36	22	61,1	81,3	85,2
64	POSTE E TELECOMUNICAZIONI	39	23	59,0	97,8	98,3
64.1	Attività postali e di corriere	10	5	50,0	99,3	99,3
64.2	Telecomunicazioni	29	18	62,1	95,4	98,0
K	ATTIV. IMMOBILIARI, NOLEGGIO, INFORMATICA, RICERCA, ALTRE ATTIV. PROFESS. E IMPRENDITORIALI	1.609	577	35,9	41,2	46,4
70	ATTIVITÀ IMMOBILIARI	24	8	33,3	27,3	35,4
70.1	Attività immobiliari su beni propri	8	3	37,5	25,8	23,1
70.2	Locazione di beni immobili propri e sublocazione	7	2	28,6	28,5	15,8
70.3	Attività immobiliare per conto terzi	9	3	33,3	27,6	47,1

Tavola A1 segue – Grado di copertura della rilevazione SCI per sezione, sottosezione, divisione, gruppi e classi di attività economica – Anno 2004

Codice di attività	ATTIVITÀ ECONOMICA	Numero imprese		Percentuali di copertura		
		Esistenti	Rispondenti	Numero imprese	Numero addetti	Valore aggiunto aziendale
71	NOLEGG. DI MACCH. E ATTREZZAT. SENZA OPERATORE E DI BENI PER USO PERSONALE E DOMESTICO	19	9	47,4	49,5	53,5
71.1	Noleggio di autovetture	11	8	72,7	80,6	78,5
71.2	Noleggio di altri mezzi di trasporto	1	0	0,0	0,0	0,0
71.3	Noleggio di altre macchine e attrezzature	5	0	0,0	0,0	0,0
71.4	Noleggio di beni per uso personale e domestico n.a.c.	2	1	50,0	13,8	28,9
72	INFORMATICA E ATTIVITÀ CONNESSE	303	130	42,9	45,6	48,3
72.1	Consulenza per installazione di elaboratori elettronici	2	1	50,0	56,9	56,9
72.2	Fornitura di software e consulenza in materia di informatica	201	86	42,8	43,8	47,3
72.3	Elaborazione elettronica dei dati	65	31	47,7	51,5	52,8
72.4	Attività delle banche di dati	6	3	50,0	50,1	57,1
72.5	Manutenzione e riparazione di macchine per ufficio e di elaboratori elettronici	7	3	42,9	57,1	64,8
72.6	Altre attività connesse all'informatica	22	6	27,3	47,4	43,4
73	RICERCA E SVILUPPO	26	12	46,2	54,7	48,6
73.1	Ricerca e sviluppo sperimentale nel campo delle scienze naturali e dell'ingegneria	24	12	50,0	57,1	51,1
73.2	Ricerca e sviluppo sperimentale nel campo delle scienze sociali e umanistiche	2	0	0,0	0,0	0,0
74	ALTRE ATTIVITÀ PROFESSIONALI ED IMPRENDITORIALI	1.237	418	33,8	40,3	45,4
74.1	Attività legali, contabilità, tenuta di libri contabili; consul. commerc.; holding	145	54	37,2	38,9	35,3
74.2	Attività in materia di architettura, di ingegneria ed altre attività tecniche	58	26	44,8	60,9	77,8
74.3	Collaudi e analisi tecniche	13	6	46,2	74,9	79,1
74.4	Pubblicità	27	13	48,1	57,9	54,5
74.5	Servizi di ricerca, selezione e fornitura di personale	66	19	28,8	30,0	33,2
74.6	Servizi di investigazione e vigilanza	159	54	34,0	39,1	40,8
74.7	Servizi di pulizia e disinfestazione	647	204	31,5	46,9	50,7
74.8	Altre attività di tipo professionale ed imprenditoriale n.c.a.	122	42	34,4	39,7	38,6
M	ISTRUZIONE	21	6	28,6	34,8	41,4
80	ISTRUZIONE	21	6	28,6	34,8	41,4
80.1	Istruzione primaria	3	1	33,3	44,8	57,2
80.2	Istruzione secondaria	4	1	25,0	26,7	26,9
80.3	Istruzione universitaria	2	0	0,0	0,0	0,0
80.4	Istruzione per gli adulti ed altri servizi di istruzione	12	4	33,3	40,7	47,1
N	SANITÀ E ALTRI SERVIZI SOCIALI	568	267	47,0	55,6	54,6
85	SANITÀ E ALTRI SERVIZI SOCIALI	568	267	47,0	55,6	54,6
85.1	Attività dei servizi sanitari	258	127	49,2	55,1	53,2
85.3	Assistenza sociale	310	140	45,2	56,1	57,1
O	ALTRI SERVIZI PUBBLICI, SOCIALI E PERSONALI	336	123	36,6	52,1	70,4
90	SMALTIMENTO DEI RIFIUTI SOLIDI, DELLE ACQUE DI SCARICO E SIMILI	181	67	37,0	44,2	51,2
90.0	Smaltimento dei rifiuti solidi, delle acque di scarico e simili	181	67	37,0	44,2	51,2
92	ATTIVITÀ RICREATIVE, CULTURALI E SPORTIVE	109	37	33,9	60,9	83,1
92.1	Produzioni e distribuzioni cinematografiche e di video	36	4	11,1	8,7	18,0

Tavola A1 – Grado di copertura della rilevazione SCI per sezione, sottosezione, divisione, gruppi e classi di attività economica – Anno 2004

Codice di attività	ATTIVITÀ ECONOMICA	Numero imprese		Percentuali di copertura		
		Esistenti	Rispondenti	Numero imprese	Numero addetti	Valore aggiunto aziendale
92.2	Attività radiotelevisive	14	8	57,1	87,4	96,8
92.3	Altre attività dello spettacolo	20	6	30,0	30,0	50,1
92.4	Attività delle agenzie di stampa	2	1	50,0	80,4	79,8
92.5	Attività di biblioteche, archivi, musei ed altre attività culturali	6	3	50,0	60,4	63,6
92.6	Attività sportive	10	4	40,0	76,9	(a)
92.7	Attività ricreative	21	11	52,4	77,1	88,7
93	SERVIZI ALLE FAMIGLIE	46	19	41,3	61,6	61,5
93.0	Servizi alle famiglie	46	19	41,3	61,6	61,5

(a) Dato non disponibile

2 - Aspetti metodologici della rilevazione sulle imprese con 1-99 addetti (*)

2.1 - Caratteristiche della rilevazione

La rilevazione sulle piccole e medie imprese e sull'esercizio di arti e professioni (Pmi) ha come campo di osservazione le imprese con 1-99 addetti appartenenti ai settori di attività economica industriale, commerciale e dei servizi alle imprese ed alle famiglie e risponde alle esigenze richieste dal regolamento comunitario sulle statistiche strutturali n° 58/97 (SBS).

L'unità di rilevazione ed analisi è l'impresa e le principali variabili rilevate riguardano gli addetti, le spese per il personale, gli investimenti, i costi ed il fatturato. L'universo oggetto di indagine è rappresentato dalle imprese attive nel 2004. L'archivio di estrazione della lista delle unità campionate è rappresentato dall'archivio Asia (Archivio statistico delle imprese attive), costruito sulla base dell'integrazione di varie fonti, di carattere sia amministrativo sia statistico.

L'indagine è campionaria ed è condotta mediante invio postale del modello di rilevazione ed il disegno di campionamento utilizzato è di tipo casuale stratificato. I dati delle imprese rispondenti sono sottoposti a revisione, a controlli di coerenza e compatibilità, a tecniche di trattamento delle mancate risposte parziali e delle osservazioni anomale.

Il riporto dei dati all'universo è realizzato:

- per categoria di attività economica della classificazione Ateco2002 (codici a quattro cifre), senza distinzione per classi di addetti ;
- per gruppo di attività economica (codici Ateco2002 a tre cifre) e fascia dimensionale delle imprese in termini di addetti (1-9, 10-19, 20-49 e 50-99 addetti per le imprese industriali e per le imprese che operano nell'intermediazione monetaria e finanziaria, nell'istruzione, nella sanità ed in altri servizi pubblici, sociali e personali; 1, 2-4, 5-9, 10-19, 20-49 e 50-99 addetti per le imprese commerciali; 1-4, 5-9, 10-19, 20-49 e 50-99 addetti per le imprese alberghiere e di ristorazione, di trasporto e comunicazione, immobiliari, di noleggio, informatiche, professionali ed imprenditoriali) ;
- per divisione di attività economica (codici Ateco2002 a due cifre) e regione.

La metodologia utilizzata si basa sugli "stimatori di ponderazione vincolata"; essa consente di calcolare pesi finali che, sotto certe ipotesi, risultano correttivi delle mancate risposte totali e della sottocopertura della lista. Essi assicurano il rispetto di uguaglianza fra alcuni totali noti dell'universo di riferimento (imprese e addetti) e le stime campionarie. La valutazione del livello di precisione delle stime è realizzata mediante l'impiego di modelli regressivi che consentono di pervenire ad una presentazione sintetica degli errori di campionamento.

Tutti gli aspetti metodologici sopra descritti sono trattati in questo capitolo e nel capitolo 3, nella quale si forniscono statistiche utili per la valutazione della strategia di campionamento adottata ed elementi per il calcolo dell'errore di campionamento, assoluto e relativo, e dell'intervallo di confidenza per ogni stima pubblicata.

(*) Giampiero Siesto, responsabile della progettazione della rilevazione, ha coordinato le diverse fasi del processo di produzione dei dati, ha curato la validazione e la diffusione dei risultati ed ha redatto i paragrafi 2.1, 2.2, 2.3, 2.7, 2.10, 2.10.1, 2.10.2, 2.10.3 e 2.11.5.2; Franco Branchi ha curato il piano per il trattamento e controllo dei dati individuali ed ha redatto i paragrafi 2.8, 2.9.1, 2.9.2 e 2.9.3; Annamaria Mazzilli, Ennio Santi, Raffaella Giovannuzzi, Lorena Ricci e Paola Mizzau hanno curato le fasi di raccolta, revisione e correzione interattiva dei modelli e i rapporti con le imprese mentre Milena Di Paola ha curato le fasi di raccolta, revisione dei modelli e il rapporto con le imprese. Piero Demetrio Falorsi, Marco Ballin e Maria Cristina Casciano hanno curato gli aspetti metodologici relativi al disegno della rilevazione, selezione del campione e riporto dei dati all'universo. Piero Demetrio Falorsi, che ha coordinato la redazione degli aspetti metodologici, ha redatto i paragrafi 2.5, 2.11.1 e 2.11.4.2; Marco Ballin ha redatto i paragrafi 2.5.3, 2.11.3 e 2.11.4.1; Maria Cristina Casciano ha redatto i paragrafi 2.4, 2.5.1, 2.5.2, 2.5.3, 2.6, 2.11.2, 2.11.5 e 2.11.5.1. Gli aspetti metodologici dell'intero lavoro sono stati curati e definiti in piena collaborazione tra gli autori citati. Il programma di controllo delle incompatibilità dei microdati è stato curato da Umberto Sansone mentre la predisposizione delle tavole è stata curata da Luciano Fanfoni e Tiziana Di Francescantonio.

2.2 - Unità di osservazione

L'unità di osservazione assunta per l'indagine è l'impresa, definita come organizzazione di un'attività economica esercitata con carattere professionale al fine della produzione di beni o per la prestazione di servizi destinabili alla vendita. In tale organizzazione il responsabile può essere una persona fisica (liberi professionisti o artigiani costituiti in imprese individuali), una persona giuridica (società di persone, società di capitali) o un insieme di persone associate (società semplice, società di fatto, associazioni di professionisti, cooperative di lavoro, eccetera).

2.3 - Campo di osservazione

Il campo di osservazione è costituito dalle imprese con 1-99 addetti, attive nel corso dell'anno solare 2004 ed esercitanti un'attività industriale, commerciale e dei servizi nei settori: estrattivo (sezione C della classificazione delle attività economiche Ateco2002⁵); manifatturiero (sezione D); della produzione e distribuzione di energia elettrica, gas e acqua (sezione E); delle costruzioni (sezione F); del commercio e riparazione (sezione G); alberghiero e della ristorazione (sezione H); di trasporto e comunicazione (sezione I); di intermediazione monetaria e finanziaria (sezione J, con l'esclusione delle divisioni 65 "intermediazione monetaria e finanziaria, escluse le assicurazioni ed i fondi pensione" e 66 "assicurazioni e fondi pensione, escluse le assicurazioni sociali obbligatorie"); immobiliare, di noleggio, informatica, professionale ed imprenditoriale (sezione K); d'istruzione (sezione M); sanitaria e in servizi sociali (sezione N) e in altri servizi pubblici, sociali e personali (sezione O, con l'esclusione della divisione 91 "attività di organizzazioni associative").

La lista delle unità appartenenti all'universo d'indagine è stata ottenuta applicando le seguenti condizioni alle imprese dell'archivio Asia:

- attive al 31.12.2003;
- numero di addetti in media annua inferiore a 100 ;
- attività economica compresa tra le divisioni 10 e 93 della classificazione Ateco2002, con esclusione delle divisioni 65, 66, 75, 91;
- esclusione delle pubbliche amministrazioni.

Occorre notare che la lista di selezione non coincide esattamente con la popolazione d'interesse, in quanto:

- non contiene le imprese nate dopo il 31.12.2003 ;
- può contenere alcune imprese cessate dopo il 31.12.2003.

Il disallineamento tra archivio di selezione e popolazione d'interesse costituisce fonte di potenziale distorsione delle stime; tuttavia la disponibilità della versione aggiornata di Asia un anno dopo il periodo di estrazione del campione rende necessario accettare in fase preliminare tale distorsione, per poi tentare comunque di ridurne l'entità in fase di stima, attraverso l'opportuna calibrazione degli stimatori.

2.4 - Domini di studio

Per *dominio di studio* si intende una sottopopolazione individuata da una partizione (detta *tipo di dominio*) della popolazione oggetto di indagine.

L'indagine sulle piccole e medie imprese ha lo scopo di fornire stime dei totali e dei valori medi per alcune variabili di interesse riferite a tre tipi di domini (Prospetto 2.1), ottenuti come combinazioni delle modalità assunte dalle seguenti variabili:

- divisione di attività economica (Ateco a 2 cifre);
- gruppo di attività economica (Ateco a 3 cifre);
- classe di attività economica (Ateco a 4 cifre);

⁵ Istat. *Classificazione delle attività economiche*. Roma: Istat, 2003. (Metodi e norme, n. 18).

- regioni amministrative (considerando separatamente le province di Trento e Bolzano);
- classi dimensionali di addetti, stabilite in modo differente a seconda della divisione di attività economica.

Prospetto 2.1 - Domini di studio e loro denominazione

Tipo di Dominio	Descrizione	N° di domini
DOM1	Classe di attività economica	464
DOM2	Gruppo di attività economica × Classi addetti	838
DOM3	Divisione di attività economica × Regione	966

Le classi dimensionali in termini di addetti sono descritte nel prospetto 2.1:

Prospetto 2.2 – Definizione delle classi dimensionali di addetti

DIVISIONI DI ATTIVITA' ECONOMICA (Ateco 2)		Classi di addetti	Etichette
Industria :	10-45	1-9	I1
		10-19	I2
		20-49	I3
		50-99	I4
Servizi:	50, 51, 52	1	SA1
		2-9	SA2
		10-19	SA3
		20-49	SA4
		50-99	SA5
Servizi:	55, 60, 61, 62, 63, 64, 67, 70, 71, 72, 73, 74	1	SB1
		2-9	SB2
		10-19	SB3
		20-49	SB4
		50-99	SB5
Servizi:	80, 85, 90, 92, 93	1-9	SC1
		10-19	SC2
		20-49	SC3
		50-99	SC4

2.5 - Disegno di campionamento

Poiché il regolamento sulle statistiche strutturali richiede stime senza limiti di fascia dimensionale delle imprese per i domini del tipo DOM1 e DOM3, lo studio del disegno di campionamento e l'allocatione del campione è avvenuta considerando congiuntamente questa indagine con quella censuaria (rilevazione sul sistema dei conti delle imprese, vedi Capitolo 1) sulle imprese con oltre 100 addetti (9.872 imprese censite).

Il disegno di campionamento adottato è ad uno stadio stratificato, con selezione delle unità con probabilità uguali; gli strati sono definiti dalla concatenazione delle modalità delle variabili 'regione', 'classi di attività economica' (Ateco a 4 cifre) e 'classi di addetti', secondo quanto descritto nei due prospetti precedenti.

Si fa notare che la stratificazione adottata costituisce la partizione minima della popolazione che permette di ottenere i domini di stima come aggregazione di strati elementari. Il fatto che tutti i domini di interesse sono di tipo stratificato⁶ presenta alcuni considerevoli vantaggi dal punto di vista della progettazione dell'indagine. In particolare, ciò permette di allocare il campione pianificando i livelli di precisione attesa delle stime su tutti i domini di interesse.

⁶ Cicchitelli, G., e A. Herze, e G.E. Montanari. "Un dominio si dice stratificato se le sue unità sono raggruppate in uno o più strati costituiti da quelle sole unità". *Il campionamento statistico* (1992): 256.

Tuttavia, come si è già accennato, al momento della progettazione del campione, l'archivio può non risultare aggiornato all'anno di riferimento della rilevazione; ciò implica che, stratificando la lista di selezione secondo i criteri illustrati sopra, si possono determinare strati teorici in cui vengono incluse imprese che nella realtà si collocano in strati differenti; pertanto, la migrazione di unità tra gli strati dovuta a questa mancata coincidenza tra lista di selezione e popolazione d'interesse fa sì che i livelli di errore sui domini siano solo in parte pianificabili.

In generale, quando un'indagine ha l'obiettivo di fornire stime per più domini di studio e per diverse variabili di interesse, il campione deve essere progettato in modo tale da assicurare simultaneamente l'accuratezza delle stime ai diversi livelli di dettaglio richiesti.

Prima di illustrare il metodo utilizzato per risolvere questo problema di allocazione multivariata e multidominio, è utile introdurre alcune notazioni simboliche.

2.5.1 - Simbologia di base

La seguente simbologia viene introdotta per definire i concetti che saranno illustrati nei successivi paragrafi:

- ${}_p y$ = la generica caratteristica di interesse p ;
- d = il generico tipo di dominio di studio, $d=1,\dots,D$;
- j_d = il generico dominio di tipo d , $j_d = 1,\dots,J_d$ con J_d = numero di domini del tipo d ;
- h = il generico strato, $h=1,\dots,H$ dove H è il numero di strati in cui è suddivisa la popolazione d'indagine;
- H_{j_d} = il numero di strati che contengono il dominio j_d ;
- N_h = il numero di unità dell'universo che appartengono allo strato h ;
- n_h = il numero di unità del campione che appartengono allo strato h ;
- ${}_p y_{hi}$ = il valore della caratteristica p osservato sulla generica unità i appartenente allo strato h ;
- ${}_p Y_h$ = il totale di popolazione della caratteristica ${}_p y$ nello strato h ;
- ${}_p \bar{Y}_h = \sum_{i=1}^{N_h} \frac{{}_p y_{hi}}{N_h}$ la media di ${}_p y$ nello strato h ; (1)
- ${}_p S_h^2 = \sum_{i=1}^{N_h} \frac{({}_p y_{hi} - {}_p \bar{Y}_h)^2}{N_h}$ la varianza di ${}_p y$ nello strato h ; (2)
- ${}_p \tilde{Y}_h = \frac{N_h}{n_h} \sum_{i=1}^{n_h} {}_p y_{hi}$ la stima diretta (o di Horvitz-Thompson) di ${}_p Y_h$ nello strato h ; (3)
- $\text{Var}({}_p \tilde{Y}_h) = N_h^2 \frac{(1-f_h)}{n_h} {}_p S_h^2$ la varianza della stima (3) nello strato h , con $f_h = \frac{n_h}{N_h}$; (4)
- ${}_p \tilde{Y}_{j_d} = \sum_{h=1}^{H_{j_d}} {}_p \tilde{Y}_h$ la stima di Horvitz-Thompson del totale di ${}_p y$ nel dominio j_d ; (5)

- $$\text{Var}({}_p\tilde{Y}_{j_d}) = \sum_{h=1}^{H_{j_d}} N_h^2 \frac{(1-f_h)}{n_h} {}_pS_h^2 \quad \text{la varianza della stima (5) nel dominio } j_d. \quad (6)$$

2.5.2 - Aspetti teorici del problema dell'allocatione multivariata

È facile verificare che la varianza (6) della stima del totale della variabile ${}_p y$ nel dominio j_d , propria di un campione stratificato con selezione delle unità senza reimmissione e con probabilità uguali, può essere decomposta come segue:

$$\text{Var}({}_p\tilde{Y}_{j_d}) = {}_pV_{j_d} + {}_pV_{j_d}^0 = \sum_{h=1}^{H_{j_d}} \frac{N_h^2}{{}_p n_h} {}_pS_h^2 - \sum_{h=1}^{H_{j_d}} N_h {}_pS_h^2 \quad (7)$$

dove ${}_pV_{j_d}^0$ è la parte di varianza non influenzata dall'allocatione⁷.

La ricerca dell'allocatione ottima prevede che sia definita una funzione C' con cui quantificare il costo della rilevazione; nel caso in esame, la funzione di costo C' è stata esplicitata come segue:

$$C' = C_o + C = C_o + \sum_{h=1}^{H_{j_d}} C_h n_h \quad (8)$$

in cui C_o rappresenta il costo fisso dell'indagine che non dipende né dalla numerosità campionaria né dall'allocatione, C il costo variabile e C_h ($h=1, \dots, H$) il costo per unità campionaria relativo allo strato h (che è stato assunto pari ad 1 in tutti gli strati).

Nel determinare la numerosità da assegnare a ciascuno strato, un approccio consiste nel fissare, per ciascuno ${}_pV_{j_d}$, un estremo superiore della varianza ${}_pV_{j_d}^*$ della stima del totale di ${}_p y$ e nel minimizzare la funzione di costo C sotto i vincoli ${}_pV_{j_d} \leq {}_pV_{j_d}^*$ ($p=1, \dots, P$; $d=1, \dots, D$; $j_d=1, \dots, J_d$).

Tale approccio è alla base della metodologia sviluppata dall'Istat per la determinazione della dimensione campionaria ottima in presenza di molteplici vincoli ed obiettivi: essa adotta infatti una generalizzazione della soluzione proposta da Bethel⁸, in cui viene definito un problema di minimo vincolato con funzione obiettivo convessa e vincoli di tipo lineare.

Il problema consiste dunque nel minimizzare la funzione obiettivo :

$$f(n_h) = \sum_{h=1}^{H_{j_d}} C_h n_h = \min \quad (9)$$

sotto un sistema di vincoli che si esplicita nella forma:

$$\sum_{h=1}^{H_{j_d}} \frac{N_h^2}{{}_p n_h} {}_pS_h^2 - \sum_{h=1}^{H_{j_d}} N_h {}_pS_h^2 \leq {}_pV_{j_d}^* \quad (p=1, \dots, P; d=1, \dots, D; j_d=1, \dots, J_d) \quad (10)$$

⁷ Falorsi, P.D., e M. Ballin, e C. De Vitiis, e G. Scepi. "Principi e metodi del software generalizzato per la definizione del disegno di campionamento nelle indagini sulle imprese condotte dall'Istat". *Statistica Applicata*, 20, n. 2 (1998).

⁸ Bethel, J. "Sample allocation in multivariate surveys". *Survey methodology*, 15 (1989): 47-57.

L'algoritmo proposto da Bethel per il calcolo dell'allocazione individua la soluzione ottima in maniera iterativa, partendo da una soluzione iniziale che coincide con la soluzione ottima nel caso univariato per la prima variabile sul primo dominio. In ciascuno dei passi successivi, la numerosità campionaria viene aumentata minimizzando la funzione obiettivo fino al soddisfacimento di tutti i vincoli. Bethel dimostra che tale algoritmo converge.

2.5.3 - Aspetti applicativi della procedura di allocazione ottima

Come già accennato, le unità di Asia facenti parte dell'universo di indagine sono state classificate secondo le modalità concatenate dei caratteri 'regione', 'classi di attività economica' e 'classi di addetti' del prospetto 2.2.

Il numero degli strati così costruiti, contenenti almeno un'impresa, è risultato pari a 25.448, con una distribuzione per classi di numerosità delle imprese negli strati illustrata nel prospetto 2.3.

Prospetto 2.3 - Distribuzione degli strati per numero di imprese

CLASSE DI AMPIEZZA DEGLI STRATI (n. di imprese)	Freq. Assoluta	Freq. Cumulata	Freq. %	Freq. % cumulata
1	4.705	4.705	18,5	18,5
2	13.975	18.680	54,9	73,4
3-5	3.440	22.120	13,5	86,9
6-10	1.624	23.744	6,4	93,3
Oltre 10	1.704	25.448	6,7	100,0

Poiché per i domini del tipo DOM1e DOM3 il regolamento richiede stime senza limiti di fasce dimensionali, oltre ai 25.448 strati campionati si sono dovuti considerare, in fase di allocazione, anche gli strati cosiddetti censiti, ossia quelli in cui si collocano le 8.872 imprese della fascia 100 addetti ed oltre soggetti alla rilevazione Sci. Nella progettazione del campione si è pertanto tenuto conto della necessità di includere tali strati completamente osservati e di non superare, per ragioni logistiche e di costo, la soglia di 120 mila unità cui inviare il questionario.

La ricerca dell'allocazione ottima è avvenuta assumendo *fatturato* e *valore aggiunto* come variabili d'interesse.

Nell'impostazione del problema di ottimizzazione, definito dalle relazioni (9) e (10), si è preferito esprimere i vincoli di precisione sulle stime di tali variabili in termini di coefficienti di variazione anziché in termini di varianze attese; in tal modo il problema di ottimo è stato risolto utilizzando come vincoli un'espressione equivalente della (10), data da:

$$\frac{\sqrt{\sum_{h=1}^{H_{jd}} \frac{N_h^2}{n_h} S_h^2 - \sum_{h=1}^{H_{jd}} N_h S_h^2}}{p \tilde{Y}_{jd}} \leq p CV_{jd}^* \quad (p=1, \dots, P; d=1, \dots, D; j_d=1, \dots, J_d) \quad (11)$$

dove si sono sostituite le quantità $p V_{jd}^*$ nella (10) con i coefficienti di variazione:

$$p CV_{jd}^* = \frac{\sqrt{p V_{jd}^*}}{d \tilde{Y}_{jd}} \quad (12)$$

Il dimensionamento del campione è stato, dunque, ottenuto imponendo alle stime dei totali di *fatturato* e *valore aggiunto* gli errori massimi pari all' 8 per cento per il primo dominio, al 9 per cento per il secondo ed al

10 per cento per DOM3 (Prospetto 2.4). Questi vincoli hanno determinato un'allocazione complessiva pari a 118.262 imprese (incluse le 9.872 della rilevazione Sci).

Prospetto 2.4 - Errore massimo imposto nella stima del fatturato e del valore aggiunto in ciascuna tipologia di dominio e dimensione del campione

ERRORE MASSIMO ATTESO	DOM1	DOM2	DOM3	Numerosità campionaria (imprese della fascia 1-99 addetti)	Numerosità complessiva (incluse le imprese censite)
$p CV_{jd}^*$	8%	9%	10%	108.390	118.262

La procedura di allocazione impiegata ha richiesto la specificazione a priori di medie e varianze per ogni strato della popolazione e per ognuna delle variabili di interesse. Tali quantità sono state stimate direttamente sulla base dei dati osservati in ciascuno strato nella precedente *indagine sulle piccole imprese e sull'esercizio di arti e professioni* (fino a 99 addetti) e nell'indagine censuaria *Sci* (100 addetti ed oltre). In quegli strati dell'archivio nei quali non è stata possibile una stima diretta per mancanza di dati rilevati con le precedenti occasioni d'indagine si sono attribuite le stime dei parametri relative ad insiemi della popolazione che contenessero tali strati.

Come già spiegato, l'errore massimo imposto rappresenta il livello teorico di errore che le stime avrebbero se valessero due condizioni: che vi sia coincidenza perfetta tra lista di selezione e popolazione di indagine; che tutte le imprese selezionate restituiscano il modello debitamente compilato.

Tuttavia entrambe le condizioni non sono realizzate, sia per il disallineamento temporale fra l'archivio di selezione e la popolazione di interesse sia per il fenomeno della mancata risposta che può provocare la caduta di interi strati e determinare, per le stime sui domini costituiti da tali strati, un livello di precisione più basso di quello pianificato. L'incremento d'errore dovuto alla mancata risposta totale può essere attenuato dall'uso degli stimatori di ponderazione vincolata.

2.6 - Selezione del campione

La procedura impiegata per la selezione del campione appartiene alla categoria di metodi, detti di *selezione coordinata*⁹, usualmente adottati per conseguire diversi obiettivi, tra cui:

- il controllo della sovrapposizione di campioni relativi ad indagini longitudinali, ossia il controllo della rotazione delle unità campionarie;
- la minimizzazione della sovrapposizione tra campioni provenienti dallo stesso archivio di estrazione e relativi ad indagini differenti.

Nel caso in esame, si è voluto fare in modo che nel campione estratto fossero presenti in percentuale minima unità già intervistate nelle rilevazioni Pmi relative ad anni precedenti o in altre rilevazioni Istat aventi come oggetto lo stesso universo d'indagine; questa operazione ha lo scopo di distribuire il più possibile la *pressione statistica complessiva*, ossia l'onere della risposta alle molteplici richieste d'informazioni che, nel corso degli anni, pervengono alle imprese da parte degli enti del Sistema Statistico Nazionale.

La tecnica impiegata, nota come *campionamento Jales*¹⁰, è basata sull'uso di *Numeri Casuali Permanenti (Ncp)*: parte cioè dal presupposto che a ciascuna unità dell'archivio di selezione sia associato un numero casuale, generato da una variabile aleatoria uniforme definita nell'intervallo [0,1]; il Ncp resta assegnato all'impresa per l'intera sua esistenza e viene riportato invariato in tutte le successive versioni dell'archivio Asia in cui l'impresa risulta essere presente.

La tecnica Jales è di tipo sequenziale: nel caso dell'estrazione di un campione stratificato, costituito da n_h unità per strato, essa prevede l'ordinamento di tutte le unità in ciascuno strato secondo l'ordine crescente

⁹ Ballin, M., e S. Loriga. "Metodi di selezione coordinata nelle indagini campionarie sulle imprese". *Rivista di statistica ufficiale*, n. 2 (2000).

¹⁰ Ohlsson, E., e B.G. Cox, e D.A. Binder, e B.N. Chinnappa, e A. Christianson, e M.J. Kott, e P.S. Colledge. "Coordination of samples using permanent random numbers". In *Business Survey Methods*. New York: Wiley, 1995.

dei Ncp e la selezione di un sottoinsieme di n_h unità contigue che hanno un Ncp immediatamente minore (o maggiore, a seconda della tecnica adottata) di un valore fissato α denominato *punto iniziale*, scelto nell'intervallo [0,1].

Il controllo della sovrapposizione tra campioni può essere effettuato applicando il metodo della traslazione alla tecnica di selezione appena illustrata. In sintesi, esso consiste nel modificare il punto iniziale α di un valore δ detto *costante di traslazione*. Dopo aver sottratto tale quantità a tutti i Ncp⁸, si procede alla selezione delle unità secondo il metodo descritto in precedenza. Qualora il numero di unità da estrarre sia maggiore del numero di unità che hanno una posizione nello strato ordinato superiore al punto in cui parte la selezione, il metodo riconsidera le prime unità della lista (e probabilmente già intervistate in altra occasione di indagine).

Quando l'obiettivo è quello della selezione casuale di un campione che presenti una sovrapposizione minima con campioni precedentemente estratti dallo stesso archivio, si procede simulando più selezioni corrispondenti a diversi valori della costante di traslazione; si sceglie infine il δ relativo al campione che presenta il minor numero di unità incluse anche nei campioni rispetto ai quali si vuole effettuare il coordinamento.

La tecnica appena descritta è stata applicata per effettuare l'estrazione delle unità campione nella presente indagine: fissando il punto iniziale $\alpha=0$ ed iterando il procedimento per diversi valori δ della costante di traslazione, si è alla fine optato per la selezione relativa al valore $\delta=0,4$. A questa corrisponde la sovrapposizione minima tra i campioni.

2.7 - Definizione degli aggregati e stima del valore aggiunto

L'obiettivo principale dell'indagine è di analizzare il ruolo che le imprese di piccola dimensione ricoprono nel sistema produttivo nazionale ed europeo. Alle imprese attive è stato chiesto di fornire dati relativamente al fatturato ed alle diverse voci dei ricavi, ai costi del processo di produzione di beni e servizi, alle giacenze, agli investimenti effettuati, al numero degli occupati, alle spese di personale e ad altre variabili come risulta dal questionario allegato in Appendice.

Sulla base delle variabili economiche rilevate è stato possibile determinare il valore aggiunto ed altri aggregati economici per le cui definizioni si rimanda al paragrafo 1.4 degli aspetti metodologici della rilevazione sulle imprese con 100 addetti ed oltre ed al glossario.

2.8 - Raccolta dei dati

La rilevazione è stata effettuata mediante l'invio postale (a mezzo Postel) del questionario Istat Ssi/Pmi a tutte le imprese appartenenti al campione, estratte secondo i criteri precedentemente descritti.

Il questionario, costituito di dodici pagine, di cui sei per la raccolta dei dati e le rimanenti sei pagine per le istruzioni necessarie per una corretta compilazione, è stato strutturato in un quadro di "notizie generali" e dieci sezioni contenenti le variabili necessarie al calcolo degli aggregati economici precedentemente descritti. Dette sezioni riguardano rispettivamente il "conto economico" (valore e costi della produzione), "l'occupazione per categoria professionale", i "costi per il personale", il "personale esterno all'impresa e i relativi costi", "l'acquisizione di immobilizzazioni nell'esercizio per tipo di beni", "altri dati", "dati regionali", le "spese per la protezione dell'ambiente", "dati retrospettivi - anno 2003" ed "altre informazioni". In quest'ultima sezione sono richieste informazioni sulla diffusione del commercio elettronico, sull'innovazione tecnologica, sulla ricerca e sviluppo, sulle relazioni fra imprese e sulla struttura di controllo dell'impresa.

Tutti i dati contabili sono richiesti in unità di Euro, Iva esclusa, relativamente all'esercizio dell'anno 2004.

Alle imprese selezionate è stato garantita l'assistenza telefonica per chiarimenti sulla compilazione del questionario.

2.9 - Trattamento dei dati

2.9.1 - Revisione e registrazione dei dati

I questionari di impresa restituiti all'Istat, sono stati sottoposti ad una attenta revisione qualitativa da parte del personale specializzato che ha interessato, in una fase antecedente la registrazione dei dati, i seguenti caratteri:

- 1) attività economica e numero di addetti, al fine di verificare l'appartenenza al campo di osservazione nell'esercizio in esame;
- 2) presenza delle risposte obbligatorie: almeno una variabile significativa relativamente alle voci dei "costi della produzione" e del "valore della produzione" ;
- 3) utilizzo della "media annua" per il numero di occupati.

Per alcuni questionari, che non superavano questa fase di revisione qualitativa, il compilatore è stato ricontattato telefonicamente.

La registrazione ha interessato solo le variabili con cifre diverse da zero (significative), ed è stata strutturata su tre tipi records tutti riportanti i dati identificativi dell'impresa (codice di impresa e caratteri di stratificazione: attività economica, regione e numero di addetti).

Per ogni impresa rispondente sono stati registrati:

- a) un tipo record "1" a campi fissi di lunghezza 320 posizioni relativo alle "notizie generali" ;
- b) un tipo record "2" a campi fissi di lunghezza 47 posizioni per ognuna delle variabili significative presenti sul questionario (sezioni da 1 a 9); per ogni impresa, in media, sono presenti 40 records su circa 220 variabili possibili ;
- c) un tipo record "3" a campi fissi di lunghezza 149 posizioni per le variabili e mutabili relative alla sezione 10 "altre informazioni".

2.9.2 - Piano di controllo , correzione automatica ed interattiva dei dati

I dati presenti sui questionari sono stati registrati su supporto magnetico in un tempo medio di circa 10 giorni lavorativi, per ciascuno degli invii in registrazione. I dati registrati sono stati sottoposti al "controllo di qualità" per accertare la rispondenza ai requisiti di qualità, nonché al controllo di leggibilità del supporto ed al conteggio delle battute utili.

A tale scopo, alcuni questionari estratti casualmente dai rispettivi pacchi sono stati nuovamente registrati ed i relativi *records* confrontati con quelli precedentemente registrati. E' risultato che tutti gli invii in registrazione presentavano, in base al campionamento effettuato, un numero di *bytes* errati al di sotto della soglia di errore prevista. L'errore di registrazione ha interessato in media lo 0,65 per mille del totale di battute utili registrate.

I records relativi alla stessa impresa (in media 42 records) sono stati compattati in un unico record a campi fissi di lunghezza 2.900 posizioni contenente tutte le variabili (significative e non) e mutabili del questionario dell'impresa: il numero di records di imprese rispondenti sono pertanto stati 50.001. Considerando, inoltre, che è stata effettuata un'integrazione delle mancate risposte totali per le imprese del campione con 20-99 addetti tramite dati di bilancio civilistici delle società di capitale depositati presso le Camere di commercio, che ha riguardato 6.079 imprese, il campione effettivo su cui sono state effettuate le stime finali è risultato pari a 56.080 imprese.

Nella costruzione del record di impresa sono stati inseriti il campo del "coefficiente di espansione all'universo", il campo del "codice di ripartizione territoriale" non presente nei dati indicativi dell'impresa, nonché i campi indicanti i tipi di incompatibilità e correzioni sull'impresa in esame.

L'esperienza maturata nelle precedenti indagini ed i suggerimenti forniti da analisti di bilancio, sono stati presi in considerazione nella metodologia di controllo e correzione dei dati elementari; ciò ha comportato la costruzione di una complessa procedura per la messa a punto quantitativa e qualitativa dei dati rilevati con i questionari di impresa. Detta procedura è articolata in varie fasi di lavoro appresso descritte:

- a) verifica dei dati indicativi dei questionari di impresa (codice di impresa e caratteri di stratificazione: attività economica, regione e numero di addetti);

b) messa a punto qualitativa dei dati elementari dei questionari di impresa, tramite un programma di "controllo e correzione automatica" che determina la correttezza del dato di ogni singolo campo oppure l'esistenza della condizione di errore.

Il controllo automatico ha riguardato, in considerazione del fenomeno oggetto di indagine, soprattutto i dati quantitativi e, in misura minima, i dati qualitativi. Questo piano di controllo è "esaustivo", nel senso che prevede tutti i possibili casi di errore nel questionario.

Una volta verificata l'esistenza della condizione di errore, nella maggioranza dei casi il campo del *record* di impresa relativo al quesito affetto da errore è stato "automaticamente" corretto. Negli altri casi sono state effettuate correzioni di tipo "interattivo", da parte del personale specializzato, secondo tre tipologie di interventi:

- *accertamenti*: interventi diretti a verificare che quanto registrato corrispondesse o meno a quanto trascritto sul questionario di rilevazione; in caso di corrispondenza si è apposto il visto di accettazione del dato mentre in caso contrario si è proceduto alla correzione interattiva;
- *forzature*: interventi di correzione automatica sulla base delle distribuzioni conosciute e delle variabili fondamentali presenti sul questionario e ritenute esatte;
- *errori*: interventi di correzione interattiva effettuati dall'esperto dell'attività economica in esame, previo ulteriore revisione del questionario di impresa.

L'accertamento, la forzatura e l'errore dipendono dalle relazioni istituite (Prospetto 2.5) per:

- valori fuori campo ;
- mancate risposte parziali o totali ;
- incongruenze logiche tra le variabili.

Tra le relazioni istituite, quelle che hanno avuto una maggiore frequenza di violazione delle regole sono risultate:

- assenza dell'indicazione dell'imprenditore alla sezione occupazione ;
- diversità della sommatoria delle variabili di costo del personale rispetto al totale del costo del personale;
- diversità della sommatoria delle variabili di spesa rispetto al totale dei costi di produzione ;
- diversità della sommatoria delle variabili di entrata rispetto al totale del valore della produzione.

Su questi ultimi due aspetti ha pesato l'articolazione dettagliata del conto economico, caratterizzato da numerosi totali parziali.

Prospetto 2.5 - Relazioni istituite per tipo di controllo

TIPO DI CONTROLLO	Numero relazioni istituite
Controlli nelle singole sezioni:	
- sui dati indicativi dell'impresa	4
- sulle notizie generali	12
- sul conto economico	16
- sulla occupazione	54
- sui costi del personale	10
- sul personale esterno all'impresa e relativi costi	11
- sull'acquisto di immobilizzazioni	30
- sugli altri dati	-
- sui dati regionali	4
- sulle spese per la protezione dell'ambiente	-
Controlli tra sezioni:	
- tra notizie generali, occupazione e spese di personale	17
- tra notizie generali e conto economico	5
- tra notizie generali e acquisto di immobilizzazioni	3
- tra notizie generali ed altri dati (in più sezioni del questionario)	3
- tra conto economico e occupazione	2
- tra conto economico e costi del personale	2
- tra conto economico, occupazione e dati retrospettivi	2
- tra conto economico e personale esterno e relativi costi	4
- tra conto economico ed altri dati	6
- tra dati regionali, occupazione e costo del personale	3
- tra spese per la protezione dell'ambiente ed acquisto di immobilizzazioni	2

Nel prospetto 2.6 sono riportate le percentuali di imprese con interventi per tipo di incompatibilità sulle principali variabili rilevate. Dall'esame delle frequenze è possibile trarre utili indicazioni sia per una migliore formulazione dei quesiti del prossimo questionario, sia per una più analitica stesura delle relazioni logiche.

Per i controlli di *range* in considerazione della natura delle variabili (voci del conto economico), il piano di controllo prevedeva sulla singola impresa l'accertamento solo per alcuni valori economici pro-capite (ad addetto). Ad esempio, sui ricavi è stato effettuato l'accertamento sul 4,6% delle imprese, sulle immobilizzazioni sul 2,1% e sulla variabile calcolata "valore aggiunto" sul 2,2% delle imprese. Altri controlli di *range* sono stati rimandati alla fase successiva di analisi delle distribuzioni di "tutte" le imprese per attività economica.

Il piano di controllo ha previsto, inoltre, sulla singola impresa l'accertamento dei dati nel caso in cui alcuni valori economici pro-capite (ad esempio il fatturato per addetto) risultassero esterni a determinati valori soglia previsti per le differenti attività economiche (controlli di range).

Per le mancate risposte parziali, il processo di correzione ha interessato in particolare le variabili "ore lavorate" (a causa della mancata indicazione, in presenza di dati sull'occupazione per categoria professionale), "retribuzioni" (a causa della mancata disponibilità per categoria professionale, in presenza di dati sull'occupazione) mentre ha riguardato un numero minore di casi l'assenza delle variabili "valore e costi della produzione" e "immobilizzazioni".

Per quanto concerne le incompatibilità logiche fra le variabili, il processo di correzione ha interessato un maggior numero di imprese per le variabili "costi del personale" (per il controllo di consistenza fra le variabili di retribuzione per categoria professionale e le corrispondenti informazioni sugli addetti e sulle ore lavorate), "costi di produzione" e "valore della produzione" (per i controlli di quadratura del conto economico dell'impresa) e, in misura minore, per le variabili "addetti" (per l'unità di misura, richiesta in media annua con un decimale) ed "acquisizioni di immobilizzazioni nell'esercizio" (per il controllo di quadratura per tipo di beni).

Per le principali variabili sommantanti del conto economico, l'analisi degli errori e forzature per classi di ampiezza delle differenze in valore assoluto ed in valore percentuale rispetto al totale della variabile, ha fornito i seguenti risultati:

- per la voce "totale A - valore della produzione" posto il 6,7% delle imprese corrette automaticamente uguale a 100, 79 di esse sono state corrette per un importo minore del 5% del totale e 72 imprese per un importo minore di 5 mila euro ;

- per la voce "totale B – costi di produzione" posto il 21,7% delle imprese corrette automaticamente uguale a 100, 70 sono state corrette per un importo minore del 5% del totale e 80 imprese per un importo minore di 5 mila euro ;

- per la voce "totale costi del personale" posto il 23,2% delle imprese corrette automaticamente uguale a 100, 46 sono state corrette per un importo minore del 5% del totale e 52 per un importo minore di 5 mila euro ;

- per la voce "totale acquisizioni di immobilizzazioni nell'esercizio" posto il 2,5% delle imprese corrette automaticamente uguale a 100, 57 sono state corrette per un importo minore del 5% del totale e 70 imprese per un importo minore di 5 mila euro.

L'analisi delle imprese per "classi di qualità" ha evidenziato la seguente distribuzione percentuale delle imprese affette da errori e/o forzature (posto uguale a 100 il totale delle imprese rispondenti):

- il 14,0% con un solo errore o forzatura;
- il 12,9% con due errori e/o forzature;
- il 26,2% da 3 a 5 errori e/o forzature;
- il 20,0% da 6 a 10 errori e/o forzature;
- il 14,1% con oltre 10 errori e/o forzature.

Le imprese rispondenti per il 58,7% sono risultate formalmente corrette già in fase di acquisizione o corrette dopo procedura di correzioni "esclusivamente automatiche" (solo forzature) mentre per il restante 41,3% delle imprese è stata necessaria la fase di correzione "interattiva" (errori e forzature) da parte degli esperti collaboratori.

Prospetto 2.6 - Principali interventi per tipo di incompatibilità (a)

TIPO DI INCOMPATIBILITÀ	Percentuali medie di Interventi	
	Errori	Forzature
A- Imprese fuori osservazione (b)	0,5	-
B- Mancate risposte parziali:		
B.1. Da tabelle decisionali:		
- Assenza variabile "costi di produzione"	0,7	-
- Assenza variabile "valore della produzione"	0,7	-
- Assenza variabile "addetti-dipendenti" (con presenza variabile "retribuzioni")	-	4,6
- Assenza variabile "retribuzioni" (con presenza variabile "addetti-dipendenti")	-	7,2
- Assenza variabile "ore lavorate" (con presenza variabili "retribuzioni-addetti")	-	14,7
- Assenza variabile "contributi" e/o "trattamento fine rapporto lavoro" (con presenza variabile "retribuzioni")	-	9,2
- Assenza variabile "costi personale esterno" (con presenza variabile "personale esterno")	-	2,0
- Assenza variabile "voci di costo" (con presenza variabile "totale costi di produzione")	-	0,1
- Assenza variabile "voci fatturato-ricavi" (con presenza variabile "totale valore della produzione")	-	0,1
- Assenza variabile "voci immobilizzazioni" (con presenza variabile "totale immobilizzazioni")	-	0,1
C-Incompatibilità logiche tra variabili:		
C.1. Dalla funzione di controllo:		
- Distribuzione "totale valore della produzione" per voci contabili di entrata	3,4	6,7
- Distribuzione "totale costi di produzione" per voci contabili di uscita	7,4	21,7
- Distribuzione "totale addetti" per categoria professionale (c)	1,5	3,7
- Distribuzione "totale costo del personale" per voci retributive	-	23,2
- Distribuzione "totale acquisizioni di immobilizzazioni" per tipo investimento	0,5	2,5
- Distribuzione "totale addetti" per regione	0,2	1,3
- Distribuzione "totale costo del lavoro" per regione	0,2	1,0
C.2. Dall'operatore logico:		
- Fatturato all'esportazione rispetto al "totale ricavi delle vendite e delle prestazioni"	-	0,4
- Altri dati rispetto ad "altri ricavi e proventi"	-	6,4
- Costi personale esterno rispetto alcuni "costi per servizi"	2,2	0,3
- Spese per la protezione dell'ambiente rispetto "acquisizioni di immobilizzazioni"	-	0,2

(a) Posto uguale a 100 il totale delle imprese rispondenti.

(b) Dette imprese (con 100 addetti ed oltre) nell'esercizio di riferimento incrementano la numerosità del campo di osservazione della relativa indagine sul "sistema dei conti economici delle imprese - 100 addetti ed oltre".

(c) Il numero di imprese con interventi su tale variabile risente dell'unità di misura che è espressa in media annua con un decimale.

2.9.3 - Analisi delle distribuzioni delle imprese

Lo studio delle distribuzioni delle imprese è stato condotto distintamente per ogni divisione di attività economica, sulla base dei seguenti valori medi e rapporti caratteristici:

- Valore aggiunto per addetto;
- Fatturato per addetto;
- Percentuale del valore aggiunto sul fatturato;
- Retribuzione lorda annua per dirigente, impiegato e quadro intermedio;
- Retribuzione lorda annua per operaio;
- Retribuzione lorda annua per apprendista;
- Retribuzione lorda annua per lavoratore a domicilio;
- Percentuale contributi sulle retribuzioni lorde;
- Percentuale costi per il personale sul valore aggiunto;
- Acquisto di immobilizzazioni nell'esercizio per addetto.

Il calcolo di alcuni indicatori di posizione e l'analisi delle distribuzioni ha consentito l'individuazione di eventuali imprese con dati "anomali", da sottoporre ad ulteriori controlli ed eventualmente a correzioni interattive.

La procedura descritta ha permesso con criteri "deterministici" la correzione dei dati osservati errati e/o l'integrazione delle mancate risposte parziali, minimizzando la differenza tra dati aggregati osservati e dati aggregati reali.

2.10 - Procedure di integrazione delle mancate risposte totali

2.10.1 - Le fasi dell'integrazione

Nella rilevazione è stata introdotta una fase di integrazione delle mancate risposte totali per le imprese del campione della fascia dimensionale 20-99 addetti utilizzando le informazioni dei bilanci civilistici delle società di capitale depositati presso le Camere di commercio. Il processo di integrazione ha permesso la ricostruzione di 6.079 imprese.

La fase di integrazione delle mancate risposte totali si articola in una fase in cui si individuano i donatori (sulla base di un criterio di vicinanza che si basa sull'attività economica, la fascia dimensionale e la localizzazione territoriale), che forniscono i dati all'impresa non rispondente a meno di un fattore di ponderazione che tiene conto delle diverse dimensioni delle imprese, e una fase in cui vengono sostituiti i dati stimati con il metodo del donatore con quelli reali dichiarati dalle imprese nei bilanci civilistici: ciò avviene per le principali voci del conto economico, ovvero il fatturato, le variazioni delle rimanenze di prodotti finiti, in corso di lavorazione e semilavorati, le variazioni dei lavori in corso su ordinazioni, gli incrementi di immobilizzazioni per lavori interni, gli altri ricavi e proventi, i costi per acquisto di materie prime, sussidiarie e di consumo, i costi per servizi, i costi per godimento di beni di terzi, i costi per il personale, le quote di ammortamento, le variazioni delle materie prime sussidiarie, di consumo e di merci, gli accantonamenti, gli oneri diversi di gestione e l'utile o la perdita di esercizio.

2.10.2 - La fase di integrazione delle mancate risposte totali tramite donatore

Preliminare a questa fase è l'individuazione delle imprese che non hanno risposto alla rilevazione, attraverso l'incrocio fra la lista di spedizione ed il file di monitoraggio degli arrivi che registra, oltre alle imprese che hanno compilato il questionario, le imprese che non sono risultate reperibili all'indirizzo (indirizzo sbagliato, impresa trasferita), le imprese che hanno modificato lo stato di attività (imprese cessate, inattive, in liquidazione, soggette ad operazioni di fusione o scorporo, eccetera), le imprese fuori campo di osservazione, eccetera.

In questo modo si individuano i casi di mancate risposta totale che in generale producono una distorsione nelle stime finali a cui si cerca di porre rimedio attraverso il processo di integrazione e, successivamente, attraverso la metodologia degli stimatori di ponderazione vincolata (Paragrafo 2.11).

La procedura individua per ogni impresa non rispondente un profilo economico ad essa applicabile, desumibile all'interno di un insieme di imprese omogenee in relazione all'attività economica esercitata (a livello di quattro cifre della classificazione ateco), alla classe dimensionale e alla localizzazione regionale. Nel caso in cui non si individuano le imprese donatrici in relazione alla stratificazione descritta, la procedura opera un progressivo collassamento degli strati fin quando non individua un gruppo di imprese donatrici: dei tre caratteri di stratificazione considerati, *attività economica*, *dimensione dell'impresa*, *localizzazione geografica*, si preferisce "perdere" più velocemente gli ultimi due nelle fasi di collassamento degli strati in quanto, rispetto a tali caratteri, l'attività economica spiega in misura maggiore l'omogeneità delle imprese dal punto di vista delle variabili economiche di interesse (ovviamente in caso di ulteriore insuccesso la procedura viene iterata per gruppi o divisioni di attività economica).

Una volta individuato il gruppo di imprese donatrici per ciascun strato, si procede all'esclusione fra queste di quelle che rispetto alla variabile fatturato per addetto presentano valori estremi, tali da essere considerate come outliers.

Si procede, quindi, all'estrazione casuale, con ripetizione e probabilità uguale dell'impresa donatrice, da ciascuno strato, individuato come descritto in precedenza, e all'integrazione di tutti i dati dell'impresa non rispondente sulla base dei dati dell'impresa donatrice, ponderati sulla base del rapporto fra gli addetti dell'impresa da integrare (desunti dall'archivio aggiornato Asia) e gli addetti dell'impresa donatrice.

In pratica si suppone che le performances siano correlate alla dimensione dell'impresa e quindi tutti i dati economici vengono modificati dal rapporto sopra citato. Per quanto riguarda la struttura occupazionale ed il costo del lavoro si procede al calcolo delle ore lavorate in media da ciascuna categoria professionale dell'impresa donatrice, della quota di occupazione femminile, delle retribuzioni medie per categoria

professionale, del peso sul costo complessivo del lavoro di tutta una serie di voci (contributi sociali a carico dell'impresa, quote accantonate al trattamento di fine rapporto, provvidenze varie al personale). Si passa, poi, alla ricomposizione della struttura occupazionale per categoria professionale ed al calcolo, sulla base degli indicatori sopra citati, di tutte le variabili del questionario ed alla fine si effettua la quadratura di tutte le variabili delle diverse sezioni del questionario.

2.10.3 - La fase di integrazione delle mancate risposte totali tramite fonte amministrativa (bilanci civilistici depositati dalle imprese presso le camere di commercio).

La fase di integrazione con dati di bilancio delle sole società di capitale consiste nella sostituzione dei dati calcolati come descritto nel paragrafo precedente, con quelli realmente dichiarati nel bilancio dalle società di capitale, tramite una scomposizione dei totali nelle diverse voci parziali ed operando, in assenza di informazioni, secondo le regole del programma di correzione delle incompatibilità oppure sulla base di alcuni rapporti fra le variabili. Alla fine si effettua la quadratura di tutte le variabili delle diverse sezioni del questionario.

Alla fine del processo di integrazione delle mancate risposte totali, si procede al calcolo di indicatori di posizione sul set completo di dati (imprese rispondenti e imprese integrate) calcolati sulla base dell'attività economica. Le variabili prese in esame sono le seguenti:

Fatturato per addetto
Valore aggiunto per addetto
Investimenti per addetto
Costo del lavoro per dipendente
Ore lavorate per dipendente
Costo del lavoro orario

2.11 - Metodo di stima

2.11.1 - Descrizione generale

Il principio su cui è fondato qualsiasi metodo di stima campionaria è quello che il sottoinsieme delle unità della popolazione incluse nel campione deve rappresentare anche il sottoinsieme complementare costituito dalle rimanenti unità della popolazione stessa. Tale principio viene realizzato attribuendo a ciascuna unità inclusa nel campione un peso, che può essere visto come numero di elementi della popolazione rappresentati da detta unità.

In generale, per ottenere la stima di un totale si devono eseguire le tre seguenti operazioni:

1. determinare il peso da attribuire a ciascuna unità inclusa nel campione;
2. moltiplicare il valore relativo ad una data variabile oggetto di indagine, rilevata sulla generica unità inclusa nel campione, per il peso attribuito alla medesima unità;
3. effettuare la somma dei prodotti di cui al punto 2.

Nella presente indagine, basata su un disegno di campionamento complesso, ad uno stadio stratificato, il peso da attribuire a ciascuna unità è ottenuto in base ad una procedura articolata in più passi:

- in primo luogo, viene calcolato un peso iniziale, definito *peso diretto* (o *peso base*), determinato in funzione del disegno di campionamento come reciproco della probabilità di inclusione dell'unità campionata;
- successivamente, vengono calcolati dei fattori correttivi del peso base, che consentono sia di correggere (almeno parzialmente) la distorsione conseguente alla mancata risposta totale, sia di rispettare la condizione di uguaglianza tra alcuni parametri noti della popolazione e le corrispondenti stime campionarie;
- infine, viene determinato un peso, noto sotto il nome di *peso finale*, espresso come prodotto del peso base per i fattori correttivi.

Nel passato le stime derivanti dalle indagini di tipo economico su base campionaria venivano costruite essenzialmente in due modi:

- come semplice aggregazione dei dati campionari rilevati con l'indagine senza tentare un riporto dei dati stessi all'universo;
- come aggregazione dei dati campionari rilevati, ponderati con i pesi diretti eventualmente corretti con l'inverso del tasso di risposta.

Negli anni più recenti, tuttavia, la ricerca sviluppata dall'Istat e da altri importanti Istituti nazionali di statistica nel campo degli stimatori ha comportato una sostanziale modifica della procedura di stima, utilizzando stimatori più complessi che rientrano nella classe nota sotto il nome di stimatori di ponderazione vincolata. Tali metodi, che sono applicabili quando esistono totali noti di variabili ausiliarie correlate alle variabili oggetto di indagine, permettono di utilizzare al meglio tutta l'informazione ausiliaria disponibile.

Gli stimatori appartenenti alla suddetta classe, sotto ipotesi piuttosto generali, sono caratterizzati dalle seguenti proprietà:

- sono più efficienti dello stimatore diretto, essendo l'efficienza dei primi tanto maggiore quanto più è alta la correlazione tra le variabili ausiliarie e le variabili oggetto di indagine;
- sono approssimativamente non distorti rispetto al disegno di campionamento;
- portano a stime dei totali noti che coincidono con i valori noti di tali totali;
- attenuano l'effetto distorsivo dovuto alla presenza di mancate risposte totali;
- attenuano l'effetto distorsivo dovuto alla sottocopertura della lista da cui è selezionato il campione.

Si ricorda, infine, che nelle indagini campionarie condotte su larga scala lo stimatore diretto è distorto e scarsamente efficiente, essendo fortemente influenzato dai fenomeni di mancata risposta e di sottocopertura della lista di selezione del campione.

2.11.2 - Simbologia e parametri oggetto di stima

Prima illustrare gli aspetti matematici del metodo di stima utilizzato, è utile esprimere con notazione insiemistica i parametri e le quantità introdotte nel paragrafo 2.5.1.

Con riferimento agli elementi della popolazione e del campione, denotiamo con:

- U = l'insieme delle imprese oggetto d'indagine, ossia le imprese operanti ed attive nel corso del 2004 ;
- α = un generico dominio di interesse ;
- U_α = l'insieme delle imprese oggetto d'indagine, ossia l'impresе operanti ed attive nel corso del 2004 , appartenenti al dominio α ;
- s^* = l'insieme delle unità selezionate nel campione (con $s^* \subseteq U_L$) ;
- s = l'insieme delle imprese selezionate nel campione e rispondenti all'indagine (con $s \subseteq s^*$) ;
- s_α = il sottoinsieme di imprese del campione s appartenenti ad U_α , con $s_\alpha = s \cap U_\alpha$;
- n_α = il numero di imprese appartenenti al campione s_α .

Inoltre, con riferimento alla generica impresa, si assume che la si possa identificare univocamente con un indice k , definito dalla concatenazione del precedente indice h di strato e i di unità.

Indichiamo pertanto con le seguenti quantità:

- y_k = il valore assunto dalla caratteristica y oggetto d'indagine sull'unità k ;
- $\lambda_k(\alpha)$ = variabile dicotomica che assume valore 1 se l'unità k appartiene al dominio α e assume valore 0 altrimenti.

Mediante la simbologia appena introdotta è possibile definire, con riferimento alla generica variabile rilevata y , i seguenti totali oggetto di stima:

a) il totale riferito alle imprese appartenenti al dominio α :

$$Y_{\alpha} = \sum_{U_{\alpha}} y_k = \sum_U y_k \lambda_k(\alpha) \quad (15)$$

in cui il simbolo $\sum_U(\cdot)$ indica, con riferimento ad un generico insieme U, la sommatoria estesa a tutti gli elementi ad esso appartenenti;

b) il totale riferito a tutte le imprese operanti sul territorio nazionale:

$$Y = \sum_U y_k \quad (16)$$

2.11.3 - Procedura di stima

In generale, la stima \tilde{Y}_{α} del totale di dominio Y_{α} definito dalla (15), ottenuta con il metodo di ponderazione vincolata, ha la seguente espressione:

$$\tilde{Y}_{\alpha} = \sum_{s_{\alpha}} y_k w_k = \sum_s y_k \lambda_k(\alpha) w_k \quad (17)$$

Analogamente, la stima \tilde{Y} del totale Y viene ottenuta come semplice somma dei totali stimati per dominio:

$$\tilde{Y} = \sum_s y_k w_k \quad (18)$$

Per quanto detto nel paragrafo 2.11.1, il peso finale w_k è ottenuto mediante il prodotto di tre fattori:

$$w_k = d_k \beta_k \gamma_k \quad (19)$$

in cui:

- $d_k = \frac{N_h}{n_h^*}$ per $k \in s_h^*$, $h=1, \dots, H$, indica il peso diretto, espresso come reciproco della probabilità di inclusione;
- β_k indica il fattore di aggiustamento per mancata risposta, determinato secondo quanto illustrato nel paragrafo 2.11.4;
- γ_k indica il fattore di aggiustamento per post-stratificazione, determinato secondo quanto illustrato nel paragrafo 2.11.5.

2.11.4 - La correzione della mancata risposta totale

2.11.4.1 - Descrizione del problema

La situazione in cui una unità statistica inclusa in una indagine non fornisce risposta ad alcun quesito previsto nel modello di rilevazione viene indicata con il termine di *mancata risposta totale*. Essa può essere determinata da cause differenti, come l'impossibilità di reperire l'unità inclusa nella popolazione oggetto di studio, l'incapacità o il rifiuto da parte di quest'ultima a rispondere, errori nella formulazione delle liste da cui si estraggono gli appartenenti al campione eccetera.

La presenza di mancate risposte totali è un problema comune sia alle indagini campionarie che a quelle censuarie; si possono applicare tecniche più o meno sofisticate per ridurne l'intensità, pur risultando difficile, se non impossibile, eliminarne totalmente la presenza.

Gli effetti più rilevanti che hanno le mancate risposte totali sui risultati finali sono due:

- in generale, riducono la quantità di informazione disponibile; nel caso di indagini campionarie, la riduzione produce un incremento dell'errore di campionamento e di conseguenza una minore precisione delle stime;
- apportano distorsioni nelle stime delle variabili oggetto di indagine quando il meccanismo di risposta è non casuale (può esservi, ad esempio, una maggiore propensione alla risposta in taluni sottogruppi della popolazione indagata, caratterizzati da particolari valori delle variabili d'indagine).

Le distorsioni che la presenza delle mancate risposte inducono sui risultati del processo di produzione dell'informazione statistica vengono prevenute od attenuate intervenendo in diverse fasi del processo. In particolare, il trattamento delle mancate risposte in fase di stima finale, come si è già accennato, prevede la riduzione dell'effetto distorsivo sulle stime attraverso l'uso di opportuni pesi correttivi da applicare al campione di rispondenti.

2.11.4.2 - La correzione per mezzo delle classi di omogeneità

L'uso dei pesi per effettuare la correzione della mancata risposta totale poggia sull'assunzione di un meccanismo aleatorio che governa il fenomeno della mancata risposta; tale meccanismo fa sì che il campione s^* si ripartisca nel sottoinsieme dei rispondenti, s , e del suo complemento, formato dalle unità non rispondenti; il peso attribuito ad ogni elemento di s sta ad indicare il numero di unità non rispondenti rappresentato da ciascuna unità che ha risposto all'indagine¹¹.

Per trattare tale problema ci si riconduce alla teoria del campionamento in due fasi¹², considerando la selezione dei non rispondenti come un'ulteriore passo di campionamento governato da un processo di tipo bernoulliano, ignoto al ricercatore.

Formalmente, sia r_k una variabile dicotomica uguale ad 1 se l'unità campionaria k (con $k \in s^*$) è rispondente ed uguale a 0 altrimenti; nell'approccio considerato si assume che le r_k siano variabili di Bernoulli mutualmente indipendenti con valore atteso pari a θ_k che indica la probabilità dell'unità k di essere inclusa in s ($\theta_k = pr(k \in s)$) ed è denominato *probabilità di risposta* di k .

Nel caso in cui le probabilità di risposta θ_k fossero note per tutti i rispondenti, una stima corretta del totale Y sarebbe ottenuta tramite lo stimatore:

$$\tilde{Y}_c = \sum_s y_k d_k \theta_k^{-1} \quad (20)$$

Essendo però le quantità θ_k generalmente ignote, occorre darne una stima introducendo un modello di risposta, ossia formalizzando una relazione funzionale tra le θ_k ed un insieme di variabili esplicative disponibili sia per i rispondenti che per i non rispondenti.

Il modello più semplice di mancata risposta è quello che assume la probabilità θ_k costante per tutte le unità della popolazione; tale ipotesi è poco realistica (a causa della effettiva eterogeneità tra le unità delle probabilità di risposta θ_k) e non è realmente utile all'eliminazione della componente distorsiva.

Viceversa, la definizione di probabilità variabili a livello di singola unità può dar luogo a stimatori poco stabili e alla generazione di valori anomali, a causa della variabilità associata alle stime delle probabilità θ_k per i singoli rispondenti.

La strategia generalmente adottata per superare tali problemi consiste in una soluzione intermedia in cui si assume un modello di mancata risposta che prevede una probabilità costante dei θ_k per sottoinsiemi di unità del campione s^* .

¹¹ Ballin, M., e P.D. Falorsi, e B. Moreschi, e A. Pavone, e A.C. Seeber, e T. Tuoto. "Aspetti metodologici connessi alla produzione delle stime per il Censimento delle Istituzioni non profit". In *Istituzioni nonprofit in Italia - I risultati della prima rilevazione censuaria - anno 1999*. Roma: Istat, 2001. (Collana Informazioni).

¹² Särndal, C.E., e B. Swensson, e J. Wretman. *Model Assisted Survey Sampling*. New York: Springer-Verlag, 1992.

Indicando con $s_{(g)}^*$, $g=1, \dots, G$, il generico sottoinsieme di una partizione di s^* , di dimensione pari a $n_{(g)}^*$, (essendo $s^* \equiv \bigcup_{g=1}^G s_{(g)}^*$; $s_{(g)}^* \cap s_{(g')}^* = \emptyset$, per $g \neq g'$) il modello si esplicita nel modo seguente:

$$\begin{aligned} pr(k \in s | s^*) &= \theta_g > 0 && \text{per ogni } k \in s_{(g)}^* \\ pr((k,l) \in s | s^*) &= pr(k \in s | s^*) pr(l \in s | s^*) && \text{per ogni } k \neq l; (k,l) \in s_{(g)}^*. \end{aligned} \quad (21)$$

Tutte le unità appartenenti allo stesso gruppo, o *classe di omogeneità*, $s_{(g)}^*$, hanno uguale probabilità di risposta, mentre gruppi distinti hanno probabilità differenti e le risposte sono indipendenti. Una volta determinata la partizione $\{s_{(g)}^*\}$, $g=1, \dots, G$, condizionando a s^* , al vettore $\mathbf{n}^* = (n_{(1)}^*, \dots, n_{(g)}^*, \dots, n_{(G)}^*)'$, contenente il numero di unità del campione appartenenti a ciascuna cella di ponderazione, ed al corrispondente vettore osservato $\mathbf{n} = (n_{(1)}, \dots, n_{(g)}, \dots, n_{(G)})'$ di unità rispondenti in ciascuna cella, una stima delle probabilità di risposta per le unità appartenenti alla cella di ponderazione g -esima è data da:

$$pr(k \in s | s^*, \mathbf{n}^*, \mathbf{n}) = \theta_k = \beta_k^{-1} = \frac{n_{(g)}}{n_{(g)}^*} \quad \text{per } k = l \in s_{(g)}; \quad (g = 1, \dots, G). \quad (22)$$

Pertanto, valendo il modello (21), uno stimatore corretto del totale Y è dato da:

$$\tilde{Y} = \sum_{g=1}^G \frac{n_{(g)}^*}{n_{(g)}} \sum_{s_{(g)}} y_k d_k = \sum_s y_k d_k \beta_k. \quad (23)$$

Nella presente indagine la definizione delle classi di omogeneità è avvenuta considerando come esplicative del fenomeno della mancata risposta quelle variabili che descrivono il settore di attività dell'impresa, la sua dimensione in termini di addetti e la sua localizzazione. Per la determinazione dei correttori si è quindi operato come segue:

1. le unità del campione sono state suddivise negli strati elementari, definiti dalla combinazione delle modalità delle variabili di archivio *regione*, *classe di attività economica (Ateco a 4 cifre)*, *classe dimensionali di addetti*;
2. a ciascuna unità del campione è stata attribuita una probabilità di risposta individuale calcolata, pari al tasso risposta osservato nello strato elementare. La presenza di poche unità rispondenti in alcuni strati ha causato la determinazione di stime poco affidabili, se non pari a 0 nel caso di caduta dell'intero strato;
3. sono state quindi definite 10 *celle di ponderazione* aventi come estremi i *decili* della distribuzione delle probabilità di risposta stimate al punto precedente;
4. alle unità *rispondenti* di ciascuna cella di ponderazione è stato attribuito un fattore correttivo β_k , pari al reciproco del tasso di risposta della cella di appartenenza.

2.11.5 - Il calcolo dei pesi finali

Per il calcolo dei pesi finali si è adottato uno stimatore di ponderazione vincolata. Si tratta di uno stimatore appartenente alla più vasta categoria degli stimatori di regressione generalizzata, i quali permettono di

migliorare l'efficienza delle stime quando si dispone, relativamente alla popolazione oggetto d'indagine, di totali noti di variabili ausiliarie correlate con le variabili d'interesse. In particolare, le variabili ausiliarie rappresentano le variabili esplicative di un modello di regressione, sottostante allo stimatore, con cui si esplicita una relazione funzionale tra le prime e le variabili d'interesse.

Per la costruzione dello stimatore è importante chiarire il concetto di *gruppo di riferimento del modello*:

con questo termine si intende una partizione $\left\{ U(p) : U_L \equiv \bigcup_{p=1}^P U(p); U(p) \cap U(p') = \emptyset, p \neq p' \right\}$ della

popolazione U_L rispetto alla quale:

- sono noti i totali di popolazione di un insieme di variabili ausiliarie;
- viene definito il modello di regressione su cui si costruisce lo stimatore.

Definiamo inoltre:

- $\mathbf{x}_k = (x_{k1}, \dots, x_{k1}, \dots, x_{km})'$ il vettore colonna di m variabili ausiliarie osservato sull'unità k ;
- $\mathbf{X}_{U(p)} = \sum_{U(p)} \mathbf{x}_k$ i totali delle m variabili ausiliarie riferiti al sottoinsieme $U(p)$ della popolazione.

Sia $s(p)$ il sottoinsieme del campione dei rispondenti, di numerosità pari a $n(p)$, relativo al gruppo p -esimo della partizione.

Per ciascun sottoinsieme $s(p)$, gli $n(p)$ pesi finali $\{(w_k = d_k \beta_k \gamma_k); \text{per } k \in s(p)\}$ sono ottenuti come soluzione di un sistema di minimo vincolato definito nel modo seguente:

$$\begin{cases} \text{Min} \left[\sum_{s(p)} c_k G(d_k \beta_k; d_k \beta_k \gamma_k) \right] \\ \sum_{s(p)} \mathbf{x}_k d_k \beta_k \gamma_k = \mathbf{X}_{U(p)} \end{cases} \quad p=1, \dots, P \quad (24)$$

dove:

- c_k indica una costante nota connessa alla dimensione dell'unità (nel caso in oggetto tale variabile è stata posta uguale al numero di addetti dell'impresa risultanti dall'archivio di selezione);
- $G(d_k \beta_k; d_k \beta_k \gamma_k)$ è la funzione di distanza logaritmica¹³ tra i *pesi finali* $\{(w_k = d_k \beta_k \gamma_k); \text{per } k \in s(p)\}$ e i *pesi diretti corretti per la mancata risposta* $\{d_k \beta_k; \text{per } k \in s(p)\}$ definita da:

$$G(d_k \beta_k; d_k \beta_k \gamma_k) = d_k \beta_k \gamma_k \ln(\gamma_k) - d_k \beta_k \gamma_k + d_k. \quad (25)$$

L'obiettivo è quello di individuare un vettore di pesi finali $\{(w_k = d_k \beta_k \gamma_k); \text{per } k \in s(p)\}$ che consenta di rispettare il sistema di vincoli (24) e che contemporaneamente modifichi il meno possibile l'insieme dei *pesi diretti corretti per la mancata risposta*.

La soluzione del sistema di minimo vincolato definito dalle equazioni precedenti viene trovata adottando il metodo dei moltiplicatori di Lagrange.

2.11.5.1 - Modalità applicative della procedura di costruzione dei pesi

Il calcolo dei pesi finali sui dati relativi all'indagine è stato effettuato impostando il problema di ottimizzazione nel modo seguente:

¹³ Falorsi, P.D., e S. Falorsi. "Un metodo di stima generalizzato per le indagini sulle famiglie e sulle imprese". *Rapporto di ricerca*, n. 13. Università di Bologna, 1995. (Quaderni CON PRI).

- il gruppo di riferimento del modello $\{U(p)\}$ è definito dalla divisione di attività economica (Ateco a 2 cifre) dell'Archivio aggiornato;
- i domini di interesse α sono rappresentati dai domini delle tre tipologie descritte nel prospetto 1;
- le variabili ausiliarie sono identificate da $x_1 = \text{numero di imprese}$ e $x_2 = \text{numero di addetti}$;
- per ogni unità, il vettore \mathbf{x}_k delle variabili ausiliarie è stato definito come segue:
 $\mathbf{x}'_k = \left((1) \mathbf{x}'_k, (2) \mathbf{x}'_k \right)$, concatenazione di due vettori $(1) \mathbf{x}'_k$ e $(2) \mathbf{x}'_k$ del tipo:
 $(1) \mathbf{x}'_k = \{\lambda_k(j_d)\}$, $(2) \mathbf{x}'_k = \{\alpha_k \lambda_k(j_d)\}$ con $d=1, \dots, 3$; $j=1, \dots, J_d$,
dove, in base alle informazioni dell'Archivio:
 $\lambda_k(j_d)$ è una variabile indicatrice che assume valore 1 se l'unità k-esima appartiene al dominio j_d e 0 altrimenti;
 α_k è il numero di addetti dell'impresa k;
- per ciascun gruppo di riferimento $\{U(p)\}$, ossia per le varie modalità della variabile Ateco2, i totali noti (calcolati sui dati d'Archivio) che compaiono a secondo membro dell'equazione nel sistema (24), sono rappresentati da:

$$X_{U(p)} = \sum_{U(p)} \mathbf{x}'_k = \sum_{U(p)} \lambda_k(j_1), \dots, \sum_{U(p)} \lambda_k(j_3), \sum_{U(p)} \alpha_k \lambda_k(j_1), \dots, \sum_{U(p)} \alpha_k \lambda_k(j_3).$$

Nel complesso, il numero di totali noti che compaiono nel problema di ottimizzazione risulta pari a 182. Nonostante l'elevato numero di vincoli da soddisfare, la procedura ha dato ottimi risultati in termini di convergenza tra stime finali e totali di archivio.

2.11.5.2 - Alcune statistiche di copertura

Nel prospetto 7 sono messi a confronto, in forma sintetica a livello di sezione di attività economica e di ripartizione territoriale, il numero delle imprese con 1-99 addetti attive nel 2004 con quelle rispondenti all'indagine. Le imprese appartenenti alle regioni Valle d'Aosta, Piemonte, Lombardia e Liguria fanno parte della ripartizione Italia nord-occidentale, le imprese del Trentino Alto Adige, Veneto, Friuli ed Emilia Romagna della ripartizione Italia nord-orientale, le imprese della Toscana, Umbria, Marche e Lazio della ripartizione Italia centrale, le imprese dell'Abruzzo, Molise, Campania, Puglia, Basilicata e Calabria della ripartizione Italia meridionale e le imprese della Sicilia e Sardegna della ripartizione Italia insulare.

Il prospetto 7 evidenzia che il grado di copertura dell'indagine è stato in media dell'1,3 per cento, con 56.080 imprese utilizzati nell'inferenza (50.001 rispondenti e 6.079 integrate dalla fonte bilanci civilistici) rispetto ad un universo di 4.257.569 imprese.

A livello di attività economica, il grado di copertura più elevato si registra per le sezioni E "produzione e distribuzione di energia elettrica, gas e acqua" (24,6 per cento), C "estrazione di minerali" (13,7 per cento), M "istruzione" (5,6 per cento), mentre risulta relativamente più basso per le sezioni H "alberghi e ristoranti" (0,4 per cento), F "costruzioni" e N "sanità e servizi sociali" (0,5 per cento). Un'analisi per ripartizione territoriale mostra che il grado di copertura è superiore alla media nell'Italia nord-orientale (1,7 per cento) e nord-occidentale (1,4 per cento), mentre risulta al di sotto della media nell'Italia centrale (1,2 per cento), meridionale ed insulare (1,0 per cento).

Il differente grado di copertura riflette le scelte metodologiche effettuate in sede di estrazione del campione, come descritto nel paragrafo 2.5, e risente anche del fenomeno della mancata risposta totale. L'effetto distorsivo delle mancate risposte totali è attenuato dall'utilizzo della metodologia di riporto dei dati all'universo basata sugli stimatori di ponderazione vincolata, descritta nel paragrafo 2.11.

Prospetto 7 - Imprese con 1-99 addetti dell'universo e del campione effettivo per ripartizione territoriale e sezione di attività economica - Anno 2004

SEZIONI DI ATTIVITA ECONOMICA	Ripartizioni territoriali					ITALIA
	Nord-occidentale	Nord-orientale	Centrale	Meridionale	Insulare	
C-estrazione di minerali						
- universo	814	714	727	674	564	3.493
- campione effettivo	122	147	89	71	49	478
- grado di copertura	15,0	20,6	12,2	10,5	8,7	13,7
D-attività manifatturiere						
- universo	156.964	125.219	106.690	92.889	37.875	519.637
- campione effettivo	6.648	6.084	4.343	3.205	1.270	21.550
- grado di copertura	4,2	4,9	4,1	3,5	3,4	4,1
E-prod. e distrib. di energia elettrica, gas e acqua						
- universo	858	513	355	372	288	2.386
- campione effettivo	172	164	106	88	58	588
- grado di copertura	20,0	32,0	29,9	23,7	20,1	24,6
F-costruzioni						
- universo	176.757	136.318	113.680	92.685	43.272	562.712
- campione effettivo	907	913	487	425	195	2.927
- grado di copertura	0,5	0,7	0,4	0,5	0,5	0,5
G-commercio e riparazioni						
- universo	318.003	235.458	257.432	310.183	137.925	1.259.001
- campione effettivo	3.014	2.699	1.815	1.573	674	9.775
- grado di copertura	0,9	1,1	0,7	0,5	0,5	0,8
H-alberghi e ristoranti						
- universo	67.549	64.879	54.188	50.134	22.282	259.032
- campione effettivo	228	322	229	204	73	1.056
- grado di copertura	0,3	0,5	0,4	0,4	0,3	0,4
I-trasporti, magazzinaggio e comunicazioni						
- universo	46.196	39.943	31.875	25.507	12.777	156.298
- campione effettivo	921	890	593	436	335	3.175
- grado di copertura	2,0	2,2	1,9	1,7	2,6	2,0
J-intermediazione monetaria e finanziaria						
- universo	20.068	12.846	14.253	10.422	4.384	61.973
- campione effettivo	144	374	162	127	40	847
- grado di copertura	0,7	2,9	1,1	1,2	0,9	1,4
K-attività immobiliari, noleggio, ricerca, professionali						
- universo	320.495	213.422	215.727	152.115	66.704	968.463
- campione effettivo	3.815	2.512	1.645	1.119	624	9.715
- grado di copertura	1,2	1,2	0,8	0,7	0,9	1,0
M-istruzione						
- universo	4.469	3.030	3.360	3.791	1.683	16.333
- campione effettivo	204	241	199	191	77	912
- grado di copertura	4,6	8,0	5,9	5,0	4,6	5,6
N-sanità e servizi sociali						
- universo	61.870	41.015	48.465	40.363	20.442	212.155
- campione effettivo	290	276	212	220	90	1.088
- grado di copertura	0,5	0,7	0,4	0,5	0,4	0,5
O-altri servizi pubblici e personali						
- universo	68.898	47.453	55.212	45.214	19.309	236.086
- campione effettivo	943	933	1.021	790	282	3.969
- grado di copertura	1,4	2,0	1,8	1,7	1,5	1,7
TOTALE ITALIA						
- universo	1.242.941	920.810	901.964	824.349	367.505	4.257.569
- campione effettivo	17.408	15.555	10.901	8.449	3.767	56.080
- grado di copertura	1,4	1,7	1,2	1,0	1,0	1,3

3 - Calcolo e presentazione sintetica degli errori campionari (*)

3.1 - Espressione della varianza campionaria

Allo scopo di derivare l'espressione della varianza della generica stima \tilde{Y}_α è necessario considerare i vari passi con cui viene costruita la stima in oggetto:

1. un passo di campionamento corrispondente alla selezione delle unità;
2. un passo di correzione per le mancate risposte totali, coincidente con il calcolo dei fattori β_k ;
3. un passo di post-stratificazione in cui vengono determinati i fattori γ_k .

Si esamini ora come questi tre passi determinano l'espressione della varianza.

Relativamente al passo 1, essendo stato adottato un disegno ad uno stadio stratificato, la varianza può essere determinata in base a quanto illustrato nei testi standard sul campionamento (Deville, J.C. e C.E. Särndal, 1992), dove le probabilità di inclusione, semplici e congiunte, sono date da:

$$\text{pr}[(k, l) \in s^*] = \pi_{kl} = \begin{cases} \pi_k = \frac{n_h^*}{N_h} = d_k^{-1} & \text{per } k = l \in U_h; & (h = 1, \dots, H) \\ \frac{n_h^*}{N_h} \frac{n_{h'}^* - 1}{N_{h'} - 1} & \text{per } k \in U_h, l \in U_{h'}, k \neq l & (h = 1, \dots, H) \\ \frac{n_h^*}{N_h} \frac{n_{h'}^* - 1}{N_{h'} - 1} & \text{per per } k \in U_h, l \in U_{h'}, h \neq h'; & [(h, h') = 1, \dots, H] \end{cases} \quad (1)$$

in cui si è indicato con U_h ($U_h \subseteq U_L$) la popolazione (riportata nella lista di campionamento) delle imprese appartenenti allo strato h .

Relativamente al punto (2), è utile rifarsi alla teoria del campionamento in due fasi; in base a tale approccio, un'espressione esplicita della varianza può essere derivata sotto le seguenti ipotesi:

1. il campionamento di seconda fase, quello cioè in cui vengono selezionati i rispondenti, è assimilabile a un campionamento di Bernoulli stratificato in cui gli strati sono dati dalle *celle di ponderazione*;
2. l'espressione della varianza è derivata in un contesto condizionale, in cui vengono fissati:
 - il campione di prima fase s^* ;
 - la struttura delle celle di ponderazione $s^*_{(g)}$ ($g=1, \dots, G$), ciascuna delle quali costituita da $n^*_{(g)}$ unità delle quali $n_{(g)}$ effettivamente rispondenti; tale struttura è riassumibile dai vettori $\mathbf{n}^* = (n^*_{(1)}, \dots, n^*_{(g)}, \dots, n^*_{(G)})'$ e $\mathbf{n} = (n_{(1)}, \dots, n_{(g)}, \dots, n_{(G)})'$.

Utilizzando tale approccio condizionato, è intuitivo dimostrare che il campionamento di *seconda fase* può essere visto come un campionamento stratificato senza reimmissione e a probabilità uguali nell'ambito di ciascuno strato, individuato dalla cella di ponderazione; di conseguenza le probabilità condizionate semplici e congiunte di risposta sono date da:

(*) Piero Demetrio Falorsi ha coordinato il lavoro scientifico e ha redatto il paragrafo 3.1; Cristina Casciano ha calcolato le varianze; Salvatore Filiberti ha redatto il paragrafo 3.2; Antonio Pavone ha redatto il paragrafo 3.3 e ha sviluppato la ricerca finalizzata all'individuazione dei modelli per la presentazione sintetica degli errori campionari.

$$\text{pr}[(k,l) \in s | s^*, \mathbf{n}^*, \mathbf{n}] = \theta_{kl} = \begin{cases} \theta_k = \frac{n_{(g)}}{n_{(g)}^*} = \beta_k^{-1} & \text{per } k=l \in s_{(g)}; & (g=1, \dots, G) \\ \frac{n_{(g)} n_{(g)}^* - 1}{n_{(g)}^* n_{(g)}^* - 1} & \text{per } k \in s_{(g)}, l \in s_{(g)}, k \neq l & (g=1, \dots, G) \\ \frac{n_{(g)} n_{(g')}^*}{n_{(g)}^* n_{(g')}^*} & \text{per } k \in s_{(g)}, l \in s_{(g')}, g \neq g'; & (g, g')=1, \dots, G \end{cases} \quad (2)$$

Relativamente al punto (3), è utile rifarsi all'importante risultato asintotico del lavoro di Deville e Särndal (1992), che dimostra che tutti gli stimatori di ponderazione vincolata tendono allo stimatore di regressione generalizzato. In base a tale risultato, si può dimostrare che la varianza dello stimatore \tilde{Y}_α può essere approssimata dall'espressione linearizzata dello stimatore di regressione generalizzata:

$$\text{Var}(\tilde{Y}_\alpha) \cong \text{Var}\left(\sum_{s_\alpha} e_k d_k \beta_k\right) = \text{Var}\left(\sum_s e_k d_k \beta_k \lambda_k(\alpha)\right), \quad (3)$$

in cui

$$e_k = y_k - \mathbf{B}'_{(p)} \mathbf{x}_k \quad \text{per } k \in U(p) \quad p=1, \dots, P \quad (4)$$

dove

$$\mathbf{B}_{(p)} = \left[\sum_{U(p)} \frac{\mathbf{x}_k \mathbf{x}'_k}{c_k} \right]^{-1} \sum_{U(p)} \frac{\mathbf{x}_k y_k}{c_k}. \quad (5)$$

Utilizzando il teorema sulla decomposizione della varianza, l'espressione (3) può esplicitarsi come segue:

$$\begin{aligned} \text{Var}\left(\sum_{s_\alpha} e_k d_k \beta_k\right) &= \\ &= \text{Var}\left[E\left(\sum_s e_k d_k \beta_k \lambda_k(\alpha) | s^*, \mathbf{n}, \mathbf{n}^*\right)\right] + E\left[\text{Var}\left(\sum_s e_k d_k \beta_k \lambda_k(\alpha) | s^*, \mathbf{n}, \mathbf{n}^*\right)\right]. \end{aligned} \quad (6)$$

Tenendo presenti la (2) e la (3), la (6) può essere espressa nella forma:

$$\begin{aligned} \text{Var}(\tilde{Y}_\alpha) &\cong \text{Var}\left(\sum_{s_\alpha} e_k d_k \beta_k\right) = \\ &= \sum \sum_U (\pi_{kl} - \pi_k \pi_l) \frac{e_k}{\pi_k} \frac{e_l}{\pi_l} + E\left[\sum_{g=1}^G \frac{n_{(g)}^* (n_{(g)}^* - n_{(g)})}{n_{(g)}} S_{\hat{e}(\alpha, g)}^2 | s^*, \mathbf{n}, \mathbf{n}^*\right] = \\ &= \sum_{h=1}^H \frac{N_h (N_h - n_h^*)}{n_h^*} S_{e(\alpha, h)}^2 + E\left[\sum_{g=1}^G \frac{n_{(g)}^* (n_{(g)}^* - n_{(g)})}{n_{(g)}} S_{\hat{e}(\alpha, (g))}^2 | s^*, \mathbf{n}, \mathbf{n}^*\right] \end{aligned} \quad (7)$$

$$S_{e(\alpha, h)}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(e_{hi} \lambda_{hi}(\alpha) - \frac{1}{N_h} \sum_{i=1}^{N_h} e_{hi} \lambda_{hi}(\alpha) \right)^2 \quad (8)$$

$$S_{\hat{e}(\alpha, (g))}^2 = \frac{1}{n_{(g)}^* - 1} \sum_{s_{(g)}^*} \left(e_k d_k \lambda_k(\alpha) - \frac{1}{n_{(g)}^*} \sum_{s_{(g)}^*} e_k d_k \lambda_k(\alpha) \right)^2. \quad (9)$$

L'espressione (7) è interessante in quanto mostra che la varianza della stima \tilde{Y}_α è ottenibile come somma di due addendi, il primo dei quali esprime la varianza della strategia di campionamento prescelta (definita dalla coppia rappresentata da disegno di campionamento stratificato e stimatore di calibrazione) in assenza di mancate risposte; mentre il secondo addendo rappresenta l'incremento di varianza dovuto al processo di mancata risposta.

La varianza della stima \tilde{Y}_α può essere derivata dall'espressione (7), ponendo la variabile $\lambda_k(\alpha)$ identicamente uguale a 1.

3.2 - Stima della varianza campionaria

Una stima corretta della (7) è data da

$$\tilde{\text{Var}}\left(\tilde{Y}_\alpha\right) = \sum_s \sum_{s'} \frac{(\pi_{kl} - \pi_k \pi_l) \tilde{e}_k \gamma_k \tilde{e}_l \gamma_l}{\pi_{kl} \theta_{kl} \pi_k \pi_l} + \sum_{g=1}^G \frac{n_{(g)}^* (n_{(g)}^* - n_{(g)})}{n_{(g)}} \tilde{S}_{\tilde{e}(\alpha), (g)}^2 \quad (10)$$

dove

$$\tilde{S}_{\tilde{e}(\alpha), (g)}^2 = \frac{1}{n_{(g)} - 1} \sum_{s(g)} \left(\tilde{e}_k d_k \gamma_k \lambda_k(\alpha) - \frac{1}{n_{(g)}} \sum_{s(g)} \tilde{e}_k d_k \gamma_k \lambda_k(\alpha) \right)^2, \quad (11)$$

$$\tilde{e}_k = y_k - \tilde{\mathbf{B}}'_{(p)} \mathbf{x}_k \quad \text{per } k \in s(p) \quad p=1, \dots, P$$

essendo

$$\tilde{\mathbf{B}}_{(p)} = \left[\sum_{s(p)} \frac{\mathbf{x}_k \mathbf{x}'_k}{c_k} w_k \right]^{-1} \sum_{s(p)} \frac{\mathbf{x}_k y_k}{c_k} w_k. \quad (12)$$

Il calcolo della varianza (10) risulta piuttosto complicato, in quanto il primo addendo è definito come somma dei prodotti di tutte le coppie di unità rispondenti. Nel caso dell'indagine in oggetto, essendo le celle di ponderazione di dimensione numerosa, le probabilità congiunte di risposta delle coppie di unità appartenenti alla stessa cella possono essere approssimate dal prodotto delle probabilità di risposta semplici, ossia:

$$\theta_{kl} = \frac{n_{(g)} n_{(g)}^* - 1}{n_{(g)}^* n_{(g)}^* - 1} \cong \left(\frac{n_{(g)}}{n_{(g)}^*} \right)^2 \quad \text{per } k \in s(g), l \in s(g), k \neq l \quad g=1, \dots, G. \quad (13)$$

Conseguentemente, una stima approssimativamente corretta della (7) può essere ottenuta dall'espressione:

$$\tilde{\text{Var}}_{\text{app}}\left(\tilde{Y}_\alpha\right) = \sum_{h=1}^H \frac{N_h (N_h - n_h^*)}{n_h^*} \tilde{S}_{e/\theta(\alpha), h}^2 + \sum_{g=1}^G \frac{n_{(g)}^* (n_{(g)}^* - n_{(g)})}{n_{(g)}} \tilde{S}_{\tilde{e}(\alpha), (g)}^2 \quad (14)$$

in cui

$$\tilde{S}_{e/\theta(\alpha), h}^2 = \frac{1}{n_h - 1} \sum_{s_h} \left(\frac{e_k}{\theta_k} \lambda_k(\alpha) - \frac{1}{n_h} \sum_{s_h} \frac{e_k}{\theta_k} \lambda_k(\alpha) \right)^2. \quad (15)$$

L'espressione (14) è stata quella effettivamente utilizzata per il calcolo delle varianze dell'indagine.

Il calcolo della varianza della stima \tilde{Y}_α può essere derivata dall'espressione (15), ponendo la variabile $\lambda_k(\alpha)$ identicamente uguale a 1.

3.3 - Presentazione sintetica degli errori campionari

Per esprimere il grado di precisione di una statistica è necessario esaminare il corrispettivo errore di campionamento. Per una generica stima \tilde{Y}_α di un totale, l'errore di campionamento assoluto è definito da:

$$\tilde{\sigma}(\tilde{Y}_a) = \sqrt{\tilde{Var}(\tilde{Y}_a)} \quad (16)$$

Se tutti i possibili campioni sono osservati in simili condizioni di campionamento, l'intervallo di confidenza di \tilde{Y}_a è definito come:

$$\tilde{Y}_a \pm k \tilde{\sigma}(\tilde{Y}_a) \quad (17)$$

dove k è una costante moltiplicativa che dipende dal livello di fiducia scelto. Con un livello di fiducia pari a 0,95, k è pari a circa 2. In tal caso si può essere confidenti che l'intervallo ottenuto in questo modo includa il vero valore del parametro della popolazione, per il 95% di tutti possibili campioni.

La stima dell'errore di campionamento è esprimibile anche in termini relativi nel modo seguente:

$$\tilde{\varepsilon}(\tilde{Y}_a) = \frac{\tilde{\sigma}(\tilde{Y}_a)}{\tilde{Y}_a} \quad (18)$$

L'errore standard solitamente si indica anche come coefficiente di variazione, ossia come percentuale del valore stimato.

L'impatto sull'errore standard del disegno campionario, rispetto al campionamento casuale semplice (CSS), è invece riassunto dalla statistica $\tilde{d}eff(\tilde{Y}_a)$ (*effetto del disegno*):

$$\tilde{d}eff(\tilde{Y}_a) = \frac{\tilde{Var}(\tilde{Y}_a)}{\tilde{Var}_{css}(\tilde{Y}_a)} \quad (19)$$

Tuttavia, nell'ambito delle rilevazioni statistiche di un istituto nazionale di statistica, riportare gli errori di campionamento non è sempre d'agevole realizzazione, in quanto, in genere, le stime sono prodotte per un ampio ammontare di variabili o caratteristiche di studio della popolazione, con un elevato numero di domini statistici. In tal caso la loro diffusione, con gli usuali metodi di stampa, appesantirebbe notevolmente le tavole di pubblicazione. Inoltre, non sarebbero disponibili gli errori relativi alle altre stime deducibili da quelle diffuse che l'utente vorrebbe poter consultare.

In tali casi, ottenere gli errori campionari in modo indiretto è da considerare una soluzione efficace. A tal fine si adotta una procedura che si articola in due passi: (i) per ciascun dominio d'interesse, alcune stime dell'indagine e le corrispondenti varianze campionarie sono calcolate in modo diretto; (ii) il legame funzionale, tra le stime riferite ai diversi domini, è modellato in modo adeguato.

L'interesse principale per questo criterio indiretto di stima della varianza campionaria risiede nell'opportunità di fornire all'utente una tavola di parametri che permetta, attraverso una formula, di attribuire un probabile errore di campionamento a qualsiasi valore assunto da una precisata variabile d'interesse. Per di più, come osserva Valliant (1987), tale criterio produce stime delle varianze più stabili e precise, rispetto a quelle ottenibili utilizzando una procedura diretta.

Nell'ambito della letteratura specializzata sull'argomento, sono state sviluppate numerose funzioni generalizzate della varianza campionaria, soprattutto per quanto riguarda le stime costituite da frequenze (assolute o relative) di variabili categoriche (vedi in tal senso Wolter 1985, capitolo 5 e Russo 1987). Nel caso di stime composte da totali di variabili quantitative, questi modelli sono di non semplice applicazione, quando è accertata un'elevata variabilità del $\tilde{d}eff(\tilde{Y}_a)$.

Scopo della presente sezione è derivare un idoneo modello predittivo per la presentazione sintetica degli errori di campionamento di variabili quantitative, quando è accertata una considerevole variabilità della statistica $\tilde{d}eff(\tilde{Y}_a)$ tra i domini (Pavone e Russo, 2004; Casciano., Falorsi., Filiberti, Pavone, Siesto, 2006; Falorsi, Filiberti, Pavone, 2006).

Il lavoro è così articolato: nel paragrafo 3.3.1 si descrive la relazione funzionale tra stime di totali e i corrispondenti errori campionari, secondo un costrutto formale; nel paragrafo 3.3.2 la modellistica è adattata alle stime ottenute dalle indagini sui Conti Economici e Finanziari delle Imprese; nel paragrafo 3.3.3 sono riportate alcune valutazioni di sintesi del livello di precisione delle stime.

3.3.1 - Concetti di base per la costruzione di funzioni generalizzate per la presentazione sintetica degli errori di campionamento

L'osservazione empirica mostra che l'errore di campionamento relativo è funzione decrescente dell'ampiezza dei totali. In Wolter (1985) sono richiamate tre funzioni regressive in grado di esibire questa proprietà:

$$\tilde{\varepsilon}(\tilde{Y}_a)^2 = \beta_1 + \frac{\beta_2}{Y_a} + u \tag{20a}$$

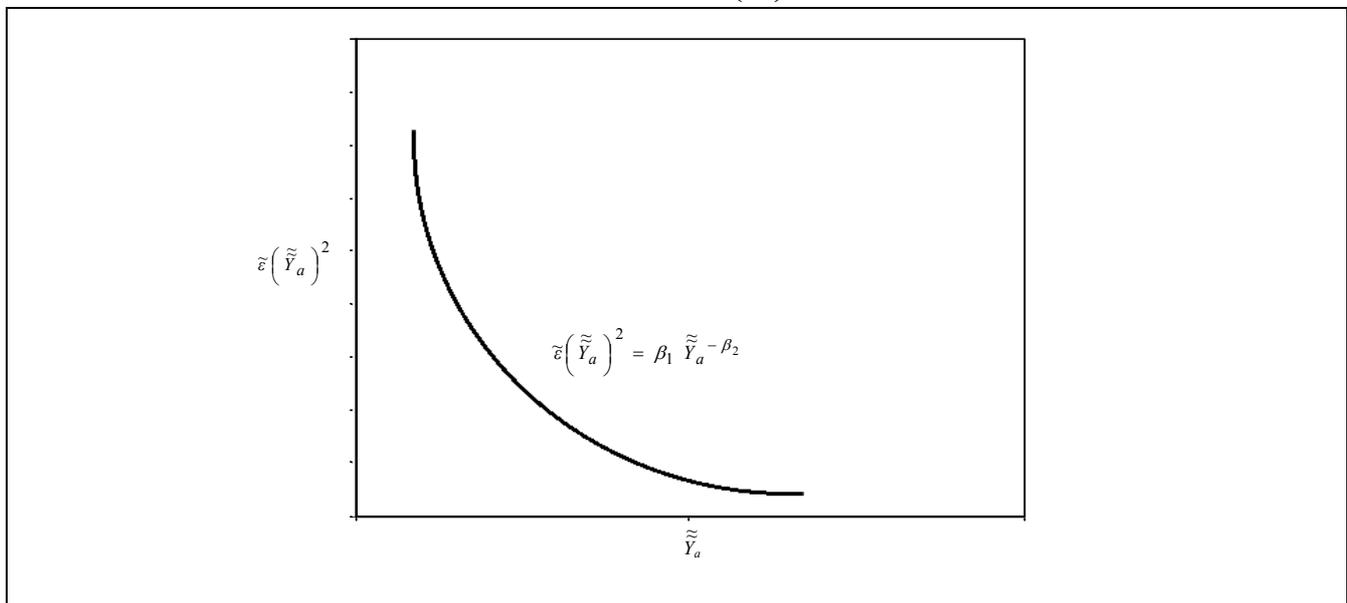
$$\tilde{\varepsilon}(\tilde{Y}_a)^2 = \beta_1 + \frac{\beta_2}{\tilde{Y}_a} + \frac{\beta_3}{\tilde{Y}_a^2} + u \tag{20b}$$

$$\ln\left(\tilde{\varepsilon}(\tilde{Y}_a)^2\right) = \ln(\beta_1) - \beta_2 \ln(\tilde{Y}_a) + u \tag{20c}$$

dove β_1, β_2 e β_3 rappresentano i coefficienti di regressione, ed u è il residuo stocastico serialmente non correlato.

Si noti che il modello (20c) implicitamente riconduce al seguente modello esponenziale (Figura 3.1):

Figura 3.1 - Relazione tra la stima \tilde{Y}_a e l'errore campionario $\tilde{\varepsilon}(\tilde{Y}_a)^2$



$$\tilde{\varepsilon}(\tilde{Y}_a)^2 = \beta_1 \tilde{Y}_a^{-\beta_2} \exp^u \tag{20d}$$

Un aspetto interessante della struttura doppio-logaritmico del modello (20c) è che il coefficiente angolare β_2 misura l'elasticità puntuale di $\tilde{\varepsilon}(\tilde{Y}_a)^2$ rispetto a \tilde{Y}_a , ossia esprime la variazione percentuale prevista di $\tilde{\varepsilon}(\tilde{Y}_a)^2$ a seguito di una data variazione percentuale del valore atteso in \tilde{Y}_a .

Senza perdita di generalità, il modello (20c) di Wolter è reso essenziale nel modo seguente:

$$\ln\left(\tilde{\varepsilon}(\tilde{Y}_a)\right) = c - \beta \ln\left(\tilde{Y}_a\right) + e. \quad (21)$$

dove $c = \frac{1}{2} \ln(\beta_1)$, $\beta = \frac{1}{2} \beta_2$ e $e = \frac{1}{2} u$.

Utilizzando semplici operatori additivi, il modello (21) può essere riformulato in termini *errore standard assoluto*, separando l'errore standard relativo nelle sue due componenti: \tilde{Y}_a e $\tilde{\sigma}(\tilde{Y}_a)$

$$\ln\left(\tilde{\sigma}(\tilde{Y}_a)\right) = c + b_1 \ln\left(\tilde{Y}_a\right) + e \quad (22)$$

dove $b_1 = (1 - \beta)$.

Tale trasformazione è agevole per il modello doppio-logaritmico (nel seguito indicato come *modello base*), non si può dire lo stesso per i modelli (20a) e (20b), in quanto un'analogia riformulazione produrrebbe effetti moltiplicativi complicando l'interpretazione dei parametri.

3.3.2 – Estensione delle funzioni generalizzate dell'errore campionario assoluto, per le indagini sui Conti economici e finanziari delle imprese

Le estensioni che si propongono al modello base sono il frutto di osservazioni empiriche sulle stime prodotte dalle rilevazioni 2004 sulle piccole e medie imprese (Pmi) e sul sistema dei conti delle imprese (Sci).

Le stime si riferiscono alle seguenti variabili economiche: *fatturato, valore aggiunto, costo del lavoro, investimenti e retribuzioni*. Queste riguardano i principali aggregati economici d'interesse statistico e sono rilasciate in conformità a tre tipologie di dominio (Prospetto 3.1).

Prospetto 3.1 – Tipologia di domini di studio della rilevazione sui conti economici e finanziari delle imprese

TIPOLOGIA DI DOMINIO	Caratteristica	Numero domini
Dom1	Classe di attività economica (a)	463
Dom2	Gruppo di attività economica × Classi addetti (b)	391
Dom3	Divisione di attività economica × Regione (c)	945
Totale numero di domini		1.799

(a) L'attività economica dell'impresa è definita secondo la classificazione Nace Rev.1.

(b) Le due classi di addetti sono (1-19) e (20 e oltre).

(c) Le regioni sono 21 comprese le province autonome.

Utilizzando un unico concatenamento a livello di dominio, le stime dei totali sono ottenute come somma delle due fonti informative. La varianza, invece, corrisponde alla sola componente Pmi, in quanto per la fonte censuaria (Sci) essa è nulla.

L'esplorazione empirica delle coppie $\left(\tilde{Y}_a, \tilde{\sigma}(\tilde{Y}_a)\right)$ ha rilevato che:

1. esiste una relazione tra il tipo di variabile economica utilizzato e il corrispondente errore assoluto, tale dipendenza produce effetti significativi solo sull'intercetta del modello;
2. le tipologie di dominio individuano nuvole di punti distinte e per lo più non coincidenti;
3. è violata l'assunzione di costanza dell'effetto del disegno (Kish, L, 1965, Russo, A, 1987), in base alla quale i modelli classici per la presentazione sintetica degli errori di campionamento sono applicabili;

Ai fini della specificazione di un appropriato modello predittivo, è adottata la seguente strategia:

1. per ciascuno delle tre tipologie di dominio è stimato un modello distinto;
2. come funzione di riferimento è utilizzato il *modello base aumentato*, specificato come segue

$$\ln\left(\tilde{\sigma}\left(\tilde{Y}_a\right)\right) = c + \sum_{j=2}^J a_j \lambda_j + b_1 \ln\left(\tilde{Y}_a\right) + e \quad (23)$$

dove λ_j rappresenta un indicatore binario delle $(J-1)$ variabili economiche e la categoria di riferimento (*fatturato*) è inclusa nella costante c ;

3. è proposta una procedura in grado di ridurre la distorsione dello stimatore OLS (*Ordinary Least Squares*) per b_1 dovuta all'omissione della variabile $\tilde{d}eff\left(\tilde{Y}_a\right)$. In accordo a tale procedura la variabilità del $\tilde{d}eff\left(\tilde{Y}_a\right)$ è principalmente spiegata dall'appartenenza della singola stima al dominio statistico.

Per ciascuno delle tre tipologie di dominio oggetto di studio, nel seguito è illustrato il metodo adottato e i risultati conseguiti.

3.3.2.1 – Tipologia di dominio Dom1 (classe di attività economica)

L'ispezione grafica evidenzia che esiste una relazione lineare tra le coppie $\left(\ln\left(\tilde{Y}_a\right), \ln\left(\tilde{\sigma}\left(\tilde{Y}_a\right)\right)\right)$, figura

3.2, nondimeno emerge la presenza di alcune stime di totali che si discostano notevolmente dalla nuvola centrale dei punti. Tali stime hanno un trascurabile errore standard, rispetto ad altre stime con analoghi livelli di totale. La spiegazione di quest'osservata anomalia dipende dalla differente composizione, nella base dei dati, tra componente censuaria e componente campionaria, con la conseguenza di produrre un'elevata variabilità della statistica $\tilde{d}eff\left(\tilde{Y}_a\right)$, il cui logaritmo è compreso tra -11,65 e 0,98.

La valutazione statistica di quanto il disegno dell'indagine impatta sull'errore standard è affidata al modello che segue ed indicato come *modello correttamente specificato*

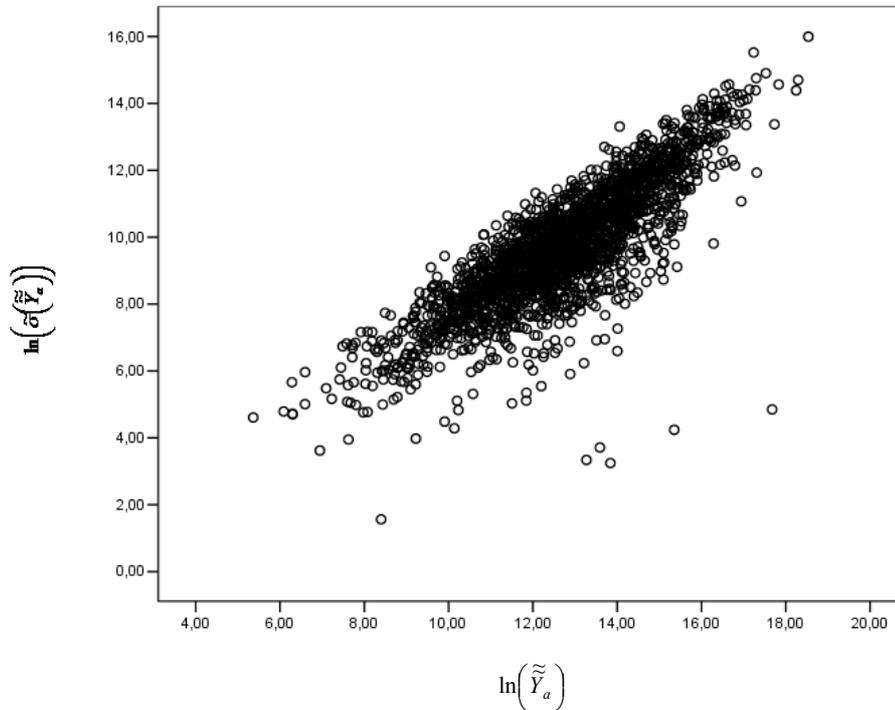
$$\ln\left(\tilde{\sigma}\left(\tilde{Y}_a\right)\right) = c + \sum_{j=2}^J a_j \lambda_j + b_1 \ln\left(\tilde{Y}_a\right) + b_2 \ln\left(\tilde{d}eff\left(\tilde{Y}_a\right)\right) + e \quad (24)$$

Il prospetto 3.2 riporta le stime dei parametri del *modello base aumentato* e del *modello correttamente specificato*. L'ordine di presentazione delle variabili predittrici si basa sul contributo che ciascuna di esse fornisce nell'incrementare la varianza spiegata. Quest'ultima è riassunta dal coefficiente di determinazione corretto R^2 , il quale misura la proporzione della variazione in $\ln\left(\tilde{\sigma}\left(\tilde{Y}_a\right)\right)$ che è spiegata dalla funzione di regressione, compensata per i gradi di libertà.

Per quanto attiene alla valutazione del contributo del predittore $\ln\left(\tilde{d}eff\left(\tilde{Y}_a\right)\right)$ al *modello correttamente specificato*, questa variabile assorbe quasi il 22 per cento della varianza spiegata. La sua omissione dal modello determinerebbe una grave distorsione dello stimatore OLS per b_1 . Infatti, se si assume che il *modello*

correttamente specificato rifletta la struttura reale dei dati, allora è possibile essere confidenti che il vero valore del parametro b_1 sia contenuto nella regione di confidenza $[0,859; 0,882]$ con probabilità pari al 95 per cento. Poiché il vero parametro b_1 è sconosciuto, si adotta la stima di ${}_{(24)}b_1 = 0,871$ come una buona approssimazione del vero valore di b_1 .

Figura 3.2 - Tipologia di dominio Dom1: diagramma di dispersione di $\ln(\tilde{\sigma}(\tilde{Y}_a))$ e $\ln(\tilde{Y}_a)$ come referenza



D'altra parte, riducendosi a considerare il solo *modello base aumentato* il medesimo intervallo rispetto a ${}_{(23)}b_1 [0,764; 0,815]$ non include il parametro ${}_{(24)}b_1$.

Prospetto 3.2 - Tipologia di dominio Dom1: stime di regressione del *modello base aumentato* e del *modello correttamente specificato*

MODELLO BASE AUMENTATO					MODELLO CORRETTAMENTE SPECIFICATO				
COEFFICIENTE DI REGRESSIONE	Stime	Errore standard	t di Student	R ² corretto	COEFFICIENTE DI REGRESSIONE	Stime	Errore standard	t di Student	R ² corretto
c	0,055	0,194	0,282		c	0,140	0,085	1,634	
b_1	0,790	0,013	60,733	0,705	b_1	0,871	0,006	150,767	0,705
-					b_2	0,790	0,008	96,069	0,944
a_4	-0,026	0,077	-0,335	0,720	a_5	-0,418	0,031	-13,282	0,946
a_5	-0,843	0,071	-11,889	0,726	a_3	-0,375	0,031	-12,221	0,949
a_3	-0,775	0,069	-11,207	0,737	a_2	-0,202	0,029	-6,954	0,949
a_2	-0,366	0,066	-5,531	0,741	a_4	-0,169	0,034	-4,977	0,950

Tuttavia, ad essere realistici, l'utente non ha alcuna conoscenza a priori di come varia il $\tilde{d}_{eff}(\tilde{Y}_a)$ tra i domini statistici di stima, ciò si caratterizza in un dilemma tra il modello predittivo correttamente specificato e

l'ottenimento di uno strumento autonomo di conversione di qualsiasi stima \tilde{Y}_a nel corrispettivo errore standard $\tilde{\sigma}(\tilde{Y}_a)$.

Una soluzione al problema consiste nell'identificare *sottoinsiemi* omogenei di domini (denominati in seguito come *cluster-dominio*), in termini del valore assunto dalla variabile $\tilde{d}eff(\tilde{Y}_a)$ e regredire tanti modelli quanti sono i gruppi sottostanti, al fine di ricondurre la stima dei parametri alle condizioni ottimalità di richieste dal modello classico di regressione lineare.

Per quanto riguarda la scelta dell'algoritmo di classificazione è indispensabile che possa garantire i seguenti requisiti:

1. *correttezza* - i *cluster-dominio* devono poter essere in grado di assorbire la maggior parte della distorsione prodotta nello stimatore OLS per b_1 , dall'omissione nel modello predittivo della variabile $\ln(\tilde{d}eff(\tilde{Y}_a))$;
2. *omogeneità* - per ciascun *cluster-dominio*, la distribuzione del $\ln(\tilde{d}eff(\tilde{Y}_a))$ deve essere disgiunta e non sovrapponibile con le distribuzioni osservate negli altri *cluster-dominio*, così che tutta la variabilità osservata possa essere attribuita alle differenze tra i gruppi;
3. *parsimonia* - il numero di *cluster-dominio* deve essere sufficientemente contenuto, al fine di continuare a soddisfare le caratteristiche sinteticità richieste da questa modalità di presentazione degli errori campionari.

Un utile strumento statistico di *clustering* è la tecnica conosciuta come alberi di classificazione (Belson, W. A., 1959; Neville, PG. 1999) tuttavia, nel caso in esame, la configurazione ottimale non può essere ottenuta in modo automatizzato, ma comporta l'uso combinato di più misure di prossimità e di procedure iterative di convergenza.

Il prospetto 3.3 riporta le statistiche descrittive per i 7 *cluster-dominio* identificati in base ai vincoli sopra riportati e ordinati rispetto ai valori medi assunti dalla variabile $\ln(\tilde{d}eff(\tilde{Y}_a))$.

Prospetto 3.3 - Tipologia di dominio Dom1: statistiche descrittive della variabile $\ln(\tilde{d}eff(\tilde{Y}_a))$ per *cluster-dominio*

CLUSTER DOMINIO	Domini	Media	Deviazione standard	Errore standard	Intervallo di confidenza 95% per la media		Minimo	Massimo
					Limite inferiore	Limite superiore		
1	169	-3,811	1,501	0,115	-4,039	-3,583	-11,650	-0,646
2	195	-2,653	0,871	0,062	-2,776	-2,530	-4,975	-0,263
3	425	-1,938	0,707	0,034	-2,005	-1,871	-3,863	0,215
4	425	-1,515	0,665	0,032	-1,579	-1,452	-4,295	0,933
5	390	-1,244	0,613	0,031	-1,305	-1,183	-2,916	0,980
6	500	-0,862	0,576	0,026	-0,912	-0,811	-3,487	0,875
7	110	-0,484	0,524	0,050	-0,583	-0,385	-1,964	0,944
Totale	2.214	-1,625	1,129	0,024	-1,672	-1,578	-11,650	0,980

Il prospetto 3.4 riporta il dettaglio della classificazione. Si rammenti che, i domini non riprodotti riguardano aggregati totalmente censiti dalla rilevazione. Per questi, i totali sono valori puntuali e non stime, da ciò deriva la loro esclusione conseguente dal modello predittivo.

In conformità a tali risultati è stato stimato il seguente modello ad intercetta variabile, indicato di seguito come *modello surrogato*.

$$\ln(\tilde{\sigma}(\tilde{Y}_a)) = c + \sum_{j=2}^J a_j \lambda_j + \sum_{i=2}^I q_i \xi_i + b_1 \ln(\tilde{Y}_a) + e \quad (25)$$

dove si è denotato con: ξ_i una variabile indicatrice che assume valore 1 se il dominio appartiene al *cluster-dominio i* e valore 0 altrimenti; q_i la specifico parametro relativo al *cluster-dominio i*, il quale esprime l'intensità del cambiamento sull'intercetta c causato da una differente media nell'effetto del disegno.

Prospetto 3.4 – Tipologia di dominio Dom1: cluster-dominio e domini corrispondenti

CLUSTER-DOMINIO	Classificazione NACE Rev.1
1	1110; 1430; 1532; 1562; 1584; 1596; 2212; 2320; 2416; 2442; 2451; 2511; 2613; 2630; 2662; 2710; 2721; 2722; 2742; 2744; 2960; 2971; 3161; 3210; 3520; 3530; 4012; 4013; 5551; 6021; 6321; 6323; 6420; 7450
2	1120; 1422; 1598; 2112; 2122; 2411; 2413; 2414; 2415; 2441; 2470; 2622; 2651; 2652; 2821; 2822; 2830; 2863; 2872; 2911; 2914; 3001; 3140; 3230; 3340; 3420; 3430; 3541; 3543; 4011; 4030; 4100; 4512; 6120; 6210; 6412; 8511; 9002; 9220
3	1412; 1450; 1512; 1520; 1533; 1542; 1551; 1552; 1571; 1582; 1583; 1585; 1591; 1716; 1721; 2020; 2121; 2123; 2125; 2213; 2221; 2412; 2420; 2430; 2452; 2466; 2512; 2513; 2521; 2522; 2625; 2626; 2640; 2661; 2663; 2665; 2734; 2743; 2751; 2752; 2753; 2871; 2873; 2912; 2913; 2931; 2941; 2942; 2943; 2952; 2955; 3002; 3110; 3120; 3130; 3150; 3220; 3320; 3511; 3550; 3613; 3640; 4022; 4524; 5139; 5186; 5211; 5212; 5261; 5552; 6022; 6312; 6322; 6340; 7110; 7222; 7240; 7310; 7430; 7460; 7470; 7486; 8531; 9003; 9271
4	1411; 1511; 1513; 1572; 1586; 1589; 1592; 1593; 1711; 1713; 1715; 1723; 1724; 1753; 1754; 1910; 1920; 1930; 2211; 2222; 2462; 2463; 2464; 2523; 2524; 2612; 2615; 2621; 2624; 2664; 2681; 2682; 2733; 2741; 2745; 2754; 2840; 2861; 2862; 2874; 2921; 2924; 2932; 2951; 2953; 2954; 2956; 2972; 3162; 3310; 3330; 3512; 3542; 3615; 3621; 3622; 3650; 3663; 3710; 4525; 5010; 5132; 5151; 5182; 5190; 5245; 5271; 5273; 5510; 5523; 6110; 6311; 7032; 7122; 7133; 7260; 7320; 7482; 8021; 8532; 9001; 9233; 9251; 9252; 9304
5	1413; 1421; 1543; 1561; 1587; 1717; 1722; 1725; 1730; 1740; 1751; 1752; 1760; 1771; 1772; 1810; 1822; 1823; 1830; 2010; 2040; 2052; 2111; 2214; 2215; 2224; 2231; 2232; 2666; 2732; 2875; 2922; 2923; 3350; 3614; 3630; 3662; 3720; 4021; 4523; 5030; 5131; 5134; 5137; 5138; 5141; 5142; 5143; 5146; 5152; 5153; 5154; 5155; 5181; 5184; 5185; 5224; 5231; 5232; 5243; 5521; 6023; 7132; 7134; 7140; 7221; 7230; 7250; 7413; 7415; 7420; 7481; 8010; 8030; 8041; 9212; 9232; 9301
6	1531; 1541; 1581; 1588; 1597; 1712; 1821; 1824; 2030; 2051; 2124; 2223; 2225; 2233; 2465; 2653; 2670; 2811; 2812; 2851; 2852; 3611; 3612; 3661; 4511; 4521; 4522; 4531; 4532; 4534; 4541; 4542; 4543; 4550; 5020; 5040; 5050; 5112; 5114; 5116; 5118; 5119; 5121; 5122; 5124; 5133; 5135; 5136; 5144; 5145; 5147; 5156; 5157; 5183; 5187; 5188; 5225; 5226; 5227; 5233; 5241; 5242; 5244; 5246; 5247; 5248; 5250; 5263; 5272; 5522; 5530; 5540; 6024; 6220; 6330; 7011; 7020; 7031; 7121; 7210; 7411; 7412; 7414; 7440; 7485; 7487; 8022; 8042; 8513; 8514; 8520; 9211; 9213; 9231; 9240; 9262; 9272; 9302; 9303; 9305
7	2461; 4533; 4544; 4545; 5111; 5113; 5115; 5117; 5123; 5125; 5221; 5222; 5223; 5262; 5274; 7012; 7123; 7131; 8512; 9234; 9253; 9261

Nel prospetto 3.5 sono riportate le stime dei parametri. L'ordine di presentazione delle variabili predittrici si basa sul contributo che ciascuna di esse fornisce nell'incrementare la varianza spiegata.

Prospetto 3.5 - Tipologia di dominio Dom1: stime di regressione del modello surrogato

COEFFICIENTI DI REGRESSIONE	Stime	Errore standard	t di Student	R ² corretto
c	-2,916	0,120	-24,359	
b_1	0,858	0,007	118,501	0,705
q_6	2,741	0,046	59,342	0,755
q_7	3,206	0,064	49,782	0,786
q_5	2,354	0,048	48,955	0,818
q_4	2,026	0,047	42,691	0,849
q_3	1,589	0,047	33,660	0,877
a_4	0,203	0,042	4,804	0,896
q_2	1,011	0,055	18,515	0,909
a_5	-0,679	0,039	-17,498	0,913
a_3	-0,633	0,038	-16,753	0,921
a_2	-0,277	0,036	-7,686	0,923

Per quanto riguarda la distorsione dello stimatore OLS per b_1 dovuta dall'omissione della variabile $\ln\left(\text{deff}\left(\tilde{Y}_a\right)\right)$, il *modello surrogato* è in grado di contenerla efficacemente in quanto ${}_{(24)}b_1$ ricade all'interno dell'intervallo di confidenza, costruito intorno alla stima di ${}_{(25)}b_1$ $[0,844; 0,872]$. Inoltre, la bontà di adattamento del *modello surrogato* R^2 , decresce di soli 2,8 punti percentuale rispetto al *modello correttamente specificato*.

3.3.2.2 - Tipologia di dominio Dom2: concatenazione delle modalità delle variabili Gruppo di attività economica e classe di addetti

Analogamente al caso precedente, la medesima metodologia è stata applicata per questa tipologia di dominio. Anche in questo caso alcune stime di totale si distinguono dalla nuvola centrale dei punti (Figura 3.3), perché esibiscono errori standard irrilevanti a parità di analoghi livelli di totale. Il logaritmo del *deff* varia tra -7,75 e 1,75.

Il prospetto 6 riporta le statistiche del *modello base aumentato* e del *modello correttamente specificato*. Come si può notare, l'introduzione dell'effetto del disegno, sintetizzato dalla variabile $\ln\left(\tilde{\text{deff}}\left(\tilde{Y}_a\right)\right)$ incrementa la varianza spiegata di quasi 23 decimi di punto.

Prospetto 3.6 - Tipologia di dominio Dom2: stime di regressione del *modello base aumentato* e del *modello correttamente specificato*

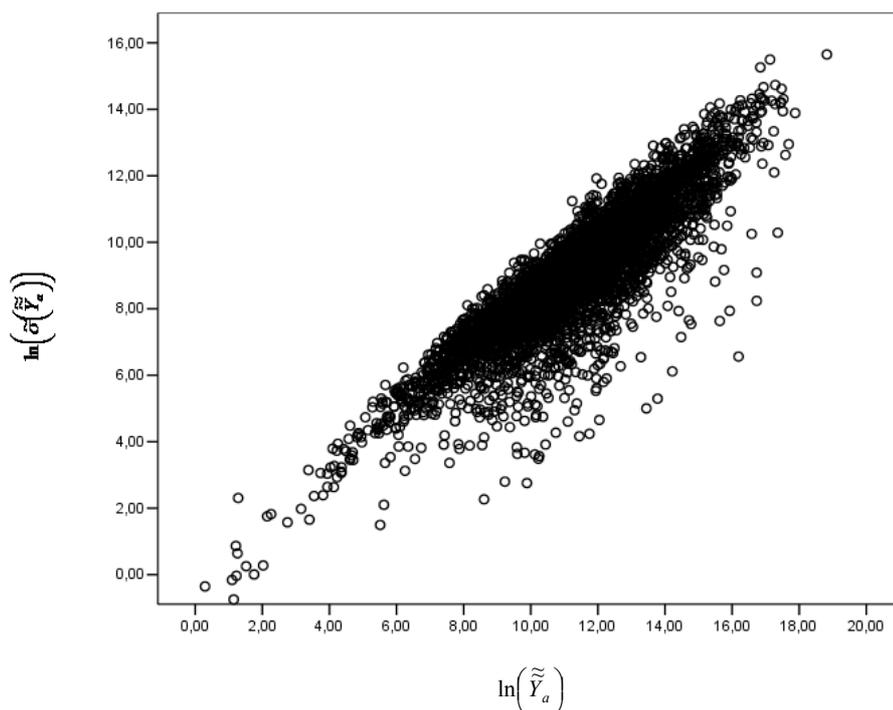
MODELLO BASE AUMENTATO					MODELLO CORRETTAMENTE SPECIFICATO				
COEFFICIENTI DI REGRESSIONE	Stime	Errore standard	t di Student	R ² corretto	COEFFICIENTI DI REGRESSIONE	Stime	Errore standard	t di Student	R ² corretto
c	0,902	0,197	4,585		c	-0,400	0,080	-5,001	
b_1	0,736	0,013	56,108	0,701	b_1	0,884	0,005	161,663	0,701
-					b_2	0,825	0,008	98,041	0,946
a_5	-1,032	0,079	-12,980	0,715	a_4	0,020	0,034	0,586	0,952
a_3	-0,946	0,078	-12,136	0,732	a_5	-0,484	0,032	-14,972	0,953
a_2	-0,506	0,075	-6,699	0,736	a_3	-0,446	0,032	-14,089	0,956
a_4	-0,309	0,086	-3,611	0,738	a_2	-0,263	0,030	-8,674	0,958

Al fine di ottenere raggruppamenti omogenei rispetto alla variabile $\ln\left(\tilde{\text{deff}}\left(\tilde{Y}_a\right)\right)$ è stata applicata la medesima procedura ideata per la tipologia di dominio Dom1.

Il prospetto 3.7 riporta le statistiche descrittive per gli 8 *cluster-dominio*, distinte per classi di addetti e ordinate rispetto alla media del logaritmo del $\tilde{\text{deff}}\left(\tilde{Y}_a\right)$.

Osservando i valori di discrepanza massimo e minimo, corrispondenti ad un intervallo di confidenza pari al 95 per cento, si nota che la classificazione ottenuta offre un significativa capacità discriminatoria tra i *cluster-dominio*, riguardo alla variabile $\ln\left(\tilde{\text{deff}}\left(\tilde{Y}_a\right)\right)$. Inoltre, in termini di effetto del disegno, appare ben chiara la distinzione tra i domini dove la componente campionaria delle unità di rilevazione è esclusiva (imprese fino a 19 addetti) e i domini composti anche da unità censite (imprese con oltre 19 addetti).

Figura 3.3 - Tipologia di dominio Dom2: diagramma di dispersione di $\ln(\tilde{\sigma}(\tilde{Y}_a))$ e $\ln(\tilde{Y}_a)$ come referenza



Prospetto 3.7 - Tipologia di dominio Dom2: statistiche descrittive della variabile $\ln(\tilde{d}_{eff}(\tilde{Y}_a))$, per classe di addetti e cluster-dominio

CLASSE DI ADDETTI	CLUSTER-DOMINIO	Domini	Media	Deviazione standard	Errore standard	Intervallo di confidenza 95% per la media		Minimo	Massimo
						Limite inferiore	Limite superiore		
>19	1	100	-3,943	1,213	0,121	-4,184	-3,702	-7,749	-0,403
	2	60	-2,843	0,765	0,099	-3,041	-2,646	-4,435	-0,837
	3	150	-2,206	0,758	0,062	-2,328	-2,083	-4,847	0,225
	4	265	-1,696	0,699	0,043	-1,780	-1,611	-3,505	0,537
	5	220	-1,206	0,655	0,044	-1,293	-1,119	-2,800	0,584
	6	115	-0,637	0,744	0,069	-0,774	-0,499	-2,219	1,752
	Totale	910	-1,850	1,212	0,040	-1,929	-1,771	-7,749	1,752
1-19	7	529	-0,443	0,342	0,015	-0,472	-0,414	-1,922	0,687
	8	405	-0,226	0,365	0,018	-0,262	-0,190	-1,404	0,774
	Totale	934	-0,349	0,368	0,012	-0,373	-0,325	-1,922	0,774
Totale		1.844	-1,090	1,165	0,027	-1,143	-1,037	-7,749	1,752

Il prospetto 3.8 riporta il dettaglio della classificazione. Anche in questo caso i domini non riportati nel prospetto riguardano domini totalmente censiti e dunque esclusi dal modello predittivo.

In base ai tali risultati è stato stimato il *modello surrogato* (Prospetto 3.9).

Per quanto riguarda la distorsione dello stimatore OLS b_1 dovuta dall'omissione della variabile $\ln(\tilde{d}_{eff}(\tilde{Y}_a))$, il *modello surrogato* è in grado di contenerla efficacemente in quanto la distribuzione dell'intervallo di confidenza, costruito sulla stima ${}_{(25)}b_1 [0,859; 0,884]$ contiene il parametro del *modello*

correttamente specificato $(_{24})b_1 = 0,884$, mentre lo stesso non si può dire per l'intervallo di confidenza costruito sulla stima di $(_{23})b_1 [0,710; 0,761]$. Inoltre, anche in questo caso la perdita di varianza spiegata per l'impiego del *modello correttamente specificato* è contenuta ed è pari all'1,1 per cento.

Prospetto 3.8 - Tipologia di dominio Dom2: cluster-dominio e domini corrispondenti

CLASSE DI ADDETTI	CLUSTER-DOMINIO	Gruppo di attività economica
	1	143; 232; 241; 244; 247; 265; 271; 296; 297; 314; 353; 401; 455; 555; 612; 632; 641; 642; 714; 745
	2	159; 183; 221; 245; 261; 272; 274; 283; 321; 334; 521; 922
	3	154; 155; 157; 158; 182; 211; 212; 262; 263; 275; 282; 291; 293; 312; 323; 342; 343; 351; 354; 363; 402; 410; 524; 611; 621; 712; 731; 803; 851; 900
>19	4	112; 144; 145; 151; 152; 153; 172; 174; 192; 202; 222; 242; 246; 251; 252; 266; 286; 287; 294; 295; 300; 311; 313; 315; 322; 331; 332; 355; 364; 366; 371; 513; 515; 518; 519; 523; 526; 552; 554; 633; 634; 713; 722; 723; 724; 742; 746; 747; 801; 923; 926; 927; 930
	5	141; 142; 156; 171; 175; 176; 177; 181; 191; 193; 203; 205; 243; 264; 268; 273; 281; 284; 292; 316; 333; 361; 362; 365; 451; 452; 453; 454; 501; 503; 504; 512; 514; 522; 551; 553; 602; 631; 726; 741; 743; 748; 804; 853
	6	173; 201; 204; 223; 267; 285; 335; 372; 502; 505; 511; 525; 527; 622; 701; 702; 703; 721; 725; 744; 802; 921; 925
	7	232; 241; 244; 271; 296; 353; 555; 612; 632; 641; 159; 221; 245; 261; 274; 283; 321; 334; 157; 211; 212; 262; 275; 282; 291; 323; 342; 343; 354; 410; 524; 621; 731; 900; 144; 145; 151; 152; 172; 192; 222; 246; 251; 252; 266; 286; 294; 295; 300; 311; 313; 315; 332; 364; 366; 371; 513; 515; 518; 519; 523; 552; 634; 713; 746; 747; 801; 927; 930; 141; 142; 156; 171; 175; 176; 181; 191; 193; 243; 264; 268; 273; 284; 292; 316; 362; 365; 503; 512; 514; 522; 551; 631; 726; 743; 201; 204; 372; 511; 525; 721; 725; 802; 925; 732; 852
1-19	8	247; 265; 297; 314; 401; 455; 642; 714; 745; 183; 272; 521; 922; 154; 155; 158; 182; 263; 293; 312; 351; 363; 402; 611; 712; 803; 851; 112; 153; 174; 202; 242; 287; 322; 331; 355; 526; 554; 633; 722; 723; 724; 742; 923; 926; 177; 203; 205; 281; 333; 361; 451; 452; 453; 454; 501; 504; 553; 602; 741; 748; 804; 853; 173; 223; 267; 285; 335; 502; 505; 527; 622; 701; 702; 703; 744; 921; 352; 403; 711; 924

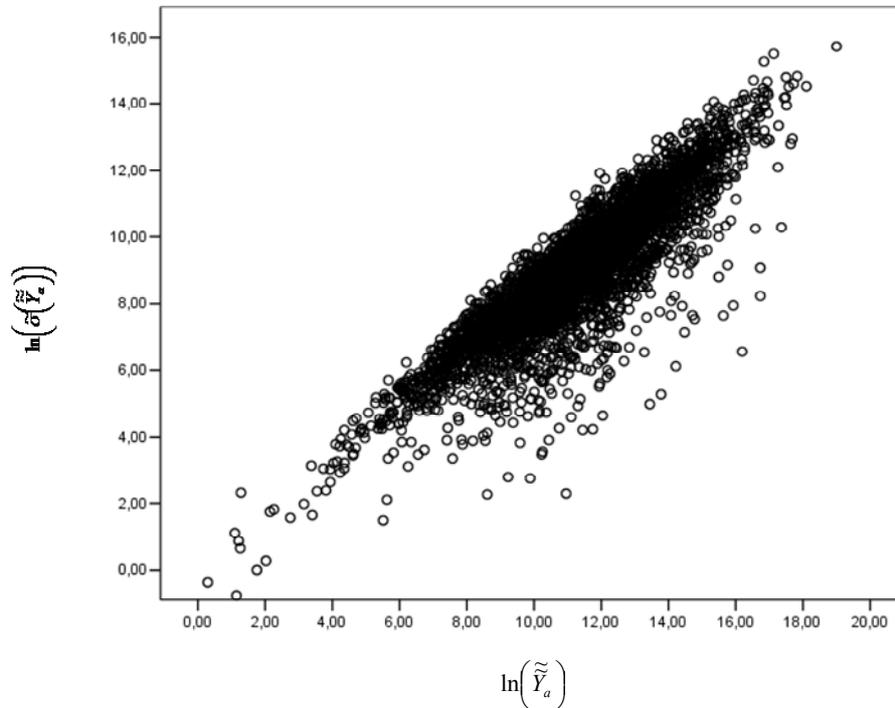
Prospetto 3.9 - Tipologia di dominio Dom2: stime di regressione del modello surrogato

COEFFICIENTI DI REGRESSIONE	Stime	Errore standard	t di Student	R ² corretto
<i>c</i>	-3,588	0,107	-33,530	
<i>b</i> ₁	0,872	0,006	139,550	0,701
<i>q</i> ₈	3,475	0,050	69,048	0,766
<i>q</i> ₇	3,043	0,049	61,548	0,834
<i>q</i> ₆	2,832	0,062	45,697	0,854
<i>q</i> ₅	2,345	0,054	43,470	0,876
<i>q</i> ₄	1,922	0,053	36,598	0,897
<i>a</i> ₄	0,148	0,039	3,790	0,915
<i>q</i> ₃	1,461	0,058	25,301	0,928
<i>q</i> ₂	0,980	0,073	13,421	0,933
<i>a</i> ₅	-0,717	0,036	-19,945	0,937
<i>a</i> ₃	-0,676	0,035	-19,196	0,945
<i>a</i> ₂	-0,328	0,034	-9,659	0,947

3.3.3 - Tipologia di dominio Dom3: concatenazione delle modalità delle variabili Divisione di attività economica e regione

Anche l'ultima tipologia di dominio mostra la consueta depressione delle stime rispetto alla nuvola centrale dei punti (Figura 3.4) a causa di una variabilità del $\tilde{d}eff(\tilde{Y}_a)$, il cui logaritmo è compreso tra -9,21 e 3,30.

Figura 3.4 - Tipologia di dominio Dom3: diagramma di dispersione di $\ln(\tilde{\sigma}(\tilde{Y}_a))$ e $\ln(\tilde{Y}_a)$ come referenza



Il prospetto 3.10 riporta le statistiche del *modello base aumentato* e del *modello correttamente specificato*. Come si può notare, l'inserimento della variabile $\ln(\tilde{d}eff(\tilde{Y}_a))$ incrementa la varianza spiegata di quasi 16,9 decimi di punto.

Prospetto 3.10 - Tipologia di dominio Dom3: stime di regressione del *modello base aumentato* e del *modello correttamente specificato*.

MODELLO BASE AUMENTATO					MODELLO CORRETTAMENTE SPECIFICATO				
COEFFICIENTI DI REGRESSIONE	Stime	Errore standard	t di Student	R ² corretto	COEFFICIENTI DI REGRESSIONE	Stime	Errore standard	t di Student	R ² corretto
<i>c</i>	0,910	0,090	10,067		<i>c</i>	0,268	0,041	6,559	
<i>b</i> ₁	0,748	0,007	114,446	0,783	<i>b</i> ₁	0,872	0,003	282,817	0,783
-					<i>b</i> ₂	0,762	0,006	133,171	0,956
<i>a</i> ₅	-0,696	0,046	-15,178	0,788	<i>a</i> ₅	-0,331	0,021	-15,936	0,957
<i>a</i> ₃	-0,622	0,045	-13,743	0,794	<i>a</i> ₃	-0,293	0,020	-14,287	0,958
<i>a</i> ₂	-0,340	0,044	-7,679	0,795	<i>a</i> ₂	-0,176	0,020	-8,816	0,959
<i>a</i> ₄	-0,301	0,048	-6,275	0,797	<i>a</i> ₄	-0,137	0,022	-6,362	0,959

Come nei casi precedenti è stata applicata la medesima procedura di ricerca del partizionamento ottimale tale da soddisfare i requisiti di *parsimonia*, *omogeneità* e *correttezza*. Il prospetto 11 riporta le statistiche

descrittive per ciascuno dei 9 *cluster-dominio*, distinte per aggregati territoriali e ordinate rispetto alla media del logaritmo del $\tilde{d}eff(\tilde{Y}_a)$.

Con riferimento alla collocazione geografica delle imprese, il primo *cluster-dominio* è costituito principalmente dalle seguenti regioni del Nord: Piemonte, Lombardia, Veneto Friuli Venezia Giulia, Emilia. A questo gruppo si aggregano anche le regioni Lazio e Campania. In questo insieme l'incidenza percentuale delle imprese censite, rispetto alla totalità delle imprese rispondenti, è pari al 19,6 per cento.

Il nono *cluster-dominio* è formato, prevalentemente, dalle regioni del Centro-Sud ad eccezione delle regioni Valle D'Aosta e della Liguria. In questo raggruppamento l'incidenza percentuale delle imprese censite è pari al 9,4 per cento delle unità rispondenti.

Prospetto 3.11 - Tipologia di dominio Dom3: statistiche descrittive della variabile $\ln(\tilde{d}eff(\tilde{Y}_a))$, per raggruppamenti territoriali e cluster-dominio

CLUSTER DOMINIO	Domini	Media	Deviazione standard	Errore standard	Intervallo di confidenza 95% per la media		Minimo	Massimo
					Limite inferiore	Limite superiore		
1	594	-3,138	1,329	0,055	-3,245	-3,031	-9,209	0,012
2	543	-1,830	0,843	0,036	-1,901	-1,759	-5,185	3,299
3	599	-1,480	0,677	0,028	-1,534	-1,426	-5,596	2,769
4	394	-1,213	0,607	0,031	-1,273	-1,153	-2,797	0,856
5	595	-0,987	0,560	0,023	-1,032	-0,941	-2,706	1,512
6	544	-0,784	0,527	0,023	-0,829	-0,740	-2,671	0,928
7	590	-0,509	0,522	0,021	-0,551	-0,467	-3,064	1,543
8	321	-0,335	0,453	0,025	-0,385	-0,285	-1,930	2,030
9	320	-0,194	0,428	0,024	-0,241	-0,147	-1,525	2,228
Totale	4.500	-1,268	1,146	0,017	-1,302	-1,235	-9,209	3,299

In conformità a tali risultati è stato stimato il *modello surrogato* (Prospetto 3.12).

Per quanto riguarda la distorsione dello stimatore OLS b_1 dovuta dall'omissione della variabile $\ln(\tilde{d}eff(\tilde{Y}_a))$, il *modello surrogato* è in grado di assorbirla a sufficienza in quanto la distribuzione dell'intervallo di confidenza, costruito sulla stima di ${}_{(25)}b_1 [0,857; 0,873]$ contiene il parametro del *modello correttamente specificato* ${}_{(24)}b_1 = 0,872$ mentre lo stesso non si può dire per il *modello base aumentato* ${}_{(23)}b_1 [0,735; 0,761]$. Inoltre, la perdita di varianza spiegata del *modello surrogato* rispetto al *modello correttamente specificato* è ridotta a 2,9 per cento.

Prospetto 3.12 - Tipologia di dominio Dom3: stime di regressione del *modello surrogato*

COEFFICIENTI DI REGRESSIONE	Stime	Errore standard	t di Student	R ² corretto
c	-2,238	0,063	-35,526	
b_1	0,865	0,004	210,351	0,783
q_9	2,768	0,039	70,688	0,801
q_7	2,289	0,032	71,806	0,819
q_8	2,501	0,038	65,471	0,839
q_6	2,049	0,032	63,478	0,857
q_5	1,837	0,031	58,495	0,875
q_4	1,666	0,035	47,452	0,887
q_3	1,433	0,031	45,775	0,904
q_2	1,064	0,032	33,189	0,921
a_4	0,050	0,028	1,768	0,926
a_5	-0,440	0,027	-16,330	0,927
a_3	-0,403	0,027	-15,189	0,930
a_2	-0,199	0,026	-7,698	0,931

Il prospetto 3.13 riporta il dettaglio dei *sottoinsiemi* omogenei di domini ottenuti dalla tecnica di classificazione descritta in precedenza. Nello specifico, la regione è indicata tra parentesi, seguita dalla divisione di attività economica. Anche in questo caso sono stati esclusi i domini completamente censiti.

Prospetto 3.13 - Tipologia di dominio Dom3: cluster-dominio e domini corrispondenti

CLUSTER-DOMINIO	Regione per Divisione di attività economica
1	(01)21; (01)24; (01)27; (01)29; (01)34; (01)35; (01)40; (01)41; (01)63; (01)64; (02)14; (02)25; (02)30; (02)40; (03)21; (03)23; (03)24; (03)26; (03)27; (03)29; (03)31; (03)32; (03)34; (03)35; (03)61; (03)63; (03)64; (05)17; (05)21; (05)23; (05)27; (05)34; (05)41; (05)62; (05)90; (06)19; (06)27; (06)29; (06)35; (06)40; (06)41; (06)62; (07)23; (07)24; (07)27; (07)32; (07)41; (07)61; (08)24; (08)26; (08)27; (08)29; (08)34; (08)35; (08)40; (08)41; (09)21; (09)24; (09)27; (09)29; (09)33; (09)34; (09)40; (09)61; (09)71; (10)26; (10)27; (10)32; (10)34; (10)35; (11)24; (11)29; (11)41; (11)61; (12)23; (12)24; (12)32; (12)33; (12)35; (12)40; (12)60; (12)62; (12)63; (12)64; (12)90; (13)21; (13)32; (13)34; (13)35; (13)41; (13)62; (13)73; (14)25; (14)27; (15)32; (15)34; (15)35; (15)41; (15)64; (16)34; (16)41; (16)61; (16)73; (17)34; (17)40; (18)14; (18)37; (19)23; (20)27; (20)41; (20)64; (21)14; (21)34; (22)18; (22)19; (22)21; (22)24; (22)34; (22)63
2	(01)17; (01)22; (01)23; (01)25; (01)30; (01)71; (01)73; (01)90; (02)32; (02)92; (03)15; (03)17; (03)22; (03)25; (03)33; (03)40; (03)51; (03)52; (03)72; (03)90; (05)19; (05)24; (05)25; (05)26; (05)28; (05)29; (05)32; (05)35; (05)36; (05)40; (05)51; (05)64; (05)71; (06)21; (06)31; (06)34; (06)63; (07)26; (07)29; (07)33; (07)35; (07)63; (08)17; (08)21; (08)23; (08)25; (08)32; (08)33; (08)51; (08)90; (09)31; (09)35; (09)90; (10)15; (10)24; (10)40; (10)41; (10)90; (11)27; (11)31; (11)32; (11)34; (11)36; (12)15; (12)21; (12)22; (12)27; (12)29; (12)34; (12)51; (12)61; (12)72; (12)92; (13)18; (13)26; (13)60; (14)24; (15)23; (15)63; (15)73; (15)90; (16)19; (16)71; (17)27; (17)63; (17)80; (18)35; (18)90; (19)14; (19)24; (19)34; (19)41; (19)61; (19)62; (19)90; (20)14; (20)40; (21)15; (21)22; (21)26; (21)27; (21)29; (21)33; (22)15; (22)25; (22)29; (22)33; (22)90
3	(01)14; (01)15; (01)18; (01)20; (01)26; (01)31; (01)61; (01)72; (01)74; (02)33; (02)60; (02)93; (03)19; (03)30; (03)37; (03)71; (03)73; (03)74; (03)80; (03)85; (03)92; (05)18; (05)31; (05)33; (05)52; (05)63; (05)72; (05)92; (06)24; (06)25; (06)26; (06)51; (06)60; (06)61; (06)64; (06)72; (06)73; (06)80; (06)85; (07)14; (07)40; (07)51; (07)62; (08)18; (08)22; (08)28; (08)36; (08)64; (08)85; (09)15; (09)17; (09)19; (09)25; (09)30; (09)32; (09)37; (09)51; (09)60; (09)63; (09)73; (09)85; (10)20; (10)21; (10)29; (10)31; (10)33; (10)37; (10)63; (11)14; (11)19; (11)21; (11)35; (11)60; (12)14; (12)17; (12)26; (12)31; (12)41; (12)71; (12)74; (13)15; (13)25; (13)27; (13)31; (14)18; (14)73; (15)14; (15)21; (15)26; (15)27; (15)33; (15)60; (15)71; (15)92; (16)25; (16)29; (16)31; (16)35; (16)36; (16)63; (16)90; (17)31; (17)32; (17)35; (17)37; (18)60; (18)72; (19)31; (19)51; (19)63; (20)63; (20)72; (21)28; (21)30; (21)31; (21)40; (21)51; (21)52; (22)17; (22)51; (22)85
4	(01)28; (01)33; (01)60; (02)19; (02)71; (03)14; (03)36; (03)41; (05)15; (05)22; (05)85; (06)15; (06)22; (06)32; (06)33; (07)15; (07)21; (07)60; (07)72; (07)90; (08)31; (08)61; (08)72; (08)73; (08)74; (09)26; (10)36; (10)50; (10)51; (10)72; (10)92; (11)20; (11)26; (11)28; (11)90; (12)25; (12)30; (12)52; (13)19; (13)22; (13)24; (13)28; (13)33; (13)37; (13)90; (14)15; (14)26; (14)74; (15)15; (15)16; (15)17; (15)22; (15)24; (15)29; (15)51; (15)52; (15)74; (16)14; (16)28; (16)52; (16)74; (16)80; (18)29; (18)55; (18)63; (18)80; (19)27; (19)60; (19)80; (20)17; (20)29; (20)50; (20)60; (20)61; (20)80; (21)25; (21)74; (22)14; (22)22; (22)26
5	(01)32; (01)36; (01)50; (01)51; (01)52; (01)80; (02)17; (02)63; (03)18; (03)20; (03)28; (03)45; (03)62; (05)20; (05)37; (05)45; (05)55; (05)74; (06)14; (06)17; (06)20; (06)28; (06)30; (06)36; (06)52; (07)25; (07)28; (07)31; (07)64; (08)19; (08)30; (08)37; (08)50; (08)52; (08)55; (08)92; (08)93; (09)14; (09)52; (09)55; (09)80; (10)14; (10)17; (10)18; (10)22; (10)52; (10)70; (10)73; (10)74; (10)85; (11)15; (11)18; (11)25; (11)33; (11)37; (11)51; (11)52; (11)72; (12)50; (12)55; (12)73; (13)14; (13)17; (13)20; (13)29; (13)36; (13)51; (13)63; (13)74; (13)85; (14)20; (14)51; (14)52; (14)90; (15)18; (15)19; (15)25; (15)28; (15)31; (15)50; (15)72; (16)24; (16)26; (16)27; (16)30; (16)37; (16)51; (16)72; (17)36; (17)51; (18)36; (18)64; (18)92; (19)21; (19)22; (19)25; (19)26; (19)37; (19)52; (20)15; (20)24; (20)26; (20)28; (20)37; (20)70; (20)71; (20)74; (21)36; (21)60; (21)63; (21)80; (21)93; (22)20; (22)40; (22)61; (22)73; (22)74; (22)80; (22)92
6	(01)37; (01)85; (02)18; (02)20; (02)51; (02)55; (02)74; (03)50; (03)55; (03)60; (03)70; (05)14; (05)30; (05)50; (05)61; (05)73; (05)80; (06)37; (06)45; (06)74; (07)17; (07)22; (07)34; (07)50; (07)52; (07)74; (07)92; (08)14; (08)60; (08)62; (08)63; (08)80; (09)18; (09)23; (09)28; (09)36; (09)45; (09)72; (10)25; (10)45; (10)93; (11)50; (11)74; (11)85; (12)18; (12)80; (12)85; (13)23; (13)52; (13)80; (14)36; (14)55; (14)63; (14)80; (14)85; (15)30; (15)37; (15)40; (15)45; (15)61; (15)80; (15)93; (16)17; (16)21; (16)50; (16)60; (16)64; (16)85; (16)92; (17)26; (17)64; (17)72; (17)92; (18)15; (18)17; (18)20; (18)24; (18)27; (18)30; (18)50; (19)15; (19)18; (19)19; (19)29; (19)35; (19)40; (19)50; (19)55; (19)72; (19)74; (19)85; (19)92; (20)20; (20)21; (20)22; (20)32; (20)36; (20)51; (20)73; (21)17; (21)24; (21)45; (21)71; (21)72; (21)85; (21)92; (22)28; (22)36; (22)72
7	(01)45; (01)55; (01)62; (01)70; (01)92; (02)15; (02)26; (02)41; (02)45; (02)50; (02)72; (02)73; (03)93; (05)60; (05)70; (05)93; (06)18; (06)50; (06)70; (06)90; (07)18; (07)36; (07)37; (07)45; (07)55; (07)85; (07)93; (08)20; (08)45; (09)22; (09)50; (09)74; (09)93; (10)30; (10)55; (10)71; (10)80; (11)22; (11)40; (11)45; (11)63; (11)71; (11)73; (11)92; (11)93; (12)19; (12)20; (12)28; (12)45; (12)70; (12)93; (13)50; (13)70; (13)92; (13)93; (14)17; (14)29; (14)33; (14)40; (14)50; (14)60; (14)70; (14)92; (14)93; (15)36; (15)55; (15)62; (16)15; (16)22; (16)33; (16)40; (16)45; (16)55; (17)20; (17)29; (17)30; (17)33; (17)60; (17)71; (17)85; (18)22; (18)25; (18)26; (18)31; (18)32; (18)51; (18)52; (18)61; (18)85; (18)93; (19)20; (19)32; (19)45; (19)64; (19)70; (20)18; (20)25; (20)30; (20)31; (20)33; (20)34; (20)35; (20)45; (20)55; (20)85; (20)93; (21)37; (21)41; (21)50; (21)55; (22)27; (22)45; (22)52; (22)55; (22)60; (22)70; (22)71; (22)93
8	(01)19; (01)93; (02)28; (02)31; (02)52; (02)64; (02)80; (06)55; (06)92; (07)19; (07)70; (07)80; (08)70; (08)71; (09)92; (10)19; (10)28; (11)17; (11)30; (11)64; (11)70; (11)80; (12)36; (13)30; (13)55; (14)14; (14)21; (14)31; (14)37; (14)45; (14)71; (14)72; (15)85; (16)20; (16)23; (16)32; (16)70; (17)14; (17)15; (17)17; (17)18; (17)24; (17)25; (17)45; (17)52; (17)93; (18)28; (18)40; (18)41; (18)70; (18)74; (19)28; (19)36; (19)71; (19)73; (19)93; (20)52; (20)90; (20)92; (21)20; (21)21; (21)70; (22)31; (22)32; (22)50
9	(02)22; (02)29; (02)36; (02)70; (02)85; (02)90; (06)71; (06)93; (07)20; (07)30; (07)71; (07)73; (09)20; (09)70; (10)60; (10)64; (11)55; (12)37; (13)40; (13)45; (13)64; (13)71; (13)72; (14)19; (14)22; (14)28; (14)35; (14)41; (14)64; (15)20; (15)70; (16)18; (16)93; (17)19; (17)21; (17)22; (17)28; (17)50; (17)55; (17)70; (17)73; (17)74; (17)90; (18)18; (18)19; (18)21; (18)33; (18)34; (18)45; (18)71; (18)73; (19)17; (19)30; (19)33; (20)19; (21)18; (21)19; (21)32; (21)64; (21)73; (21)90; (22)35; (22)37; (22)64

3.3.4 - Valutazione dell'errore campionario

Dalla componente deterministica del *modello surrogato* è possibile riformulare la relazione matematica in termini errore relativo

$$\tilde{\varepsilon}\left(\tilde{Y}_a\right)=\exp\left(c+\sum_{j=2}^J a_j \lambda_j+\sum_{i=2}^I q_i \xi_i+\left(b_1-1\right) \ln\left(\tilde{Y}_a\right)\right) \quad (26)$$

Dopo alcuni passaggi algebrici, dall'equazione (26) si ottiene il valore \tilde{Y}_a^* della stima corrispondente ad uno specifico valore dell'errore relativo $\tilde{\varepsilon}\left(\tilde{Y}_a\right)$

$$\tilde{Y}_a^*=\exp\left[\frac{\ln\left(\tilde{\varepsilon}\left(\tilde{Y}_a\right)\right)-\left(c+\sum_{j=2}^J a_j \lambda_j+\sum_{i=2}^I q_i \xi_i\right)}{b_1-1}\right] \quad (27)$$

La determinazione della quantità \tilde{Y}_a^* è di notevole interesse pratico, in quanto tutte le stime con valori superiori ad \tilde{Y}_a^* presentano un'elevata probabilità di essere caratterizzate da un errore relativo, $\tilde{\varepsilon}\left(\tilde{Y}_a\right)$, minore; di converso, tutte le stime con valore inferiore ad \tilde{Y}_a^* presentano un'elevata probabilità di essere qualificate da un errore relativo maggiore.

Nei prospetti 3.14, 3.15 e 3.16 sono riportati i valori dei totali \tilde{Y}_a^* corrispondenti ad errori relativi in percentuale, $\tilde{\varepsilon}\left(\tilde{Y}_a\right) \times 100$, pari rispettivamente al 5 per cento, 10 per cento e 15 per cento, per le tre tipologie di dominio, per i *cluster-dominio* e per le 5 variabili economiche oggetto di statistiche.

In generale si può affermare che i primi *cluster-dominio* riportano stime molto accurate, perché corrispondono a domini dove la componente censuaria della rilevazione Sci risulta dominante; di converso, negli ultimi *cluster-dominio* è la componente campionaria della rilevazione Pmi ad essere predominante, ciò spiega i differenziali di livello osservati per ciascun $\tilde{\varepsilon}\left(\tilde{Y}_a\right)$.

Per meglio chiarire il significato statistico di quanto affermato, si consideri, ad esempio, il prospetto 3.14 corrispondente alla tipologia di dominio Dom1. Nell'ambito del *cluster-dominio* 1, si consideri la cella corrispondente all'incrocio della colonna *fatturato* e la riga relativa ad un $\tilde{\varepsilon}\left(\tilde{Y}_a\right) \times 100$ del 5 per cento: il valore riportato nella cella in questione indica che le stime superiori a 1.750 euro presentano un errore relativo inferiore al 5 per cento; mentre tutte le stime inferiori a 1.750 euro hanno un errore relativo superiore al 5 per cento.

Per confronto, si consideri la cella riferita al *fatturato* con un $\tilde{\varepsilon}\left(\tilde{Y}_a\right) \times 100$ pari al 5 per cento, ma relativa al *cluster-dominio* 7: le stime con un errore relativo inferiore al 5 per cento sono superiori a 11.516.400 milioni di euro; mentre tutte le stime inferiori a 11.516.400 milioni di euro avranno un errore relativo più elevato.

Per analogia, il medesimo ragionamento può essere effettuato per gli altri *cluster-dominio* e per le altre tipologie di dominio.

Prospetto 3.14 - Tipologia di dominio Dom1: aggregati economici in euro, corrispondenti ad alcuni valori tipici degli errori relativi

<i>CLUSTER-DOMINIO</i>	<i>Soglia di</i> $\tilde{\varepsilon}(\tilde{Y}_a) \times 100$	Fatturato	Valore aggiunto	Costo del lavoro	Investimenti	Retribuzioni
1	5	1.750	248	20	7.349	15
	10	13	2	0	55	0
	15	1	0	0	3	0
2	5	2.180.704	308.331	25.085	9.155.356	18.186
	10	16.444	2.325	189	69.036	137
	15	943	133	11	3.958	8
3	5	128.680.998	18.194.285	1.480.240	540.247.767	1.073.162
	10	970.315	137.193	11.162	4.073.719	8.092
	15	55.625	7.865	640	233.533	464
4	5	2.795.132.790	395.205.541	32.152.902	11.734.943.566	23.310.596
	10	21.076.601	2.980.034	242.448	88.486.932	175.773
	15	1.208.254	170.836	13.899	5.072.670	10.076
5	5	28.211.077.397	3.988.781.548	324.516.964	118.439.954.746	235.272.190
	10	212.724.643	30.077.268	2.447.009	893.092.338	1.774.062
	15	12.194.818	1.724.233	140.279	51.198.103	101.701
6	5	433.346.498.183	61.271.127.356	4.984.860.667	1.819.339.932.045	3.613.983.911
	10	3.267.634.121	462.012.794	37.588.168	13.718.669.159	27.251.119
	15	187.322.923	26.485.703	2.154.808	786.447.046	1.562.219
7	5	11.516.400.603.823	1.628.310.949.868	132.475.173.173	48.349.871.476.537	96.043.435.621
	10	86.839.016.173	12.278.221.796	998.924.412	364.580.515.698	724.212.169
	15	4.978.200.650	703.870.845	57.265.114	20.900.224.809	41.516.747

Prospetto 3.15 - Tipologia di dominio Dom2: aggregati economici in euro, corrispondenti ad alcuni valori tipici degli errori relativi

<i>CLUSTER-DOMINIO</i>	<i>Soglia di</i> $\tilde{\varepsilon}(\tilde{Y}_a) \times 100$	Fatturato	Valore aggiunto	Costo del lavoro	Investimenti	Retribuzioni
1	5	10	1	0	31	0
	10	0	0	0	0	0
	15	0	0	0	0	0
2	5	20.641	1.605	107	65.423	77
	10	93	7	0	295	0
	15	4	0	0	13	0
3	5	875.125	68.035	4.521	2.773.740	3.270
	10	3.943	307	20	12.498	15
	15	167	13	1	530	1
4	5	31.880.741	2.478.508	164.705	101.047.189	119.137
	10	143.651	11.168	742	455.308	537
	15	6.093	474	31	19.313	23
5	5	857.668.897	66.677.844	4.430.960	2.718.413.323	3.205.072
	10	3.864.564	300.443	19.965	12.248.880	14.442
	15	163.927	12.744	847	519.573	613
6	5	38.158.588.497	2.966.567.194	197.138.041	120.945.058.963	142.597.028
	10	171.938.518	13.367.034	888.283	544.965.499	642.527
	15	7.293.281	567.002	37.679	23.116.326	27.255
7	5	198.605.287.380	15.440.191.927	1.026.050.987	629.486.811.193	742.179.539
	10	894.894.182	69.571.853	4.623.276	2.836.400.240	3.344.182
	15	37.959.587	2.951.096	196.110	120.314.317	141.853
8	5	5.748.802.501.483	446.929.762.775	29.699.936.782	18.221.042.362.864	21.483.031.235
	10	25.903.489.165	2.013.817.706	133.824.738	82.102.067.918	96.800.241
	15	1.098.773.210	85.422.042	5.676.573	3.482.602.369	4.106.069

Prospetto 3.16 - Tipologia di dominio Dom3: aggregati economici in euro, corrispondenti ad alcuni valori tipici degli errori relativi

<i>CLUSTER-DOMINIO</i>	<i>Soglia di</i> $\tilde{\varepsilon}(\tilde{Y}_d) \times 100$	Fatturato	Valore aggiunto	Costo del lavoro	Investimenti	Retribuzioni
1	5	272.089	62.274	13.774	394.093	10.481
	10	1.614	369	82	2.338	62
	15	80	18	4	116	3
2	5	711.260.065	162.787.436	36.006.601	1.030.186.530	27.397.105
	10	4.219.103	965.634	213.587	6.110.934	162.516
	15	210.194	48.108	10.641	304.445	8.096
3	5	10.936.972.675	2.503.165.612	553.669.794	15.841.072.041	421.282.452
	10	64.876.718	14.848.457	3.284.298	93.967.206	2.498.993
	15	3.232.135	739.745	163.623	4.681.413	124.499
4	5	61.347.454.238	14.040.707.826	3.105.633.832	88.855.432.945	2.363.049.329
	10	363.905.226	83.287.677	18.422.221	527.079.025	14.017.305
	15	18.129.631	4.149.363	917.789	26.258.893	698.337
5	5	217.138.671.816	49.696.938.310	10.992.358.424	314.502.874.373	8.363.988.354
	10	1.288.038.735	294.795.860	65.205.260	1.865.590.690	49.614.106
	15	64.169.639	14.686.627	3.248.503	92.943.076	2.471.757
6	5	1.043.841.893.178	238.906.067.431	52.843.116.945	1.511.896.858.586	40.207.860.555
	10	6.191.936.152	1.417.160.133	313.458.588	8.968.378.141	238.507.869
	15	308.480.094	70.602.422	15.616.397	446.801.464	11.882.379
7	5	6.133.246.058.393	1.403.727.620.030	310.487.479.794	8.883.371.619.042	236.247.178.694
	10	36.381.628.528	8.326.732.099	1.841.771.885	52.695.020.426	1.401.387.946
	15	1.812.520.014	414.834.882	91.756.431	2.625.247.495	69.816.658
8	5	29.435.671.775.423	6.736.997.846.151	1.490.141.999.617	42.634.521.548.242	1.133.835.874.787
	10	174.608.627.442	39.963.006.653	8.839.331.109	252.902.510.463	6.725.768.902
	15	8.698.940.775	1.990.943.020	440.372.386	12.599.514.655	335.075.455
9	5	212.638.131.972.573	48.666.891.248.082	10.764.524.539.813	307.984.308.587.513	8.190.631.564.916
	10	1.261.342.110.647	288.685.753.378	63.853.778.140	1.826.923.394.389	48.585.775.325
	15	62.839.622.981	14.382.223.307	3.181.172.903	91.016.684.807	2.420.526.341

Avvertenze

Nelle tavole sono adoperati i seguenti segni convenzionali:

- Linea (-): 1) quando il fenomeno non esiste;
2) quando il fenomeno esiste e viene rilevato, ma i casi non si sono verificati.
- Asterisco (*): quando il fenomeno esiste ma i dati non sono pubblicati per garantire il rispetto del segreto statistico.
- Simbolo (a): rapporto non significativo.

Bibliografia

- Ballin, M., e P.D. Falorsi, e B. Moreschi, e A. Pavone, e A.C. Seeber, e T. Tuoto. “Aspetti metodologici connessi alla produzione delle stime per il Censimento delle Istituzioni non profit”. In *Istituzioni nonprofit in Italia -I risultati della prima rilevazione censuaria - anno 1999*. Roma: Istat, 2001. (Collana Informazioni).
- Ballin, M., e S. Loriga. “Metodi di selezione coordinata nelle indagini campionarie sulle imprese”. *Rivista di statistica ufficiale*, n. 2 (2000).
- Belson, W. A. “Matching and Prediction on the Principle of Biological Classification”. *Applied Statistics*, 8 (1959): 65-75.
- Bethel, J. “Sample allocation in multivariate surveys”. *Survey methodology*, 15 (1989): 47-57.
- Biggs, D., e B. de Ville, e E. Suen. “A method of choosing multiway partitions for classification and decision trees”. *Journal of Applied Statistics*, 18 (1991): 49-62.
- Cicchitelli, G., e A. Herze, e G.E. Montanari. “Un dominio si dice stratificato se le sue unità sono raggruppate in uno o più strati costituiti da quelle sole unità”. *Il campionamento statistico* (1992): 256.
- Cochran, W.G., e J. Wiley. *Sampling Techniques*. New York, 1977.
- Deville, J.C., e C.E. Särndal. “Calibration Estimators in Survey Sampling”. *Journal of the American Statistical Association*, 87 (1992): 376-382.
- Falorsi, P.D., e M. Ballin, e C. De Vitiis, e G. Scepi. “Principi e metodi del software generalizzato per la definizione del disegno di campionamento nelle indagini sulle imprese condotte dall’Istat”. *Statistica Applicata*, 20, n. 2 (1998).
- Falorsi, P.D., e S. Falorsi. “Un metodo di stima generalizzato per le indagini sulle famiglie e sulle imprese”. *Rapporto di ricerca*, n. 13. Università di Bologna, 1995. (Quaderni CON PRI).
- Falorsi P.D., Filiberti P.D., Pavone A. “The new strategy for the concise presentation of sampling errors in the Italian Structural Business Statistics Survey” *Statistical Methods and Applications, Statistical Methods and Applications Journal of the Italian Statistical Society* vol 15.2, (2006): 243-265.
- Casciano C., Falorsi P.D., Filiberti S., Pavone A, Siesto G. “Principi e metodi per il calcolo delle stime finali e la presentazione sintetica degli errori di campionamento nell’ambito delle rilevazioni strutturali sulle imprese” *Rivista di Statistica Ufficiale* n°1/2004 (2006). 67-102.
- Istat. *Classificazione delle attività economiche*. Roma: Istat, 2003. (Metodi e norme, n. 18).
- Istat. Relazione finale del gruppo di lavoro sulla *Definizione delle metodologie applicabili alle rilevazioni ed indagini sulle statistiche economiche circa il problema delle mancate risposte parziali e totali*. Mimeo.
- Kish, L. *Survey Samplin*. New York: John Wiley, 1965.
- Neville, PG. *Decision Trees for Predictive Modeling*. “SAS Technical Report”. The SAS Institute, 1999.

Ohlsson, E., e B.G. Cox, e D.A. Binder, e B.N. Chinnappa, e A. Christianson, e M.J. Kott, e P.S. Colledge. "Coordination of samples using permanent random numbers". In *Business Survey Methods*. New York: Wiley, 1995.

Pavone, A., A. Russo. "Generalized Variance Function: Theory and Empirics". Atti della XLII Riunione Scientifica, Società Italiana di Statistica, 2004.

Russo, A. *Sulla Presentazione degli Errori di Campionamento mediante Modelli: Il Metodo dei Modelli Regressivi*. Roma, 1987. (Quaderni di Discussione, n. 4).

Särndal, C.E., e B. Swensson, e J. Wretman. *Model Assisted Survey Sampling*. New York: Springer-Verlag, 1992.

Valliant, R. "Generalized Variance Functions in Stratified Two-Stage Sampling". *Journal of the American Statistical Association*, 82 (1987): 499-508.

Wolter, K. M. *Introduction to Variance Estimation*. New York: Springer-Verlag, 1985.

Glossario

Acquisti: comprendono gli acquisti di materie prime, materie di consumo, merci da rivendere senza trasformazione, servizi e di beni di godimento di terzi. Gli acquisti sono richiesti al netto dell'Iva fatturata dai fornitori o gravante sui prodotti importati, delle merci rese ai fornitori e degli sconti ed abbuoni da questi concessi ma al lordo di qualsiasi onere sostenuto fino all'immissione dei prodotti nei magazzini dell'impresa (trasporti pagati a terzi, dazi doganali, imballaggi da non rendere eccetera).

Addetto: la persona occupata in un'unità giuridico-economica, come lavoratore indipendente o dipendente (a tempo pieno, a tempo parziale o con contratto di formazione lavoro), anche se temporaneamente assente (per servizio, ferie, malattia, sospensione dal lavoro, Cassa integrazione guadagni eccetera). Comprende il titolare/i dell'impresa partecipante/i direttamente alla gestione, i cooperatori (soci di cooperative che come corrispettivo della loro prestazione percepiscono un compenso proporzionato all'opera resa ed una quota degli utili dell'impresa), i coadiuvanti familiari (parenti o affini del titolare che prestano lavoro manuale senza una prefissata retribuzione contrattuale), i dirigenti, quadri, impiegati, operai e apprendisti.

Attività economica: la produzione di beni o servizi che ha luogo quando risorse quali lavoro, impianti e materie prime concorrono all'ottenimento di beni o alla prestazione di servizi. Un'attività economica è caratterizzata dall'uso di fattori della produzione, da un processo di produzione e da uno o più prodotti ottenuti (merci o prestazione di servizi). Ai fini della produzione dell'informazione statistica, le attività economiche sono classificate secondo una nomenclatura internazionale che a livello europeo è denominata Nace Rev. 1 (per la classificazione Ateco versione 1991).

Attività economica (classificazione della Ateco 91): la classificazione che distingue le unità di produzione secondo l'attività da esse svolte e finalizzata all'elaborazione di statistiche di tipo macroeconomico, aventi per oggetto i fenomeni relativi alla partecipazione di tali unità ai processi economici. La classificazione comprende 874 categorie, raggruppate in 512 classi, 22 gruppi, 60 divisioni, 17 sezioni e 16 sottosezioni.

Dipendente (addetto): la persona che svolge la propria attività lavorativa in un'unità giuridica-economica e che è iscritta nei libri paga dell'impresa o istituzione. Sono considerati tra i lavoratori dipendenti:

- i soci di cooperativa iscritti nei libro paga;
- i dirigenti, i quadri, gli impiegati e gli operai, a tempo pieno o parziale;
- gli apprendisti;
- i lavoratori a domicilio iscritti nei libro paga;
- i lavoratori stagionali;
- i lavoratori con contratto di formazione lavoro.

Fatturato: comprende i ricavi delle vendite e delle prestazioni. In particolare le vendite di prodotti fabbricati dall'impresa, gli introiti per lavorazioni eseguite per conto terzi, gli introiti per eventuali prestazioni a terzi di servizi non industriali (commissioni, noleggi di macchinari eccetera), le vendite di merci acquistate in nome proprio e rivendute senza trasformazione, le commissioni, provvigioni ed altri compensi per vendite di beni per conto terzi, gli introiti lordi del traffico, le prestazioni di servizi a terzi. Il fatturato viene richiesto al lordo di tutte le spese addebitate ai clienti (trasporti, imballaggi, assicurazioni e simili) e di tutte le imposte indirette (fabbricazione, consumo eccetera) ad eccezione dell'Iva fatturata ai clienti, al netto degli abbuoni e sconti accordati ai clienti e delle merci rese; sono esclusi anche i rimborsi di imposte all'esportazione, gli interessi di mora e quelli sulle vendite rateali. Il valore dei lavori eseguiti nel corso dell'esercizio da parte delle imprese di costruzione e cantieristiche sono conglobati nel valore complessivo del fatturato.

Fatturato esportato: la quota del fatturato dovuto all'esportazione di beni o servizi.

Giacenze (variazione di): la differenza fra la consistenza delle rimanenze di magazzino, a fine ed inizio esercizio, dei prodotti finiti, prodotti in corso di lavorazione, semilavorati, lavori in corso su ordinazione, materie prime e sussidiarie di consumo e merci da rivendere senza trasformazione.

Impresa: l'unità giuridico-economica che produce beni e servizi destinabili alla vendita e che, in base alle leggi vigenti o a proprie norme statutarie, ha facoltà di distribuire profitti realizzati ai soggetti proprietari siano essi privati o pubblici. Il responsabile è rappresentato da una o più persone fisiche, in forma individuale o associata, o da una o più persone giuridiche. Tra le imprese sono comprese: le imprese individuali, le società di persone, le società di capitali, le società cooperative, le aziende speciali di comuni o province o regioni. Sono considerate imprese anche i lavoratori autonomi e i liberi professionisti.

Impresa esportatrice: l'impresa che segnala la presenza di fatturato dovuto all'esportazione di beni o servizi.

Indipendenti (addetti): sono rappresentati da a) imprenditori, titolari, liberi professionisti, lavoratori autonomi, purché partecipino direttamente alla gestione dell'impresa e non si servano di un gestore o coadiutore o di altra persona diversamente nominata. Nel caso di società, sono la persona o le persone fisiche che risultano tali dagli atti amministrativi della società stessa (amministratore unico, consigliere delegato eccetera); b) soci di cooperativa di produzione e di lavoro i quali, come corrispettivo dell'opera prestata, non percepiscono una remunerazione regolata dai contratti di lavoro ma bensì un compenso proporzionato alla prestazione nonché una quota parte degli utili dell'impresa. In tale categoria non sono compresi i soci semplicemente iscritti o conferenti; c) coadiuvanti familiari (parenti o affini dell'imprenditore, titolare eccetera che prestano lavoro manuale senza una prefissata retribuzione contrattuale).

Investimenti fissi lordi: misurano le acquisizioni di capitali fissi effettuate nel corso dell'anno e comprendono anche il valore dei beni capitali prodotti dall'azienda per uso proprio e delle riparazioni e manutenzioni straordinarie eseguite dall'impresa stessa sugli impianti aziendali.

Margine operativo lordo: rappresenta il surplus ed è calcolato sottraendo il costo del lavoro al valore aggiunto generato dall'attività produttiva dopo aver remunerato il lavoro dipendente.

Ore effettivamente lavorate: comprendono le ore di lavoro ordinarie, straordinarie, festive e notturne con esclusione delle ore pagate ma non lavorate per ferie, malattia eccetera.

Produzione di capitali fissi per uso proprio: comprende il valore delle immobilizzazioni realizzate dall'impresa nel corso dell'esercizio per uso proprio ed impiegando la manodopera interna all'impresa e quello delle riparazioni e manutenzioni straordinarie agli impianti aziendali utilizzando manodopera dell'impresa stessa.

Redditività lorda: misurata dalla quota di valore aggiunto assorbita dal margine operativo lordo. Tale indicatore si ottiene depurando il margine operativo lordo dalla componente di remunerazione dei lavoratori indipendenti assimilabile al "reddito da lavoro" dell'imprenditore.

Retribuzioni lorde: comprendono i salari, gli stipendi e le competenze accessorie, in denaro e in natura, al lordo delle trattenute erariali e previdenziali, corrisposte ai lavoratori dipendenti direttamente e con carattere di periodicità, secondo quanto stabilito dai contratti, dagli accordi aziendali e dalle norme di legge in vigore.

Ricavi: vedi Fatturato

Spese per il personale: comprendono tutte le voci che costituiscono la retribuzione lorda del personale dipendente (dirigenti, quadri, impiegati, operai, commessi, apprendisti e lavoranti a domicilio): paga base,

indennità di contingenza ed altre indennità similari per la parte non conglobata, interessenze, lavoro straordinario, compensi per ferie e festività, gratifiche natalizie, mensilità oltre la dodicesima ed altre analoghe erogazioni e corresponsioni in natura. Sono inoltre incluse le spese per contributi sociali al netto di eventuali fiscalizzazioni, le provvidenze varie, le quote accantonate nell'esercizio per provvedere alla successiva corresponsione delle indennità di fine rapporto lavoro e le spese sociali varie (nidi di infanzia, colonie marine e montane eccetera).

Valore aggiunto: rappresenta l'incremento di valore che l'attività dell'impresa apporta al valore dei beni e servizi ricevuti da altre aziende mediante l'impiego dei propri fattori produttivi (il lavoro, il capitale e l'attività imprenditoriale). Tale aggregato è ottenuto sottraendo l'ammontare dei costi al totale dei ricavi: i primi comprendono i costi per acquisti lordi, per servizi vari e per godimento di servizi di terzi, le variazioni delle rimanenze di materie e di merci acquistate senza trasformazione e gli oneri diversi di gestione; i secondi contengono il valore del fatturato lordo, le variazioni delle giacenze di prodotti finiti, semilavorati ed in corso di lavorazione, gli incrementi delle immobilizzazioni per lavori interni ed i ricavi accessori di gestione.

Unità funzionale: l'unità che all'interno di una impresa raggruppa l'insieme delle parti che concorrono all'esercizio di un'attività economica a livello di classe (quattro cifre) della classificazione Nace Rev.1. Si tratta di un'entità che corrisponde a un sistema di informazioni che consentono di fornire o calcolare per ogni unità di attività economica almeno il valore della produzione, i consumi intermedi, le spese per il personale, il risultato di gestione, l'occupazione e gli investimenti fissi lordi.