

# Sesta Conferenza nazionale di statistica

Roma, 6-7-8 novembre 2002

## Il fattore territorio nelle statistiche economiche: fonti amministrative e rilevazioni statistiche

M. Ballin, M. Calzaroni, C. Casciano, G. Garofalo, L. Martino, R. Monducci, G. Siesto  
(ISTAT)

**Sintesi:** La domanda di informazioni proveniente da utenti, istituzionali e non, circa un sempre maggiore dettaglio territoriale nella produzione di statistiche ufficiali, unitamente alla esigenza di tenere conto di due vincoli quali la necessità di contenere il fastidio statistico e i costi per produrre tali informazioni, impongono la ricerca sistematica di nuovi approcci utili a perseguire tali obiettivi. Per quanto riguarda le statistiche economiche sulle unità produttive, le strategie seguite dall'Istat prevedono, da un lato, un utilizzo sempre maggiore delle fonti amministrative dagli enti che acquisiscono, per propri fini istituzionali, informazioni con elevato dettaglio territoriale; dall'altro, la ricerca di soluzioni metodologiche ed organizzative adeguate ad incrementare la qualità delle stime territoriali contenendo i costi di rilevazione/elaborazione. Qui di seguito vengono illustrati i problemi, di carattere concettuale, metodologico ed organizzativo, connessi all'adozione di questi approcci ed alcune soluzioni innovative per il miglioramento della qualità dei dati territoriali, con specifico riferimento alla costruzione dell'archivio statistico delle imprese, alla stima dei risultati economici delle imprese industriali e dei servizi, alle rilevazioni congiunturali e strutturali sul settore agrario..

**Parole chiave:** fonti amministrative, unità locali, piccole e medie imprese, aziende agricole, qualità.

### 1. La dimensione territoriale nelle fonti amministrative: Problemi concettuali e di misura.<sup>1</sup>

I problemi di produzione e di qualità delle statistiche economiche a livello locale sono ben noti, essi sono legati ai costi relativamente maggiori che la produzione di statistiche a livello locale deve sopportare, ma anche, se non soprattutto, alle difficoltà oggettive di rilevazione. La disomogeneità dell'unità utilizzata per la raccolta dei dati a livello locale, la difficoltà nella osservazione di tale unità e nella stima di variabili<sup>2</sup>

---

<sup>1</sup> A cura di M. Calzaroni e G. Garofalo.

<sup>2</sup> Le difficoltà di stima delle variabili sono ben presenti nell'ambito dei regolamenti comunitari basti pensare che il sistema dei conti regionali, nell'ambito degli aggregati per branca di attività, si limita a sole

significative, la disponibilità solo parziale di informazioni a carattere amministrativo, fanno sì che il problema della qualità delle statistiche economiche a livello locale deve essere in qualche maniera “relativizzato” nella considerazione che statistiche a livelli territoriali più ristretti producono qualità decrescente nelle informazioni.

A questo va aggiunto che alcuni fenomeni quali la terziarizzazione dell’economia (i “servizi” sono meno legati al concetto di luogo fisico rispetto ad altri settori produttivi), la riduzione dei costi e dei tempi di trasporto (che riduce i problemi collegati alle barriere naturali “locali”<sup>3</sup>), lo sviluppo della *e-economy* (che al luogo fisico sostituisce un luogo virtuale) rendono sempre meno lineare il legame fra il soggetto economico, la sua residenza e il territorio dove avviene la sua attività di produzione. In termini più generali l’ambito territoriale in cui l’attività economica viene effettivamente svolta è sempre più indipendentemente dalla residenza sia del fattore sia dell’unità che la produce.

### 1.1. Concetti e definizioni

L’unità statistica di osservazione e di analisi a cui si riferiscono tutti gli aggregati microeconomici a livello territoriale è l’unità locale. Essa è definita dal regolamento europeo n° 696/93<sup>4</sup> nel seguente modo: *“l’unità locale corrisponde ad una impresa o parte di essa (fabbrica, negozio, officina, miniera, ufficio, deposito, ...) situata in un luogo topograficamente identificato. In tale luogo, o a partire da esso, sono esercitate attività economiche per le quali una o più persone lavorano (eventualmente a tempo parziale) per conto di una stessa impresa”*.

La definizione europea mette in evidenza tre elementi fondamentali che permettono di considerare un “luogo” come unità statistica: la localizzazione fisica, lo svolgimento di una attività, la presenza del fattore lavoro. Di seguito per ciascuno di questi elementi verranno evidenziati i problemi concettuali e le problematiche connesse alla loro corretta interpretazione

L’elemento fondamentale della definizione è il concetto di unità legale come *“luogo topograficamente identificato”*. Tale concetto deve essere inteso in senso stretto: due unità di una stessa impresa che hanno localizzazioni diverse devono essere considerate due unità locali<sup>5</sup>.

Se l’impresa (intesa come organizzazione di fattori di produzione per la realizzazione di beni e servizi) è un “concetto” univoco ben delineato – e direttamente osservabile per tramite dell’unità legale - quello di unità locale ha sempre presentato tutta una serie di eccezioni: attività con localizzazioni a tempo parziale, attività che si muovono nel territorio, attività senza localizzazione. E’ ben nota la difficoltà “censuaria” di individuare le cosiddette attività non estensibili, svolte presso la residenza dell’imprenditore, o le unità del settore dell’edilizia caratterizzate da una instabilità in

---

cinque variabili (valore aggiunto, redditi da lavoro dipendente, occupati totale, lavoratori dipendenti, investimenti fissi) che tengono conto dei reali elementi “osservazionali” - cfr SEC 95 par 13.14. -

<sup>3</sup> Non è più vero che “un’impresa produttrice di confezioni di lana trova una convenienza economica ad operare in prossimità di imprese produttrici di filati”, G Alvaro....

<sup>4</sup> Council Regulation on “Statistical units for the observation and analysis of the production system in the Community”.

<sup>5</sup> Nelle note esplicative del regolamento si specifica anche che il confine dell’unità deve essere identificato dal confine del “sito” in cui essa si colloca, ove questo sito si estende in più aree amministrative la localizzazione è determinata per convenzione dall’indirizzo postale.

termini di localizzazione. In relazione a questi problemi nella definizione comunitaria unità locale non è solo il luogo dove si svolge l'attività ma è anche il *“luogo dove si organizza l'attività”*. L'introduzione del concetto di *“organizzazione”* risolve solo per alcune attività economiche – trasporti, servizi alle comunità, distribuzione delle fonti energetiche - i problemi evidenziati. Da un lato, infatti, per tutta una serie di attività (libero professionali e di lavoro autonomo) l'unica informazione disponibile e *“reale”* è la *“residenza (anagrafica o fiscale)”* dell'operatore economico. D'altro lato la terziarizzazione e l'informatizzazione dell'economia ha determinato lo sviluppo di processi, ancora non ben quantificati, di commercializzazione e di produzione di servizi che in quanto tali sono *“senza territorio”*.

Le considerazioni brevemente sviluppate evidenziano che nella produzione di statistiche economiche a livello locale si utilizza come unità di analisi una entità (l'unità locale) non omogenea concettualmente, essendo un insieme di entità differenti: luoghi fisici dove si svolge la produzione, residenza dell'operatore economico, luogo da cui si organizza l'attività o, infine, attività la cui suddivisione in termini territoriale o non esiste o è difficile da individuare.

Altro tema messo in luce dalla definizione è che nell'unità locale deve essere presente (e rilevabile) un input di lavoro significativo, definito almeno pari all'equivalente annuo di una persona che lavora a tempo parziale. Questo implica che tutte le unità *“non presidiate”* non sono da considerarsi come unità locali. A parte le difficoltà di individuare se una unità sia presidiata o meno, per la quantificazione dell'occupazione in ambito locale si pongono due problemi di una certa rilevanza sia concettuale che pratica:

- la localizzazione del lavoro *“indipendente”* in presenza di una attività imprenditoriale che si svolge in più unità. Il rischio della duplicazione di posizioni lavorative è ben evidente<sup>6</sup>; d'altro lato se la ripartizione di tale lavoro nelle varie unità è sicuramente complessa, l'assegnazione di tutto il lavoro indipendente presso la *“sede amministrativa”* dell'impresa può produrre distorsioni<sup>7</sup> nella stima dell'occupazione a livello territoriale;

- la stima della cosiddetta *“area grigia”* di occupazione (collaboratori, lavoro interinale, ecc.), la cui incidenza è accresciuta negli ultimi anni. Si tratta di occupazione mobile nel territorio o meglio occupazione, per le caratteristiche di prestazione del lavoro, a cui è difficile assegnare un attributo quale la posizioni sul territorio, in qualche maniera siamo presenti a *“occupazione senza territorio”*. D'altra parte le *“regole”* attualmente fissate in ambito nazionale ed internazionale obbligano a non collocare tale forza lavoro presso l'impresa presso cui prestano la loro attività ma, ad esempio, presso l'impresa interinale. Utilizzare la residenza del collaboratore, o dell'impresa interinale, può falsare in maniera notevole le stime a livello territoriale (soprattutto sub-provinciale), fino a determinare la soppressione di unità locali se *“presidiate”* solo da questo tipo di occupazione. Del resto l'assegnazione di tale occupazione nel luogo dove svolge effettivamente l'attività (ad es. unità locale dell'impresa che la utilizza) è operazione complessa, data la necessità di collegare fra loro le posizioni individuali a quelle

---

<sup>6</sup> Nell'ambito della rilevazione censuaria l'individuazione di tali duplicazioni (imprenditori, soci, familiari dichiarati in più unità locali) è una operazione particolarmente complessa.

<sup>7</sup> Soprattutto nelle imprese di media dimensione del settore del commercio e dei servizi dove possono coesistere sia la presenza di più unità locali sia la presenza di un significativo apporto di lavoro indipendente alle attività.

dell'unità locale<sup>8</sup> dove viene prestato il servizio, e può portare a duplicazione di occupazione in presenza di attività svolte in periodi brevi per più imprese.

La definizione individua l'unità locale come luogo in cui si effettua una attività economica. Un ulteriore problema sorge quindi dal trattamento delle unità territoriali dedicate alle attività "ausiliarie". Una attività ausiliaria, o ancillare, consiste in una attività di supporto (marketing, amministrazione, trasporto, immagazzinamento,...) esercitata all'interno di una impresa al fine di creare le condizioni idonee all'esercizio dell'attività dell'impresa stessa. In quanto tale il prodotto dell'attività ausiliaria non è destinato al mercato e gli input (lavoro, macchinari, immobili,...) sono considerati input dell'attività da essa supportata. Per tale ragione l'attività ancillare è classificata secondo il codice di attività economica (prevalente o secondaria) da essa supportata. Il problema delle attività ancillari si presenta soprattutto nella grandi imprese dove, spesso, l'ambito territoriale dell'unità ancillare è differente da quello dell'unità servita. A queste si applica la seguente definizione: "le attività ausiliarie devono essere attribuite alla regione in cui sono esercitate e restano incluse nella stessa branca di attività dell'unità di cui sono al servizio"<sup>9</sup>. In tal modo vengono classificate, ad es. come attività manifatturiere, attività dei servizi. Se a livello dell'intera economia nazionale, e in termini di analisi macroeconomica, questo approccio è del tutto corretto poiché, evitando la registrazione di processi spuri di terziarizzazione, garantisce corrette analisi sull'equilibrio dei vari mercati; in un ambito geografico limitato, dove maggiore è l'esigenza di legare la struttura economica alla realtà socio-demografica, può produrre analisi economiche (ad es. utilizzando indici di specializzazione, di vocazione territoriale, di dotazione) fortemente inconsistenti. Anche in questo caso la risposta che si sta cercando di dare nell'ambito della realizzazione di ASIA Unità Locali è la possibilità di avere una descrizione non unica delle unità locali: una che rispetta le regole di classificazione ed una che permetta di descrivere l'effettiva attività svolta in una unità locale, soprattutto se ancillare, in modo da descrivere specifici ambiti territoriali, si pensi ai SLL, anche secondo la effettiva attività svolta e le reali caratteristiche della forza lavoro presente.

## **1.2. Le fonti amministrative e le stime territoriali**

L'uso di dati amministrativi a fini statistici è sicuramente la strada più innovativa che l'Istat, almeno per quello che riguarda le statistiche economiche, ha intrapreso negli ultimi anni: numerose sono le esperienze già realizzate (l'archivio ASIA, l'utilizzo dei dati dell'Inps per l'analisi congiunturale dell'occupazione, l'uso dei bilanci civilistici per le analisi strutturali) o in fase avanzata di realizzazione (l'uso dei dati fiscali infrannuali per le analisi congiunturali). Tali esperienze delineano un quadro "generalizzato" e non "episodico" nell'utilizzo delle fonti amministrative, strettamente connesso allo sviluppo di una "nuova metodologia statistica" che garantisce normalizzazione e integrazione fra i dati di fonti amministrative e fra

---

<sup>8</sup> Nell'ambito dell'archivio ASIA è in fase di sviluppo la costruzione di un archivio satellite in cui, per ciascuna impresa, si cerca di stimare il totale di input di lavoro ottenuto aggiungendo, al numero di dipendenti ed indipendenti, una componente di occupazione definita "esterna" e composta da lavoratori interinali, collaboratori coordinati e continuativi, ecc..

<sup>9</sup> Eurostat, metodologia dei conti regionali

questi e i dati statistici. Questo significa definire regole di traduzione dai concetti e classificazioni giuridico-amministrative ai concetti e classificazioni statistiche.

I dati amministrativi, opportunamente trattati, rappresentano sempre più la base informativa generale, il *framework*, mentre l'indagine statistica sta assumendo il ruolo di strumento per:

- approfondire particolari tematiche;
- raccogliere le informazioni per le unità produttive più rilevanti economicamente;
- validare sia le informazioni amministrative sia le metodologie realizzate per il loro trattamento.

Le potenzialità (e le difficoltà) nell'uso per fini statistici di informazioni residenti in basi di dati amministrative sono, però, differenti per differenti tipologie di unità di rilevazione. Se, infatti, il loro utilizzo è ormai diffuso e consolidato per la raccolta di informazioni dove l'unità di analisi sono le "imprese" o i "gruppi d'impresa", per altre tipologie di unità, quali le "unità locali", le attività sono ancora in una fase sperimentale. La causa di questo ritardo va ricercata, essenzialmente, nelle caratteristiche delle fonti amministrative:

1. *La mancanza di un codice identificativo univoco*, quale è il codice fiscale per le imprese, determina problemi nell'integrazione di informazioni residenti in più archivi amministrativi e fra queste e quelle raccolte da indagini statistiche. Per risolverli è necessario lo sviluppo di metodologie di *linkage* che utilizzano come chiavi di accoppiamento stringhe di caratteri alfanumerici. L'utilizzo di queste metodologie, il cui sviluppo nell'istituto è comunque in una fase molto avanzata di sperimentazione<sup>10</sup>, deve comunque fare i conti con il bilanciamento fra "quantità" dei *link* cercati e "qualità" dei *link* ottenuti<sup>11</sup>;
2. *La qualità dei dati registrati* in fonti non statistiche è strettamente connessa con "l'importanza" che le amministrazioni e i soggetti danno all'informazione raccolta. Le informazioni sono "migliori" dove sono maggiormente funzionali ai soggetti che le dichiarano e/o alle amministrazioni che le raccolgono. In questo senso la qualità delle informazioni sulle imprese (nel senso di unità giuridiche su cui pesano, ad esempio, vincoli fiscali, previdenziali o finanziari) è sicuramente maggiore di quella sulle unità locali, dove i vincoli amministrativi sono sicuramente minori e meno stringenti;
3. *Le informazioni disponibili* per le unità locali sono limitate. All'ovvia mancanza di dati di tipo contabile, si accompagna, spesso, la carenza di informazioni sulla "dimensione" delle unità. L'informazione sugli "addetti all'unità locale", pur teoricamente disponibile in alcune fonti, è raccolta in maniera del tutto parziale e non controllata. Questa situazione implica che una variabile fondamentale, anche per operare stime su altri aggregati economici a livello locale (ad. es. il valore

---

<sup>10</sup> Sono già utilizzate in campi quali l'individuazione dei flussi demografici "reali" di imprese, la codifica automatica dell'attività economica e delle natura giuridica, la standardizzazione degli indirizzi (bibliografia, xxx).

<sup>11</sup> Intesa come "link" che rappresentano accoppiamenti di unità realmente uguali.

della produzione), o deve essere rilevata da indagini statistiche o deve essere stimata, utilizzando variabili ausiliarie e/o opportuni modelli statistici.

Pur in presenza di tali problemi, l'Istat sta approfondendo l'analisi sulle informazioni disponibili sulle unità locali in varie fonti amministrative e ha iniziato lo sviluppo di opportune metodologie di integrazione e di stima. L'obiettivo principale è quello di supportare le attività censuarie nella definizione di un universo di unità locali da considerarsi come *benchmark* e di garantire un aggiornamento continuo di tale "lista", cioè di realizzare ASIA unità locali, che è una delle priorità individuata nell'attività dell'Istat di questi anni.

In considerazione delle problematiche presenti nella realizzazione di stime territoriali, è utile sottolineare il potenziale informativo di una "nuova" fonte amministrativa. Il cosiddetto "contatore Inail", cioè la denuncia nominativa degli assicurati – DNA, riunisce in se tutte le caratteristiche necessarie per essere fortemente utilizzata nel processo di stima delle unità locali. Da essa, infatti, si possono ricavare, già integrate, informazioni sull'impresa, sul lavoratore, sul luogo dove questi svolge la sua attività ed anche, con un maggiore margine di errore, il tipo di attività svolta, utilizzando l'informazione circa il tipo di rischio di infortunio per cui è assicurato. Unitamente ad esperti dell'Inail è iniziato un lavoro di analisi di tali informazioni che, a patto di superato problemi di qualità non di poco conto, hanno confermato l'enorme potenziale di tale fonte per "dare una mano" nella realizzazione di stime con un dettaglio territoriale che può arrivare ad ambiti infracomunali.

E' necessario infine sottolineare come i problemi messi in evidenza non sono superabili, almeno nel breve periodo, senza integrare le informazioni desunte con dalle fonti amministrative con informazioni direttamente acquisite da indagini statistiche. Se, come detto, il censimento rappresenta il *benchmark*, da cui partire, è da prevedere che, almeno per le grandi imprese, sia necessario effettuare indagini specifiche per verificare la qualità delle liste di unità locali ottenute tramite le fonti amministrative. Una indagine progettata per avere un elevato profilo di qualità su un numero limitato di imprese è l'investimento minimo necessario per arrivare a definire e validare una metodologia efficace e che in prospettiva consenta di evitare la realizzazione di un nuovo censimento dell'industria, ottenendo quindi, nel medio periodo, il massimo di riduzione possibile sia del fastidio statistico sia dei costi di rilevazione.

### **1.3. Prime valutazioni sulla qualità di alcune fonti amministrative**

Le fonti attualmente disponibili presso l'Istituto e utilizzabili al fine di individuare le unità locali delle imprese e i loro principali caratteri sono brevemente descritte qui di seguito:

CCIAA – Le Camere di Commercio, nell'ambito della loro attività di gestione del Registro delle Imprese e del Repertorio Economico Amministrativo (REA), registrano le unità locali di tutti i soggetti iscritti nei due archivi. Ogni soggetto è obbligato a denunciare, presso la Camera competente in ambito provinciale e per tramite di un apposito modello, l'apertura, modifica o cessazione delle unità locali (diverse dalla Sede). Ai fini camerali è considerata unità locale "l'impianto operativo

o amministrativo-gestionale, ubicato in luogo diverso dalla sede, nel quale l'impresa esercita stabilmente una o più attività economiche. La diversificazione dell'ubicazione può essere determinata anche dalla sola variazione del numero civico o del numero di interno nell'ambito dello stesso fabbricato". A livello di unità locale devono essere dichiarati, oltre ai caratteri di localizzazione, l'attività economica prevalente e secondaria svolta, il tipo di unità locale (deposito, miniera, agenzia, officina, ecc.) e il numero di addetti, indipendenti e dipendenti, che prestano lavoro solo nell'unità locale;

IRAP - L'Imposta Regionale sulle Attività Produttive, è una imposta locale introdotta nel 1997. Il presupposto dell'imposta è l'esercizio abituale di un'attività economica autonomamente organizzata e diretta alla produzione e/o scambio di beni o alla prestazione di servizi. In tal senso l'imposta si applica nei confronti di imprenditori individuali, società, enti commerciali e non, esercenti arti e professioni, Stato e altre pubbliche amministrazioni. La base imponibile è il valore della produzione netta derivante dall'attività esercitata nel territorio della Regione. La produzione netta è data dalla differenza fra la somma delle voci del valore della produzione e la somma dei costi sostenuti. Alla base imponibile calcolata si applica una aliquota determinata a livello regionale che può variare secondo il settore di attività.

MUD – Il Modello unico di dichiarazione Ambientale è utilizzato per denunciare i rifiuti derivanti da un'attività economica (e diversi da quelli raccolti dal Comune dove l'attività è svolta) prodotti, smaltiti, avviati al recupero e trasportati. La dichiarazione deve essere effettuata da tutte le imprese (o gli Enti) e per tutte le unità locali in cui vengono prodotti rifiuti definiti pericolosi. Sono esonerati dalla presentazione del MUD le attività artigianali con meno di 3 dipendenti, le attività di demolizione e costruzione e le attività di commercio e di servizio sole se non producono rifiuti pericolosi. Le informazioni sono quindi limitate a un sottoinsieme di unità locali, per il quale vengono raccolte informazioni identificative che vanno dalla localizzazione, all'attività economica svolta al numero dei dipendenti.

SEAT / Pagine gialle<sup>12</sup> - I dati disponibili sono il risultato dell'incrocio fra differenti informazioni su utenti telefonici disponibili in differenti Data Base di diverse società del gruppo TELECOM. In particolare l'Istat acquista le seguenti informazioni: localizzazione, attività economica, n° di telefono fisso (relative agli utenti che utilizzano le pagine gialle per la loro pubblicità integrate con informazioni aggiuntive di contratto) numero di linee telefoniche, URL, e-mail, cellulari, ecc.. I dati di localizzazione sono relativi al luogo fisico dove è presente l'allaccio telefonico, in pratica l'unità locale.

Da quanto esposto risulta evidente come gli universi di riferimento delle fonti descritte siano diversi, mentre per le unità registrate si utilizzano, quantomeno, definizioni non omogenee. E' anche da sottolineare che la rilevanza "amministrativa" di tali informazioni è diversa da fonte a fonte, determinando, come già accennato, una diversa qualità delle informazioni acquisite.

---

<sup>12</sup> Pur non essendo una fonte amministrativa l'importanza e il suo utilizzo in Istat sono tali da poterla inserire in questa analisi.

Di seguito sono riportati risultati dei primi confronti effettuati fra le diverse fonti descritte. I confronti, effettuati su sottoinsiemi definiti dall'accoppiamento di codici fiscali, utilizzano come archivio di riferimento quello camerale in quanto contenente (per definizione e in pratica) l'universo di imprese e di unità locali più ampio. In questa prima fase i confronti sono effettuati utilizzando unità fittizie quali l'unità regionale e quella provinciale<sup>13</sup>. Questo ci permette di utilizzare tutte le fonti descritte e di confrontare la copertura relativa in termini di presenza di unità locali per ambiti territoriali definiti.

L'obiettivo degli schemi presentati nelle figure 1, 2 e 3 è quello di determinare una prima quantificazione dei problemi che possono sorgere nel confrontare le informazioni residenti in diverse fonti amministrative. Negli schemi vengono messi in evidenza, in maniera sintetica, due aspetti fondamentali:

- l'area problematica, individuata da tutte le imprese che presentano nelle due fonti confrontate un numero diverso di unità e/o differenti localizzazioni nelle unità stesse;
- il numero complessivo di unità differenti (localizzate in diverse province o regioni) nelle fonti considerate.

Gli elementi generali che si possono trarre sono descritti brevemente di seguito

1. Come era ipotizzabile i problemi di coerenza fra le varie fonti aumentano se si prendono in considerazione ambiti territoriali più ristretti: se nel confronto, a livello regionale, fra CCIAA e IRAP l'area problematica è relativa al solo 2% delle imprese accoppiate, considerando i confronti fra CCIAA e MUD e fra CCIAA e SEAT, la percentuale di imprese che presentano problemi in termini di localizzazione a livello di provincia è maggiore risultando pari, rispettivamente, al 6,4% e al 5,3%. Inoltre se gli ultimi due confronti si effettuano a livello regionale (piuttosto che provinciale) le percentuali che si ottengono, per l'area di imprese con problemi, sono del tutto simili alla percentuale ottenuta nel confronto CCIAA/IRAP.
2. Pur presentando percentuali relativamente basse il numero delle imprese da "verificare", perché con informazioni discordanti, è comunque consistente: oltre 100 mila nel confronto fra CCIAA e SEAT, 54 mila in quello con l'IRAP e circa 21 mila in quello con il MUD. In particolare in quest'ultimo caso il valore ottenuto sembra particolarmente elevato in considerazione del numero dei confronti possibili. Si ricorda che il MUD è l'unica fonte che presenta il valore dei dipendenti per ciascuna unità registrata, la presenza di problemi non marginali nella localizzazione delle unità determina una maggiore difficoltà nell'utilizzo corretto e completo di tale variabile..
3. La bassa percentuale di imprese con localizzazione differente è in realtà dovuta al fatto che la maggioranza delle imprese è localizzata in un solo ambito territoriale. Se si prendono in considerazione le sole imprese in cui almeno una fonte, fra la coppia di fonti confrontate, è plurilocalizzata (pluriprovinciale o

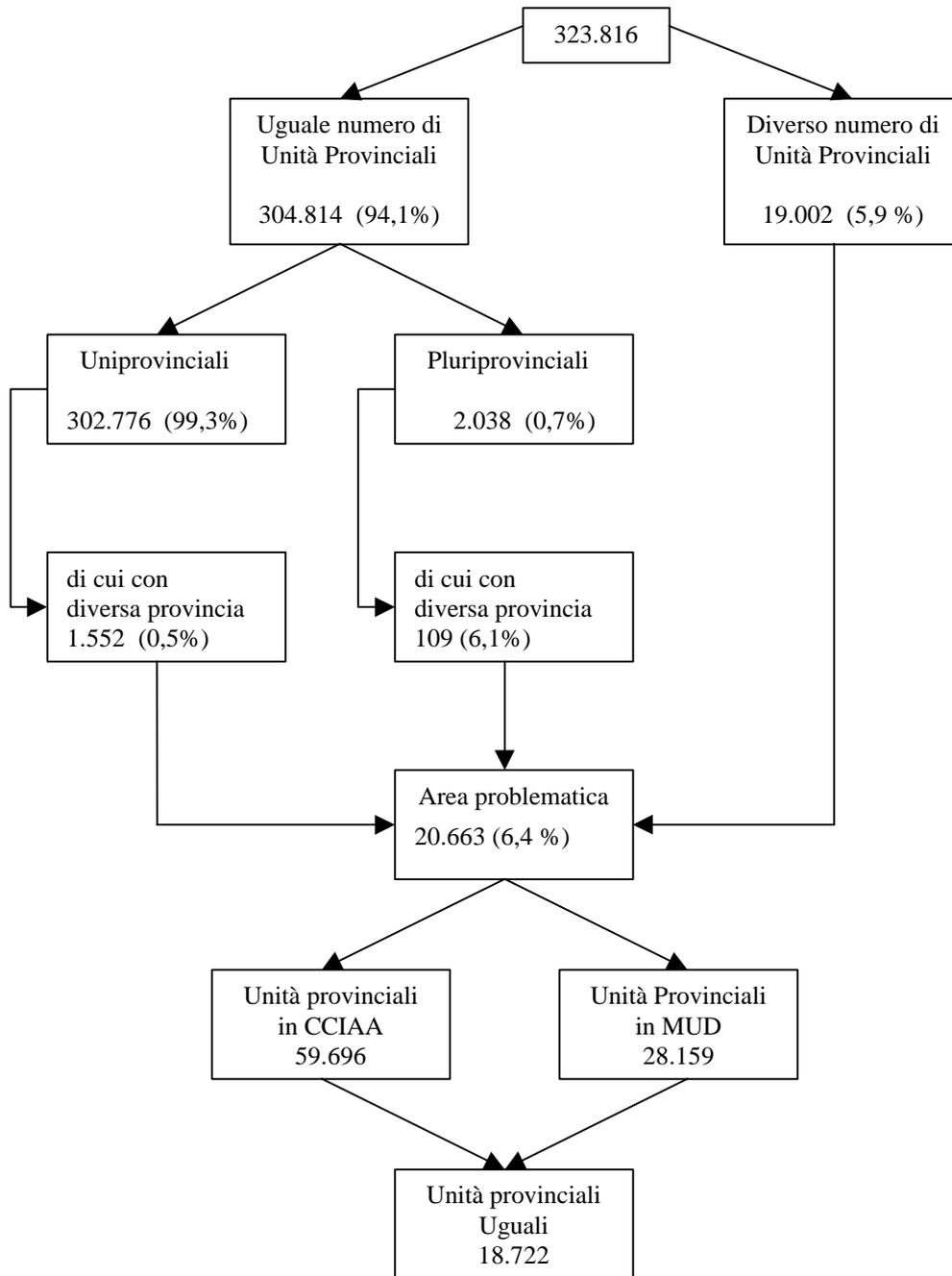
---

<sup>13</sup> Tutte le unità che fanno riferimento, in ciascuna fonte, ad un unico codice fiscale sono state aggregate per ambiti territoriali di provincia di regione.

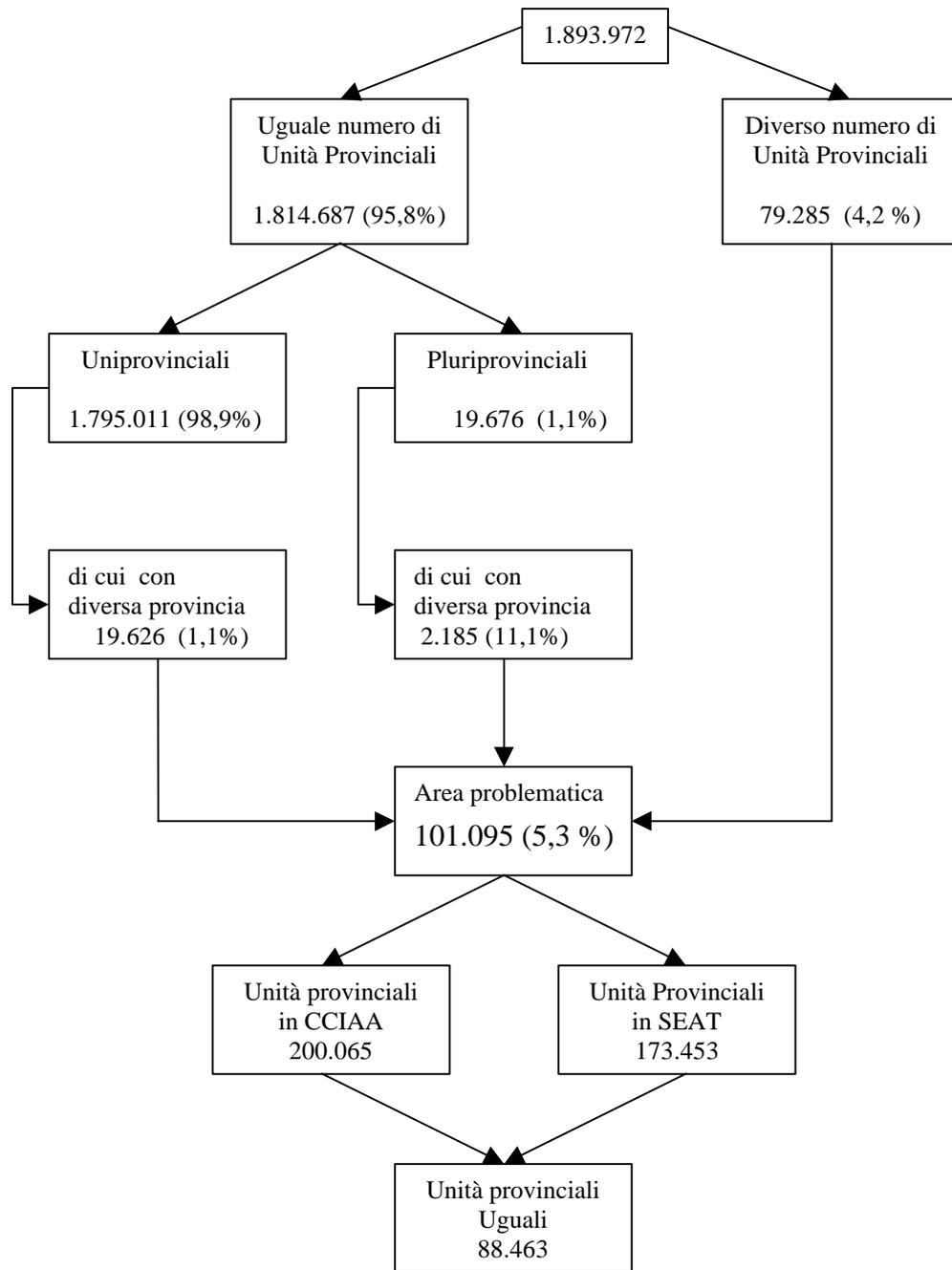
pluriregionale), i risultati sono notevolmente diversi. Infatti praticamente tutte le imprese con localizzazione plurime (Il 91% nel confronto CCIAA/MUD, 82% in quello CCIAA/SEAT e il 95% in quello CCIAA/IRAP) presentano problemi determinati o dal diverso numero di unità registrate o dalla differente localizzazione delle unità nelle fonti.

4. Come si è detto la fonte CCIAA presenta l'universo più ampio (sia in termini di imprese che di unità locali registrate). Nei casi in cui si presenta un diverso numero di unità fra le fonti confrontate, l'aspettativa era di avere un numero maggiore di unità nella CCIAA. In realtà questo è vero solo nel confronto con il MUD (il 4% dei casi presenta un numero di unità MUD maggiore della CCIAA), non lo è per le altre fonti. Infatti nel 31,5% dei casi la SEAT presenta un numero maggiore di unità rispetto alla CCIAA; mentre nel 17,4% l'IRAP presenta un numero maggiore di unità regionali.
5. In termini di unità territoriali il numero di unità che l' "area problematica" individua è pari, nel confronto CCIAA/MUD, a circa 60 mila unità registrate nella CCIAA e di 28 nel MUD. Il sottoinsieme comune nelle due fonti è pari a circa 19 mila unità; il che comporta che per circa 60 mila unità deve essere effettuata una attività di verifica e validazione delle informazioni. Nello stesso modo nel confronto CCIAA/SEAT le unità provinciali da verificare superano le 200.000, mentre nel caso CCIAA/IRAP le unità regionali non concordi assommano a circa 80 mila.

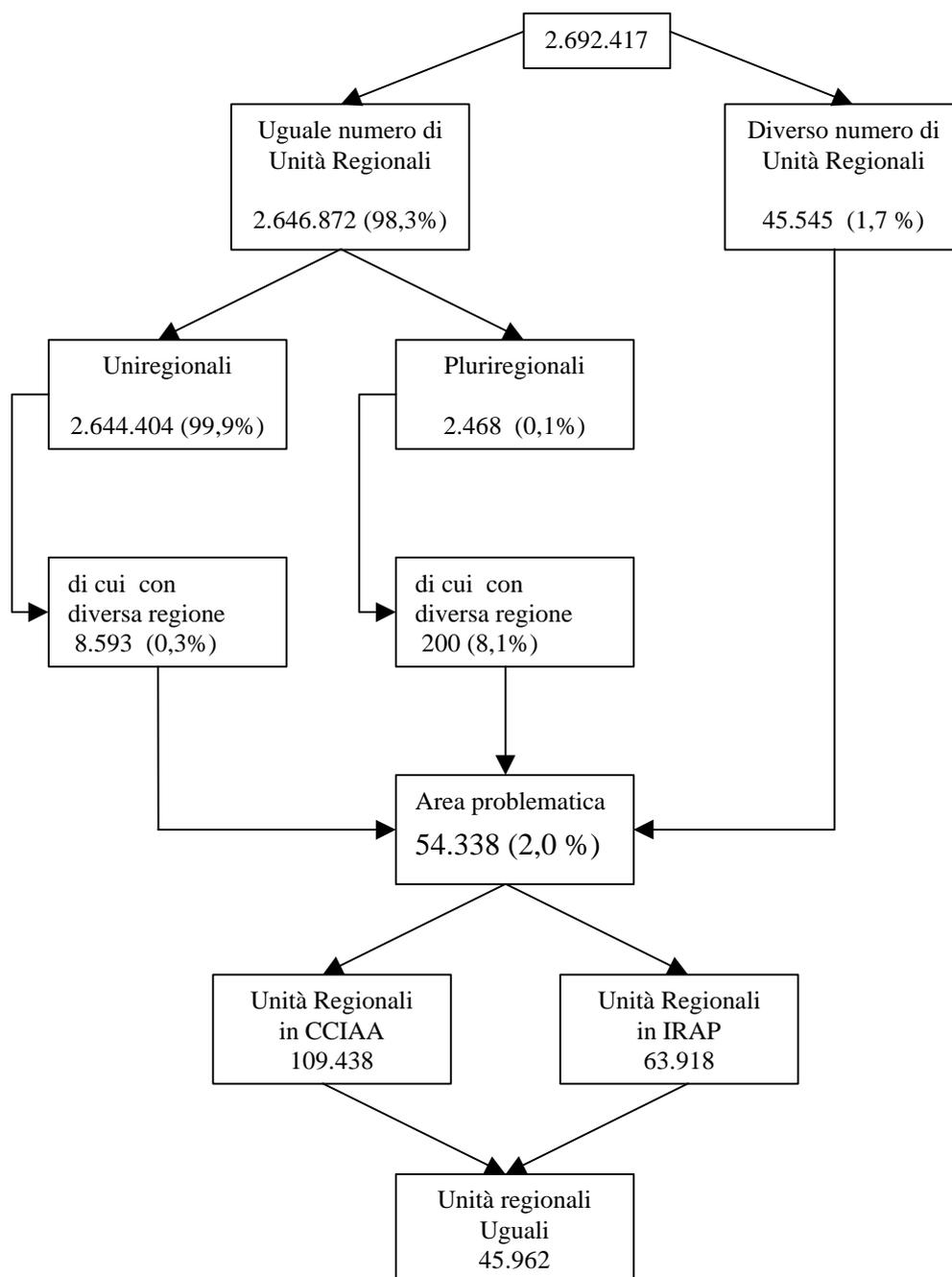
Fig. 1 Confronto CCIAA / MUD – Analisi per unità provinciali



**Confronto CCIAA / SEAT – Analisi per unità provinciali**



Confronto CCIAA / IRAP – Analisi per unità regionali



## **2. Le problematiche di stima di dati territoriali nella rilevazione campionaria sui risultati economici delle piccole e medie imprese <sup>14</sup>**

Le informazioni di carattere strutturale più recenti sui principali aggregati economici delle imprese italiane dell'industria e dei servizi confermano una rilevante presenza di imprese di piccole dimensioni ed una relativa scarsità di grandi imprese. Nel 2000 nelle microimprese (quelle con meno di 10 addetti, pari a poco meno di 3,9 milioni di unità) si concentra il 48,5% degli addetti, il 24,2% dei dipendenti, il 30,3% del fatturato e il 32,3% del valore aggiunto. Le oltre 190.000 imprese con 10 e più addetti risultano anch'esse fortemente caratterizzate da un'elevatissima numerosità della fascia di imprese di piccole dimensioni. Le imprese con 100 e più addetti sono infatti circa solo 9.000, ed assorbono il 24,5% del totale degli addetti (il 37,7% dei dipendenti) e realizzano il 38,6% del valore aggiunto complessivo.

Queste caratteristiche strutturali differenziano nettamente il nostro Paese dalle altre economie europee, caratterizzate da una minore polverizzazione del sistema delle imprese.

Dal punto di vista della statistica ufficiale, emergono almeno due problemi connessi alla misurazione - in un contesto statistico armonizzato a livello comunitario - dei diversi aspetti caratteristici di un sistema produttivo così strutturato: oneri statistici relativamente superiori a quelli di paesi con una maggiore concentrazione di imprese nelle classi dimensionali medie e grandi; la produzione statistica è sempre più sottoposta al rispetto delle normative comunitarie che, sotto forma soprattutto di regolamenti del Consiglio dell'Unione europea, hanno determinato una crescita notevole degli obblighi di produzione e diffusione di statistiche strutturali ad un elevato livello di disaggregazione settoriale e dimensionale.

In questo contesto, l'esigenza di misurare i diversi aspetti del sistema delle imprese ad un sufficiente dettaglio territoriale sub-nazionale risulta vincolata dalla necessità di contenere le dimensioni campionarie delle indagini entro numerosità ragionevoli sotto il profilo dell'impegno finanziario ed organizzativo da applicare alle rilevazioni, e dal rispetto prioritario delle esigenze conoscitive definite a livello europeo.

L'esigenza di migliorare i termini del *trade off* esistente tra le esigenze di dettaglio settoriale e dimensionale degli indicatori strutturali sul sistema delle imprese e quelle relative alla misurazione degli aspetti territoriali della loro struttura e *performance* ha stimolato innovazioni metodologiche, organizzative e di impianto delle rilevazioni strutturali sulle imprese, con particolare riferimento a quelle finalizzate alla misurazione dei risultati economici. In particolare, l'accelerazione dell'utilizzo integrato di dati statistici raccolti attraverso rilevazioni dirette e di dati amministrativi (bilanci civilistici e dati fiscali) consente di aumentare notevolmente la precisione delle stime territoriali senza incrementare la numerosità campionaria delle rilevazioni.

### **2.1 Le rilevazioni sui risultati economici delle imprese**

#### *2.1.1 Il contesto di riferimento*

Le rilevazioni annuali sui risultati economici delle imprese sono finalizzate al soddisfacimento del Regolamento Ue n° 58/97 per le statistiche strutturali (*SBS - Structural Business Statistics*).

---

<sup>14</sup> A cura di C. Casciano, R. Monducci, G. Siesto.

L'impianto delle rilevazioni statistiche condotte dall'Istat per la stima degli aggregati economici è basato su due rilevazioni integrate: la prima, campionaria, si riferisce alle imprese con 1-99 addetti (rilevazione PMI - Piccole e Medie Imprese); la seconda, totale, copre le imprese della fascia dimensionale superiore (rilevazione SCI - Sistema dei Conti delle Imprese); L'universo di riferimento è quello fornito annualmente dall'Archivio Statistico delle Imprese Attive (ASIA), realizzato dall'Istat sulla base di un'ampia base informativa derivante sia da rilevazioni statistiche sia da fonti amministrative.

### 2.1.2 Il territorio come dominio di stima

Il regolamento sulle statistiche strutturali prevede la produzione, entro 18 mesi dalla fine dell'anno di riferimento dei dati, delle stime dei totali e delle medie per alcune variabili di interesse riferite a tre tipi di sottopopolazioni (dette *domini di stima*) descritte nel tabella 1:

**Tabella 1 : Domini di stima e loro denominazione**

Tipo di Dominio	Descrizione	N° di domini
DOM1	Classe di attività economica (NACE rev.1 a 4 cifre)	461
DOM2	Gruppo di attività economica (NACE rev.1 a 3 cifre) × Classi di addetti	1.047
DOM3	Divisione di attività economica (NACE rev.1 a 2 cifre)× Regione (NUTS II)	984

Come si vede, le esigenze poste dal Regolamento SBS riguardano, relativamente alle stime territoriali, la produzione di variabili per l'incrocio tra settore di attività economica, definito al livello delle due cifre della classificazione Nace.rev.1, e regione. Si tratta quindi di stimare i principali aggregati economici per poco meno di mille combinazioni di settori/regioni.

## 2.2 La rilevazione campionaria sulle piccole e medie imprese

### 2.2.1 Principali caratteristiche della rilevazione

Il disegno campionario utilizzato per la rilevazione PMI è ad uno stadio stratificato con selezione delle unità con probabilità uguali; gli strati sono definiti dalla concatenazione delle modalità delle variabili '*classi di attività economica*', '*classi di addetti*' e '*regione*'; la stratificazione adottata costituisce la partizione minima della popolazione che permette di ottenere i domini di stima come aggregazione degli strati elementari; tale definizione degli strati ha importanti vantaggi nella progettazione dell'indagine: in particolare, permette l'allocazione del campione fissando a priori i livelli di precisione attesa delle stime su tutti i domini di interesse, in modo da garantire simultaneamente l'accuratezza delle stime ai diversi livelli di dettaglio richiesti.

Determinata la numerosità del campione nei singoli strati per mezzo di tecniche di allocazione multivariata, in fase di estrazione delle unità campionarie dalla lista di riferimento si adottano metodi di selezione coordinata: nell'ottica della riduzione dell'onere statistico complessivo, tali procedure consentono di minimizzare la presenza,

nell'indagine corrente, di imprese già intervistate nelle rilevazioni PMI precedenti o in altre rilevazioni Istat aventi come oggetto lo stesso universo di indagine.

Il campione della rilevazione PMI riferito all'anno 2000 è costituito da circa 120.000 imprese, pari a circa il 3% del complesso delle imprese italiane dello specifico campo di osservazione.

### 2.2.2 Metodologia di stima

Il riporto all'universo dei dati della rilevazione PMI è realizzato per fascia dimensionale delle imprese, classe di attività economica e regione amministrativa. Per quanto riguarda la stima delle variabili di interesse, 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 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 la 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 finale*, espresso come prodotto del peso base per i fattori correttivi.

Il calcolo dei fattori correttivi della mancata risposta totale avviene stimando un modello di risposta, ossia formalizzando una relazione funzionale tra le probabilità di risposta individuali  $\pi_k$  ed un insieme di variabili esplicative disponibili sia per i rispondenti che per i non rispondenti.

Se correttamente esplicitato, l'introduzione di tale modello permette l'eliminazione della componente distorsiva determinata dalla riduzione dell'informazione dovuta alla mancata risposta totale. Il modello di mancata risposta adottato assume una probabilità di risposta  $\pi_k$  costante per sottoinsiemi di unità del campione, detti *classi di omogeneità*.

Per il calcolo dei pesi finali si adotta 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.

## 2.3 Aspetti territoriali della qualità dei dati

### 2.3.1 I tassi di risposta

Negli ultimi tre anni la rilevazione PMI ha evidenziato un significativo e progressivo recupero del tasso di risposta. Questo è stato ottenuto attraverso due principali strumenti: il primo è il coordinamento dei campioni riferiti a ciascun anno, allo scopo di evitare per quanto possibile un eccesso di pressione statistica sulle singole unità; il secondo è il ricorso a solleciti telefonici effettuati tramite *call center*, dopo una fase di sollecito postale. Nonostante queste iniziative, permane bassa la propensione delle imprese a rispondere ai questionari statistici, soprattutto se complessi e densi di

richieste di misurazioni quantitative come quello della rilevazione PMI: le imprese che hanno inviato il questionario compilato, con dati utili per la rilevazione relativa all'anno 2000, sono state 54.623. Tenendo conto anche delle imprese cessate, inattive, fallite, in liquidazione, fuori campo di osservazione e quelle per le quali non è stato possibile instaurare un contatto (imprese trasferite, sconosciute, non trovate all'indirizzo, ecc.) il tasso di risposta si colloca sul 52,8%. Dal punto di vista territoriale, la tabella 2 evidenzia una maggiore propensione alla risposta delle imprese localizzate nel Nord-est, una propensione che cresce all'aumentare della dimensione dell'impresa, mentre il settore di attività economica non sembra discriminare in misura significativa le imprese.

**Tabella 2 :** *Tassi di risposta per ripartizione geografica, classe di addetti delle imprese e macrosettore di attività economica – Anno 2000*

Rilevazione PMI – Anno 2000	Tassi di risposta (%)
Italia	52,8
Nord-ovest	57,4
Nord-est	61,0
Centro	51,9
Meridione	41,3
Isole	41,2
1 - 9 addetti	51,8
10 - 19 addetti	54,7
20 - 99 addetti	55,7
Industria in senso stretto	51,3
Costruzioni	53,0
Commercio	55,5
Servizi	52,8

L'analisi degli effetti differenziali delle variabili di stratificazione (settore di attività economica, territorio, dimensione della impresa) sulla propensione alla risposta è stata effettuata anche stimando un modello di regressione logistica sulla base dei dati relativi alla rilevazione PMI. Ciò allo scopo di evidenziare effetti specifici associati alla localizzazione delle imprese, al di là delle specificità settoriali e dimensionali.

Le imprese costituenti il campione teorico di circa 120.000 unità a cui è stato inviato il modello sono state classificate secondo una variabile binaria  $Y$  con valori  $\{1, 0\}$  per rappresentare rispettivamente l'evento "impresa rispondente" ed il suo complementare.

Sia  $\underline{V}$  l'insieme delle variabili esplicative considerate (che nel caso di variabili di tipo qualitativo sono dette *fattori*); sia  $\mathbf{p}(\underline{v})$  la probabilità di un "successo", ossia dell'evento  $\{Y = 1\}$ , in corrispondenza di una particolare combinazione  $\underline{v}$  delle modalità o *categorie* dei fattori  $\underline{V}$ ; la relazione di dipendenza di  $\mathbf{p}(\underline{v})$  da  $\underline{V}$  formalizzata dal modello di regressione logistica utilizzato per l'analisi è del tipo seguente:

$$\text{logit}[\mathbf{p}(\underline{v})] = \log\left(\frac{\mathbf{p}(\underline{v})}{1 - \mathbf{p}(\underline{v})}\right) = \mathbf{a} + \sum_i \mathbf{b}_i x_i + \sum_i \mathbf{g}_i w_i + \sum_i \mathbf{d}_i w_i \quad (1)$$

in cui  $\mathbf{a}$  rappresenta un effetto medio generale; i parametri  $\mathbf{b}_i$ ,  $\mathbf{g}_i$  e  $\mathbf{d}_i$  misurano l'effetto netto (di incremento o decremento, rispettivamente se il segno del parametro è positivo o negativo) della categoria  $i$ -esima rispettivamente dei fattori *dimensione dell'impresa*, *sezione di attività economica* e *regione* sul *logit* della probabilità dell'evento "impresa rispondente".

La stima dei parametri del modello ha previsto la ricerca della suddivisione in classi delle modalità delle variabili considerate che portasse a risultati soddisfacenti dal punto di vista della bontà di adattamento del modello, valutata in base al *test* del rapporto della massima verosimiglianza ed al *plot* dei residui di Pearson; in particolare, partendo dal modello corrispondente alla suddivisione in classi più sottile delle modalità di  $\mathbf{v}$ , si è proceduto ad aggregazioni successive fino a definire un insieme di categorie che portasse a stime dei parametri significativamente non nulli. La definizione delle categorie a cui si è pervenuti per le tre variabili è riportata in nota <sup>15</sup>.

Dato un fattore, si può valutare l'importanza relativa delle sue categorie sulla probabilità di successo dell'evento rispetto ad una categoria di riferimento per mezzo dei rapporti tra le *odds*. Le *odds* di un "successo", cioè dell'evento  $\{Y = 1\}$  sono date dal rapporto tra la probabilità di un evento e la probabilità dell'evento complementare.

Si può facilmente verificare che il rapporto tra le *odds* delle categorie  $i$  e  $j$  di un certo fattore  $V$  è pari all'esponenziale della differenza tra i rispettivi parametri; ad esempio, nel caso di confronto tra le regioni  $i$  e  $j$ , vale la relazione:

$$\left[ \frac{\mathbf{p}(z_i = 1)}{1 - \mathbf{p}(z_i = 1)} \right] / \left[ \frac{\mathbf{p}(z_j = 1)}{1 - \mathbf{p}(z_j = 1)} \right] = \exp\{\mathbf{d}_i - \mathbf{d}_j\} \quad (2)$$

Vi è pertanto indipendenza di  $Y$  da  $V$  se  $\mathbf{d}_i = \mathbf{d}_j$  per ogni  $i$ , ossia l'effetto marginale di  $i$  e di  $j$  sul *logit* dell'evento  $\{Y = 1\}$  non cambia se il rapporto tra le rispettive *odds* è pari ad 1. Viceversa, tanto maggiore è la differenza tra  $\mathbf{d}_i$  e  $\mathbf{d}_j$  tanto maggiore è l'effetto moltiplicativo di  $i$  rispetto a  $j$  sulle probabilità di un "successo".

Nella tabella 3 le regioni sono state ordinate in senso crescente secondo il valore del rapporto tra *odds*, calcolato ponendo a denominatore della formula (2) le *odds* della Lombardia, assunta quindi come categoria base. Come si può notare, rispetto alla categoria di riferimento le regioni *Veneto* e *Friuli Venezia Giulia* hanno un rapporto pressochè pari ad 1, manifestando così un effetto marginale della stessa entità di quello della Lombardia. Analogamente, apportano un contributo di pari entità alle *odds* stimate di successo i seguenti gruppi di regioni: *Molise* e *Campania*, *Lazio* e *Valle d'Aosta, Trentino Alto Adige* e *Piemonte*; in tutti i casi si tratta di una riduzione delle *odds* rispetto alla categoria base. Ad esempio, il valore minimo del rapporto (2), osservato per la Calabria e pari a 0.33, indica che le imprese di tale regione hanno una probabilità di rispondere all'indagine nettamente più bassa di quelle della Lombardia, in quanto le *odds* dell'evento "impresa rispondente" risulta, per le prime, 0.33 volte le *odds* delle

<sup>15</sup> Per la variabile *dimensione dell'impresa* le categorie considerate corrispondono alle classi 1, 2-9, 10-19, 20-99 addetti; per la variabile *settore di attività economica* le categorie sono rappresentate dalle sezioni C-F (industria in senso stretto), G (commercio), H (alberghi e ristoranti), I-O (altri servizi); per la variabile *territorio* le categorie sono rappresentate da tutte le regioni considerate separatamente, con la sola eccezione dell'aggregazione di Umbria e Abruzzo.

seconde. All'estremo opposto si colloca l'Emilia Romagna, per la quale il valore superiore a 1 della sua *odds* rispetto a quella base indica una maggiore probabilità che le imprese dell'Emilia Romagna siano rispondenti rispetto a quelle della Lombardia. L'esercizio evidenzia quindi l'esistenza di profonde differenziazioni territoriali nei tassi di risposta, anche una volta tenuto conto delle specificità settoriali e dimensionali che caratterizzano le strutture produttive locali.

**Tabella 3** : *Rapporti tra odds e relativi intervalli di confidenza*

Regione	Rapporti tra <i>odds</i>	Inf	sup
Calabria	0.33	0.30	0.35
Basilicata	0.41	0.37	0.44
Sicilia	0.43	0.40	0.46
Molise	0.46	0.41	0.51
Campania	0.46	0.43	0.49
Sardegna	0.57	0.53	0.61
Valle d'Aosta	0.60	0.54	0.67
Lazio	0.60	0.57	0.63
Puglia	0.62	0.58	0.66
Umbria e Abruzzo	0.64	0.60	0.68
Liguria	0.75	0.70	0.81
Toscana	0.83	0.79	0.88
Marche	0.85	0.80	0.92
Trentino Alto Adige	0.91	0.86	0.97
Piemonte	0.92	0.87	0.97
Friuli Venezia Giulia	0.99	0.92	1.07
Veneto	1.04	0.98	1.09
Emilia Romagna	1.20	1.14	1.26

### 2.3.2 I coefficienti di variazione delle stime

Sulla base delle evidenze precedenti, i tassi di risposta alla rilevazione PMI segnalano un "fattore di rischio" per la qualità dei dati stimati, derivante da un chiaro *pattern* territoriale nell'adempimento agli obblighi di risposta alla rilevazione. Come si è già illustrato, la metodologia di riporto all'universo adottata consente, entro certi limiti, di correggere gli effetti negativi dell'eterogeneità dei tassi di risposta. In effetti, l'analisi dei coefficienti di variazione delle stime riportati nella tabella 4, relativi in questo caso alla sola variabile *valore aggiunto*, evidenziano eterogeneità territoriali solo parzialmente riconducibili a differenze nei tassi di risposta della rilevazione ed alla differente dimensione della popolazione di riferimento. Infatti, se l'errore relativo si colloca mediamente intorno al 4%, esso presenta punte massime per le regioni Valle d'Aosta e Basilicata, valori più elevati della media per Friuli Venezia Giulia, Trento e Bolzano; risulta invece nettamente più contenuto per Lombardia, Emilia Romagna e Toscana.

**Tabella 4:** *Coefficienti di variazione stimati a livello di regione per la variabile valore aggiunto - Dati dell'indagine PMI - anno 2000*

Regione	Coefficienti di variazione
Piemonte	2.25
Valle d'Aosta	8.27
Lombardia	1.68
Trentino Alto Adige	2.81
Veneto	2.32
Friuli Venezia Giulia	4.84
Liguria	3.28
Emilia Romagna	1.94
Toscana	1.95
Umbria	2.94
Marche	3.37
Lazio	4.35
Abruzzo	4.09
Molise	5.36
Campania	3.07
Puglia	2.75
Basilicata	7.19
Calabria	4.26
Sicilia	3.03
Sardegna	4.46

## **2.4 Iniziative per il miglioramento della qualità dei dati, con particolare riferimento alle stime territoriali**

### *2.4.1 Il costo del miglioramento della qualità dei dati territoriali nell'attuale contesto di indagine*

Un primo aspetto da considerare è quello relativo al miglioramento della qualità delle stime territoriali nell'attuale contesto di indagine.

Nella tabella 5 è riportata in grassetto la numerosità del campione relativo alla rilevazione PMI - 2001, ottenuta considerando censite (e rilevate attraverso l'indagine SCI e l'utilizzo dei dati amministrativi relativi ai bilanci civilistici) le imprese della fascia "99 addetti ed oltre"; tale numerosità rappresenta la dimensione minima del campione necessaria per stimare i totali di *fatturato*, *valore aggiunto* ed *addetti* con un livello massimo dell'errore relativo atteso (espresso in termini di coefficienti di variazione) su ciascun tipo di dominio pari ai valori elencati in colonna. Questi valori rappresentano il livello teorico di errore che le stime produrrebbero sotto due condizioni: coincidenza perfetta tra lista di selezione e popolazione di indagine; assenza del fenomeno della mancata risposta totale. I dati riportati di seguito forniscono una misura della "sensibilità" del campione, ossia della variazione della sua numerosità dovuta ad una diminuzione del 10% degli errori attesi sui domini; ciò permette di valutare il costo, in termini di dimensione campionaria, del miglioramento della precisione attesa delle stime delle variabili di interesse.

Come si può osservare sulla base dei valori riportati nell'ultima colonna, portare l'errore massimo atteso tra il 5 ed il 6% sui tre domini comporterebbe un aumento della numerosità campionaria minima pari al 67%, corrispondente a circa 80.000 unità in più da contattare rispetto a quelle coinvolte nella rilevazione PMI riferita al 2001. In particolare, per quanto riguarda il dominio DOM3, relativo all'incrocio settore di attività/regione, l'errore massimo passerebbe dal 10% al 6,6%.

**Tabella 5 :** *Variazione della dimensione del campione per una diminuzione del 10% dell'errore relativo massimo atteso sui domini di stima  
Indagine PMI 2001- fascia campionata: 1-99 addetti*

Domini di stima	Errore relativo massimo atteso				
DOM1	0.08	0.072	0.065	0.058	0.052
DOM2	0.07	0.063	0.057	0.051	0.046
DOM3	0.10	0.09	0.081	0.073	0.066
Dimensione complessiva	<b>124.965</b>	144.332	163.699	183.066	202.433
<i>Unità campionate</i>	<b>115.871</b>	135.238	154.605	173.972	193.339
<i>Unità censite</i>	<b>9.094</b>	9.094	9.094	9.094	9.094

#### 2.4.2 L'utilizzo dei dati amministrativi

L'Istat, con la ristrutturazione delle rilevazioni PMI e SCI - la rilevazione campionaria PMI ha allargato il campo di osservazione dalla fascia dimensionale 1-19 addetti a 1-99 addetti e la rilevazione totale SCI ha contestualmente ridotto l'universo di riferimento dalle imprese con 20 e più addetti a quelle con 100 e più addetti - ha eliminato il fastidio statistico ad oltre 70.000 imprese ed ha avviato un progressivo sfruttamento del complesso delle informazioni di natura amministrativa esistenti sul sistema delle imprese, allo scopo di integrare e/o sostituire le rilevazioni dirette per quel che riguarda le informazioni economiche già raccolte dalla Pubblica amministrazione (variabili di bilancio di fonte Camere di Commercio; dati relativi all'occupazione ed alle retribuzioni di fonte Inps, dati fiscali di fonte Ministero dell'Economia).

A partire dal 2000 come anno di riferimento dei dati i risultati economici delle imprese con almeno 100 addetti, campo di osservazione della rilevazione SCI, sono stati ottenuti integrando le mancate risposte totali alla rilevazione e le unità soggette a trasformazioni societarie non presenti nella lista di avvio dell'indagine con i bilanci civilistici<sup>16</sup> acquisiti dalle Camere di commercio (fonte BIL)<sup>17</sup>.

<sup>16</sup> L'adeguatezza dei bilanci a rappresentare molte delle variabili contenute nel questionari Istat (circa il 65% delle variabili contenute nel questionario SCI) è stata testimoniata dai risultati del progetto relativo all' "Analisi di fattibilità sull'utilizzo dei bilanci a fini statistici" (la relazione conclusiva è stata presentata nel mese di giugno 1998)

<sup>17</sup> Le statistiche sulle imprese si riferiscono a tutte le imprese attive nei settori dell'industria e dei servizi; i bilanci civilistici si riferiscono ai soggetti obbligati per legge al deposito del bilancio stesso presso le CCIAA. Le società di capitali nel campo di osservazione delle rilevazioni strutturali sui risultati economici delle imprese sono poco meno di 500.000 ed assorbono un'occupazione di poco inferiore al 50% di quella totale, con quote occupazionali fortemente crescenti al crescere della dimensione media delle imprese

L'avanzamento del progetto di uso di dati amministrativi per la stima dei risultati economici delle imprese<sup>18</sup> prevede anche la ristrutturazione della rilevazione PMI, con riferimento a due principali segmenti dimensionali delle imprese: imprese con 20-99 addetti e imprese con 1-19 addetti.

Per quanto riguarda le imprese con 20-99 addetti, a partire dal 2001 come anno di riferimento dei dati (la diffusione dei risultati è prevista per luglio 2003) l'utilizzo dei bilanci civilistici verrà esteso anche alla stima dei risultati economici delle imprese con 20-99 addetti, che rappresentano un importante segmento della rilevazione PMI. Ai bilanci civilistici si aggiungono, per le imprese con forma giuridica diversa da quella di società di capitale, i dati fiscali delle imposte dirette (fonte UNICO)<sup>19</sup>. L'universo delle imprese con 20-99 addetti è composto da circa 64.000 imprese, che occupano poco meno di 2,4 milioni di addetti, pari al 16% del totale. Analogamente a quanto effettuato per la rilevazione SCI, relativa alle imprese con 100 e più addetti, i dati riferiti al 2001 verranno prodotti sulla base di una ricostruzione dei principali risultati economici di tutte le imprese dell'universo con 20-99 addetti derivanti dalle tre fonti disponibili (campione PMI, fonte BIL, fonte UNICO), integrate con il registro ASIA. Per le variabili d'impresa non presenti nelle fonti BIL e UNICO, ma rilevate attraverso il questionario della rilevazione PMI, verrà utilizzata un'imputazione da donatore, basata sui rispondenti al campione PMI relative allo specifico campo di osservazione. Dal punto di vista della qualità dei dati territoriali il miglioramento atteso è significativo, essendo connesso alla modificazione della natura della fonte da campionaria a censuaria con tasso di copertura sostanzialmente totale, almeno per le principali variabili economiche.

Per quanto riguarda le imprese con 1-19 addetti, la strategia è quella di selezionare un campione notevolmente allargato (campione C.1, con oltre 300.000 imprese), rispetto a quello della rilevazione PMI (campione C.2, con circa 100.000 imprese), che ne rappresenta un sottoinsieme. Per le imprese del campione C.1 si estraggono le principali variabili economiche dalla fonte fiscale UNICO, che consentono – opportunamente trattate sulla base del *linkage*, a livello di dato individuale, con i rispondenti al campione PMI e con il registro ASIA – di disporre di una base informativa notevole sotto il profilo della qualità delle stime. I livelli stimati sulla base del campione C.1 delle principali variabili economiche per settore, dimensione e regione rappresentano i *totali noti* per le successive stime, basate sulle sole imprese rispondenti al campione C.2, finalizzate a stimare un ulteriore, ampio, *set* di variabili quantitative e qualitative.

---

<sup>18</sup> Il progetto è stato presentato come *best practice* nell'ambito della riunione dei direttori delle statistiche economiche dei paesi Ue (DeGregorio e Monducci, 2002).

<sup>19</sup> Per quanto riguarda i dati fiscali, due sono le novità che di recente hanno ampliato le possibilità di utilizzo dei dati delle imposte dirette e indirette a fini statistici. Sul piano normativo, anche se, fino dal 1992, una convenzione stipulata tra l'Istat ed il Ministero delle finanze riconosceva un comune interesse su questo argomento e definiva le linee di un piano di lavoro solo parzialmente realizzato, è con la l. 681/97 che all'Istat viene riconosciuto il diritto ad un pieno accesso a tutte le fonti amministrative. In questo quadro, la convenzione tra i due soggetti, stipulata il 2 agosto 2000 e avente come oggetto la realizzazione di un sistema di interscambio di informazioni statistiche, assume una rilevante importanza. Sul piano tecnologico e della disponibilità reale delle informazioni, lo straordinario sviluppo, intervenuto negli ultimi anni, della trasmissione telematica delle dichiarazioni fiscali, che pone ormai il nostro paese in una posizione avanzata a livello mondiale, rende possibile un drastico abbattimento dei tempi di disponibilità delle informazioni fornite dai contribuenti.

Come si può notare dall'esame della tabella 6, campionare nella sola fascia dimensionale 1-19 addetti circa 300.000 unità (campione C.1) per le quali si dispone, come già descritto, dei principali aggregati economici da fonte amministrativa, porterebbe al 3 % l'errore massimo relativo atteso delle stime d'interesse a livello territoriale.

**Tabella 6 :** *Variazione della dimensione del campione C.1 per una diminuzione del 10% dell'errore relativo massimo atteso sui domini di stima*

Domini di stima	Errore relativo massimo atteso				
	0.05	0.045	0.041	0.036	0.033
DOM1	0.05	0.045	0.041	0.036	0.033
DOM2	0.04	0.036	0.032	0.029	0.026
DOM3	0.03	0.027	0.024	0.022	0.020
Dimensione complessiva	367.867	424.873	481.879	538.885	595.891
<i>Unità campionate</i>	295.423	352.429	409.435	466.441	523.447
<i>Unità censite</i>	72.444	72.444	72.444	72.444	72.444

La strategia di utilizzo integrato di dati statistici raccolti attraverso rilevazione campionaria e dati amministrativi consente quindi di aumentare notevolmente la precisione delle stime territoriali senza incrementare la numerosità campionaria delle rilevazioni.

### **3. La dimensione territoriale nelle principali indagini congiunturali e strutturali del settore agrario<sup>20</sup>**

Gli elementi che concorrono a determinare la qualità dell'informazione statistica sono la *rilevanza, l'accuratezza, la tempestività* nella diffusione, *la chiarezza e l'accessibilità, la comparabilità* nel tempo e nello spazio, *la congruenza* (Collins & Sykes, 1999). La garanzia di questi requisiti presuppone la progettazione di rilevazioni statistiche integrate che concorrano in tempi ragionevolmente brevi a fornire un quadro completo del fenomeno indagato. Il territorio rappresenta un fondamentale fattore di integrazione tra fonti diverse e, più in generale, un elemento di notevole importanza ai fini statistici poiché, da un lato, determinando una distribuzione spaziale dei fenomeni, consente analisi più dettagliate degli stessi, dall'altro, ad esso sono legate direttamente alcune branche dell'attività economica e quindi la rilevanza che ciascuna indagine può rivestire per gli utenti provenienti da realtà territoriali differenti.

Nell'attività agricola, in particolare, questo legame con il fattore territorio è particolarmente eclatante sotto entrambi i punti di vista. Non a caso negli ultimi anni la progettazione delle indagini Istat in questo settore ruota sempre più intorno al fattore territorio. Due esempi di questo "approccio territoriale" sono rappresentati dalla rilevazione areale delle superfici agricole e forestali e dall'indagine sulla struttura e

<sup>20</sup> A cura di M. Ballin e L. Martino

produzione delle principali produzioni legnose agrarie che verranno illustrate brevemente nel seguito.

### **3.1 L'indagine areale sulle superfici e produzioni agricole e forestali**

L'indagine areale sulle superfici agricole e forestali adotta il territorio come unità statistica di rilevazione. La popolazione oggetto di osservazione è costituita dall'intero territorio nazionale che viene idealmente ripartito in aree o punti di dimensione e forma prefissata. I vantaggi nell'adozione di questo approccio sono molteplici. In primo luogo, il riferimento territoriale consente di ottenere dati integrati sulle superfici occupate dai diversi usi del suolo (agricoli e non) congruenti con la superficie nazionale totale, fornendo così informazioni fondamentali per la gestione del territorio (estensione delle aree agricole, aree boschive, aree urbane etc.). Sotto il profilo metodologico vanno citati invece i vantaggi relativi alla possibilità di estrarre campioni significativi senza dover disporre di liste esaustive delle unità della popolazione, nonché della drastica riduzione dei dati mancanti che, in questo tipo di rilevazione, si limitano ai punti per i quali è stato impossibile raggiungere fisicamente il punto di osservazione o riconoscere l'uso del suolo in atto. Infine, la possibilità di ricorrere a cornici di campionamento miste che prevedano l'utilizzo congiunto di unità fisiche georeferenziate e unità amministrative (aziende agricole) localizzate geograficamente apre nuove prospettive sulla possibilità di ampliare quantitativamente e qualitativamente il panorama informativo disponibile.

Attualmente la rilevazione italiana sulle superfici e produzioni su base territoriale è oggetto di una sperimentazione congiunta da parte del Ministero delle Politiche Agricole e Forestali e dell'Istat. Elementi salienti della sperimentazione, che ha avuto inizio con l'anno 2002 e dovrà concludersi alla fine del 2003, sono rappresentati da:

- ◆ adozione di unità campionarie rappresentate da porzioni di territorio puntiformi su tutto il territorio nazionale;
- ◆ adozione di uno schema di campionamento a due fasi stratificato;
- ◆ adozione di metodi di campionamento ottimo secondo criteri multivariati;
- ◆ adozione di stimatori di calibrazione che utilizzano quali variabili ausiliarie le immagini satellitari;
- ◆ integrazione delle stime ottenute su campione areale con le stime ottenute per via estimativa;
- ◆ coinvolgimento delle regioni nell'attività di pianificazione del campione ottimo e dei controlli di qualità delle rilevazioni in campo.

Un ulteriore aspetto di interesse è rappresentato dalla possibilità di integrare le metodologie e i risultati della rilevazione italiana con quelli di indagini analoghe a livello comunitario (vedi LUCAS).

Il disegno di campionamento adottato è stratificato a due fasi. Il campione di prima fase, di tipo sistematico su base spaziale, è stato ottenuto ricoprendo l'intero territorio nazionale con un griglia quadrata di punti distanti 500 metri l'uno dall'altro. Tutti i punti sono stati fotointerpretati e classificati secondo una nomenclatura gerarchica a due livelli. Il campione di seconda fase è stato ottenuto ripartendo il campione di prima fase in domini (regionali o provinciali a seconda dei finanziamenti disponibili) e stratificandolo in base a categorie colturali molto aggregate (seminativi, foraggiere permanenti, coltivazioni permanenti). Nelle regioni per le quali si disponeva di una stima delle variabilità per provincia e categorie colturali (regioni rilevate per punti nel

2001), si è fatto ricorso ad un disegno campionario di seconda fase basato sull'*allocazione ottima multivariata* (Bethel 1989) fissando in modo opportuno e diversificato da regione a regione le variabili delle quali fissare l'errore atteso e l'entità dell'errore stesso. In ogni combinazione strato (tipologie colturali)/dominio (regioni o province) è stato comunque garantito un tasso di campionamento minimo del 5% per evitare di ottenere strati troppo poco rappresentati o addirittura vuoti (eventualmente a seguito di mancate risposte).

La raccolta dei dati avviene con osservazione diretta sul campo da parte di rilevatori specializzati. Anche se sul piano teorico i punti sono privi di dimensione, per effettuare le rilevazioni, ad essi viene assegnata una dimensione operativa areale costituita da un cerchio di raggio 3m centrato sul punto di riferimento (~30m<sup>2</sup>). In presenza di consociazioni, tale superficie di osservazione viene estesa a circa 700m<sup>2</sup> (cerchio di raggio 15m). Se sono presenti differenti utilizzi del suolo, ciascuno viene attribuito al punto con la rispettiva proporzione (criterio *pro-rata*).

Oltre a registrare la tipologia colturale su un'apposita scheda in base ad una determinata codifica, per alcune colture specifiche, il rilevatore ha anche il compito di fornire una propria stima di resa sulla base di elementi visibili (stato fenologico, ecc.). Per semplificare la pianificazione degli spostamenti ed il raggiungimento dei punti, i rilevatori vengono dotati di carte tecniche regionali e di GPS palmari.

Per garantire il contenimento dell'errore di misura, oltre che di quello campionario, i dati vengono sottoposti a controlli di qualità a due livelli:

- ◆ sul campo, allo scopo di verificare eventuali errori di posizionamento (errata individuazione del punto da parte dei rilevatori) o di individuazione delle tipologie colturali in atto;
- ◆ in fase di elaborazione: i dati presenti nelle schede di rilevazione sono inseriti in una apposita banca dati e sottoposti ad una serie di controlli volti a verificarne la congruenza e ad evidenziare possibili anomalie nella rilevazione).

I controlli sul campo riguardano circa il 3% dei punti osservati da ciascun rilevatore.

La selezione dei punti da controllare non è casuale. Vengono scelti preferenzialmente i punti di osservazione non acquisiti mediante GPS o per i quali vi è una distanza elevata fra punto di campionamento e punto di osservazione registrato mediante GPS.

I controlli effettuati in fase di elaborazione sono implementati in un piano di compatibilità che rileva alcune incongruenze quali codice di uso del suolo non esistente, somma delle percentuali di copertura dei singoli usi del suolo diversa dal 100%, assenza di note descrittive quando esplicitamente richiesto dalle specifiche di rilevazione, coerenza fra indicazioni presenti nelle note descrittive e le altre informazioni registrate nel punto.

### **3.2 Il ruolo delle amministrazioni regionali nella progettazione dell'indagine sulle principali produzioni legnose agrarie**

Con l'indagine sulle principali produzioni legnose agrarie è stata sperimentata una strategia progettuale che mira a sostituire la ricerca di un disegno d'indagine "ottimo per l'Istat" con un insieme di "disegni localmente ottimi" (ovvero ottimi a livello regionale), che nel loro complesso permettono comunque all'Istituto il rispetto dei vincoli posti dalla legislazione comunitaria e nazionale.

La procedura adottata per collezionare l'insieme dei “disegni localmente ottimi” si articola sostanzialmente in tre passi:

1. il Servizio Agricoltura formula un insieme di proposte, coerenti con le esigenze nazionali, su una particolare fase o aspetto dell'indagine;
2. tali proposte vengono inviate alle amministrazioni regionali che indicano quali tra esse siano più adatte alla propria realtà (ovviamente ciascuna amministrazione può, limitatamente al territorio di propria competenza, modificare le soluzioni proposte dal servizio agricoltura o formularne di alternative);
3. il Servizio Agricoltura definisce una strategia combinando le indicazioni provenienti dalle diverse amministrazioni, riservandosi di apportare quelle modifiche necessarie per cogliere gli obiettivi indicati dalla specifica normativa dell'indagine.

Questa strategia è stata applicata nella scelta della lista di selezione delle unità, nella scelta di alcuni quesiti del questionario e nella definizione del disegno di campionamento. Nel seguito vengono brevemente descritte le procedure per la definizione della lista di selezione e del disegno di campionamento.

### 3.2.1 Lista di selezione delle unità

La direttiva 2001/109/CE prevede che, nel caso l'indagine sia condotta su un campione casuale di aziende agricole, la selezione delle unità debba avvenire su una lista contenente aziende che nel loro complesso rappresentano almeno il 95% della superficie piantata ad alberi da frutto a livello nazionale. Tale norma permette implicitamente l'adozione di regole di *cut off*, ovvero, di regole per l'esclusione dalla lista di campionamento delle unità di piccole dimensioni.

Poiché la distribuzione delle aziende secondo la loro dimensione è molto differenziata tra le regioni, la scelta della regola di *cut off*, anche se sostanzialmente ininfluenza a livello nazionale, può modificare in modo significativo i risultati dell'indagine a livello locale. Per questo ad ogni regione sono state suggerite 8 possibili regole di *cut off* descrivendo per ciascuna di esse le caratteristiche provinciali della lista in termini di copertura sia rispetto al numero di aziende e sia rispetto alla superficie dedicata alle produzioni legnose agrarie. Le indicazioni provenienti dalle diverse regioni si sono concentrate su due di queste regole.

La prima di queste prevedeva che nella lista di selezione fossero incluse tutte le aziende che complessivamente hanno dichiarato al censimento almeno 20 are dedicate alle produzioni legnose agrarie.

La seconda regola prevedeva tre passi: il primo passo definiva le produzioni “caratteristiche regionali”; il secondo passo le soglie dimensionali (in termini di superficie per ciascuna produzione caratteristica) tali che l'insieme delle aziende con una superficie maggiore od uguale a tale soglia rappresentassero almeno il 95% della superficie dedicata a ciascuna produzione caratteristica per regione; infine, il terzo passo definiva la lista di selezione come l'insieme di tutte le aziende la cui dimensione superava per almeno una coltivazione legnosa la corrispondente soglia.

Grazie al contributo delle regioni dalle oltre 350.000 aziende con produzioni legnose agrarie (frutta fresca) si è scesi ad una lista di circa 225.000 aziende che, nel complesso, coprono oltre il 97% della superficie dedicata alle specie indicate dalla normativa comunitaria.

### 3.2.2. Il disegno campionario

Le alternative sottoposte alle regioni per la determinazione del disegno di campionamento hanno riguardato i seguenti tre aspetti: l'individuazione delle aziende rilevanti da includere con certezza nel campione, la stratificazione, gli errori di campionamento.

L'individuazione delle imprese rilevanti avviene attraverso la determinazione delle cosiddette soglie di censimento. Queste sono dei valori di una o più variabili d'interesse (nel nostro caso superfici dedicate alle legnose agrarie).

Utilizzando queste soglie si definisce la popolazione censita come l'insieme delle aziende che hanno almeno una superficie maggiore a tale soglia.

La determinazione numerica di tali soglie avviene utilizzando un algoritmo presente in letteratura. Il principio su cui esso lavora è squisitamente algebrico ma richiede alcuni input fondamentali da parte di chi pianifica l'indagine. In particolare, ciascuna regione ha avuto modo di indicare l'insieme delle produzioni legnose agrarie rispetto alle quali era opportuno individuare le aziende di grande dimensione nella propria realtà. Grazie alle loro indicazioni è stato possibile individuare un insieme formato da poco più di 5000 aziende (2.2% della lista di selezione) che nel loro complesso rappresentano mediamente il 20% della superficie dedicata a ciascuna produzione legnosa agraria.

Per quanto riguarda la stratificazione alle diverse regioni sono state sottoposte due possibili alternative. La prima stratificazione proposta è stata definita concatenando:

1. provincia;
2. classe di superficie dedicata alle principali coltivazioni regionali;
3. classe di superficie dedicata alle rimanenti coltivazioni.

La seconda è stata ottenuta sostituendo la variabile provincia con la variabile regione.

Per definire le classi di superficie le regioni sono state inizialmente suddivise in due gruppi. Nel primo gruppo sono state incluse le regioni in cui vi sono prevalentemente aziende di piccola dimensione (Valle d'Aosta, Liguria, Toscana, Sardegna). Nel secondo gruppo sono state incluse le rimanenti. Per ciascuna azienda è stata definita:

- la superficie principale come somma delle superfici dedicate alle principali coltivazioni legnose della propria regione;
- la superficie non principale (superficie dedicata alle rimanenti produzioni legnose agrarie).

Le superfici che hanno concorso alla definizione della superficie principale non necessariamente coincidevano con le superfici caratteristiche impiegate nella fase di identificazione delle aziende rilevanti.

Nelle regioni del primo gruppo (Valle d'Aosta, Liguria, Toscana, Sardegna) sono state definite tre classi sia per la superficie principale sia per la superficie non principale. Nelle regioni del secondo gruppo sono state definite 4 classi per ciascun tipo di superficie.

Ciascuna regione ha quindi potuto indicare delle preferenze tra: l'adozione di una stratificazione provinciale o regionale, il numero delle classi di superficie (sia principale, sia non principale), le coltivazioni da considerare come caratteristiche.

Infine, ciascuna regione ha avuto modo di indicare la precisione con cui desiderava stimare le superfici dedicate a ciascuna produzione legnosa agraria (a livello regionale per le regioni che hanno adottato la stratificazione regionale e a livello provinciale per le altre).

È il caso di segnalare che, data la differente importanza che rivestono le produzioni legnose agrarie nelle diverse provincie, alcune regioni hanno suggerito di adottare una stratificazione ibrida mantenendo tra loro distinte le provincie in cui la produzione di frutta è di grande rilevanza e aggregando le altre provincie.

La numerosità campionaria determinata da questo disegno di campionamento è circa 25.000 aziende ovvero 5.000 aziende meno di quanto era stato preventivato dal protocollo d'intesa siglato tra l'Istat, il Mipaf e le Regioni.

#### **4. Alcune considerazioni conclusive**

Da quanto esposto è evidente che l'utilizzo sempre maggiore delle fonti amministrative e la ricerca di soluzioni metodologiche ed organizzative adeguate ad incrementare la qualità delle stime territoriali, contenendo i costi di rilevazione/elaborazione, risultano senz'altro percorribili. I risultati già ottenuti, e quelli previsti a breve termine, confermano la possibilità di aumentare il dettaglio territoriale, e la relativa qualità, delle informazioni fornite dalla statistica ufficiale.

Dal punto di vista della domanda di informazione, ciò presuppone un collegamento sempre più stretto tra l'Istituto nazionale di Statistica e le amministrazioni, centrali o locali, interessate a disporre di dati statistici secondo le esigenze specifiche di ogni amministrazione.

Dal lato dell'offerta, le innovazioni di processo e di prodotto necessarie al miglioramento della qualità dei dati territoriali presuppongono in primo luogo investimenti adeguati nel campo della ricerca metodologica integrata con gli aspetti di produzione del dato. Inoltre, per quanto riguarda l'uso dei dati amministrativi, si tratta di superare i limiti derivanti dalla quasi completa assenza del punto di vista statistico soprattutto nella fase di definizione dell'informazione raccolta dagli enti; ciò determina una distanza a volte insormontabile tra le regole amministrative e quelle statistiche, tale da renderle difficilmente utilizzabili. Inoltre, nelle variazioni (se non cancellazioni) di queste da parte dell'amministratore, sia esso centrale o locale, non si tiene conto della perdita di informazioni, a volte richieste ed utilizzate dallo stesso amministratore.

Relativamente al collegamento con le amministrazioni locali, si rende necessario: i) definire dei canali (delle strutture) che consentano di gestire in modo organico le richieste provenienti da queste amministrazioni; ii) sviluppare una cultura statistica all'interno delle stesse in modo che, sia nella fase di acquisizione sia in quella di richiesta ed analisi dei dati, ci sia la possibilità di un intervento delle strutture statistiche presenti in tali istituzioni. Queste condizioni sono funzionali al miglioramento della qualità delle informazioni sia nella fase di definizione delle regole con cui queste sono raccolte sia nella fase di analisi ed integrazione con altre informazioni parte integrante del sistema statistico nazionale ed internazionale.

#### **Riferimenti bibliografici**

Agresti A. (1990), *Categorical data analysis*, Wiley, New York.

Bacchini F., Iannaccone R., (2002), *Rappresentatività dei dati tratti dall'archivio Casse edili*, documento di lavoro Istat.

- Bethel J. (1989), *Sample allocation in multivariate surveys*, Survey methodology, 15, pp.47-57.
- M. Calzaroni e G. Garofalo (2002), "Enterprise, employment, labour input: definitions and measurement problems, ICIS 2002, Copenhagen, August 2002;
- Cicchitelli G., Herzel A., Montanari G.E. (1992), *Il campionamento statistico*, Il Mulino.
- Cochran W.G. (1977), *Sampling Techniques*, J.Wiley, New York.
- Collins M. and Sykes W. (1999): Extending the definition of survey quality. *Journal of Official Statistics*, 15, 1, 57-66.
- DeGregorio C., Monducci R. (2002), *SBS data for Italy: integrated use of administrative sources and survey data. New experiences and medium and long term strategies*, lavoro presentato alla riunione del BSDG (Business Statistics Directors Group), Lussemburgo, giugno.
- Deville J. C., Särndal C. E., (1992), Calibration Estimators in Survey Sampling, *Journal of the American Statistical Association*, 87, 376-382.
- Eltinge J. L., Yansaneh I. (1997), *Diagnostics for formation of nonresponse adjustment cells, with an application to income nonresponse in the U.S. Consumer Expenditure Survey*, Survey Methodology, vol.23 n°1, pp.33-40, Statistics Canada
- Falorsi P.D., Ballin M., De Vitiis C., Scepi G. (1998), *Principi e metodi del software generalizzato per la definizione del disegno di campionamento nelle indagini sulle imprese condotte dall'Istat*, *Statistica Applicata*, vol.20 n° 2.
- Falorsi P.D., Falorsi S. (1995), *Un metodo di stima generalizzato per le indagini sulle famiglie e sulle imprese*, Rapporto di ricerca n.13. Quaderni CONPRI dell'Università di Bologna, gennaio.
- G. Garofalo e C. Viviano (2000), "The problem of links between legal units: statistical techniques for enterprise identifications and the analysis of continuity", *Quaderni di ricerca, Istat*, 2000/1 (paper presented at the 12<sup>th</sup> Roundtable on Business Survey Frames, Helsinki, September 1998).
- G.Garofalo – A. Paggiaro– N. Torelli- C. Viviano (2000), "A record linkage procedure for the management and the analysis of the Italian Business Register", *Conference proceedings ( su cd-rom) di ICES II, Buffalo, Giugno 2000*.
- G.Garofalo (2001), "To exploit administrative sources: a Framework of concepts", 15<sup>th</sup> International Roundtable on Business Survey Frames, Washington D.C, 2001
- Istat (1991), *Classificazione delle attività economiche*, Metodi e norme, serie C-n.11.
- Johston J. (1984), *Econometric methods*, McGraw, Auckland.
- Ohlsson E. (1995), *Coordination of samples using permanent random numbers*. In Cox B.G., Binder D.A., Chinnappa B.N., Christianson A., Colledge M.J., Kott P.S. (eds.) *Business Survey Methods*. Wiley, New York.
- Eurostat (1997), Regolamento (CE) n°58/97 del Consiglio del 20.12.1996 sulle statistiche strutturali.
- Eurostat (1999), Regolamento (CE) n°1618/99 della Commissione del 23.7.1999 concernente i criteri di valutazione della qualità delle statistiche strutturali.
- Särndal, Swensson, Wretman (1992), *Model assisted survey sampling*, Springer Verlag, USA.
- Wolter K.M. (1985), *Introduction to Variance Estimation*, Springer Series in Statistics, Springer-Verlag, USA.