\\ 394 \\

Struttura e cambiamento nelle relazioni tra le imprese metalmeccaniche nella provincia di Modena

II. Distribuzioni degli addetti e pesi per le stime dei parametri

di
Michele Lalla

Dicembre 2001

Università degli Studi di Modena e Reggio Emilia
Dipartimento di Economia Politica
Via Jacopo Berengario 51
41100 Modena (Italia)
e-mail: lalla@unimo.it

Riassunto

Per condurre un'indagine sulla struttura e sul cambiamento nelle relazioni tra le imprese del settore metalmeccanico, si è costruito un archivio integrando quelli esistenti presso alcune amministrazioni. Le distribuzioni degli addetti, risultanti dal nuovo archivio, sono stati confrontati con quelli preesistenti, settore per settore, applicando il test di Kolmogorv-Smirnov corretto per le numerose ripetizioni dell'applicazione al fine di proteggersi dall'errore di prima specie. Il nuovo archivio è stato utilizzato per estrarre un campione di imprese da intervistare, stratificando le unità statistiche per classe di dimensione (otto classi) e per (sub)settore (dieci domini di studio). Si descrivono, poi, gli elementi essenziali della procedura di campionamento adottato e le problematiche connesse alle difficoltà di rilevazione dei dati: imprese che non appartengono alla popolazione obiettivo o che sono cessate, imprese non rintracciabili, imprese che hanno subíto cambiamenti interni e esterni. Si presentano, poi, i calcoli dei pesi e si valutano le precisioni effettive risultanti dopo la rilevazione e si determinano anche i pesi normalizzati all'unità, da utilizzare nella verifica di ipotesi per non alterare la numerosità campionaria e incorporare nei dati la struttura del piano di campionamento.

Summary

A database had been set up merging information incorporated in different administrative data set. The purpose was to conduct a survey on the structure and changes in the relationships between enterprises operating in the metal, machinery, and transportation industries. The distributions of the number of employees for each sub-sector of those industries, obtained from the new database, were compared with those obtained from the previous administrative archives. The Kolmogorov-Smirnov test was used to ascertain the differences observed between the various distributions, taking into account the repetition of the application of the test to protect against an error of type I. The new database was used to draw a sample of enterprises, stratified by class interval of employees and by sub-industries. Afterwards, the fundamental elements of the adopted sampling procedure and the problems relative to data collection were described: companies carrying out economic activities differing from the object of interest, dead or untraceable companies, companies subjected to internal or external changes. The strategy to obtain the weights and the precision of the estimates resulting from the data actually surveyed were presented. Moreover, in order to maintain the sample size constant and to incorporate the sampling design into the weights a set of normalized weights was calculated.

1. Introduzione

L'indagine sulla «Struttura e cambiamento nelle relazioni tra le imprese metalmeccaniche nella provincia di Modena» intende accertare l'influenza della ristrutturazione delle grandi e medie imprese metalmeccaniche sulle imprese di subfornitura locale rispetto alle modifiche delle specializzazioni presenti nell'area, del ricorso a subforniture esterne all'area, delle tipologie dei prodotti coinvolti, e delle aree stesse (Russo, Giardino, 2000). Per conseguire tali obiettivi, si è deciso di costruire un archivio adatto e molto accurato, integrando le informazioni contenute in alcuni archivi disponibili, costruiti per fini amministrativi; infatti, la procedura di integrazione di informazioni contenute in archivi costruiti da diversi enti amministrativi per conseguire le loro finalità è ritenuta, in genere, efficace per migliorare la qualità delle informazioni contenute negli archivi amministrativi (Martini, 1990; Abbate e Baldassarini, 1994). Tra i numerosi archivi di natura amministrativa sul sistema delle imprese, disponibili presso la Camera di Commercio Industria e Artigianato (CCIAA) di Modena, si sono considerati il Registro delle Imprese e il Repertorio Economico Amministrativo (REA): la loro integrazione ha generato una base di dati, denominata «Archivio CCIAA». Tra gli archivi della sede provinciale dell'INPS, si è utilizzato quello dell'Osservatorio sulle imprese (la prima delle tre parti), che contiene i dati richiesti dai modelli amministrativi (DM10) presentati mensilmente alle imprese per il versamento dei contributi previdenziali dei propri dipendenti, denominato «Archivio INPS». L'integrazione dell'archivio CCIAA e dell'archivio INPS ha generato un archivio denominato «Archivio UNI-MEC», che è stato costruito anche con accertamenti mirati per le imprese che non erano presenti in entrambi le basi di dati e rappresentava una base campionaria soddisfacente per gli obiettivi dell'indagine.

Per quanto concerne l'attività delle imprese, si erano individuati 10 aggregati ai quali ci si riferirà con il termine «comparto» (Russo, Giardino, 2000). In relazione alla costituzione del campione, i domini di studio si sono ottenuti disaggregando le imprese, oltre che per comparto, anche per la loro dimensione, misurata in termini di addetti e suddivisa in classi, in quanto è una caratteristica fondamentale per l'analisi e si suppone correlata con diverse altre variabili rilevanti per lo studio del settore. Le classi utilizzate per il numero di addetti erano otto: da 1 a 5, da 6 a 9, da 10 a 19, da 20 a 49, da 50 a 99, da 100 a 249, da 250 a 499, 500 e oltre. Si era interessati a verificare, quindi, se la distribuzione degli addetti nell'archivio ricostruito era conforme alla distribuzione relativa agli archivi di partenza, per aggregato, ma per questa verifica non era necessario utilizzare la suddivisione del numero di addetti in classi. Si deve súbito osservare che non è necessario verificare tale ipotesi, perché la bontà dell'archivio ricostruito è data essenzialmente dalla correttezza e accuratezza della procedura utilizzata e dei controlli eseguiti. Si è interessati a controllare, invece, la coerenza dell'«Archivio UNI-MEC» ricostruito *ad hoc* con le distribuzioni ottenute dall'«Archivio AIDI» dell'Unioncamere e dal censimento intermedio del 1996 condotto dall'ISTAT. La verifica della coerenza tra le distribuzioni degli addetti relativa ai vari archivi fornisce semplicemente un'indicazione sulle differenze ottenute: se non si rifiuta l'ipotesi nulla di uguaglianza si può concludere che gli errori rilevati, ammesso che non se ne siano commessi degli altri, possono considerarsi quasi come effetti del caso e, in un certo senso, si compensano. Diversamente, si deve concludere che si è alterata la distribuzione dimensionale degli addetti tra i due archivi. Tale argomento costituisce la prima parte del presente lavoro.

Nella conduzione delle indagini e, in particolare, durante la fase di rilevazione delle unità statistiche campionarie si incontrano numerose difficoltà, specie quando le unità statistiche sono entità complesse come le imprese. Le unità selezionate dalla lista (o base campionaria) possono risultare: *non appartenenti* alla popolazione di riferimento, $\wp$; *cessate* (o morte) perché hanno interrotto la loro attività; *non rintracciabili* perché potrebbero avere cessato l'attività, essere emigrate o traslocate in altra sede, avere registrato erroneamente le loro generalità negli archivi, avere attuato modifiche rilevanti nella struttura e nella ragione sociale. Le modifiche possono alterare le caratteristiche della popolazione di riferimento e, pertanto, anche i mutamenti rilevati tramite il campione dànno una indicazione dei cambiamenti avvenuti in $\wp$, ma influenzano alcuni aspetti del campionamento (come le probabilità di selezione o il calcolo dei pesi) e possono distorcere le stime. Le evoluzioni delle unità statistiche nel tempo generano modifiche che si possono distinguere o classificare in due fenomeni distinti: (a) la *divisione* dell'impresa esistente, che genera spesso la nascita di nuove imprese (una unità cessa e se ne costituiscono due o piú, una unità cessa e se ne costituisce una nuova, un socio rileva parte di un'impresa e costituisce una nuova unità o un socio si separa e l'altro socio resta e cambia nome); la *trasformazione* dell'impresa esistente, che produce un cambio radicale della sua fisionomia e struttura con integrazione di nuovi soci, separazione dei vecchi soci, introduzione di nuove tecnologie, acquisizione di unità esistenti (fusioni), cambio del ramo di attività. La dinamica della popolazione $\wp$ nel tempo è caratterizzata: sia dalle imprese di *nuova costituzione* (o nate), che hanno iniziato l'attività nel lasso di tempo (superiore anche a un anno) intercorrente tra la data di creazione della lista e la data di rilevazione effettiva delle imprese; sia dall'*immigrazione* di imprese, assimilabile alla nuova costituzione; sia dall'*emigrazione* di imprese, assimilabile alla cessazione. In aggiunta a questi problemi strutturali, si hanno anche le *mancate collaborazioni* all'indagine che possono essere: *parziali*, quando le imprese intervistate rispondono solo a alcune parti o domande del questionario; o *totali*, quando le imprese si rifiutano di rispondere o di partecipare all'indagine.

L'evoluzione della popolazione $\wp$ e le risposte mancanti incidono nella determinazione delle stime dei suoi parametri e, in particolare, nella precisione delle stime e nel calcolo dei pesi per riportare i valori campionari a quelli relativi a $\wp$. L'obiettivo della seconda parte del lavoro è presentare gli aspetti teorici e pratici inerenti a questi problemi nell'indagine sulla «Struttura e cambiamento nelle relazioni tra imprese meccaniche» che intende accertare l'influenza della ristrutturazione delle grandi e medie imprese meccaniche sulle imprese di subfornitura locale rispetto alle modifiche delle specializzazioni presenti nell'area, del ricorso a subforniture esterne all'area, delle tipologie dei prodotti coinvolti, e delle aree stesse (Russo, Giardino, 2000).

La prima parte del lavoro descrive, pertanto, l'applicazione di alcuni test di conformità sulle distribuzioni degli addetti delle imprese del settore metalmeccanico della provincia di Modena, derivate da fonti diverse, ossia da archivi di natura amministrativa di provenienza differente. L'esposizione è sviluppata in due sezioni: nel paragrafo 2 si espone la teoria del test di Kolmogorov-Smirnov e la procedura di applicazione; nel paragrafo 3 si riportano gli esiti dei calcoli eseguiti.

La seconda parte del lavoro illustra, invece, alcune considerazioni sui pesi e sulla precisione delle stime dei parametri della popolazione di riferimento. La struttura

dell’esposizione è composta da cinque paragrafi, che sviluppano i diversi argomenti inerenti alla procedura di campionamento. Nel paragrafo 4 si descrivono sinteticamente le problematiche del campionamento e la strategia adottata, con le precisioni effettive risultanti dopo la rilevazione. Nel paragrafo 5 si illustrano gli aspetto teorici della determinazione dei pesi in presenza delle difficoltà di rilevazione empiriche. Nel paragrafo 6 si presenta una possibile normalizzazione all’unità dei pesi, per «incorporare» il piano di campionamento nel processo di stima senza alterare la numerosità campionaria. Nel paragrafo 7 si riportano le tabelle relative all’indagine in oggetto con qualche breve commento.

Le conclusioni, per la prima e per la seconda parte, seguono nel paragrafo 8.

I. Distribuzioni degli addetti in alcuni archivi amministrativi

2. Il test di Kolmogorov-Smirnov

Il test di Kolmogorov-Smirnov (per un solo campione) è un test di adattamento concepito per stabilire se vi è coerenza tra la distribuzione di probabilità empirica derivata dai dati osservati inerenti al carattere *Y* e una distribuzione di probabilità ipotizzata, nel caso di un solo campione. In altri termini, consente di verificare se il carattere *Y* può essere descritto da una determinata distribuzione teorica nella popolazione da cui è stato estratto il campione; svolge, pertanto, la stessa funzione del test del chi-quadrato seguendo un procedimento diverso che risulta più efficace.

Il test di Kolmogorov-Smirnov per due campioni indipendenti è un test di adattamento concepito per stabilire se vi è coerenza tra le distribuzioni di probabilità empiriche derivate dai dati osservati inerenti al carattere *Y* nei due campioni e una distribuzione di probabilità ipotizzata. In altri termini, consente di verificare se il carattere *Y* può essere descritto dalla stessa distribuzione in entrambi i gruppi e la popolazione da cui sono stati estratti i due campioni, quindi, è la stessa; svolge sempre una funzione analoga al test del chi-quadrato per una tabella di contingenza (di dimensione *K* per 2, con $K>2$) seguendo un procedimento simile al test per un singolo campione, ma risulta più efficace del chi-quadrato.

Il test si basa sul confronto tra le funzioni di ripartizione empiriche derivate dai due campioni, mentre nel caso di un singolo campione il confronto avviene con una ipotizzata funzione di ripartizione che caratterizza la distribuzione di probabilità del carattere *Y*. In particolare, si considera la differenza massima osservata tra le due funzioni di ripartizione e si valuta la probabilità di significatività, ossia si valuta se la differenza è più grande di quanto ci potrebbe aspettare per effetto del caso nell'ipotesi che i due campioni siano stati estratti dalla stessa popolazione.

2.1. I requisiti del test

Gli elementi essenziali per l'applicazione del test sono tre: il livello di misura delle variabili implicate nel test, il modello statistico, e il sistema di ipotesi da verificare.

Il *livello di misura* del carattere *Y* deve essere almeno qualitativo ordinale per il quale la proprietà che rappresenta è continua, ma si esprime con un ordinamento semplice solo per l'imprecisione dello strumento di misura. Il carattere *X* distingue i due gruppi e, quindi, è dicotomo.

Il *modello statistico* assume che il carattere *Y* sia la variabile casuale che rappresenta le modalità del carattere

$$P\left(Y \le y \middle| X = x_i\right) = \begin{cases} F_1(y) & \text{per} \quad x_1 = 1 \\ F_2(y) & \text{per} \quad x_2 = 2 \end{cases}$$

dove $F_i(y)$ è la funzione di ripartizione, indicizzata dal valore del carattere *X* che può assumerne solo due: $\{x_1, x_2\}$.

Il *sistema di ipotesi* da verificare concerne la validità del modello statistico relativo alla distribuzione di Y che assume una funzione di ripartizione identica nei due gruppi: H_0: $F_1(y) = F_2(y)$ contro H_1: $F_1(y) \neq F_2(y)$.

2.2. Scelta della statistica del test

Sia ($y_{(1)}$, $y_{(2)}$, …, $y_{(n_1)}$) il vettore ordinato delle osservazioni del gruppo 1, per cui si ha ($y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(n_1)}$). Sia ($y_{(1)}$, $y_{(2)}$, …, $y_{(n_2)}$) il vettore ordinato delle osservazioni del gruppo 2 per cui si ha ($y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(n_2)}$). La funzione di ripartizione (empirica) per ciascun gruppo sarà data da

$$S_{j;n_1}\left(y_{(i)}\right) = \sum_{j=1}^{i} \frac{1}{n}\ 1_{\{y_{(1)},\Lambda\,,y_{(n_1)}\}}\left(y_j\right) = \frac{i}{n}$$

per j=1,2 e dove $1_{\{y_{(1)},\Lambda\,,y_{(n1)}\}}\left(y_j\right)$ è la funzione indicatrice (o caratteristica) che vale uno quando l'argomento y_j è uguale a uno dei valori specificati nell'insieme degli indici di 1 e zero altrimenti. Il test di Kolmogorov-Smirnov per due campioni può essere unilaterale, se si considerano solo le differenze positive; e bilaterale, se si considerano le differenze positive e negative; pertanto, le statistiche diventano due:

$$D^{+}_{n_1,n_2} = \max_{-\infty<y<\infty}\left\{S_{1;n_1}(y) - S_{2;n_2}(y)\right\}$$

$$D_{n_1,n_2} = \max_{-\infty<y<\infty}\left|S_{1;n_1}(y) - S_{2;n_2}(y\right|.$$

In questo àmbito si considera solo la statistica bi-direzionale D_{n_1,n_2}. Sia n= $n_1 n_2/(n_1+n_2)$. Sia $Q_{n_1,n_2}(l)$ la funzione di ripartizione della variabile casuale $D_{n_1,n_2}\sqrt{n}$. Allora,

$$Q_{n_1,n_2}(l) = P\left(D_{n_1,n_2}\sqrt{n} \leq l\right) = \begin{cases} P\left(D_{n_1,n_2} \leq \dfrac{l}{\sqrt{n}}\right) & \text{per} \quad l > 0 \\ 0 & \text{per} \quad l \leq 0 \end{cases}.$$

Siano $S_{1;n_1}(y)$ e $S_{2;n_1}(y)$ le due funzioni di ripartizione empiriche dei due caratteri relativi a due campioni casuali estratti dalla stessa popolazione, in cui la Y è continua con funzione di ripartizione $F(y)$ e la X è discreta individuando i due gruppi; allora:

$$\lim_{\substack{n_1\to\infty \\ n_2\to\infty}} Q_{n_1,n_2}(l) = \begin{cases} \displaystyle\sum_{k=-\infty}^{\infty}(-1)^k e^{-2k^2 l^2} & \text{per} \quad l > 0 \\ 0 & \text{per} \quad l \leq 0 \end{cases}$$

Ora, si conoscono gli elementi essenziali per procedere nell'applicazione del test. Il test unilaterale si usa per accertare se i valori dei dati osservati in uno dei due campioni sono stocasticamente più grandi dei valori della popolazione da cui è stato estratto l'altro gruppo; per esempio, per decidere se i punteggi conseguiti da un gruppo sono superiori

a quelli ottenuti dell'altro gruppo (per posizione, per dispersione, per asimmetria, e per appiattimento). Il test bilaterale, invece, è sensibile a qualunque tipo di differenza in piú o in meno (sempre per posizione, per dispersione, per asimmetria, e per appiattimento).

2.3. Livello di significatività del test e dimensione del campione

Si fissa il livello di significatività del test, a , e il tipo di test (unilaterale o bilaterale) in base ai quali si determina il valore l_c (valore critico) tale che

$$P\left(D_{n_1,n_2} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \geq l_c \right) = a \ .$$

Per *piccoli campioni*, quando n_1 e n_2 sono minori o uguali a 25, la determinazione di l_c deve avvenire in modo esatto e si devono, perciò, eseguire i calcoli utilizzando la funzione di ripartizione sopra riportata o ricorrere all'uso delle tavole. Per *grandi campioni*, quando n_1 e n_2 sono maggiori di 25, la determinazione di l_c può avvenire utilizzando la distribuzione asintotica (approssimata) della variabile casuale D_{n_1,n_2} nel caso di un test bilaterale.

Nel test unilaterale si utilizza la statistica $D^+_{n_1,n_2}$ per costruire la statistica X^2 di Pearson che si distribuisce approssimativamente come un chi-quadrato con 2 gradi di libertà (Goodman, 1954). Per il test bilaterale, invece, si confronta la statistica D_{n_1,n_2} con $\frac{l_c}{\sqrt{n}} = l_c \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$ dove i valori di l_c sono: *1,224* per un livello di significatività del 10%; *1,358* per un livello di significatività del 5%; e *1,628* per un livello di significatività dell'1%.

2.4. Calcolo della statistica-test

Siano $(y_{(1)}, y_{(2)}, \ldots, y_{(n_1)})$ e $(y_{(1)}, y_{(2)}, \ldots, y_{(n_2)})$, i vettori osservati e ordinati nei due gruppi. In base a essi e al tipo di test (unilaterale o bilaterale), si calcola il valore osservato della statistica desiderata: D_{n_1,n_2} o $D^+_{n_1,n_2}$. In particolare, si può calcolare, per esempio,

$$l_{\text{oss}} = D_{n_1,n_2} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \ .$$

2.5. Decisione sull'accettabilità dell'ipotesi nulla

Si è in grado, ora, di decidere se le ipotesi sono coerenti con i dati. Per i piccoli campioni si usano le tavole, dalle quali si desume la probabilità di significatività, e i dati non sono coerenti con l'ipotesi nulla quando essa è inferiore 0,05. Per i grandi campioni si confronta il valore osservato, l_{oss}, di l con il valore critico $l_{c;a}$, determinato secondo la distribuzione di probabilità sopra riportata in base al livello di significatività fissato

a : se $l_{\text{oss}} \geq l_{c;a}$, i dati non sono coerenti con H_0 e non si può accettare H_0, quindi, al livello di significatività a ; se $l_{\text{oss}} < l_{c;a}$, i dati sono coerenti con H_0 e si può accettare H_0, quindi, al livello di significatività a .

2.6. Potenza e efficienza del test di Kolmorogov-Smirnov

La potenza e l'efficienza del test sono molto elevate, pari circa al 95%, per piccoli campioni rispetto al test t di Student. Al crescere della dimensione del campione la potenza e l'efficienza decrescono leggermente. Il test di Kolmogorov-Smirnov è, tuttavia, più potente del test del chi-quadrato e del test della mediana. Per piccoli campioni è anche più potente del test di Wilcoxon-Mann-Whitney (Siegel, Castellan, 1988; Landenna, Marasini, 1990).

2.7. Il caso di k campioni indipendenti

La popolazione delle imprese è stata suddivisa per comparto; in particolare, si sono individuati K=10 gruppi, che si possono considerare anche indipendenti (Russo, Giardino, 2000). In ogni caso, l'interesse oggetto del test prevede l'esecuzione di K confronti simultanei. In generale, siano $(Y_1,\Lambda\ ,Y_K)$ variabili casuali assolutamente continue con funzioni di ripartizione $(F_1(y),\Lambda\ ,F_K(y))$, relative alle K popolazioni cui si riferiscono.

L'ipotesi nulla che occorre verificare è $H_o : F_1(y) = \Lambda\ = F_K(y)$ contro l'ipotesi alternativa H_1: almeno una $F_i(y)$ è diversa dalle altre per i=1,…,K. Tra le procedure disponibili si ricordano il test di Birnbaum-Hall (1960) e il test di Kiefer (1959).

Il test di Birnbaum-Hall è una generalizzazione del test di Kolmogorov-Smirnov. Si devono calcolare, perciò, le K funzioni di ripartizione $(S_1(y),\Lambda\ ,S_K(y))$ empiriche e eseguire $K(K-1)/2$ confronti che generano le statistiche di interesse date dai moduli delle differenze $\left|S_i(y) - S_j(y)\right|$ al variare degli indici i,j=1,…,K. Tra queste differenze si sceglie quella massima

$$c_1 = \max_{\substack{-\infty<y<\infty \\ i<j}} \left|S_i(y) - S_j(y)\right|.$$

Il valore c_1 è la realizzazione della variabile casuale C_1 che, sotto l'ipotesi nulla H_0, ha una distribuzione nota. L'applicazione del test non è agevole perché bisogna costruirsi la distribuzione di volta in volta; Birnbaum e Hall hanno disposto una tavola per K=3 e n_1=n_2=n_3=m e per diversi valori di m.

Anche il test di Kiefer è una variante del test di Kolmogorov-Smirnov, per verificare l'ipotesi nulla $H_o : F_1(y) = \Lambda\ = F_K(y)$ contro l'ipotesi alternativa H_1: almeno una $F_i(y)$ è diversa dalle altre per i=1,…,K. Si devono calcolare, perciò, le K funzioni di ripartizione $(S_1(y),\Lambda\ ,S_K(y))$ empiriche relative ai K campioni e la funzione di ripartizione empirica $S(y)$ relativa al totale n= n_1+…+n_K risultati osservati. La statistica per eseguire il test è data da

$$c_3 = \max_{-\infty<y<\infty} \left\{ \sum_{k=1}^{K} n_k \left[S_i(y) - S_j(y) \right]^2 \right\},$$

che è la realizzazione della variabile casuale C_3 della quale Kiefer ha fornito distribuzione asintotica e ha anche disposto una tavola che si può utilizzare per determinare la probabilità di significatività.

Il confronto tra i *K* gruppi non deve essere eseguito, però, per tutte le possibili combinazioni perché le distribuzioni possono essere ovviamente diverse da un gruppo (settore) all'altro. Per ogni valore di *k* si deve eseguire un confronto tra due funzioni di ripartizione che dovrebbero derivare dalla stessa popolazione. Per proteggersi dall'errore di prima specie si può usare il metodo di Bonferroni che si basa sulla disuguaglianza di Bonferroni (Miller, 1966, pp. 7-8; Feller, 1968, pp. cap. 4). A un insieme di *m* ipotesi sottoposte a verifica a un livello di significatività α, si assegna un saggio di errore α/m e l'ipotesi nulla di uguaglianza delle funzioni di ripartizioni relative ai due gruppi è incoerente con i dati quando la probabilità di significatività è minore di α/m.

3. Risultati empirici

Gli esiti dell'esecuzione dei diversi test di Kolmogorov-Smirnov —applicati per confrontare le distribuzioni degli addetti, per codice di attività economica e per classe di dimensione, estratte dall'«Archivio UNI-MEC» ricostruito *ad hoc* con le distribuzioni ottenute dall'«Archivio ISTAT»— sono riportati nella Tabella 1. Si sono utilizzate le distribuzioni degli addetti relative alla suddivisione in classi seguenti: *k*=1, da 1 a 5 addetti; *k*=2, da 6 a 9 addetti; *k*=3, da 10 a 19 addetti; *k*=4, da 20 a 49 addetti; *k*=5, da 50 a 99 addetti; *k*=6, da 100 a 249 addetti; *k*=7, da 250 a 499 addetti; *k*=*K*=8, da 500 addetti e oltre. Solo per l'attività concernente i «Prodotti finiti in metallo» si rileva una differenza statisticamente significativa; in effetti, in questo settore di attività, anche il numero di imprese risulta assai diverso. Le classi in cui si osservano le differenze maggiori tra le due distribuzioni empiriche sono, in genere, la prima.

Tabella 1 - *Test di Klmogorov-Smirnov per verificare l'uguaglianza tra le distribuzioni dell'«Archivio UNI-MEC» con la distribuzione ottenuta dal censimento intermedio del 1996 condotto dall'ISTAT (disponibile solo in classi)*

Settori	n_1	n_2	y_{D_k}	D_{n_1,n_2}	l_{oss}	$P<$
DJ27 Produzione di metalli e loro leghe	47	45	*k*=3	0,052	0,249	1,000
DJ28.1 +28.2 +28.3 Carpenteria metallica	487	525	*k*=1	0,042	0,673	
DJ28.40 Fucinatura, imbutitura, stampaggio, e profilatura metalli	87	76	*k*=1	0,025	0,160	1,000
DJ28.51 Trattamento e rivestimento dei metalli	108	117	*k*=1	0,041	0,310	
DJ28.52 Lavori di meccanica generale, conto terzi	787	736	*k*=2	0,012	0,230	
DJ28.6+28.7+Resto DJ28 Prodotti finiti in metallo	150	256	*k*=1	0,141	**1,374**	**0,000**
DK29 Fabbricazione di macchine e app. mecc.	1336	1322	*k*=1	0,007	0,174	1,000
DL31 Fabbricazione di macchine e app. elettrici	258	270	*k*=2	0,058	0,667	
DL30+32+33 Fabbric. macch. el. e app. el. e ott.	536	514	*k*=1	0,040	0,641	
DM34+35 Fabbricazione di mezzi di trasporto	104	63	*k*=3	0,087	0,546	

Gli esiti dell'esecuzione dei diversi test di Kolmogorov-Smirnov —applicati per confrontare le distribuzioni degli addetti, per codice di attività economica e per classe di dimensione, estratte dall'«Archivio UNI-MEC» ricostruito *ad hoc* con le distribuzioni ottenute dall'«Archivio AIDI»— sono riportati nella Tabella 2. In tal caso si sono utilizzate le distribuzioni "originarie", in cui le distribuzioni sono costruite sul numero effettivo di addetti; sicché le differenze massime si osservano per un valore dato degli addetti nell'insieme $\{1,2,\Lambda\ ,n,\Lambda\}$. Nella Tabella 2, si vede che le differenze emergono in valori della dimensione intorno a 5 addetti, eccetto per il settore di lavorazione «Produzione di metalli e loro leghe». Anche in questi confronti, la differenza piú consistente si osserva per l'attività concernente i «Prodotti finiti in metallo»; ma anche nei settori «Lavori di meccanica generale per conto terzi» e «Fabbricazione di macchine e apparecchi meccanici» si hanno differenze che mostrano l'incoerenza dell'ipotesi di uguaglianza tra le distribuzioni empiriche, anche se la probabilità di significatività è vicina al valore critico (*borderline*). Per eseguire un confronto analogo a quello eseguito tra gli archivi «UNI-MEC» e «ISTAT», sia nell'«archivio UNI-MEC» e sia nell'«archivio AIDI», si sono raggruppati in classi le distribuzioni degli addetti secondo la suddivisione in classi precedente: k=1, da 1 a 5 addetti; k=2, da 6 a 9 addetti; k=3, da 10 a 19 addetti; k=4, da 20 a 49 addetti; k=5, da 50 a 99 addetti; k=6, da 100 a 249 addetti; k=7, da 250 a 499 addetti; k=K=8, da 500 addetti e oltre. I risultati del confronto sono riportati nella Tabella 2bis, dove si può osservare che restano pressoché immutati, rispetto a quelli della Tabella 2 —eccetto il settore di attività «Fabbricazione di macchine e apparecchi meccanici», nel quale l'ipotesi di uguaglianza tra le distribuzioni empiriche non mostra coerenza con i dati dei due archivi—. Le classi in cui si osservano le differenze maggiori tra le due distribuzioni empiriche sono ancora, in genere, la prima.

Tabella 2 - *Test di Klmogorov-Smirnov per verificare l'uguaglianza tra le distribuzioni dell'«Archivio UNI-MEC» ricostruito ad hoc con le distribuzioni ottenute dall'«Archivio AIDI» dell'Unioncamere (raggruppata in classi)*

Settori	n_1	n_2	y_D	D_{n_1,n_2}	l_{oss}	$P<$
DJ27 Produzione di metalli e loro leghe	47	56	15	0,0771	0,390	0,998
DJ28.1 +28.2 +28.3 Carpenteria metallica	487	493	2	0,6504	1,018	0,251
DJ28.40 Fucinatura, imbutitura, stampaggio, e profilatura metalli	87	83	7	0,0637	0,415	0,995
DJ28.51 Trattamento e rivestimento dei metalli	108	105	6	0,0477	0,349	1,000
DJ28.52 Lavori di meccanica generale, conto terzi	787	788	5	0,0729	**1,447**	***0,030***
DJ28.6+28.7+Resto DJ28 Prodotti finiti in metallo	150	327	2	0,4018	**4,075**	**0,000**
DK29 Fabbricazione di macchine e app. mecc.	1336	1547	2	0,0535	**1,433**	***0,033***
DL31 Fabbricazione di macchine e app. elettrici	258	296	4	0,0956	1,122	0,161
DL30+32+33 Fabbric. macch. el. e app. el. e ott.	536	566	3	0,0576	0,956	0,321
DM34+35 Fabbricazione di mezzi di trasporto	104	89	4	0,1161	0,804	0,537

Tabella 2bis - *Test di Klmogorov-Smirnov per verificare l'uguaglianza tra le distribuzioni dell'«Archivio UNI-MEC» ricostruito ad hoc con le distribuzioni ottenute dall'«Archivio AIDI» dell'Unioncamere*

Settori	n_1	n_2	y_{D_k}	D_{n_1,n_2}	l_{oss}	$P<$
DJ27 Produzione di metalli e loro leghe	47	56	k=1	0,023	0,117	1,000
DJ28.1 +28.2 +28.3 Carpenteria metallica	487	493	k=1	0,039	0,616	
DJ28.40 Fucinatura, imbutitura, stampaggio, e profilatura metalli	87	83	k=2	0,056	0,366	
DJ28.51 Trattamento e rivestimento dei metalli	108	105	k=2	0,015	0,112	1,000
DJ28.52 Lavori di meccanica generale, conto terzi	787	788	k=1	0,073	**1,447**	***0,042***
DJ28.6+28.7+Resto DJ28 Prodotti finiti in metallo	150	327	k=1	0,238	**2,412**	**0,000**
DK29 Fabbricazione di macchine e app. mecc.	1336	1547	k=1	0,036	0,960	
DL31 Fabbricazione di macchine e app. elettrici	258	296	k=1	0,080	0,944	
DL30+32+33 Fabbric. macch. el. e app. el. e ott.	536	566	k=1	0,032	0,530	
DM34+35 Fabbricazione di mezzi di trasporto	104	89	k=1	0,067	0,461	

Si noti che le probabilità di significatività riportate nelle tabelle non sono state neanche corrette secondo il criterio di Bonferroni. Infatti, le probabilità, riportate in grassetto e corsivo nelle tabelle, non sarebbero significative dopo l'applicazione della correzione di Bonferroni.

II. Pesi e precisione delle stime dei parametri

4. Strategia di campionamento e errore relativo delle stime

Il campione è stato costruito con il metodo del campionamento stratificato perché è il piú idoneo e usuale a indagare la realtà economica delle imprese. La stratificazione è stata condotta secondo due caratteri distintivi delle unità statistiche oggetto di indagine: il comparto del settore meccanico (indice *i*) e il numero di addetti delle imprese (indice *j*). Nella determinazione della dimensione del campione per strato si sarebbe potuto adottare l'allocazione proporzionale che è autoponderante e possiede alcuni vantaggi, consistenti in una varianza ridotta (rispetto al campione casuale semplice), semplicità, e robustezza dei risultati (Kish, 1990, 1992). Si è preferito determinare la dimensione del campione strato per strato («singola allocazione per strato») perché il campione cosí estratto è assimilabile a una specie di allocazione ottimale; infatti, la variabilità per comparto, rispetto a una data classe di dimensione degli addetti, è quasi costante; allora, produce stime della media o del totale con una varianza che è fissata e nota per strato; e, presumibilmente, è ancora piú piccola delle stime conseguite con un campione autoponderante.

4.1. Stratificazione per comparto e dimensione

La suddivisione per comparto rappresenta una stratificazione quasi naturale perché raggruppa le unità statistiche in base alla loro attività. Si sono costituite dieci classi (Russo e Giardino, 2000): (1) DJ27 produzione di metalli e loro leghe; (2) DJ28.1 +28.2 +28.3 carpenteria metallica; (3) DJ28.40 fucinatura, imbutitura, stampaggio, e profilatura metalli; (4) DJ28.51 trattamento e rivestimento dei metalli; (5) DJ28.52 lavori di meccanica generale per conto terzi; (6) DJ28.6+28.7+Resto DJ28 prodotti finiti in metallo; (7) DK29 fabbricazione di macchine e apparecchi meccanici; (8) DL31 fabbricazione di macchine e apparecchi elettrici n.c.a.; (9) DL30+32+33 fabbricazione di macchine elettriche e apparecchi elettrici e ottici; (10) DM34+35 fabbricazione di mezzi di trasporto.

Il numero degli addetti dell'impresa, *Y*, era una variabile oggetto di stima e, quindi, rappresentava la caratteristica ideale per la stratificazione (Cochran, 1977, p. 101). La suddivisione in classi di addetti è stata fissata, per convenzione, come segue: la prima classe comprende le imprese da 1 a 5 addetti, la seconda da 6 a 9 addetti, la terza da 10 a 19 addetti, la quarta da 20 a 49 addetti, la quinta da 50 a 99 addetti, la sesta da 100 a 249 addetti, la settima da 250 a 499, e l'ottava da 500 al massimo numero di addetti. Nella prime due classi vi sono imprese in cui il numero di addetti è spesso costituito da soci e/o famigliari e sono ritenute "stabili" rispetto a alcuni caratteri economici e aziendali, mentre le altre sono pressoché uguali a quelle dell'ISTAT (1986) (Brusco, Giovannetti, e Malagoli, 1979; Jalla, 1981).

4.2. Determinazione dell'ampiezza del campione

L'ampiezza totale del campione si può determinare considerando sia la precisione desiderata delle stime, sia i costi da sostenere per condurre l'indagine. La precisione

desiderata delle stime si calcola, allora, in base al numero degli addetti, *Y*, l'unico carattere oggetto di stima che era disponibile nella lista e rilevante anche per le altre grandezze dell'impresa (Marzi, 1990): fatturato, tipo di impresa, tipo di prodotto, piazza di destinazione, tipo di cliente, e cosí via. Infatti, le loro stime con i dati dell'indagine risultano piú affidabili e precise se quelle grandezze sono correlate con la dimensione stessa dell'impresa. L'ampiezza del campione, quindi, è data dal minimo tra quella determinata dalla precisione desiderata e quella determinata dalle risorse finanziarie disponibili (Cochran, 1977):

$$n = \min \left\{ \frac{\left(z_{a/2} S / (r\bar{Y})\right)^2}{1 + \left(z_{a/2} S / (r\bar{Y})\right)^2 / N}, \frac{C_{\max} - c_0}{c_1} \right\}. \qquad (1)$$

Nel primo termine tra parentesi (ampiezza risultante dalla precisione desiderata), $\bar{Y}$ e S sono media e deviazione standard degli addetti, rispettivamente, N è il numero totale di imprese nella popolazione, r è l'errore relativo desiderato nella stima della media o del totale della Y, e $z_{a/2}$ è il valore dell'ascissa della distribuzione normale corrispondente al livello di significatività α. Nel secondo termine (ampiezza risultante dalle risorse disponibili), $C_{\max}$ è l'ammontare totale di risorse disponibili, c_0 è l'ammontare dei costi fissi, c_1 è il costo di rilevazione per impresa, mentre la funzione di costo è lineare con c_1 costante.

Per gli obiettivi dell'indagine sulla «Struttura e cambiamento nelle relazioni tra imprese meccaniche» si voleva controllare la precisione delle stime strato per strato; pertanto, si è applicata l'espressione precedente all'interno di ogni strato, utilizzando soltanto il primo termine tra parentesi del secondo membro.

4.3. Allocazione del campione tra gli strati

Il numero di imprese, n_{ij}, da intervistare negli strati *ij* (determinati dall'*i*-esimo comparto e dalla *j*-esima classe di dimensione), è stato determinato con l'applicazione dell'espressione (1) in ogni strato, controllando la dimensione totale del campione attraverso la variazione della precisione relativa delle stime per classe di dimensione. Iterando la procedura di calcolo per diversi valori dell'errore relativo si sono fissati gli errori relativi per classe di dimensione: r=30% nella classe di addetti 1-5; r=15% nella classe di addetti 6-9; r=15% nella classe di addetti 10-19; r=10% nella classe di addetti 20-49; r=10% nella classe di addetti 50-99; mentre nelle classi rimanenti si sono selezionate tutte le imprese e l'errore relativo è, in tal caso, zero (Russo e Giardino, 2000). Nelle prime cinque classi, in cui si è eseguito il campionamento, la variabilità per comparto è quasi costante sicché la ripartizione delle unità campionarie tra i vari comparti è assai simile all'allocazione proporzionale di Neyman. Si noti che la dimensione dell'impresa è un carattere che presenta una forte asimmetria positiva (tante aziende con pochi addetti e poche con tanti addetti) e la sua variabilità aumenta all'aumentare dell'intensità media dello stesso; allora, si possono trarre notevoli benefici dall'allocazione di Neyman (Cicchitelli, Herzel, e Montanari, 1992, p. 324).

4.4. Errore relativo effettivo

Le difficoltà di rilevazione incontrate, che sono comuni a tutte le indagini empiriche, hanno ridotto la dimensione campionaria desiderata (programmata) per strato; di

conseguenza è aumentata l'entità dell'errore relativo delle stime. Per valutare l'ordine di grandezza dell'errore campionario, a posteriori, si può ricavare r dal primo termine del secondo membro della (1). L'errore relativo «a posteriori», $r_{p;ij}$, considerando le imprese partecipanti, $n_{p;ij}$, nello strato ij è pari a

$$r_{p;ij}^2 = \frac{1}{n_{p;ij}} \left(\frac{z_{a/2} S_{ij}}{\bar{Y}_{ij}} \right)^2 \left(\frac{N_{ij} - n_{p;ij}}{N_{ij} - 1} \right), \tag{2}$$

dove $\bar{Y}_{ij}$ indica la media del carattere Y nello strato ij, S_{ij} indica la corrispondente deviazione standard, e N_{ij} indica l'entità della popolazione di riferimento. Se nel calcolo di S_{ij} si usa l'espressione «corretta per i gradi libertà», ossia si calcola la media dei quadrati degli scarti dalla media dividendo per $(N_{ij} - 1)$, allora l'espressione precedente diventa

$$r_{p;ij}^2 = \frac{1}{n_{p;ij}} \left(\frac{z_{a/2} S_{ij}}{\bar{Y}_{ij}} \right)^2 \left(\frac{N_{ij} - n_{p;ij}}{N_{ij}} \right). \tag{3}$$

5. Fattori di espansione

Sia data una popolazione $\wp$ di N unità. Sia Y la variabile in oggetto con distribuzione statistica incognita e valori $(Y_1, Y_2 \Lambda , Y_N)$. Si voglia stimare il totale di Y, dato da $Y = \sum_{i=1}^{N} Y_i$ in base al campione osservato $(y_1, y_2 \Lambda , y_n)$, con l'eventuale uso di variabili ausiliarie. Nello schema adottato, y_1 indica il valore osservato di Y nell'unità ottenuta dalla prima estrazione, y_2 indica il valore osservato di Y nell'unità ottenuta dalla seconda estrazione e cosí via fino all'n-esima estrazione. Gli stimatori che si considerano, in genere, sono lineari del tipo

$$T = \sum_{i=1}^{n} w_i y_i \tag{4}$$

dove le quantità w_i sono dette *pesi* che non dipendono dal numero d'ordine delle osservazioni, ma possono dipendere dal tipo di campionamento adottato e dall'etichetta che individua l'unità statistica selezionata (Cicchitelli, Herzel, Montanri, 1992). Piú in particolare, si consideri un piano di campionamento probabilistico che genera un campione di n unità estratte senza ripetizione (reimmissione). Siano $(y_1, y_2 \Lambda , y_n)$ le osservazioni campionarie e siano $(p_1, p_2 \Lambda , p_n)$ le probabilità di inclusione del primo ordine delle unità della popolazione di riferimento $\wp$ dalla quale si sono estratte le osservazioni; allora, lo stimatore corretto del totale, $\hat{Y}$, è dato da

$$\hat{Y} = \sum_{i=1}^{n} \frac{1}{p_i} y_i , \tag{5}$$

noto come stimatore di Horvitz-Thompson (Horvitz e Thompson, 1952). Si tratta, quindi, di uno stimatore ottenuto dalla combinazione lineare delle osservazioni campionarie con pesi pari a $1/p_i$ $(i = 1,2,\Lambda , n)$ dipendenti dalle etichette delle unità cui si riferiscono le osservazioni, ossia dal piano di campionamento adottato. L'allocazione, adottata nell'indagine sulla «Struttura e cambiamento nelle relazioni tra imprese meccaniche», richiede l'uso di pesi diversi per ciascun dominio di studio (o strato), a

differenza del campione autoponderante. L'espressione per stimare il totale del carattere *Y* si ottiene adattando l'equazione (5) al contesto del piano di campionamento specifico

$$\hat{Y} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{n_{ij}} \frac{1}{p_{ij}} y_{ijk} \ . \tag{6}$$

La statistica definita dalla (6) è detta anche *stimatore per espansione* perché nel caso di un campionamento casuale semplice o autoponderante diventa semplicemente il prodotto della corrispondente grandezza campionaria moltiplicata per la frazione di campionamento: $\hat{Y} = (N/n)\sum_{ijk} y_{ijk} = N\ \bar{y}$. La frazione di campionamento è, in totale, n/N; quindi, si trattano i dati come se ogni unità del campione rappresenti N/n unità della popolazione. Tale fattore, N/n, è detto anche *coefficiente di espansione*. Nel caso in oggetto, all'interno di ogni strato (o dominio di studio) si ha un peso che corrisponde proprio a questa rappresentazione, dato dall'inverso della probabilità di selezione del primo ordine $1/p_{ij}$; pertanto, all'interno di ogni strato il fattore di espansione o peso è

$$w_{ij} = \frac{1}{p_{ij}} = \frac{N_{ij}}{n_{ij}} \tag{7}$$

dove, come si è detto, l'indice *i* denota il comparto $(i = 1, \Lambda\ I = 10)$ e l'indice *j* denota la classe di dimensione $(j = 1, \Lambda\ J = 8)$.

Per la determinazione dei pesi, che riportano la popolazione alla data di riferimento della lista, occorre considerare: la non appartenenza alla popolazione di riferimento; la cessazione; e la non rintracciabilità che può includere sia la mortalità, sia l'emigrazione, sia la divisione e sia la fusione con cambio di nome. Il trattamento delle imprese che rientrano in tali categorie può seguire diverse strategie alternative nel calcolo dei pesi.

(a) Si ignorano e si trattano come non rispondenti, ma ne consegue una possibile sovrastima della popolazione di riferimento.
(b) Si assume che nella lista vi siano imprese estranee alla popolazione di riferimento, perché non classificate correttamente nell'attività dichiarata alle istituzioni che generano gli archivi.

L'entità degli errori è sembrata trascurabile sicché la stima risente della rarità degli eventi e della scarsa numerosità per strato del campione; allora, si rischia di enfatizzare l'effetto nelle stime e ottenere una considerevole sottostima dei valori della popolazione. Nel calcolo dei pesi per strato si utilizza, infatti, la stima della popolazione effettiva, N_E; ma l'esigua numerosità delle unità campionarie per strato comporta un impatto rilevante appena si scopre anche una sola unità statistica in ogni strato. Tuttavia, a causa della specificità delle attività considerate e dell'accuratezza con cui è stata condotta la rilevazione, si è preferito correre il rischio di una sottostima e di apportare la correzione:

$$N_E = p_{ij} N_{ij} = \frac{n_{\in;ij}}{n_{c;ij}} N_{ij} \tag{8}$$

dove N_{ij} è l'entità della popolazione di riferimento ottenuta dalla lista nello strato *ij*; p_{ij} è la proporzione di imprese appartenenti alla popolazione di riferimento nello strato *ij* (stimata dal rapporto tra le imprese contattate e appartenenti alla popolazione di riferimento, $n_{\in;ij}$, e il numero totale di imprese contattate, $n_{c;ij}$).

Per il calcolo dei pesi finali occorre considerare la probabilità di rintracciare una impresa e la probabilità di appartenere alla popolazione di riferimento:

$$w_{e;ij} = \frac{1}{p_{ij}} \quad \frac{1}{p_{r;ij}} \quad \frac{1}{p_{\in;ij}} = \frac{N_{ij}}{n_{ij}} \quad \frac{n_{ij}}{n_{c;ij}} \quad \frac{n_{c;ij}}{n_{\in;ij}} \tag{9}$$

dove $p_{r;ij}$ è la probabilità che un'impresa sia rintracciata, $p_{\in;ij}$ è la probabilità che un'impresa appartenga alla popolazione e sono stimate, rispettivamente, da $n_{c;ij}/n_{ij}$ e $n_{\in;ij}/n_{c;ij}$. I problemi non terminano con queste considerazioni perché l'impresa, contattata e appartenente alla popolazione di riferimento, potrebbe non partecipare all'indagine. Per determinare il peso occorre introdurre, quindi, anche la probabilità di partecipare all'indagine, $p_{p;ij}$, che si può stimare con

$$p_{p;ij} = \frac{n_{p;ij}}{n_{\in;ij}} \tag{10}$$

dove $n_{p;ij}$ indica il numero di imprese che partecipano all'indagine. Il peso per le imprese partecipanti all'indagine diventa, allora

$$w_{p;ij} = \frac{1}{p_{ij}} \quad \frac{1}{p_{r;ij}} \quad \frac{1}{p_{\in;ij}} \quad \frac{1}{p_{p;ij}} = \frac{N_{ij}}{n_{ij}} \quad \frac{n_{ij}}{n_{c;ij}} \quad \frac{n_{c;ij}}{n_{\in;ij}} \quad \frac{n_{\in;ij}}{n_{p;ij}} \tag{11}$$

che corrisponde, banalmente, al rapporto tra il numero di imprese della popolazione nello strato *ij* (eventualmente stimato) e il numero di imprese partecipanti all'indagine

$$w_{p;ij} = \frac{N_{ij}}{n_{p;ij}} = \frac{1}{p^{*}_{p;ij}} \tag{12}$$

dove $1/p^{*}_{p;ij}$ può interpretarsi come una «pseudo-probabilità» di selezione o probabilità di rilevare effettivamente i dati dell'unità statistica perché deriva dalla probabilità di inclusione modificata o corretta per le difficoltà incontrate e che sarà utile in questa forma solo per determinare l'espressione di normalizzazione a uno dei pesi (v. *infra*); infatti, è in questa forma espressiva che si utilizzerà per ricavarli. Naturalmente, questa è la soluzione piú semplice per compensare le stime dalle difficoltà delle indagini e dalle non risposte; altre strategie piú sofisticate e complesse, che non si possono spesso applicare alle indagini su larga scala, si trovano in Rubin (1977). Gli stimatori diventano non lineari e le varianze possono aumentare (Kish, 1990); inoltre, le correzioni apportate non correlano con le variabilità negli strati e incrementano generalmente la varianza (Bethlehem e Keller, 1987; Potter, 1990).

6. Normalizzazione dei pesi all'unità

Per eseguire test statistici e/o stimare i parametri di modelli rappresentativi della realtà indagata non si può pesare con $w_{p;ij}$, dato dall'equazione (12) perché esso altera la numerosità campionaria e, quindi, le probabilità di significatività relative alle ipotesi da sottoporre a verifica. In pratica, quindi, per rimediare a tali inconvenienti è utile «scalare» i pesi in modo che la loro somma sia uguale all'unità, anche se i totali non sono cosí riportati alla popolazione di riferimento (Verma, 1995). Per incorporare la struttura del campione nella determinazione degli stimatori e non alterare la numerosità

campionaria, si può utilizzare un insieme di pesi che, partendo da $w_{p;ij}$, mantengano inalterate le caratteristiche del campione, ossia soddisfacciano i seguenti due vincoli:

$$\text{(a)} \quad \sum_{i=1}^{I}\sum_{j=1}^{J} w^{*}_{p;ij} = 1 \qquad\qquad \text{(b)} \quad \sum_{i=1}^{I}\sum_{j=1}^{J} w^{*}_{p;ij} n_{ij} = n\,.$$

Per soddisfare entrambi i criteri si può utilizzare un peso dato dal rapporto tra i pesi «originari», $1/p^{*}_{p;ij}$, e un peso medio, $1/\bar{p}^{*}_{p;ij}$, in modo da soddisfare le condizioni (a) e (b). Si tratta di grandezze che figurano al denominatore sicché si può usare la media secondo il criterio del Chisini (1926), usando come aggregazione la funzione somma delle quantità inverse perché tutte positive (sono «pseudo-probabilità»). La media secondo Chisini di una variabile Y è quel valore intermedio $\bar{Y}$ compreso tra il minimo, $x_{(1)}$, e il massimo, $x_{(n)}$, di un insieme di osservazioni che, rispetto a una funzione sintetica delle osservazioni, ne lascia invariato il valore:

$$f(y_1, y_2, \Lambda\, , y_n) = f(\bar{Y}, \bar{Y}, \Lambda\, , \bar{Y}).$$

La definizione comporta la trasferibilità del carattere Y perché il valore $\bar{Y}$ uguaglia la funzione $f(\cdot)$ quando si sostituiscono le osservazioni con il valore costante $\bar{Y}$. Si richiede, pertanto, di specificare la $f(\cdot)$ in base alla natura del carattere (additiva, moltiplicativa, inversa, e cosí via) e alla sua trasferibilità (Piccolo, 1998, pp. 78-92). Nel caso in oggetto, si definisce la funzione $f(\cdot)$ come somma degli inversi dei valori osservati $f(y_1, y_2, \Lambda\, , y_n) = \sum_{i=1}^{n} \frac{1}{y_i}$ da cui si ottiene, adattando i simboli agli strati *ij*:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{n_{ij}} \frac{1}{p^{*}_{ij}} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{n_{ij}} \frac{1}{\bar{p}^{*}} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{n_{ij}}{\bar{p}^{*}} \quad \Leftrightarrow \quad \bar{p}^{*} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}}{\sum_{i=1}^{I}\sum_{j=1}^{J} \frac{n_{ij}}{p^{*}_{ij}}}$$

dove $\bar{p}^{*}$ è la media armonica delle probabilità di selezione per i vari strati *ij*. Il peso normalizzato a uno per ogni strato *ij* sarà dato dal rapporto tra i pesi effettivi finali $w_{p;ij}$ e il peso medio dato dall'inverso della media armonica, $1/\bar{p}^{*}$. Allora, il peso normalizzato a uno, $w^{*}_{p;ij}$, che rispetta entrambi i vincoli (a) e (b) diventa

$$w^{*}_{p;ij} = \frac{\bar{p}^{*}}{p^{*}_{ij}} = \frac{1}{p^{*}_{ij}} \; \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}}{\sum_{i=1}^{I}\sum_{j=1}^{J} \frac{n_{ij}}{p^{*}_{ij}}}\,. \qquad (13)$$

Si può mostrare che i pesi $w^{*}_{p;ij}$ sono dati dal rapporto tra i pesi degli strati rispetto alla popolazione totale di riferimento e i pesi degli strati nel campione rispetto alla dimensione totale del campione: $w^{*}_{p;ij} = W_{ij} / w_{p;ij} = (N_{ij}/N) : (n_{p;ij}/n)$; infatti,

$$w^{*}_{p;ij} = \frac{N_{ij}}{n_{p;ij}} \; \frac{n}{N} = \frac{1}{p^{*}_{ij}} \; \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} n_{p;ij}}{\sum_{i=1}^{I}\sum_{j=1}^{J} N_{ij} \frac{n_{p;ij}}{n_{p;ij}}} = \frac{1}{p^{*}_{ij}} \; \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} n_{p;ij}}{\sum_{i=1}^{I}\sum_{j=1}^{J} \frac{n_{p;ij}}{p^{*}_{ij}}}\,.$$

Si osserva che i pesi $w^{*}_{p;ij}$ alterano completamente la struttura delle dimensioni campionarie per strato rispetto al campione effettivo.

7. Determinazione empirica dei pesi per strato e errore relativo delle stime

Nella Tabella 3 si sono riportati i dati della popolazione $\wp$ per comparto e per classe di dimensione. Nella Tabella 4 si sono riportati il numero di imprese mancanti alla rilevazione, IMR, e gli errori relativi risultati nella stima degli addetti per strato ij, $r_{p;ij}$, per comparto e per classe di dimensione. Naturalmente, nelle classi di dimensione relative a imprese con 100 e piú addetti, nelle quali bisognava rilevare l'intera popolazione degli strati, l'errore relativo è stato calcolato ugualmente dove era possibile, ma sarebbe opportuno e raccomandabile impegnare il massimo degli sforzi per convincere le imprese a collaborare al fine di non inficiare o distorcere le stime complessive della popolazione di riferimento. Si può osservare che dove il numero di imprese da rilevare è stato raggiunto, l'errore relativo corrisponde al valore fissato; per esempio: nel settore (6) e classe di dimensione 6-9 addetti, l'errore relativo è pari al 15%; nel settore (5) e classe di dimensione 50-99 addetti, l'errore relativo è zero perché sono state rilevate tutte le imprese della popolazione $\wp$. Si può ancora osservare che l'errore relativo per il totale del settore meccanico, derivante dall'aggregazione dei comparti, si è ottenuto applicando l'equazione (3) ai dati del totale per classe di dimensione; pertanto, la dimensione del campione aumenta piú consistentemente dell'entità della popolazione $\wp$ e l'errore relativo risulta molto piú piccolo della precisione per strato perché l'aggregato presenta una perdita di informazione e, quindi, un aumento della precisione dei risultati del campione. Infine, si noti che l'aggregazione per classe di dimensione (totale di riga) richiede un procedimento differente da quello usato nell'aggregazione per comparto (totale di colonna) nel calcolo dell'errore relativo perché i diversi errori relativi per classi di dimensione devono essere riportati prima agli errori commessi nella stima del totale per strato; poi, l'errore deve essere sommato strato per strato e riportato al totale da stimare; mentre nel piano di campionamento l'errore relativo è stato mantenuto costante per comparto. Infatti, i valori dell'errore relativo risentono meno dell'aumento della numerosità del campione e rappresentano una specie di media degli errori relativi per classi di dimensione.

Nella Tabella 5 si sono riportati il numero di imprese non appartenenti alla popolazione di riferimento, $n_{\notin;ij}$, e il numero di imprese stimato o effettivo, N_E, per comparto e per classe di dimensione. Si può osservare che il numero di imprese, che non appartengono alla popolazione, non sembra del tutto casuale, ma potrebbe dipendere dal settore e dalla dimensione dell'impresa. Il campionamento indipendente, con precisione fissata per strato, consente di applicare agevolmente la correzione, ma rimane fortemente limitato dalla scarsa numerosità della popolazione $\wp$ e della dimensione del campione.

Tabella 3 — *Numero di imprese e deviazione standard per comparto e per classe di dimensione*

CA	A1-5		A6-9		A10-19		A20-49		A50-99		A100-249		A250-499		A>=500		Tot.	
Sett	NI	DS	NI	DS	NI	DS	NI	DS	NI	DS	NI	DS	NI	DS	NI	DS	NI	DS
S1	12	1,56	11	1,09	16	3,30	5	7,47	2	7,07	1	0	0	0	0	0	**47**	29,44
S2	292	1,46	82	1,13	86	2,86	22	8,54	2	16,97	1	0	2	8,48	0	0	**487**	18,41
S3	23	1,26	13	1,12	36	2,81	14	9	1	0	0	0	0	0	0	0	**87**	11,17
S4	50	1,30	24	1,13	20	3,33	12	9,90	2	23,33	0	0	0	0	0	0	**108**	13,37
S5	427	1,37	149	1,17	150	2,75	54	7,55	5	15,37	2	4,95	0	0	0	0	**787**	10,42
S6	79	1,28	28	1,20	30	2,92	10	6,86	2	33,23	1	0	0	0	0	0	**150**	14,30
S7	757	1,33	177	1,17	171	2,56	146	7,87	48	14,66	28	42,16	8	59,64	1	0	**1336**	57,67
S8	144	1,42	35	1,08	41	2,85	24	7,26	12	15,05	1	0	1	0	0	0	**258**	24,75
S9	374	1,22	45	1,16	57	2,90	43	8,65	9	12,49	4	29,57	3	105,63	1	0	**536**	40,23
S10	48	1,59	17	1,03	20	2,36	6	4,45	5	6,77	5	41,61	2	6,36	1	0	**104**	184,55
Tot.	**2206**	1,36	**581**	1,15	**627**	2,77	**336**	8,01	**88**	14,44	**43**	39,64	**16**	69,32	**3**	688,47	**3900**	49,24

Legenda	**S1**	(1) DJ27 produzione di metalli e loro leghe
	S2	(2) DJ28.1 +28.2 +28.3 carpenteria metallica
	S3	(3) DJ28.40 fucinatura, imbutitura, stampaggio, e profilatura metalli
	S4	(4) DJ28.51 trattamento e rivestimento dei metalli
	S5	(5) DJ28.52 lavori di meccanica generale per conto terzi
	S6	(6) DJ28.6+28.7+Resto DJ28 prodotti finiti in metallo
	S7	(7) DK29 fabbricazione di macchine e apparecchi meccanici
	S8	(8) DL31 fabbricazione di macchine e apparecchi elettrici n.c.a.
	S9	(9) DL30+32+33 fabbricazione di macchine elettriche e apparecchi elettrici e ottici
	S10	(10) DM34+35 fabbricazione di mezzi di trasporto.

Tabella 4 — *Numero di imprese mancanti alla rilevazione, IMR, e errori relativi nella stima degli addetti,* $r_{p;ij}$ *, per comparto e per classe di dimensione*

CA	A1-5		A6-9		A10-19		A20-49		A50-99		A100-249		A250-499		A>=500		Tot.	
Sett	IMR	$r_{p;ij}$	IMR	$r_{p;ij}$	IMR	$r_{p;ij}$	IMR	$r_{p;ij}$	IMR	$r_{p;ij}$	IMR	$r_{p;ij}$	IMR	$r_{p;ij}$	IMR	$r_{p;ij}$	IMR	$r_{p;ij}$
S1			2	0,200	3	0,231	4	0	1	0,164	1	0	0	0	0		**11**	**0,111**
S2			1	0,180	4	0,242	7	0,187	1	0,381	1	0	2	0	0		**16**	**0,142**
S3			1	0,191	1	0,171	1	0,115	1	0,000	0	0	0	0	0		**4**	**0,133**
S4			1	0,166	3	0,206	5	0,245	1	0,468	0	0	0	0	0		**10**	**0,225**
S5			2	0,224	4	0,226	10	0,153	4	0	1	0,063	0	0	0		**21**	**0,154**
S6			0	0,150	5	0,406	4	0,205	1	0,623	1	0	0	0	0		**11**	**0,245**
S7			1	0,184	3	0,214	11	0,141	4	0,128	17	0,132	4	0,120	0		**40**	**0,123**
S8			3	0,295	2	0,176	7	0,192	4	0,209	0	0	0	0	0		**16**	**0,156**
S9			1	0,178	4	0,242	12	0,190	3	0,171	4	0	1	0,224	1		**26**	**0,142**
S10			2	0,288	3	0,309	2	0,182	1	0,129	3	0,291	1	0,027	1		**13**	**0,104**
Tot.			**14**	**0,063**	**32**	**0,070**	**63**	**0,063**	**21**	**0,080**	**28**	**0,113**	**8**	**0,102**	**2**		**168**	**0,065**

Tabella 5 — *Numero di imprese non appartenenti alla popolazione di riferimento, $n_{\notin;ij}$, e numero di imprese stimato o effettivo, N_E, per comparto e per classe di dimensione*

CA	A1-5		A6-9		A10-19		A20-49		A50-99		A100-249		A250-499		A>=500	
Sett	$n_{\notin;ij}$	N_E	$n_{\notin;ij}$	N_E	$n_{\notin;ij}$	N_E	$n_{\notin;ij}$	N_E	$n_{\notin;ij}$	N_E	$n_{\notin;ij}$	N_E	$n_{\notin;ij}$	N_E	$n_{\notin;ij}$	N_E
S1				10		16,0		5		2		1		0		0
S2				82		86,0		22		2		1		2		0
S3				13		36,0		14		1		0		0		0
S4				24		20,0		12		2		0		0		0
S5				149		150,0		54	1	4		2		0		0
S6				28		30,0		10		2		1		0		0
S7				177	1	142,5	1	140		48		28		8		1
S8				35	1	35,1		24		12		1		1		0
S9				45		57,0	5	32	2	6	2	2		3	1	0
S10				17	1	15,0		6		5		5		2		1
Tot.				**580**	**3**	**588**	**6**	**319**	**3**	**84**	**2**	**41**	**0**	**16**	**1**	**2**

Nella Tabella 6 si sono riportati il numero di imprese effettivamente rilevate, IER, e i pesi (per strato e arrotondati per comodità a due cifre decimali), $w_{p;ij}$, per comparto e per classe di dimensione. Si deve osservare che l'uso dei pesi per stimare le quantità riferite alla popolazione ha come conseguenza che i totali per cella delle elaborazioni (nelle tabelle di contingenza e in altri tipi di distribuzione) sono numeri decimali arrotondati a interi secondo la convenzione usuale, sicché *i totali marginali non coincidono sempre con la somma dei contenuti delle celle*, ma può presentare differenze di qualche unità, che variano secondo il numero di celle.

Tabella 6 — *Numero di imprese effettivamente rilevate, IER, e pesi (per strato), $w_{p;ij}$, per comparto e per classe di dimensione*

CA	A1-5		A6-9		A10-19		A20-49		A50-99		A100-249		A250-499		A>=500	
Sett	IER	$w_{p;ij}$	IER	$w_{p;ij}$	IER	$w_{p;ij}$	IER	$w_{p;ij}$	IER	$w_{p;ij}$	IER	$w_{p;ij}$	IER	$w_{p;ij}$	IER	$w_{p;ij}$
S1			1	10,0	3	5,33	0	0,00	1	2,00	0	0	0	0	0	0
S2			3	27,3	3	28,67	7	3,14	1	2,00	0	0	0	0	0	0
S3			2	6,5	5	7,20	9	1,56	0	0	0	0	0	0	0	0
S4			3	8,0	4	5,00	4	3,00	1	2,00	0	0	0	0	0	0
S5			2	74,5	3	50,00	11	4,91	0	0	1	2,00	0	0	0	0
S6			4	7,0	1	30,00	4	2,50	1	2,00	0	0	0	0	0	0
S7			3	59,0	3	47,50	14	10,01	9	5,33	11	2,55	4	2,00	1	1,00
S8			1	35,0	5	7,03	5	4,80	3	4,00	1	1,00	1	1,00	0	0
S9			3	15,0	3	19,00	8	4,03	3	2,00	0	0	2	1,50	0	0
S10			1	17,0	1	15,00	2	3,00	2	2,50	2	2,50	1	2,00	0	0
Tot.			**23**		**31**		**64**		**21**		**15**		**8**		**1**	

Nella Tabella 7 sono riportati gli errori relativi per comparto "a posteriori", dopo la rilevazione per tenere conto delle imprese che non hanno partecipato o non stono state rintracciate. Nella tabella 8 sono riportati, invece, le imprese effettivamente rilevate (IER), i pesi, e gli errori relativi "a posteriori" di due classi di dimensioni aggregate per fornire una idea della grandezza della precisione attesa dal campione. Questi ultimi sono stati calcolati sia considerando la classe aggregata come fosse un unico dominio di studio, ignorando che all'origine le imprese sono state selezionate considerando una determinata struttura degli strati e non ignorando tale struttura.

Tabella 7 — *Errori relativi per comparto e per classe di dimensione*

CA	**A1-5**	**A6-9**	**A10-19**	**A20-49**	**A50-99**	**A100-249**	**A250-499**	**A>=500**	**Totale**	**Err.Rel.**
Sett	Err.Rel.	Err.Rel.	Err.Rel.	Err.Rel.	Err.Rel.	Err.Rel.	Err.Rel.	Err.Rel.		
S1		0,301	0,231		0,164	0	0	0	5	0,119
S2		0,180	0,242	0,187	0,381	0	0	0	14	0,142
S3		0,191	0,171	0,115		0	0	0	16	0,133
S4		0,166	0,206	0,245	0,468	0	0	0	12	0,225
S5		0,224	0,226	0,153	0,000	0,063	0	0	17	0,154
S6		0,150	0,406	0,205	0,623	0	0	0	10	0,245
S7		0,184	0,214	0,141	0,128	0,132	0,120	0	45	0,123
S8		0,295	0,176	0,192	0,209	0	0	0	16	0,156
S9		0,178	0,242	0,190	0,171	0	0,224	0	19	0,142
S10		0,288	0,309	0,182	0,129	0,291	0,027	0	9	0,104
Totale		0,065	0,070	0,063	0,080	0,113	0,102	0	163	0,066

Tabella 8 — *Numero di imprese effettivamente rilevate, IER, pesi (per strato),* $w_{p;ij}$ *, e errori relativi per comparto e per alcune classi di dimensione*

CA	**A>= 100**				**6<=A<=99**			
Sett	IER	PESI	Err.Rel.[(a)]	Err.Rel.[(b)]	IER	PESI	Err.Rel.[(a)]	Err.Rel.[(b)]
S1	0	0,0000	**1,000**	0,000	5	6,6000	**0,664**	0,162
S2	0	0,0000	**1,000**	0,000	14	13,7143	0,360	0,220
S3	0	0,0000	0,000	0,000	16	4,0000	0,278	0,141
S4	0	0,0000	0,000	0,000	12	4,8333	0,450	0,251
S5	1	2,0000	0,063	0,063	16	22,2969	0,359	0,188
S6	0	0,0000	**1,000**	0,000	10	7,0000	**0,518**	0,336
S7	16	2,3125	0,411	0,103	29	17,5055	0,323	0,157
S8	2	1,0000	0,000	0,000	14	7,5816	0,477	0,205
S9	2	2,5000	0,782	0,117	17	8,2500	0,382	0,198
S10	3	2,6667	**1,292**	0,075	6	7,1667	**0,646**	0,222
Totale	24	2,4583	0,368	0,081	139	11,3007	0,143	0,069

[(a)]L'errore relativo è calcolato considerando la classe come un solo dominio di studio, senza piano di campionamento.

[(b)]L'errore relativo è calcolato considerando la classe composta di domini di studio, con piano di campionamento.

6. Conclusioni

I confronti eseguiti sulle distribuzioni degli addetti mostrano che il lavoro di ricostruzione dei dati, che ha prodotto l'«archivio UNI-MEC» costruito *ad hoc*, sembra garantire una accettabile omogeneità delle distribuzioni degli addetti dell'«archivio UNI-MEC» con quelle derivate dall'«archivio ISTAT» e dall'«archivio AIDI». Il lavoro svolto si può ritenere soddisfacente perché le differenze osservate sono suscettibili di essere "spiegate" sia con le procedure usate, sia per le differenze esistenti negli archivi iniziali che rispecchiano le peculiarità della loro costruzione e finalità. Si è utilizzato, pertanto, l'«archivio UNI-MEC» per costruire il campione di imprese da intervistare.

A questo stadio dell'indagine si può ancora sostenere, per quanto riguarda i pesi e le precisioni delle stime (seconda parte del lavoro), che occorre concentrare tutti gli sforzi per rilevare le imprese mancanti nei diversi strati; soprattutto, si deve mirare l'azione al fine di rilevare tutte le imprese nelle classi designate a contenere tutte le imprese della popolazione. Per quanto concerne la sostituzione delle imprese che si rifiutano di collaborare, si è operato con l'accortezza di non ricorrere a tale pratica perché se da un lato si migliora la precisione delle stime, dall'altro si consegue un aumento della distorsione, perché le imprese piú disponibili a collaborare potrebbero avere caratteristiche distintive che inficiano o distorcono le stime dei parametri della popolazione.

Nelle elaborazioni dei dati, che non coinvolgono verifiche di ipotesi, si possono usare i pesi $w_{p;ij}$ che riportano le stime della popolazione obiettivo perché sono piú immediati e leggibili. Nella verifica di ipotesi si dovrebbero usare i pesi $w^{*}_{p;ij}$ anche se alterano assai la struttura del campione effettivo per strato, ma è un modo per tenere conto delle differenze introdotte dal piano di campionamento nella diversa numerosità per strato.

Bibliografia

Abbate C., Baldassarini A. (1994). Contenuto informativo degli archivi Inps e confronto con altre fonti sul mercato del lavro, *Economia e Lavoro*, XXVIII (2), pp. 115-133.

Bethlehem J.G., Keller W.J. (1987). Linear weighting of sample survey data, *Journal of Official Statistics*, **3**, pp. 141−153.

Birnbaum Z.W., Hall R.A. (1960). Small sample distributions for multiple-sample statistics of the Smirnov type, *The Annals of Mathematical Statistics*, 31, pp.710-720.

Brusco S., Giovannetti E., Malagoli W. (1979). *La relazione tra dimensione e saggio di sviluppo nelle imprese industriali: una ricerca empirica*, Studi e Ricerche dell'Istituto Economico, n. 5, Modena.

Chisini O. (1929). Sul concetto di media, *Periodico di matematiche*, **9** (4).

Cicchitelli G., Herzel A., Montanari G.E. (1992). *Il campionamento statistico*, il Mulino, Bologna.

Cochran W. G. (1977). *Sampling Techniques*, John Wiley & Sons, 3.rd ed., New York.

Feller W. (1968). *An Introduction to Probability Theory and Its Application*, vol. 1, 3rd ed., John Wiley & Sons, New York.

Horvitz D.G., Thompson D.J. (1952). A Generalization of Sampling Without Replacement from a finite Universe, *Journal of the American Statistical Association*, **47**, pp. 663–685.

ISTAT (1986). Annuario di statistiche industriali, **26**, ISTAT, Roma.

Jalla E. (1981). L'industria manifatturiera italiana: avvio a un nuovo criterio di analisi statistica, *L'industria*, **II** (1), 88−116.

Kiefer J. (1959). K-sample analogues of the Kolmogorv-Smirnov and Cramér-von Mises tests, *The Annals of Mathematical Statistics*, 30, 420-447.

Kish L. (1990). Weighting: why, when, and how, *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 121−130.

Kish L. (1992). Weighting for unequal P_i, *Journal of Official Statistics*, **8**, 2, pp. 121−130.

Landenna G., Marasini D. (1990). *Metodi statistici non parametrici*, Bologna: il Mulino.

Martini M. (1990). I dati amministrativi come fonte di informazione statistica sulle imprese, *Economia e Lavoro*, XXIV, pp. 45-58.

Marzi G. (1990). Il problema della dimensione d'impresa: una nuova definizione, *L'industria*, **XI** (2), 317−328.

Miller R.G. Jr. (1966). *Simultaneous Statistical Inference*, McGraw Hill, New York.

Potter F.J. (1990). A study of procedures to identify and trim extreme sampling weights, *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 121−130.

Rubin D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, New York.

Russo M., Giardino R. (2000). Struttura e cambiamento nelle relazioni tra le imprese meccaniche. **I**. La popolazione di imprese meccaniche della provincia di Modena: procedure impiegate per integrare le informazioni amministrative del Registro Imprese e dell'Inps, *Materiali di discussione*, Dipartimento di Economia Politica, Università degli Studi di Modena e Reggio Emilia, Modena, pp. 1−32.

Piccolo D. (1998). *Statistica*, il Mulino, Bologna.

Verma V. (1995). *Weighting for Wave 1*, Working Group "European Community Household Panel", Doc. PAN 36/95, Statistical Office of the European Communities, Luxembourg.

Siegel S., Castellan N.J. Jr. (1988). *Nonparametric Statistics for the Behavioral Science*, McGraw-Hill, New York.