



ISTITUTO DI STUDI E ANALISI ECONOMICA

**L'inchiesta ISAE sugli investimenti
delle imprese manifatturiere ed estrattive:
aspetti metodologici e risultati**

di

Tatiana Cesaroni

ISAE, Piazza dell'Indipendenza, 4, 00185 Roma
Università degli Studi di Roma "La Sapienza"

Marco Malgarini

ISAE, Piazza dell'Indipendenza, 4, 00185 Roma

Gaia Rocchetti

ISAE, Piazza dell'Indipendenza, 4, 00185 Roma e
Scuola Superiore di Studi Universitari e di Perfezionamento Sant'Anna, Pisa

Working paper n. 50
Aprile 2005

La serie “Documenti di Lavoro” dell’Istituto di Studi e Analisi Economica ospita i risultati preliminari di ricerche predisposte all’interno dell’ISAE: La diffusione delle ricerche è autorizzata previo il parere favorevole di un anonimo esperto della materia che qui si ringrazia. Le opinioni espresse nei “Documenti di Lavoro” riflettono esclusivamente il pensiero degli autori e non impegnano la responsabilità dell’Ente.

La serie è destinata agli esperti e agli operatori di politica economica, al fine di formulare proposte e suscitare suggerimenti o critiche.

The Series “*Documenti di Lavoro*” of the *Istituto di Studi e Analisi Economica* – Institute for Studies and Economic Analyses (ISAE) hosts the preliminary results of the research projects carried out within ISAE. The diffusion of the papers is subject to the favourable opinion of an anonymous referee, whom we would like to thank. The opinions expressed are merely the Authors’ own and in no way involve the ISAE responsibility.

The series is meant for experts and policy-makers with the aim of submitting proposals and raising suggestions and criticism.

Stampato presso la sede dell’Istituto

ISAE - Piazza dell’Indipendenza, 4 – 00185 Roma.

Tel. +39-06444821; www.isae.it

SINTESI

L'ISAE conduce dagli anni '80, nell'ambito del programma armonizzato della Commissione Europea, un'inchiesta sull'attività d'investimento delle imprese manifatturiere ed estrattive. L'indagine, effettuata sullo stesso campione utilizzato mensilmente dall'Istituto per l'usuale inchiesta sull'andamento congiunturale del settore, fornisce informazioni sia di carattere qualitativo, concernenti la destinazione della spesa per investimenti e le motivazioni sottostanti le decisioni delle imprese, sia di carattere quantitativo, sull'effettivo ammontare degli investimenti realizzati. La raccolta di informazioni di tipo quantitativo, da un punto di vista metodologico, impone una maggiore cautela rispetto alle inchieste puramente qualitative, sia nella fase di controllo della qualità, sia in quella di elaborazione dei dati. Nel passato, nella fase di controllo dei microdati ci si limitava all'utilizzo di procedure di tipo deterministico; in questo lavoro, vengono introdotte anche tecniche di controllo di tipo statistico, per il trattamento, rispettivamente, dei dati mancanti e di quelli anomali. Inoltre, vengono riviste rispetto al passato le procedure di aggregazione dei dati elementari. La presentazione della nuova banca dati ISAEINV, che raccoglie i principali risultati dell'inchiesta sugli investimenti delle imprese manifatturiere ed estrattive per gli anni 2000-2004, conclude il lavoro.

Parole chiave: Inchieste congiunturali; Investimenti; Dati mancanti; Dati anomali; Aggregazione.

Classificazione JEL: C42, C81, E22

SOMMARIO

L'inchiesta ISAE sugli investimenti delle imprese manifatturiere ed estrattive fornisce informazioni di tipo qualitativo riguardanti la destinazione della spesa effettuata dalle imprese, le motivazioni sottostanti eventuali revisioni in corso d'anno dei programmi di spesa e i fattori che influenzano le decisioni di spesa. Alle imprese interpellate viene inoltre chiesto di indicare, in termini quantitativi, l'ammontare della spesa effettuata nell'anno precedente e nell'anno in corso, e quello degli investimenti programmati per l'anno successivo. La presenza di informazioni quantitative, accanto a quelle più usuali di tipo qualitativo, implica la necessità di un più severo controllo di qualità/coerenza delle risposte e di tecniche di aggregazione più sofisticate rispetto a quelle abitualmente utilizzate nelle inchieste esclusivamente qualitative. Nel passato, il controllo di qualità/coerenza dei dati d'impresa effettuato dall'ISAE si basava esclusivamente su procedure deterministiche, volte ad accertare la presenza di dati mancanti e anomali; si procedeva quindi all'aggregazione dei dati d'impresa ricavando la stima del totale degli investimenti per l'intera industria come sommatoria semplice dei dati elementari; per i dati qualitativi, la metodologia d'aggregazione seguita era basata su un sistema di doppia ponderazione analogo a quello seguito per l'inchiesta mensile sull'industria manifatturiera.

Il lavoro confronta, in primo luogo, il campione effettivo su cui si sono basate le inchieste ISAE negli anni 2000-2004 con il campione teorico e con l'Universo di riferimento per l'indagine; vengono quindi proposte alcune modifiche relative alle tecniche utilizzate per il controllo di qualità dei risultati e per la loro aggregazione. Più nel dettaglio, per quanto riguarda il trattamento dei dati anomali, nel lavoro viene proposto di utilizzare tecniche di stime robuste per lo stimatore del dato quantitativo riguardante l'ammontare degli investimenti effettivamente realizzati dalle imprese; per quanto riguarda invece il trattamento dei dati mancanti, nel lavoro vengono proposte tecniche di imputazione di medie condizionate, basate su un semplice modello di regressione che usa le informazioni contenute nei questionari compilati dalle imprese durante l'indagine. Infine, si sono aggiornate anche le procedure di aggregazione dei dati qualitativi, sulla base della considerazione che nella maggior parte dei casi il peso da utilizzare dovrebbe essere costituito dalle informazioni fornite dalle stesse imprese circa l'ammontare degli investimenti effettivamente realizzati o programmati nel periodo di riferimento.

ABSTRACT

The Joint harmonized EU Investment survey, carried out for Italy by the Institute for Studies and Economic Analysis (ISAE), provides quantitative information about investment plans and the structure of investment and qualitative data on the factors influencing firm' behavior. As shown in the recent OECD' Business Tendency Surveys Handbook (OECD, 2003), in dealing with business surveys a special attention should be devoted to the control for the reliability of the results and to the appropriate use of weighting in processing survey' data. From a methodological standpoint, this is particularly important for surveys providing quantitative as well as purely qualitative data, as is the case for the Investment survey. In this respect, the paper proposes deterministic and statistically-based methods for the treatment of missing data and outliers, focussing especially on the quantitative question on investment plans. After a careful deterministic control, based on the possibility of re-interviewing non-responding firms, a fairly simple econometric model is used to estimate "residual" missing data, based on information on sector, geographical partition and number of employee of the firm, and its investment plans for the previous years (when available). The robustness of sample estimates in case of extreme observation in the tails of the distribution is then tested with the use of the winsorized mean. In processing the results, estimate of investment plans at industry level is then derived using sample weights, based, for each stratum, on the ratio between the Universe Population and that of the sample. Aggregation of qualitative data concerning investment structure and factors influencing firm' behaviour is then based on appropriate, variable-specific, size weights, i.e. the investment plan of the firm itself. This appears to be a fairly more appropriate choice with respect to the more usual value added-based weight. A presentation of the resulting data-bank, together with some consideration on possible future research, concludes the work.

Key Words: Business surveys; Investment; Missing data; Outliers; Aggregation.

JEL Classification: C42; C81; E22

INDICE

1	INTRODUZIONE	Pag.	7
2	IL QUESTIONARIO	"	8
3	UNIVERSO DI RIFERIMENTO, CAMPIONE TEORICO E CAMPIONE EFFETTIVO	"	11
	3.1 Universo di riferimento e campione teorico	"	11
	3.2 Il campione effettivo: le inchieste dall'aprile 2000 all'aprile 2004		15
4	I CONTROLLI DI QUALITÀ	"	18
	4.1 I dati anomali	"	18
	4.2 I dati mancanti	"	27
5	LE PROCEDURE DI AGGREGAZIONE	"	37
	5.1 L'aggregazione dei dati quantitativi	"	37
	5.2 L'aggregazione dei dati qualitativi	"	40
6	CONSIDERAZIONI CONCLUSIVE	"	46
	APPENDICE:		
	La Banca dati ISAEINV (novembre 2000 -novembre 2004)		48
	BIBLIOGRAFIA	"	53

1 INTRODUZIONE¹

Dagli anni ottanta l'ISAE conduce ogni sei mesi, nell'ambito del progetto armonizzato della Commissione europea, un'inchiesta sugli investimenti effettuati dalle imprese manifatturiere ed estrattive. Il campione su cui viene condotta l'inchiesta è lo stesso dell'indagine mensile, stratificato secondo la struttura dimensionale e la collocazione settoriale e spaziale delle imprese. L'indagine è condotta per via postale (anziché telefonica, come nel caso dell'indagine mensile): ciò, assieme alla relativa complessità del questionario, implica un tasso di risposta notevolmente inferiore rispetto a quello dell'inchiesta mensile.

L'indagine fornisce informazioni di tipo qualitativo riguardanti la destinazione della spesa effettuata dalle imprese (per sostituzione/rinnovo delle apparecchiature, per razionalizzazione o ampliamento della capacità produttiva, per la tutela ambientale), le motivazioni sottostanti eventuali revisioni in corso d'anno dei programmi di spesa (per vincoli di finanziamento, variazioni della domanda, ecc.) e i fattori che influenzano le decisioni di spesa. Alle imprese interpellate viene inoltre chiesto di indicare, in termini quantitativi, l'ammontare della spesa effettuata nell'anno precedente e nell'anno in corso, e quello degli investimenti programmati per l'anno successivo; a partire da queste informazioni, si ricava una stima del tasso di crescita degli investimenti, regolarmente pubblicata dall'ISAE e considerata un utile supporto all'analisi e alla previsione dell'andamento del processo di accumulazione.

La presenza di informazioni quantitative, accanto a quelle più usuali di tipo qualitativo, implica la necessità di un più severo controllo di qualità/coerenza delle risposte e di tecniche di aggregazione più sofisticate rispetto a quelle abitualmente utilizzate nelle inchieste esclusivamente qualitative². Nel passato, il controllo di qualità/coerenza dei dati d'impresa si basava esclusivamente su procedure deterministiche, volte ad accertare la presenza di dati mancanti e anomali, ovvero, rispettivamente, di questionari parzialmente incompleti e di

1 Una versione in lingua inglese del presente lavoro è stata presentata alla 27^a Conferenza CIRET, Varsavia, Settembre 2004. Pur essendo il lavoro frutto di una riflessione comune, T. Cesaroni ha curato i par. 4.1 e 5, M. Malgarini i par. 1, 2 e 6, Gaia Rocchetti i par. 3 e 4.2; gli autori desiderano ringraziare un anonimo *referee*, nonché Bianca Maria Martelli e tutti i partecipanti ad un seminario ISAE tenuto nel giugno 2003, per gli utili commenti e suggerimenti. Ogni errore resta come sempre di loro esclusiva responsabilità.

2 Nell'inchiesta sulle imprese manifatturiere ed estrattive, una volta effettuato un controllo di coerenza di tipo deterministico sulle risposte fornite dalle singole imprese, si procede, per ogni singola domanda qualitativa, alla elaborazione delle frequenze percentuali attraverso un sistema di "doppia ponderazione" (cfr. Malgarini, Margani e Martelli, 2005).

risposte apertamente non coerenti, rispetto ad altre contenute nello stesso questionario, o rispetto alle indicazioni fornite da una specifica impresa nelle precedenti rilevazioni. Si procedeva quindi all'aggregazione dei microdati: per i dati quantitativi, il totale degli investimenti per l'intera industria era ottenuto come sommatoria semplice dei dati d'impresa; per i dati qualitativi, la metodologia seguita era basata su un sistema di doppia ponderazione, analogo a quello seguito per l'inchiesta mensile.

In questo lavoro, si propone di modificare le tecniche utilizzate, sia per quanto riguarda il controllo di qualità, sia per le procedure di aggregazione. I risultati ottenuti sono quindi raccolti in una nuova banca dati, denominata ISAEINV, relativa alle inchieste effettuate a partire dalla primavera del 2000. Nel seguito, il paragrafo 2 descrive il questionario ed il processo di rilevazione dei dati, il paragrafo 3 illustra l'Universo di riferimento ed il campione teorico estratto, in comparazione con il campione effettivo di imprese che hanno risposto all'indagine sugli investimenti nel periodo aprile 2000-aprile 2004. Il paragrafo 4 descrive i controlli di qualità adottati dall'ISAE per i dati anomali (par. 4.1) e quelli mancanti (par. 4.2), illustrando sia le tecniche deterministiche adottate storicamente dall'Istituto, sia le nuove procedure statistiche. Il paragrafo 5 presenta quindi le diverse procedure di aggregazione utilizzate per i dati quantitativi (par. 5.1) e per quelli qualitativi (par. 5.2); alcune considerazioni sui risultati ottenuti e sulle future prospettive di ricerca concludono il lavoro.

2 IL QUESTIONARIO

L'indagine ISAE sugli investimenti delle imprese manifatturiere ed estrattive ha cadenza bi-annuale ed è condotta, per via postale, nell'ambito del progetto armonizzato dell'Unione Europea³: i questionari sono spediti, rispettivamente, a metà febbraio e a metà settembre, e pervengono all'ISAE per fine marzo e fine ottobre; i risultati vengono pubblicati ad inizio maggio e inizio dicembre. I questionari sono indirizzati a tutte le 4.000 imprese del campione ISAE utilizzato per l'indagine mensile sulle imprese manifatturiere ed estrattive; rispondono però esclusivamente le imprese che hanno effettivamente svolto, almeno in uno dei 4 anni oggetto d'indagine, attività di investimento.

Il questionario originale della Commissione Europea mira a raccogliere informazioni sul tasso di crescita degli investimenti, sui piani di investimento

3 Si veda Commissione Europea (2002).

delle imprese manifatturiere, sulla destinazione degli investimenti effettuati e sui fattori influenzanti le decisioni di spesa. In particolare, per quanto riguarda i piani di investimento, la Commissione richiede informazioni circa la variazione degli investimenti da un anno all'altro, espressa a prezzi correnti. Il questionario proposto su base semestrale dall'ISAE differisce da quello della Commissione, per l'introduzione di alcune domande aggiuntive e perché richiede alle imprese di indicare l'esatto ammontare degli investimenti effettuati in anni contigui; l'informazione relativa al tasso di variazione (quella cioè inviata alla Commissione) viene ricavata direttamente dalle indicazioni circa gli investimenti effettuati dalle imprese.

Nel dettaglio, i questionari somministrati dall'ISAE differiscono tra l'inchiesta "primaverile" e quella "autunnale" (cfr. Tab. 1).

Tab. 1 Indagine semestrale ISAE sugli investimenti

	Domande presenti nel questionario di aprile		Domande presenti nel questionario di ottobre	
		Periodo di riferimento		Periodo di riferimento
Numero di addetti	SI	t	SI	t
<i>Investimenti effettuati o programmati</i>	SI	t-2, t-1, t	SI	t-1,t, t+1
Variazione dei prezzi	SI	t-1/t-2, t/t-1	SI	t/t-1, t+1/t
<i>Destinazione della spesa</i>	SI	t-1 e t	SI	t e t+1
Utilizzazione degli impianti	SI	t-1	NO	
Variazione capacità produttiva	SI	t-1/t-2	NO	
Tendenza investimenti	SI	t+1	SI	t+2
Leasing	SI	t-1	NO	
Modifiche ai piani	NO		SI	t
<i>Fattori influenzanti</i>	NO		SI	t e t+1
Ripartizione territoriale investimenti	NO		SI	t e t+1
Ripartizione per categorie di beni	NO		SI	t e t+1
Incorporazione di strutture produttive preesistenti	NO		SI	t e t+1

NB: In corsivo le domande comprese anche nel questionario della Commissione Europea.

E' possibile individuare all'interno dei questionari 4 sezioni principali⁴. nella prima, comune ad entrambe le rilevazioni, è compresa una domanda di carattere strutturale sul numero degli addetti dell'impresa interpellata, oltre ad alcune informazioni sulle caratteristiche anagrafiche (ragione sociale, settore, regione, provincia di appartenenza).

La seconda sezione riporta le domande di tipo quantitativo, relative all'ammontare degli investimenti effettuati dall'impresa e alla variazione dei prezzi dei beni di investimento. Nel questionario di marzo, viene richiesto di indicare l'ammontare di investimenti effettuati nei due anni precedenti la rilevazione (ad esempio, nella rilevazione del marzo 2004 viene richiesto di indicare l'ammontare degli investimenti effettuati nel 2002 e nel 2003) e quelli previsti per l'anno in corso; nel questionario di ottobre, le imprese devono invece indicare l'ammontare degli investimenti effettuati nell'anno precedente, quelli dell'anno in corso e quelli previsti per l'anno successivo. Tutte le indicazioni sono, a partire dall'inchiesta del marzo 2002, espresse in migliaia di euro (precedentemente, in milioni di lire). La domanda sulla variazione del prezzo dei beni di investimento si riferisce all'anno precedente e all'anno in corso (nel questionario di marzo) e all'anno in corso e all'anno successivo (nel questionario di ottobre).

La terza parte del questionario comprende una domanda (di tipo qualitativo) sulla tendenza degli investimenti e due domande concernenti la destinazione della spesa nell'anno precedente e nell'anno in corso (nel questionario di marzo) e nell'anno in corso e nell'anno successivo (nel questionario di ottobre): in particolare, la spesa per investimenti può essere destinata alla sostituzione/rinnovo delle attrezzature fuori uso o obsolete, alla razionalizzazione (a sua volta suddivisa in tre sottotipologie), all'ampliamento della capacità produttiva e alle spese per la tutela dell'ambiente. Alle imprese viene richiesto di indicare la quota percentuale di investimenti destinata alle principali macrotipologie di spesa e, all'interno di queste, se la spesa è rivolta all'una o all'altra sottotipologia, o a più di una (con possibilità, quindi, in questo caso di fornire una risposta multipla). In questa sessione, viene anche richiesto all'impresa di indicare se la spesa per investimenti prevista per l'anno successivo (nel questionario di marzo) o per il secondo anno successivo (nel

4 Nel questionario di ottobre, si aggiunge una quinta sezione, con domande riguardanti la ripartizione territoriale della spesa effettuata o da effettuare nell'anno corrente o nell'anno successivo, la ripartizione per categoria della spesa e l'incorporazione nell'anno in corso o nell'anno successivo di strutture produttive esistenti da parte dell'impresa interpellata (con un'indicazione anche, in caso di risposta positiva, dell'incidenza delle spese di incorporazione sul totale delle spese per investimenti effettuate).

questionario autunnale, cioè, ad esempio ad ottobre 2004 si chiede un'indicazione sul 2006) sarà superiore, uguale od inferiore rispetto all'anno in corso.

La quarta sezione si differenzia tra il questionario inviato a marzo e quello di ottobre; i questionari "primaverili" sono composti di ulteriori 3 domande, riguardanti il grado di utilizzo degli impianti, la variazione della capacità produttiva e il ricorso al leasing nell'anno appena trascorso. Nel questionario "autunnale", invece, viene richiesta una valutazione sulle modifiche apportate in corso d'anno ai piani d'investimento: in particolare, viene richiesto di indicare se gli investimenti sono stati rivisti al rialzo o al ribasso, e, quindi, se tali revisioni sono state determinate da motivi legati alle possibilità di finanziamento, a modifiche nella struttura della domanda, a misure di politica economica, a variazioni nei prezzi dei beni di investimento, a motivi amministrativi interni o, infine, ad altri motivi non altrimenti specificati. Viene infine richiesta una valutazione su alcuni fattori influenzanti la spesa per investimenti, riferita all'anno in corso e, in previsione all'anno successivo.

3 UNIVERSO DI RIFERIMENTO, CAMPIONE TEORICO E CAMPIONE EFFETTIVO

3.1 Universo di riferimento e campione teorico

L'Universo statistico di riferimento è costituito, dal 1999, dalle imprese con 10 o più addetti operanti nelle divisioni ATECO C e D, come risultanti dall'archivio ISTAT denominato ASIA, relativo all'anno 1999. La disponibilità di un archivio particolarmente ricco e aggiornato contenente dati a livello d'impresa ha determinato la scelta dell'impresa stessa, anziché dell'Unità Locale, come Unità Statistica di riferimento. Va rilevato peraltro che, da un punto di vista strettamente teorico, sarebbe stato preferibile scegliere l'Unità Locale, così da poter raccogliere informazioni il più possibile attendibili anche a livello territoriale; si è preferito tuttavia privilegiare la fonte d'informazione più aggiornata ed accurata a livello nazionale, considerando peraltro che spesso, nel passato, a livello d'Unità Locale le risposte erano in ogni caso fornite facendo riferimento alle informazioni provenienti dalle "sedi centrali" delle imprese interessate. Nel selezionare il campione si è optato per un'allocatione ottimale delle unità di rilevazione all'interno degli strati (*optimum allocation to*

strata); in pratica, tale procedura consiste nel sovracampionare gli strati che sono caratterizzati da una maggiore variabilità interna, e di sottocampionare invece quelli in cui la variabilità è più bassa (su tutti questi aspetti, per una descrizione più dettagliata si vedano Malgarini, Margani e Martelli, 2005 e Martelli, Rocchetti, 2004). Il campione è in particolare stratificato secondo tre criteri: geografico (19 regioni, aggregando Piemonte e Val d'Aosta), dimensionale (imprese piccole – da 10 a 49 addetti; medie – da 50 a 249 addetti; grandi – 250 addetti e oltre) e settoriale (22 settori di attività economica, i cosiddetti “Rami ISCO”). La classificazione ISCO-ISAE è basata sulla NACE 1970; le differenze più rilevanti con l'attuale classificazione delle attività economiche adottata a livello europeo (l'ATECO2002) riguardano l'accorpamento dei mobili (compresa nella voce residuale “Altre manifatturiere” nell'ATECO2002) nel ramo legno-mobili, e una classificazione leggermente diversa dei vari comparti del settore meccanico (cfr. Tab. 2).

Complessivamente, si hanno quindi 19 (regioni) x 3 (dimensioni) x 22 (settori) = 1.254 celle di campionamento, a fronte di circa 4.000 interviste mensili. La numerosità di imprese per cella è dunque in media molto ridotta, e ciò può potenzialmente indurre effetti dannosi sulla variabilità delle stime. A questo proposito, va però considerato che alcuni degli strati nel campione potrebbero essere vuoti, sia perché nessuna impresa è stata selezionata in quello strato campionario, sia perché sono vuoti i relativi strati nell'universo delle imprese manifatturiere. Di conseguenza, il numero medio di imprese in ogni strato campionario è più elevato di quanto si possa evincere dal rapporto tra numerosità del campione e numero di strati.

D'altra parte, anticipando alcune delle considerazioni contenute nel par. 3.2, l'inchiesta ISAE sugli Investimenti è caratterizzata da un tasso di risposta decisamente inferiore a quello che si osserva nella corrispondente inchiesta mensile. Segue che, nel caso dell'inchiesta sugli Investimenti, potrebbero presentarsi distorsioni più gravi di quelle eventualmente riscontrabili nell'inchiesta mensile, in quanto vi saranno, generalmente, molti più strati vuoti; come vedremo oltre (cfr. par. 5), per rendere minima questa potenziale distorsione, gli strati elementari sono opportunamente uniti tra loro nella fase di aggregazione dei risultati elementari. Va infine ricordato che, rispetto al disegno teorico di campionamento, il campionamento effettivo tiene conto dell'esigenza di garantire la presenza nel campione delle aziende “fedeli”, cioè di quelle che rispondono più assiduamente ai questionari mensili dell'ISAE; ciò implica, in sede di aggiornamento periodico del campione, il mantenimento di tutte le aziende che rispondono con regolarità, integrate da nuove imprese negli strati sottorappresentati. Inoltre, nel campione ISAE alcuni strati risultano sovracampionati rispetto al campione teorico, sulla base di interventi *ad hoc*,

Tab. 2 Raccordo tra la classificazione “Rami ISCO” e l’ATECO2002

	Rami di attività – Classificazione ISCO	ATECO2002
1	TESSILI	17.1+17.2+17.3+17.5+17.6+17.7+36.63.2
2	INDUSTRIE CALZATURE E ABBIGLIAMENTO	17.4+18+19.30.1+36.63.3
3	LEGNO SUGHERO E MOBILIO IN LEGNO	20+36.11+36.12.2+36.13+36.14+36.15
4	CARTA, CARTOTECNICA, GRAFICHE EDITORIALI	21+22
5	CUOIO	19.10+19.20+19.30.2
6	TRASFORMAZIONE MATERIE PLASTICHE	25.2
7	DERIVATI DEL PETROLIO	23.2
8	SIDERURGIA, TRASFORMAZIONE METALLI FERROSI E NON FERROSI	27.1+27.2+27.3+27.4+28.4+28.51
9	MATERIALE DA COSTRUZIONE, VETRO E CERAMICA	26
10	CHIMICHE	24.1+24.2+24.3+24.4+24.5+24.6
11	PRODUZIONE FIBRE CHIMICHE (ARTIFICIALI E SINTETICHE)	24.7
12	COSTRUZIONE ATTREZZATURE METALLICHE	27.5+28.1+28.2+28.3+28.52+28.61+28.62.1 +28.63+28.7+29.21+29.54.3+29.72+36.12.1
13	COSTRUZIONE MACCHINE NON ELETTRICHE E MATERALE MECCANICO	28.62.2+29.1+29.22+29.23+29.24+29.3+29.4 +29.51+29.53+29.54.1+29.54.2+29.55+29.56 +29.6
14	COSTRUZIONE MACCHINE PER UFFICIO, MACCHINE E IMPIANTI PER L'ELABORAZIONE DATI	30
15	COSTRUZIONI ELETTRICHE ED ELETTRONICHE	29.71+31+32
16	AUTOVEICOLI E PARTI STACCATE	34+29.52
17	ALTRI MEZZI DI TRASPORTO (COMPRESI I MOTORI)	35
18	GOMMA	19.30.3+25.1
19	STRUMENTI DI PRECISIONE, OTTICA, OROLOGERIA E SIMILI	33
20	INDUSTRIE ESTRATTIVE	10+11+12+13+14
21	INDUSTRIE ALIMENTARI	15
22	INDUSTRIE MANIFATTURIERE VARIE	36.2+36.3+36.4+36.5+36.61+36.62+36.63.1 +36.63.4+36.63.5+36.63.6+37
	TOTALE MANIFATTURIERE ED ESTRATTIVE	10-37

basati su analisi economiche e di settore: ad esempio, la soglia minima di intervista al Sud è stata aumentata, in alcuni strati, rispetto alla numerosità teorica, così da poter garantire la presenza di un adeguato numero di imprese anche negli strati in cui, sulla base del campionamento teorico “puro”, sarebbe risultato allocato un numero particolarmente esiguo di imprese. In pratica, sulla

base di una serie di informazioni collaterali sono effettuate calibrazioni che portano ad ottenere un nuovo campione teorico lievemente modificato rispetto a quello originale. Tali modifiche hanno l'obiettivo di ridurre l'errore campionario, aumentando il campionamento in strati potenzialmente sottorappresentati e probabilmente caratterizzati da varianza elevata (cfr. ancora Martelli e Rocchetti, 2004).

La composizione del campione teorico risultante dalla procedura sopra descritta è confrontata con quella dell'Universo di riferimento nella tabella 3;

Tab. 3 **Universo e campione teorico**

Settori	Universo		Campione teorico		n/N (%)
	Numerosità (N)	Percentuale	Numerosità (n)	Percentuale	
Tessile, abbigliamento, calzature, cuoio	20.582	22,1	828	20,5	4,02
Carta, stampa, editoria	5.033	5,4	214	5,3	4,26
Chimica, gomma, plastica	7.295	7,8	379	9,4	5,19
Estrazione e prima lavorazione minerali metalliferi e non metalliferi	9.445	10,1	549	13,6	5,81
Meccanica	32.427	34,8	1.109	27,5	3,42
Mezzi di trasporto	2.177	2,3	184	4,6	8,45
Manifatturiere varie	9.800	10,5	440	10,9	4,49
Alimentari	6.474	6,9	337	8,3	5,20
Totale	93.233	100,0	4.040	100,0	4,33

l'ultima colonna fornisce la frazione di campionamento sondata per ognuno dei settori considerati. Gli 8 settori di attività economica presentati nella tabella sono stati ottenuti aggregando i dati relativi ai 22 RAMI ISAE descritti nella tabella 2⁵. Il metodo di campionamento sopra descritto ha condotto ad una sovrarappresentazione dei settori alimentare, estrattivo e siderurgico, mezzi di trasporto e chimico; sono invece lievemente sottorappresentati il tessile, abbigliamento, cuoio e calzature e la meccanica. Al sovracampionamento corrisponde una maggiore variabilità di strato in termini di addetti delle imprese "sondate".

5 Per comodità di presentazione dei risultati, si sono nel seguito considerati 8 settori principali, costruiti come aggregazione dei 22 Rami ISAE; in particolare, il settore "Tessile, abbigliamento, calzature, cuoio" è dato dall'aggregazione dei rami 1, 2 e 5; il settore "Carta, stampa editoria" è il Ramo 4; il settore "Chimica, gomma, plastica" è l'aggregazione dei Rami 6, 7, 10, 11 e 18; il settore "Estrazione e prima lavorazione dei minerali metalliferi e non metalliferi" aggrega i Rami 8, 9 e 20; il settore della "Meccanica" i Rami 12, 13, 14, 15 e 19; il settore dei "Mezzi di trasporto" i Rami 16 e 17; il settore delle "Manifatturiere varie" i Rami 3 e 22; il settore "Alimentare" comprende infine il solo Ramo 21.

3.2 Il campione effettivo: le inchieste dall'aprile 2000 all'aprile 2004

Il campione descritto è quello utilizzato sia per l'inchiesta congiunturale mensile, sia per l'inchiesta bi-annuale sugli investimenti. Nel caso dell'inchiesta mensile, l'utilizzo per le interviste della tecnologia C.A.T.I. (Computer Aided Telephonic Interviews) garantisce un tasso di risposta *ex-post* prossimo al 100%. La tecnica di rilevazione telefonica consente, infatti, la "sostituzione" in tempo reale delle imprese non rispondenti con altre imprese estratte dallo stesso strato, garantendo in questo modo in ogni rilevazione che il campione effettivo riproduca in modo quasi esatto quello teorico⁶. L'inchiesta sugli investimenti è, come detto precedentemente, condotta tramite l'invio postale dei questionari, e quindi non è possibile garantire a priori la corretta composizione del campione dei rispondenti. Inoltre, il questionario è per sua natura più complesso rispetto a quello mensile, sia perché richiede di fornire indicazioni di carattere quantitativo oltre a quelle qualitative, sia perché richiede una conoscenza approfondita da parte del rispondente del fenomeno oggetto di studio; va inoltre considerato che non tutte le imprese manifatturiere effettuano investimenti in ogni anno e che la stessa frequenza ridotta dell'invio dei questionari (due volte l'anno, invece che su base mensile) può generare difficoltà nelle risposte, per scarsa abitudine, maggiore difficoltà ad individuare all'interno dell'impresa la persona che risponde, ecc. Alla luce di tali considerazioni, non sorprende che il tasso di risposta all'indagine sia notevolmente inferiore rispetto a quello dell'inchiesta mensile; nelle inchieste analizzate in questo lavoro il numero delle imprese rispondenti ha oscillato tra le 927 e 723 unità, con un tasso di risposta dunque compreso tra il 23% e 18% del campione.

Dato il basso tasso di risposta all'indagine, per valutare l'attendibilità generale e la confrontabilità intertemporale dei risultati è importante verificare la stabilità nel tempo del campione; a tale proposito la tabella 4 presenta la struttura del campione effettivo dei rispondenti in termini di composizione settoriale, sia guardando alla distribuzione in termini di numero di imprese sia a quella degli addetti delle imprese stesse, relativamente alle indagini condotte tra l'aprile 2000 e l'aprile 2004. La distribuzione delle imprese per settore è piuttosto stabile tra una rilevazione e l'altra: i comparti più rappresentati sono quelli della meccanica (con una quota compresa tra il 26,5 e il 31,1%), del tessile, abbigliamento, cuoio e calzature (che ha rappresentato tra il 14,1 e il

6 Più precisamente, l'inchiesta mensile presenta un tasso di sostituzione mensile pari a circa il 2%, ossia circa il 2% delle imprese complessivamente intervistate viene ogni mese rinnovato, sostituendo le imprese "silenti" con altre imprese estratte in modo da rispettare strettamente lo schema campionario sopra delineato.

17,2% del campione) e dell'estrazione e prima lavorazione dei minerali (tra il 12,1 e il 14,9% del campione). I tre settori sopra menzionati sono anche, nello stesso ordine, quelli che assumono maggiore rilevanza nella composizione del campione teorico di riferimento. Il settore tessile è però leggermente sottorappresentato rispetto al campione teorico, a favore delle industrie meccaniche e delle manifatturiere varie. A livello dimensionale, in tutti i campioni circa il 57% delle imprese ha fino a 49 addetti; il 30% circa dei rispondenti è costituito da imprese con un numero di addetti compreso tra 50 e 249 e poco più del 12% dalle grandi imprese. La variabilità della composizione campionaria, in termini sia di settore sia di dimensione, è invece maggiore se si guarda alla composizione del campione secondo gli addetti delle imprese rispondenti; le maggiori oscillazioni si hanno per il settore meccanico (tra il 19,4 e il 35,6% del totale) e per i mezzi di trasporto (tra il 29,6 e il 40% del totale). In tutte le inchieste, la maggior parte degli addetti è concentrata, in effetti, in questi due settori con percentuali che oscillano cumulativamente dal 58 al 66 per cento.

Guardando ai dati delle tabelle 3 e 4, ad un basso tasso di risposta si associa quindi una composizione del campione piuttosto stabile in termini di numero di imprese partecipanti alla rilevazione. Il campione non è però altrettanto stabile dal lato della distribuzione degli addetti, essenzialmente a causa di un alto tasso di sostituzione tra una rilevazione e l'altra tra imprese piccole e medie: in termini di addetti, infatti, le imprese di grande dimensione rappresentano sempre oltre l'80% del campione, mentre la frazione delle piccole imprese è compresa tra il 6 e il 4,5% e quella delle medie imprese tra il 15,4 e il 10,9%. Complessivamente, la relativa stabilità del campione e le similitudini tra la struttura dei campioni teorico ed effettivo fanno ritenere i risultati dell'indagine sufficientemente rappresentativi dei sottostanti andamenti effettivi.

Anticipando, per concludere, alcune considerazioni che saranno svolte più nel dettaglio nel par. 5, va considerato però che le nuove procedure di aggregazione si basano su una struttura di ponderazione basata su un "fattore di scala", costituito ove possibile dagli stessi investimenti dichiarati dalle imprese. L'utilizzo come peso di un "fattore di scala" come gli investimenti appare appropriato anche alla luce di recenti contributi di letteratura (cfr. OCSE, 2003), che raccomandano, ove possibile, l'utilizzo di pesi "specifici" per la variabile analizzata. D'altra parte, la variabilità della distribuzione degli addetti espone, in questo contesto, al rischio di una elevata variabilità dei pesi stessi. In presenza di tale *trade-off*, si è deciso, operativamente, di dare priorità alla scelta appropriata del peso, assumendo quindi il rischio di una sua maggiore variabilità⁷.

7 Si ringrazia un anonimo *referee* per aver sollevato questo punto.

Tab. 4 Il campione effettivo: le inchieste ISAE dall'aprile 2000 all'aprile 2004

	2000		2001		2002		2003		2004
	Apr.	Nov.	Apr.	Nov.	Apr.	Nov.	Apr.	Nov.	Apr.
Imprese (valori percentuali sul totale imprese)									
Tessile, abbigliamento, calzature, cuoio	17,2	14,1	16,4	16,6	16,0	15,6	14,7	15,1	14,8
Carta, stampa, editoria	5,9	7,1	5,1	5,5	5,0	5,3	4,4	5,6	5,8
Chimica, gomma, plastica	9,2	11,3	10,5	10,5	10,4	9,9	8,9	11,2	11,2
Estrazione e prima lavorazione minerali metalliferi e non metalliferi	13,5	14,9	12,1	12,9	12,6	12,4	14,1	12,5	13,5
Meccanica	28,0	26,5	28,8	27,6	29,3	30,6	31,1	30,6	29,5
Mezzi di trasporto	5,2	5,2	5,1	4,7	4,9	5,1	5,0	5,0	4,9
Manifatturiere varie	12,4	12,0	12,3	12,8	12,3	12,5	13,3	11,9	11,8
Alimentari	8,6	8,9	9,7	9,4	9,5	8,5	8,6	8,1	8,4
Totale	927	807	846	830	881	767	723	784	830
Piccole imprese (0-49 addetti)	61,5	58,0	56,4	56,3	54,9	55,7	58,5	53,1	59,0
Medie imprese (50-249 addetti)	27,8	29,1	29,2	30,4	32,6	31,7	30,4	33,0	30,7
Grandi imprese (oltre 250 addetti)	10,7	12,9	14,4	13,3	12,5	12,6	11,1	13,9	10,2
Addetti delle imprese rispondenti (valori percentuali sul totale addetti)									
Tessile, abbigliamento, calzature, cuoio	6,5	4,6	6,2	7,3	5,7	6,9	4,4	4,9	4,5
Carta, stampa, editoria	2,9	4,1	1,5	1,7	3,6	2,5	2,3	3,2	3,5
Chimica, gomma, plastica	17,4	13,7	9,1	11,0	9,1	9,1	8,4	9,9	7,4
Estrazione e prima lavorazione minerali metalliferi e non metalliferi	6,9	9,2	9,6	7,6	7,7	7,9	7,9	6,7	7,0
Meccanica	22,9	19,4	29,9	28,1	28,7	33,7	31,6	27,4	35,6
Mezzi di trasporto	35,5	40,0	35,5	33,4	33,2	29,6	34,2	36,2	30,6
Manifatturiere varie	3,4	2,5	2,7	3,2	3,5	3,5	3,4	3,7	3,6
Alimentari	4,5	6,5	5,5	7,6	8,5	6,9	7,7	8,2	7,9
Totale									
Piccole imprese (0-49 addetti)	6,0	5,2	4,5	4,9	5,0	5,4	5,4	5,1	6,0
Medie imprese (50-249 addetti)	13,3	11,9	10,9	12,7	14,1	13,5	13,4	15,4	15,3
Grandi imprese (oltre 250 addetti)	80,8	82,9	84,6	82,4	80,9	81,2	81,2	79,5	78,8

4 I CONTROLLI DI QUALITA'

Un elemento importante da analizzare, nel valutare la qualità dei risultati di un'indagine, è la loro affidabilità, ovvero l'accuratezza con la quale si misura la variabile oggetto di studio (OCSE, 2003, p. 25). Nel caso dell'indagine sugli investimenti, tale problema è particolarmente rilevante, in quanto vengono raccolte anche informazioni quantitative oltre a quelle qualitative. L'ISAE effettua di solito per tutte le indagini, ed in particolare per quella sugli investimenti, una serie di controlli di qualità, sino al 2002 basati esclusivamente su procedure di tipo deterministico (cioè sul giudizio dell'analista); con questo lavoro, ci si propone di utilizzare anche procedure di tipo statistico, in particolare per quanto riguarda il trattamento dei dati mancanti (derivanti da questionari che pervengono incompleti in qualche loro parte) e anomali (ossia, dati che in un qualche modo si discostano notevolmente dal resto delle osservazioni).

4.1 I dati anomali

E' piuttosto usuale⁸ che in indagini di tipo campionario alcune informazioni raccolte si presentino come "anomale" rispetto al complesso delle informazioni rilevate. Più precisamente, da un punto di vista teorico gli *outliers* possono essere distinti, sulla base della classificazione adottata da Chambers (1986), in "non rappresentativi", ossia conseguenza di errori di rilevazione o di inserimento dei dati nel database, e "rappresentativi", derivanti cioè dalla realizzazione di fenomeni a bassa probabilità. Nel primo caso, i dati, essendo frutto di errori, non contribuiscono in alcun modo a fornire informazioni sulla variabilità delle restanti unità della popolazione non campionate; è pertanto necessario procedere ad una loro identificazione e, ove possibile, eliminazione o correzione (Smith, 1987). Nel secondo caso, i dati sono correttamente osservati e contengono dunque informazioni relative al resto della popolazione; tuttavia, anche la presenza di *outliers* rappresentativi può distorcere le stime campionarie, in particolare nel caso in cui la stima del parametro di riferimento (in questo caso, degli investimenti aggregati per il totale del settore manifatturiero) è ottenuta "riportando" all'Universo i singoli dati, con conseguente "espansione" del dato d'impresa. Per tale ragione si rende necessario procedere al loro trattamento con metodi statistici (cfr. par. 4.1.2)

8 Cfr. Banca d'Italia (1996).

volti a ridurre l'influenza sulle stime campionarie, rendendole più stabili (Battipaglia, 2000).

4.1.1 Il controllo deterministico

Per l'identificazione degli *outliers*, le procedure di controllo deterministico adottate dall'ISAE prevedono di procedere in primo luogo a verificare l'esistenza di incongruenze "interne" tra le risposte fornite: ad esempio, si valuta se l'ammontare di investimenti dichiarati/programmati per un certo anno t è coerente con quelli dichiarati/programmati per un altro anno compreso nello stesso questionario. In seguito, viene verificata la coerenza "longitudinale" tra questionari contigui: ad esempio, se in due questionari consecutivi l'informazione rispetto allo stesso anno t risulta notevolmente diversa, si verifica se la revisione effettuata (al rialzo o al ribasso) è coerente con il resto delle informazioni di natura qualitativa e quantitativa contenute nello stesso questionario. In questo modo, è possibile determinare se i valori anomali individuati sono non rappresentativi o rappresentativi. Nel primo caso, si provvede a correggere l'eventuale errore riscontrato in fase di inserimento dei dati, spesso procedendo con la tecnica della reintervista, ove si ritenga che l'errore sia generato da una non corretta risposta dell'impresa. Nel secondo caso, trattandosi di valori correttamente osservati, è possibile prendere in esame la possibilità di sottoporre i dati a procedure statistiche, al fine di ridurre la variabilità generata dalla loro presenza.

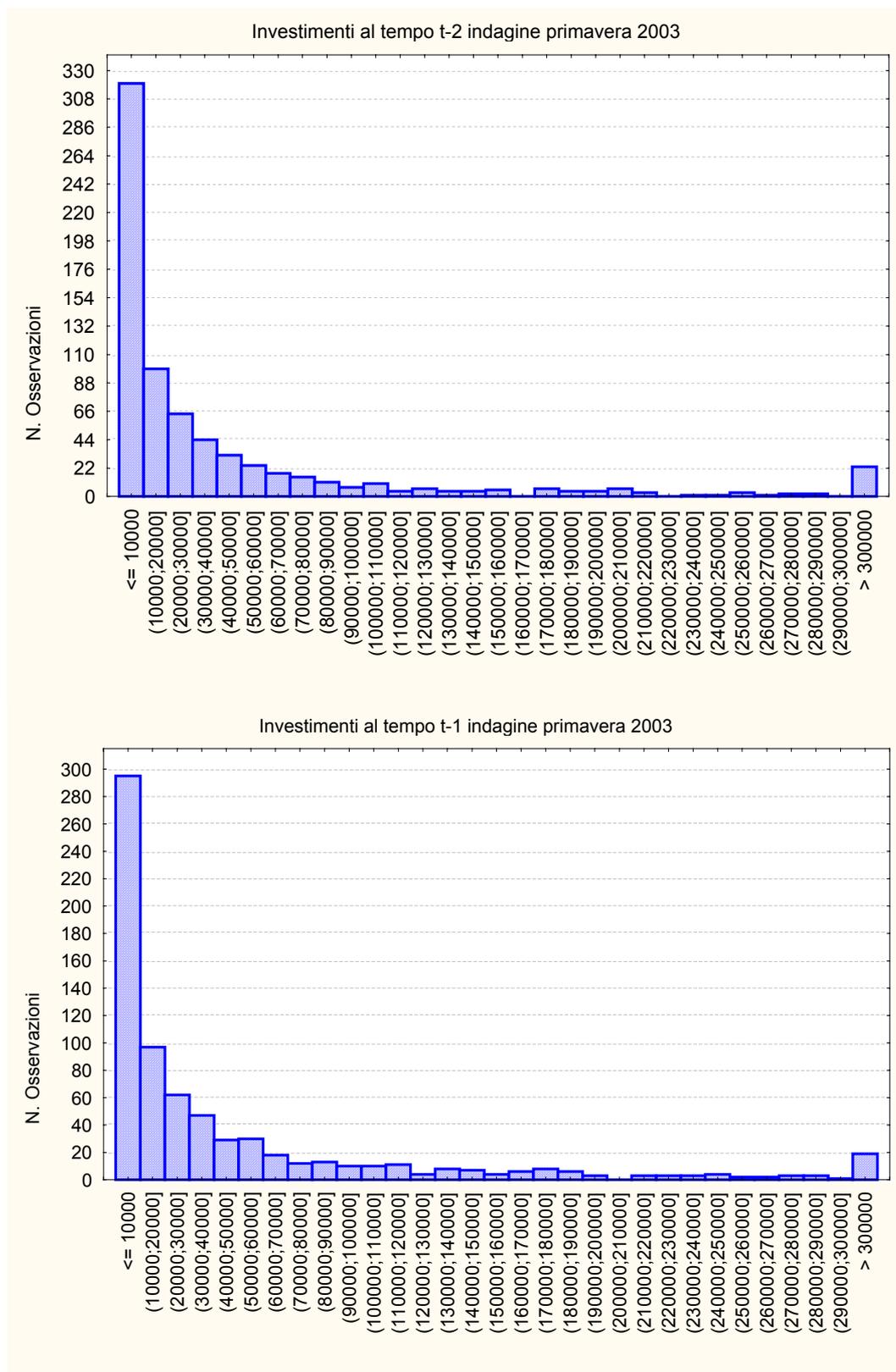
4.1.2 Le procedure statistiche

Una volta che i dati sono stati controllati per la presenza di eventuali *outliers* di tipo non rappresentativo, essi sono usualmente inclusi nel database. E' però possibile che a questo punto "valori estremi" di risposta siano ancora presenti nel data-set; una verifica di tale ipotesi viene dall'analisi della distribuzione delle risposte fornite dalle imprese. A tal proposito, si può osservare (Fig. 1) che le distribuzioni campionarie degli investimenti dichiarati e programmati, rispettivamente, per gli anni 2001, 2002 e 2003 mostrano una forte asimmetria positiva, con le prime due classi di investimento⁹ comprendenti circa il 70% delle osservazioni. Inoltre, nell'ultima classe aperta di tali distribuzioni sono presenti osservazioni caratterizzate da valori particolarmente elevati.

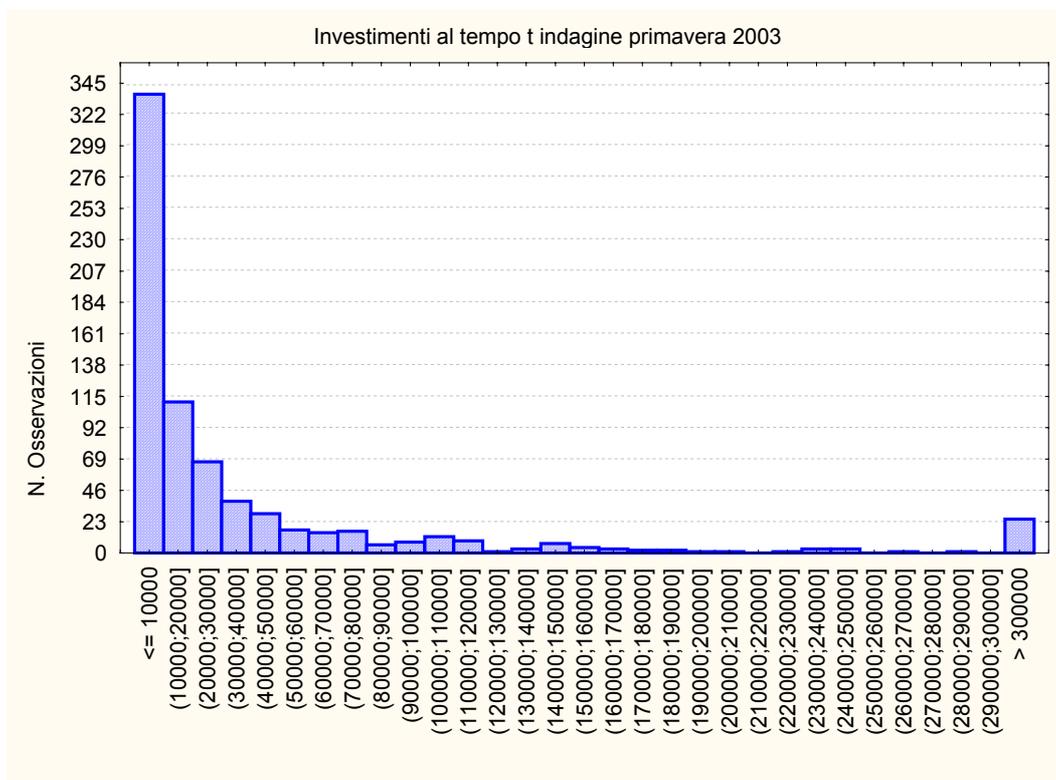
⁹ L'ampiezza delle classi di investimento è fissa. L'ultima classe è stata lasciata aperta.

Fig. 1

**Distribuzione empirica pesata degli investimenti
dichiarati e programmati (t-2, t-1, t)
(indagine ISAE, primavera 2003)**



segue Fig. 1 **Distribuzione empirica pesata degli investimenti dichiarati e programmati (t-2, t-1, t) (indagine ISAE, primavera 2003)**

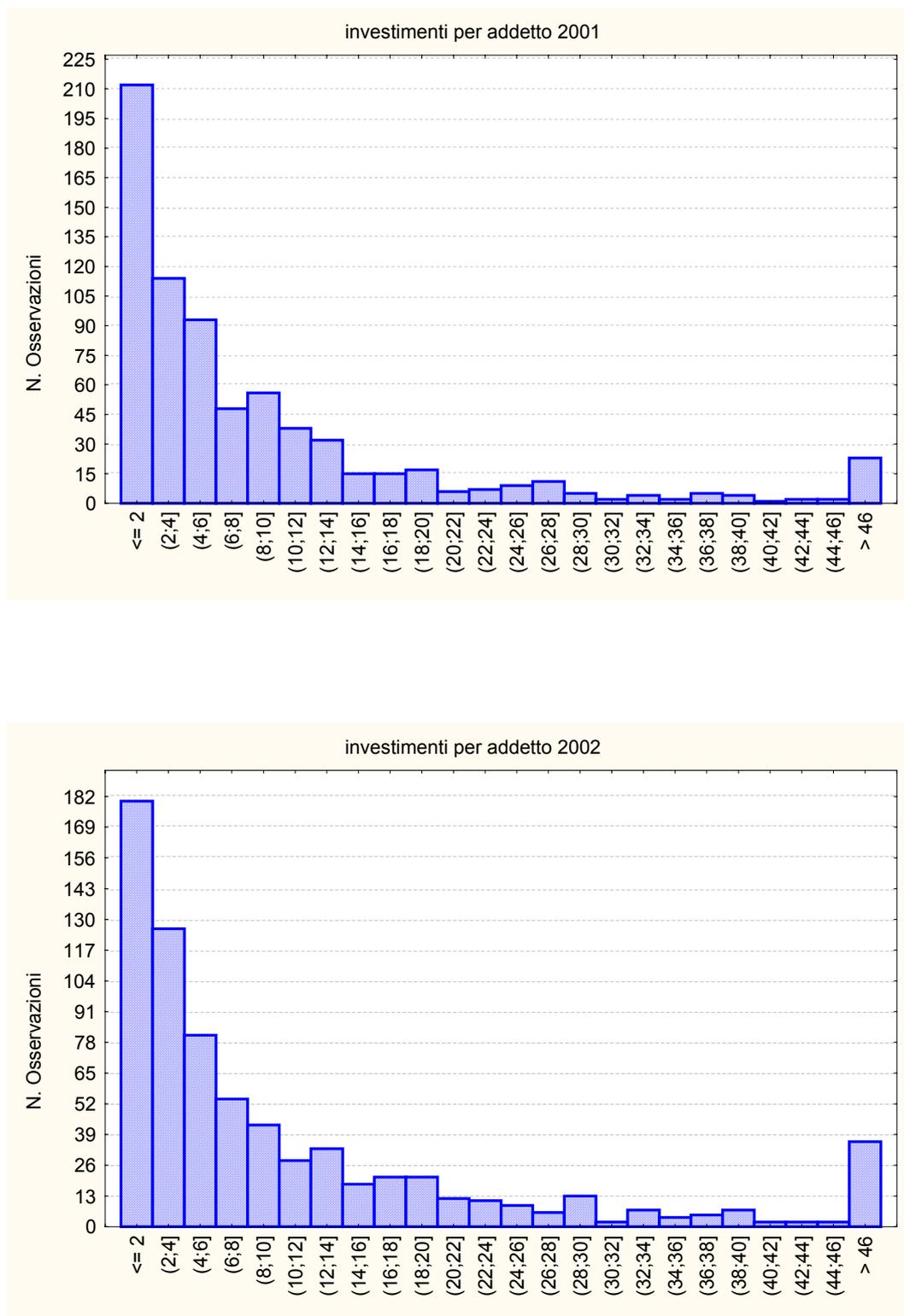


Indicazioni analoghe vengono anche dall'analisi dei dati relativi agli investimenti per addetto (Fig. 2), ad indicare che probabilmente il comportamento "asimmetrico" delle risposte delle imprese non è funzione esclusiva del "size", non dipende cioè soltanto dalla composizione del campione e dalla presenza al suo interno di imprese di grande dimensione, ma è una caratteristica delle risposte fornite dalle imprese. In altri termini, l'asimmetria evidenziata è probabilmente causata dalla presenza di osservazioni "estreme", interpretabili come valori anomali di tipo rappresentativo, nel senso sopra descritto. Tale diagnosi è confermata anche dall'analisi della distribuzione delle variazioni assolute dei livelli degli investimenti (Fig. 3), da cui si evince che tale distribuzione presenta, come era plausibile attendersi osservando le distribuzioni dei livelli degli investimenti, un certo grado di simmetria e valori anomali in entrambe le code della distribuzione¹⁰.

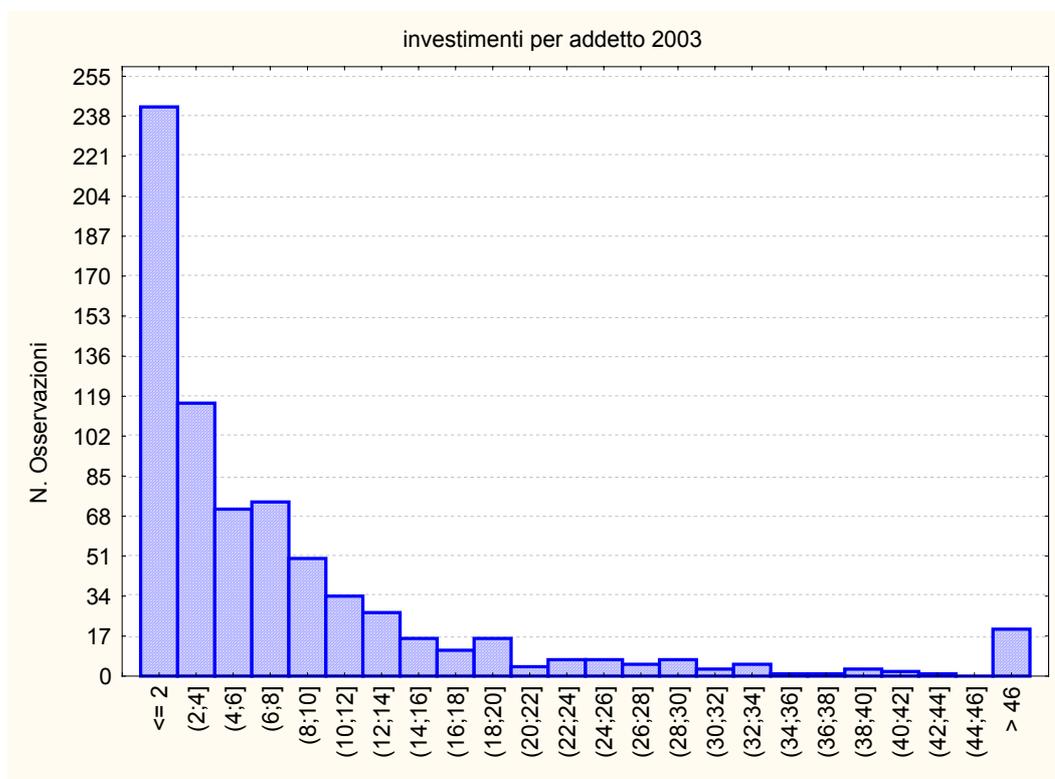
10 Mentre gli *outliers* nelle distribuzioni dei livelli degli investimenti sono originati, come già ribadito, da quelle imprese che effettuano un elevato ammontare di investimenti, quando si considerano le variazioni assolute dei livelli essi derivano piuttosto da quelle imprese i cui investimenti subiscono una forte variazione al rialzo o al ribasso rispetto all'anno precedente.

Fig. 2

**Distribuzione empirica degli investimenti per addetto
dichiarati nel 2001 e 2002 e programmati per il 2003
(indagine ISAE, primavera 2003)**



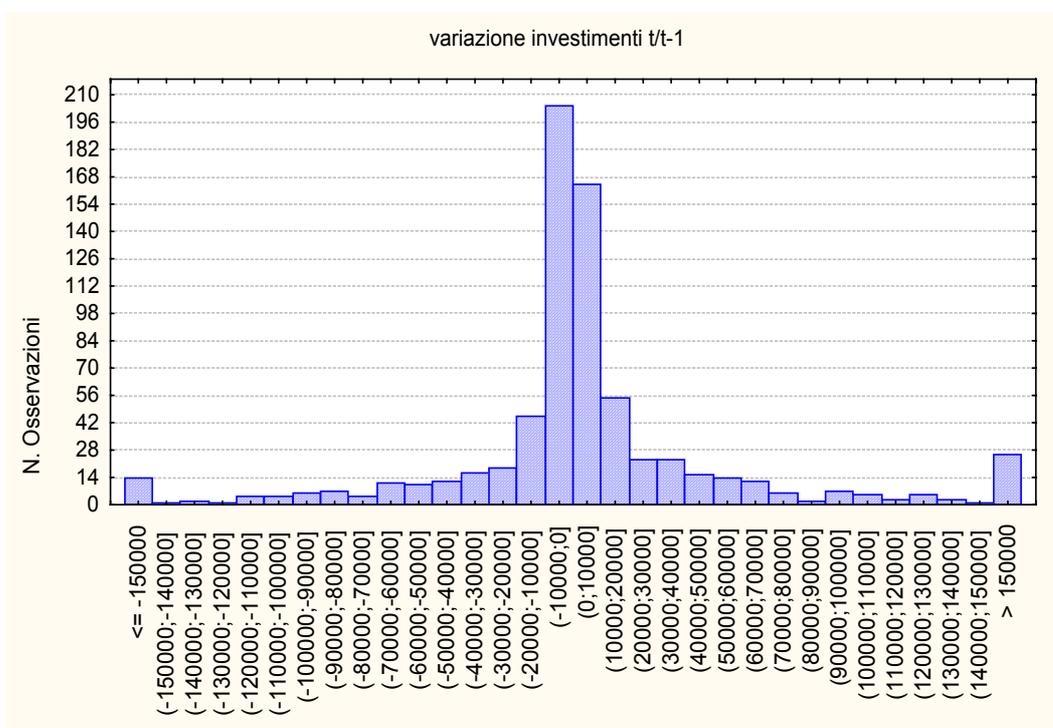
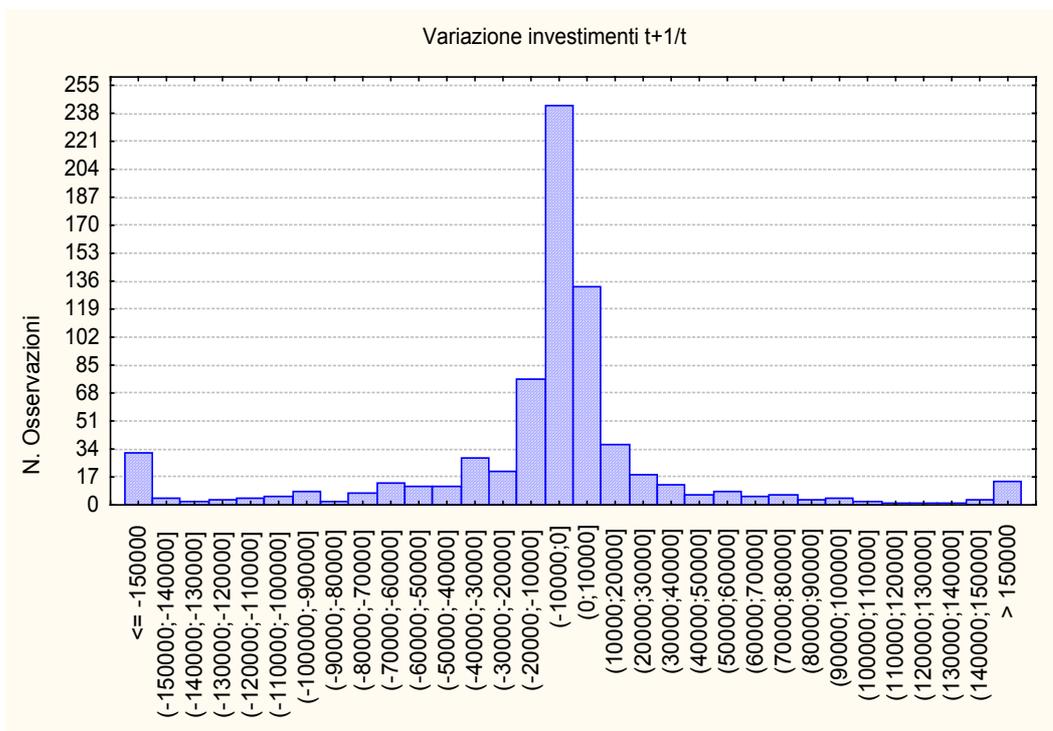
segue Fig. 2 **Distribuzione empirica degli investimenti per addetto dichiarati nel 2001 e 2002 e programmati per il 2003 (indagine ISAE, primavera 2003)**



L'analisi dimostra dunque che osservazioni estreme, considerabili di tipo rappresentativo, sono ancora presenti nel database, anche dopo i controlli di tipo deterministico mirati a correggere gli *outlier* non rappresentativi; è dunque opportuno valutare la possibilità di correggere le stime del parametro di riferimento utilizzando appropriati metodi statistici.

Per rilevazioni simili, da un punto di vista teorico, (si veda a tal proposito Battipaglia 2000), si sono utilizzate procedure di stima "robuste" dei dati aggregati; in particolare, sono stati proposti diversi approcci finalizzati a produrre stimatori robusti per dati derivanti da indagini campionarie (approccio *model based*, cfr. ad esempio Chambers, 1986 ed approccio *design based*, cfr. Hidioglou e Srinath, 1981); ad oggi tuttavia si dispone di pochi risultati relativi all'applicazione di tali tecniche ai dati provenienti da indagini. Inoltre, il problema della scelta del tipo di stimatore da utilizzare in questi casi non è sempre risolvibile in modo generale, dipendendo da fattori quali le finalità che ci si propone di raggiungere ed il tipo di informazione che si desidera estrarre dai dati (cfr. Gismondi, 2002). Ad ogni modo, è possibile dare indicazioni di massima su quali dovrebbero essere le proprietà desiderabili che uno stimatore per dati provenienti da indagini campionarie deve possedere: tra

Fig. 3 La distribuzione della variazione assoluta degli investimenti
(indagine ISAE, primavera 2003)



le prime vi è innanzi tutto la robustezza ovvero, lo stimatore deve risultare poco sensibile ai dati anomali; altre caratteristiche desiderabili riguardano la semplicità nell'implementazione e nell'interpretazione dei risultati. Lo stimatore dovrebbe essere caratterizzato inoltre da una bassa varianza e una bassa distorsione (Cfr Hulliger, 1999). Gli stimatori robusti maggiormente utilizzati nella letteratura corrente sono costituiti da mediane, medie trimmed e medie winsorizzate. Nel seguito vengono effettuate alcune considerazioni sull'uso di medie trimmed e medie winsorizzate per valutare e ridurre l'impatto degli *outliers* sui dati ISAE.

Le medie trimmed vengono calcolate escludendo le k osservazioni più piccole e le k osservazioni più grandi dal computo della media, eliminando in tal modo l'influenza dei valori estremi della distribuzione. In altre parole tali medie eliminano tutte le osservazioni che eccedono un particolare valore soglia, che può essere fissato a priori (in questo caso si escludono tutte le osservazioni che eccedono un particolare valore in entrambi i lati della distribuzione - ad esempio tutte le osservazioni aventi un valore oltre la soglia ± 10.000), o individuato decidendo di eliminare una frazione fissa di osservazioni (ad esempio, il 5% delle osservazioni in entrambi i lati della distribuzione). In formula, si ha:

$$(1) \quad y_k = \frac{1}{n - 2k} \sum_{i=k+1}^{n-k} y_i$$

Nel caso della winsorizzazione, le k osservazioni più piccole vengono sostituite dalla $k+1$ -esima osservazione più piccola, mentre le ultime $n-k$ osservazioni più grandi vengono sostituite dalla $n-k-1$ -esima osservazione che le precede¹¹.

In formula si ha:

$$(2) \quad y_k = \frac{1}{n} \left\{ (k+1)y_{(k+1)} + \sum_{i=k+2}^{n-k-1} y_{(i)} + (k+1)y_{(n-k)} \right\}$$

La winsorizzazione, rispetto alle medie trimmed ha il vantaggio di conservare tutte le osservazioni, attribuendo però un peso diverso a quelle che si collocano agli estremi della distribuzione. Ciò risulta particolarmente utile nel caso in cui si disponga di un campione di numerosità non particolarmente elevata, in quanto consente di considerare comunque tutti i dati campionari e la

11 Alternativamente i valori agli estremi della distribuzione potrebbero essere sostituiti oltre che con il valore assunto dalla soglia anche con valori intermedi compresi tra essi e la soglia di *cut off*.

relativa informazione ad essi associata; per tale motivo, la winsorizzazione costituisce uno strumento ampiamente utilizzato nel caso di indagini campionarie dagli analisti per limitare l'impatto dei dati anomali sulle stime.

Per produrre stime robuste del tasso di variazione degli investimenti tramite la winsorizzazione è possibile (cfr. ancora Battipaglia, 2000) utilizzare due diversi approcci (entrambi riferiti alle distribuzioni delle variazioni degli investimenti anziché ai livelli). Nel primo essa viene applicata ad un solo lato della distribuzione delle variazioni relative degli investimenti I_i^t/I_i^{t-1} che è asimmetrica e che contiene pertanto valori estremi soltanto nella coda destra; nel secondo, essa si applica in modo simmetrico ad entrambi i lati della distribuzione delle variazioni relative degli investimenti $I_i^t - I_i^{t-1}$ dato che in quest'ultima i valori anomali sono presenti in entrambe le code. Poiché l'utilizzo della winsorizzazione asimmetrica introduce per costruzione un grado di distorsione nelle stime, nell'esercizio svolto si è ritenuto opportuno applicarla in modo simmetrico, secondo la formula:

$$(3) \quad \Delta I_t^{wins} = \frac{\sum_{i=1}^n w_i [I_i^t - I_i^{t-1}]^{wins}}{\sum_{i=1}^n w_i^s I_i^{t-1}}$$

dove ΔI_t^{wins} è il tasso di variazione degli investimenti, w_i^s è il coefficiente di espansione all'universo (cfr. anche par. 5) che permette di tenere conto della diversità esistente tra la struttura del campione e quella dell'universo di riferimento e $[I_i^t - I_i^{t-1}]^{wins}$ è la distribuzione delle variazioni assolute degli investimenti winsorizzata secondo una data soglia di *cut off*.

La tabella 5 riporta le stime del tasso di variazione degli investimenti calcolato utilizzando la (3) ed applicando diversi livelli di *cut off* (2, 5 e 10%) al totale delle osservazioni per le indagini condotte tra il novembre 2000 e l'aprile 2004. I risultati sono comparati con quelli ottenuti calcolando il tasso di variazione sulla distribuzione degli investimenti non trattata (e dunque comprensiva di eventuali *outliers*). La winsorizzazione introduce rilevanti cambiamenti nella stima del tasso di crescita degli investimenti; le differenze sono rilevanti confrontando i tassi "originali" con quelli winsorizzati, ma i risultati non sono particolarmente sensibili alle diverse soglie di *cut off* applicate. Quest'ultimo aspetto è dovuto essenzialmente al fatto che la maggior parte dei valori estremi risultano concentrati nell'ultimo 2% della coda della distribuzione:

se si fossero applicati diversi tagli nell'intorno del 2% si sarebbe sicuramente registrata una maggiore sensibilità dei risultati alla scelta della soglia.

Tab. 5 Tassi di variazione degli investimenti winsorizzati

Tasso di variazione degli investimenti	Tasso di <i>Cut Off</i>	Questionario primaverile				Questionario autunnale				
		2001	2002	2003	2004	2000	2001	2002	2003	2004
t+1/t per novembre t/t-1 per aprile	Nessuno	-8,9	-21,8	-11,0	3,0	2,7	-18,4	-9,4	0,4	-2,1
	2%	-1,1	-5,0	-11,6	-4,1	0,2	-7,3	-10,3	-2,5	-3,6
	5%	-2,1	-3,6	-12,4	-4,2	-1,4	-6,0	-10,4	-4,0	-2,4
	10%	-2,4	-3,0	-11,5	-2,3	-0,6	-3,6	-9,5	-3,0	-2,1
t/t-1 per novembre t-1/t-2 per aprile	Nessuno	20,2	-10,1	-7,8	-15,7	8,6	-13,2	-6,8	-18,9	-1,2
	2%	5,7	-1,8	6,2	-14,2	5,2	0,3	-3,0	-22,8	-6,3
	5%	5,3	-1,1	7,4	-13,5	4,3	0,0	-3,0	-22,5	-6,7
	10%	5,0	-0,5	7,9	-13,5	3,5	-0,2	-2,4	-21,3	-5,9

Va peraltro considerato che l'esercizio svolto rappresenta un'esplorazione iniziale del possibile impatto dei tagli sulle stime: l'individuazione del *cut off* ottimo, come indicato in Battipaglia (2000), dovrebbe infatti essere basata su criteri di minimizzazione dell'errore quadratico medio. Tuttavia, anche sulla base dei risultati preliminari presentati, emerge con chiarezza che per valutare al meglio i risultati dell'indagine occorre tenere conto dell'influenza dei valori anomali sulle stime del tasso di variazione degli investimenti. Per tale ragione l'ISAE a partire dall'indagine dell'autunno 2004 pubblica stime del tasso di crescita degli investimenti basate sulle medie winsorizzate descritte in questo paragrafo, applicando una soglia di *cut off* pari al 10%; nel prossimo futuro, si procederà a tentare di individuare la soglia di *cut off* ottimale sulla base degli appropriati criteri statistici.

4.2 I dati mancanti

Quando si parla di dati mancanti ci si riferisce in questo contesto a questionari che pervengono incompleti in qualche loro parte, o perché le imprese non sono in grado di rispondere ad una data domanda, o per disattenzione da parte del rispondente. In altri termini, nel lavoro si trattano esclusivamente i casi di non risposta ad alcune delle domande del questionario

(mancata risposta parziale); nel presente lavoro non viene, invece, trattato il problema della mancata risposta di una parte consistente del campione.

4.2.1 Il controllo deterministico

Per l'individuazione dei dati mancanti, in una prima fase si effettua un controllo manuale durante l'inserimento dei questionari nel database; una volta individuati i questionari caratterizzati da mancate risposte, si utilizza la tecnica campionaria della reintervista, attraverso la quale l'azienda viene ricontattata al fine di ottenere spiegazioni sulla mancata risposta, provvedendo eventualmente ad integrare l'informazione originale. Nel caso in cui l'azienda non sia in grado di fornire informazioni relative al dato mancante (perché ad esempio questo non è disponibile), il questionario non viene eliminato, ma ci si limita a considerare i dati in esso presenti, rimandando l'eventuale reintegro dei dati mancanti all'utilizzo di procedure statistiche.

A questo proposito, recentemente, il questionario ISAE è stato perfezionato prevedendo, per la domanda sull'ammontare degli investimenti effettuati e/o programmati, la possibilità di barrare una casella "nessun investimento", posta accanto all'indicazione dell'ammontare delle spese in beni capitali. Nel passato, infatti, nella fase di inserimento delle informazioni contenute nei questionari, la mancata indicazione di una risposta da parte delle imprese era considerata come un'indicazione di nessun investimento realizzato, anche in mancanza di una esplicita indicazione in tal senso da parte dell'impresa, generando possibili sottostime della spesa per investimenti. Per questa ragione, nella ricostruzione della base dati ISAEINV, si è provveduto a ricontrollare l'intero database delle inchieste effettuate tra il 2000 e il 2002, allo scopo di giungere ad una più corretta individuazione dei dati mancanti, da considerarsi strettamente distinti da indicazioni di "nessun investimento" fornite dalle imprese¹².

4.2.2 Le procedure statistiche

L'analisi dettagliata dei risultati delle inchieste ISAE sugli investimenti mostra che l'utilizzo delle sole procedure di tipo deterministico (basate come visto essenzialmente sulla tecnica della reintervista) non sono sufficienti a risolvere completamente il problema della presenza di dati mancanti. A tale

12 Per il database relativo ad aprile 2000 non è stato tuttavia possibile distinguere i dati mancanti dall'investimento nullo. Si ringraziano Paola Bellincampi e Massimo di Tommaso per il prezioso e puntuale lavoro di controllo effettuato sui questionari.

tabella 7, il tasso di non risposta differisce sensibilmente tra strati¹³, diventando particolarmente elevato in alcune celle. Segue da queste considerazioni la necessità di tentare di minimizzare, almeno per quanto riguarda le variabili quantitative, il numero di dati mancanti nel database, usando specifiche tecniche statistiche di imputazione degli stessi.

Tab. 7 Dati mancanti per settore, classe di ampiezza e ripartizione geografica (percentuale delle risposte)

Aprile 2004	t-2/t-1		t-1/t		t/t+1	
	% dati mancanti (*)	% pesate dati mancanti (**)	% dati mancanti (*)	% pesate dati mancanti (**)	% dati mancanti (*)	% pesate dati mancanti (**)
Tessile, abbigliamento, calzature, cuoio	0,81	0,84	0	0	4,88	5,19
Carta, stampa, editoria	4,17	3,57	0	0	8,33	11,98
Chimica, gomma, plastica	3,23	2,98	0	0	3,23	2,12
Estrazione e prima lavorazione minerali metalliferi e non metalliferi	1,79	2,05	0	0	4,46	5,67
Meccanica	1,63	1,43	0,82	0,7	2,04	2,28
Mezzi di trasporto	0	0	0	0	2,44	5,34
Manifatturiere varie	2,04	2,92	0	0	0	0
Alimentari	5,71	7,16	2,86	3,87	2,86	3,87
Piccole imprese (0-49 addetti)	2,86	2,24	0,82	0,59	4,29	4,06
Medie imprese (50-249 addetti)	1,18	1,2	0	0	1,18	1,22
Grandi imprese (oltre 250 addetti)	1,18	1,64	0	0	2,35	1,52
Nord Ovest	1,79	0,70	0,45	0,36	2,24	3,85
Nord Est	1,29	1,83	0	0	2,27	2,05
Centro	3,59	4,81	0	0	5,99	6,90
Sud	3,05	3,02	2,29	2,84	3,05	2,82

(*) $(n''/n) \times 100$

$$(**) M = \frac{\sum_{i=1}^{n''} \frac{1}{f_i}}{\sum_{i=1}^n \frac{1}{f_i}}$$

13 A titolo esemplificativo i tassi sono riportati per la sola indagine dell'aprile 2004; simili risultati, disponibili su richiesta, si ottengono per le altre indagini.

In questo lavoro, si propone di utilizzare un metodo di imputazione di medie condizionate, basato su un semplice modello di regressione che usa le informazioni contenute nei questionari compilati dalle imprese durante l'indagine. Nel dettaglio, per stimare l'informazione mancante relativa alle spese di investimento di una generica impresa, si adottano modelli leggermente diversi per stimare le spese di investimento in t-2, t-1 e t (indagine primaverile) oppure quelle in t-1, t e t+1 (indagine autunnale).

I modelli adottati si basano su alcune assunzioni preliminari:

- i. la presenza di dati mancanti è casuale e non dipende da particolari caratteristiche delle imprese del campione; più precisamente, si assume che la probabilità di risposta dell'impresa non dipenda dal valore dell'investimento omesso (ovvero i dati mancanti devono essere *missing at random* – MAR): sia Y una matrice $n \times k$ di dati relativa alle risposte fornite da n unità statistiche a k domande, ed R una matrice di variabili indicatrici che assumono valore 1 se l'impresa ha risposto alla domanda e 0 altrimenti. Dire che i dati sono MAR equivale ad assumere che:

$$P(R | Y_{\text{oss}}, Y_{\text{miss}}) = P(R | Y_{\text{oss}})$$

essendo Y_{oss} le variabili completamente osservate (ovvero quelle relative alle domande a cui hanno risposto tutte le imprese) e Y_{miss} quelle che presentano dati mancanti;

- ii. tutte le variabili esplicative contenute nei modelli adottati devono essere completamente osservabili, ovvero, non devono contenere esse stesse dei dati mancanti.

In particolare, nei modelli adottati si utilizzano le seguenti variabili:

- i. per la stima degli investimenti realizzati nel periodo t-1 dalla generica impresa i , si utilizza come fattore di scala il numero dei dipendenti dell'impresa stessa. Questa variabile è sempre completamente osservabile. Inoltre, un'analisi preliminare della matrice di correlazione tra le variabili quantitative rilevate dall'indagine mostra l'esistenza, in tutti i data-set, di una forte correlazione positiva tra numero di addetti ed investimenti effettuati;

- ii. per la stima degli investimenti effettuati in t, o programmati per l'anno t+1, i modelli includono, inoltre, informazioni sugli investimenti effettuati dall'impresa, rispettivamente, per l'anno t-2 (nel caso in cui si debbano stimare dati mancanti per l'investimento in t-1), t-2 e t-1 (per stimare gli investimenti mancanti relativi alla data t), e t-1, t (per stimare i dati mancanti per l'anno t+1); si suppone, infatti, che gli investimenti dichiarati dall'impresa in anni contigui siano tra loro autocorrelati;
- iii. in aggiunta a queste variabili, anche in considerazione della distribuzione delle mancate risposte per strato mostrata nella tabella 7, si includono nella regressione delle variabili indicatrici che permettono di individuare il settore¹⁴, la ripartizione geografica e la classe di ampiezza dell'impresa, al fine di calcolare, in ogni strato, la media condizionata con un più elevato grado di precisione.

E' opportuno sottolineare, prima di procedere con la discussione sui modelli, che in questa fase dell'analisi l'obiettivo principale è quello di imputare correttamente il valore dell'investimento mancante a livello di impresa - e stimare quindi, a partire dagli investimenti delle singole unità rispondenti, il tasso di variazione aggregato – piuttosto che preservare alcune proprietà generali della distribuzione degli investimenti (come, ad esempio, la variabilità totale nel campione della spesa in beni capitali). Si è optato, di conseguenza, per un metodo di imputazione deterministico, non includendo nel calcolo della media condizionata alcun effetto aleatorio. L'analisi sulla bontà del metodo adottato è stata svolta, quindi, sulla base delle considerazioni appena presentate.

I modelli utilizzati

I modelli descritti dalle relazioni (5) – (10) sono stati utilizzati per stimare, rispettivamente, i dati mancanti per l'anno t-2, t-1 e t per l'indagine primaverile e t-1, t e t+1 per quella autunnale:

$$(5) \quad I_i^{t-2} = const + \gamma_r + \alpha_d + \delta_s + Num_i + \delta_s Num_i + \varepsilon_{isr}$$

$$(6) \quad I_i^{t-1} = const + \gamma_r + \alpha_d + \delta_s + Num_i + I_{isr}^{t-2} + \delta_s Num_i + \delta_s I_{isr}^{t-2} + \varepsilon_{isr}$$

$$(7) \quad I_i^t = const + \gamma_r + \alpha_d + \delta_s + Num_i + I_{isr}^{t-2} + I_{isr}^{t-1} + \delta_s Num_i + \delta_s I_{isr}^{t-2} + \delta_s I_{isr}^{t-1} + \varepsilon_{isr}$$

$$(8) \quad I_i^{t-1} = const + \gamma_r + \alpha_d + \delta_s + Num_i + \delta_s Num_i + \varepsilon_{isr}$$

14 Per quanto riguarda le *dummy* di settore, sulla base di risultati preliminari, si è optato per l'inclusione nei modelli di indicatrici relativi agli 8 settori ISAE presentati in tabella 3 (cfr. nota 5).

$$(9) \quad I_i^t = const + \gamma_r + \alpha_d + \delta_s + Num_i + I_{isr}^{t-1} + \delta_s Num_i + \delta_s I_{isr}^{t-1} + \varepsilon_{isr}$$

$$(10) \quad I_i^{t+1} = const + \gamma_r + \alpha_d + \delta_s + Num_i + I_{isr}^{t-1} + I_{isr}^t + \delta_s Num_i + \delta_s I_{isr}^{t-1} + \delta_s I_{isr}^t + \varepsilon_{isr}$$

Nei modelli (5) – (10), I_i^t sono le spese per investimento della generica impresa i nell'anno t , Num_i è il numero di addetti dell'impresa, γ_r , α_d e δ_s sono le variabili *dummy* che individuano rispettivamente le 4 ripartizioni geografiche, le 3 classi di ampiezza e gli 8 settori ISAE, ε_{isr} è un disturbo aleatorio *i.i.d* con distribuzione normale.

I modelli presentati sono stati stimati utilizzando sia il metodo dei minimi quadrati ordinari (OLS), sia quello dei minimi quadrati ponderati (WOLS), ottenuto applicando il metodo dei minimi quadrati alle osservazioni moltiplicate per l'inverso della loro probabilità di inclusione nel campione, $1/f_i^{15}$. Il modello è stato sottoposto a verifica, utilizzando i consueti test di significatività congiunta. Nel dettaglio, si è adottato lo stimatore robusto della matrice di varianze e covarianze dei coefficienti stimati, proposto da White (1980), al fine di costruire opportunamente gli intervalli di confidenza delle stime, tenendo debitamente conto della presenza di eteroschedasticità. Lo stimatore robusto di White è, inoltre, fortemente raccomandato qualora si utilizzino i metodi di stima pesati, dal momento che l'introduzione dei pesi di strato introduce eteroschedasticità (si vedano a questo proposito Winship e Radbill, 1994).

Nelle tabelle 8 e 9 vengono presentati i valori dell' R^2 e gli esiti dei test Wald di significatività congiunta applicati ai modelli (5) – (7) per l'indagine di aprile 2004, e (8) - (10) per l'indagine di novembre 2003¹⁶, stimati utilizzando sia i minimi quadrati ordinari che il metodo pesato. Si osserva che, in generale, i valori dell' R^2 sono piuttosto elevati e i modelli sono, quindi, in grado di spiegare una gran parte della varianza del fenomeno in esame. Gli esiti dei Wald test mostrano invece che le *dummy* geografiche non sono significative in entrambi i data-set e, per l'indagine di novembre 2003, nemmeno le *dummy* relative alla classe di ampiezza passano i test. Inoltre, il confronto tra i risultati ottenuti tramite i minimi quadrati ordinari e quelli pesati, mostra l'esistenza di alcune differenze nell'esito dei test nei due distinti data-set, soprattutto per il modello (5) utilizzato in aprile 2004 e il modello (10) adottato per novembre 2003.

¹⁵ I metodi "pesati" sono spesso adottati quando si opera con dati campionari per i quali la probabilità di inclusione varia tra unità (Winship e Radbill, 1994).

¹⁶ Dati simili sono disponibili, su richiesta, per tutti i data-set dal novembre 2000.

Tab. 8

Stime per l'indagine di novembre 2003

	Regressione OLS						Regressione WOLS					
	Modello (8)		Modello (9)		Modello (10)		Modello (8)		Modello (9)		Modello (10)	
	Wald	P	Wald	P	Wald	P	Wald	P	Wald	P	Wald	P
γ_r	1,88	0,132	1,45	0,226	0,38	0,771	1,13	0,337	0,69	0,556	1,99	0,114
α_d	0,20	0,818	0,48	0,619	1,18	0,307	0,07	0,932	0,14	0,866	1,70	0,184
δ_s	1,85	0,075	1,75	0,121	2,33	0,074	1,63	0,125	0,75	0,584	0,81	0,522
$\delta_s Num$	11,63	0,000	7,95	0,000	2,06	0,046	9,33	0,000	5,27	0,000	1,39	0,205
$\delta_s I^{t-1}$	-	-	16,60	0,000	1,79	0,086	-	-	20,13	0,000	2,26	0,028
$\delta_s I^t$	-	-	-	-	4,05	0,000	-	-	-	-	5,43	0,000
<i>Num</i>	4,63	0,032	-	-	-	-	7,11	0,008	-	-	-	-
<i>Num + Inv_t-1</i>	-	-	36,50	0,000	-	-	-	-	58,12	0,000	-	-
<i>Num + Inv_t-1 + Inv_t</i>	-	-	-	-	20,71	0,000	-	-	-	-	11,55	0,000
<i>N</i>	725		694		626		725		694		626	
<i>K (inclusa intercetta)</i>	21		29		37		21		29		37	
<i>F (k, n)</i>	164		564		6836		124		588		4573	
<i>R²</i>	0,9580		0,9941		0,9784		0,9607		0,9933		0,9523	

Tab. 9

Stime per l'indagine di aprile 2004

	Regressione OLS						Regressione WOLS					
	Modello (5)		Modello (6)		Modello (7)		Modello (5)		Modello (6)		Modello (7)	
	Wald	P	Wald	P	Wald	P	Wald	P	Wald	P	Wald	P
γ_r	0,45	0,720	0,29	0,830	0,75	0,522	0,84	0,471	0,09	0,964	1,24	0,296
α_d	7,48	0,001	0,40	0,669	1,96	0,141	3,77	0,023	0,11	0,896	4,02	0,018
δ_s	3,75	0,001	1,71	0,129	0,62	0,599	1,16	0,324	1,28	0,269	1,59	0,135
$\delta_s Num$	7,52	0,000	16,72	0,000	2,28	0,027	3,39	0,001	8,03	0,000	3,51	0,001
$\delta_s I^{t-1}$	-	-	57,13	0,000	0,69	0,677	-	-	23,87	0,000	0,42	0,887
$\delta_s I^t$	-	-	-	-	1,04	0,404	-	-	-	-	2,21	0,032
<i>Num</i>	1,62	0,204	-	-	-	-	6,66	0,010	-	-	-	-
<i>Num + Inv_t-1</i>	-	-	15,72	0,000	-	-	-	-	26,29	0,000	-	-
<i>Num + Inv_t-1 + Inv_t</i>	-	-	-	-	4,60	0,003	-	-	-	-	5,85	0,000
<i>N</i>	755		751		665		755		751		665	
<i>K (inclusa intercetta)</i>	21		29		37		21		29		37	
<i>F (k, n)</i>	3074		.		.		5222		.		.	
<i>R²</i>	0,9379		0,9978		0,9958		0,9341		0,9961		0,9950	

L'analisi dei casi completi

Come già osservato in precedenza, i modelli (5) - (10) sono stati utilizzati allo scopo di stimare i dati mancanti a livello di impresa per poi calcolare, da questi, il tasso di variazione aggregato. Per valutare l'utilità del metodo proposto si è ritenuto opportuno, quindi, confrontare il tasso stimato eliminando dal data-set le imprese che non hanno risposto alla domanda sugli investimenti effettuati e/o programmati (la cosiddetta *complete case analysis*), con quello ottenuto stimando i dati mancanti con il metodo proposto. Per questo scopo, è stata condotta una simulazione per verificare se il tasso stimato sui "casi completi" differisse sostanzialmente da quello ottenuto adottando il metodo di imputazione; abbiamo voluto appurare, inoltre, quale tra i due metodi di imputazione, quello basato sui minimi quadrati ordinari e quello pesato, fornisse stime più accurate del tasso di variazione.

Nel dettaglio, è stato preliminarmente stimato il tasso di variazione r_C su un data-set privo di dati mancanti, ottenuto eliminando tutte le informazioni delle imprese contraddistinte da mancata risposta su almeno una delle variabili quantitative; il tasso r_C è stato quindi considerato come tasso di variazione "vero" ottenuto dai dati disponibili. Nel seguito, sono stati selezionati dal data-set tre sotto-campioni, uno relativo ad imprese con dato mancante dell'investimento in $t-1$ ($t-2$ per l'inchiesta di aprile), uno per quelle con dato mancante in t ($t-1$ per aprile) e l'ultimo per quelle con mancata risposta in $t+1$ (t in aprile), e per le imprese di ogni sottocampione si è sostituito il dato effettivo con una *flag* indicante la presenza di mancata risposta. I dati mancanti così generati sono stati successivamente sostituiti da quelli stimati, alternativamente, con il metodo dei minimi quadrati ordinari (OLS) e quello dei minimi quadrati pesati (WOLS).

Infine, si è stimato:

1. il tasso di variazione r_{CC} ottenuto con l'analisi dei casi completi (ovvero non considerando nel computo del tasso gli investimenti mancanti);
2. il tasso di variazione r_{OLS} ottenuto stimando i dati mancanti con il metodo OLS;
3. il tasso di variazione r_{WOLS} ottenuto stimando i dati mancanti con il metodo WOLS.

La procedura è stata ripetuta più volte estraendo, ad ogni esecuzione, i tre diversi sotto-campioni di imprese a cui sostituire il dato effettivo con l'indicazione di dato mancante. La tabella 10 mostra, per l'indagine di aprile 2004, la differenza tra il tasso "vero" r_C e la media ottenuta su 10 iterazioni indipendenti dei tassi r_{CC} , r_{OLS} e r_{WOLS} .

Tab. 10 Differenza tra stima del tasso ottenuto alternativamente con l'analisi dei casi completi o con i metodi di imputazione OLS e WOLS (risultati ottenuti su 10 iterazioni indipendenti; indagine di aprile 2004)

	r_C	CC		WOLS		OLS	
		$m(r_{CC})$	$r_C - m(r_{CC})$	$m(r_{WOLS})$	$r_C - m(r_{WOLS})$	$m(r_{OLS})$	$r_C - m(r_{OLS})$
t/t-1							
Piccole imprese	-24,02	-22,70	-1,32	-23,30	-0,72	-24,44	0,42
Medie imprese	-18,41	-16,01	-2,40	-17,80	-0,61	-18,89	0,48
Grandi imprese	-8,34	-8,49	0,15	-8,41	0,07	-8,47	0,13
Totale	-15,09	-14,29	-0,80	-14,79	-0,30	-15,43	0,34
t+1/t							
Piccole imprese	-16,46	-17,77	1,31	-14,39	-2,07	-12,39	-4,07
Medie imprese	-10,66	-12,81	2,15	-10,72	0,06	-10,64	-0,02
Grandi imprese	17,53	16,74	0,79	17,84	-0,31	17,89	-0,36
Totale	3,16	2,06	1,10	3,88	-0,72	4,49	-1,33

Possiamo osservare che l'analisi dei casi completi introduce maggior distorsione nelle stime di quanto non accada imputando i dati mancanti con i modelli di regressione (5) – (10). Inoltre, il metodo pesato fornisce, in media, stime più accurate di quanto non faccia quello basato sui minimi quadrati ordinari. Si può concludere, quindi, che sebbene i modelli proposti, applicati a titolo esemplificativo sui data-set di novembre 2003 e aprile 2004, presentino alcuni problemi di specificazione, la simulazione effettuata mostra che l'imputazione dei dati mancanti permette di ottenere una stima del tasso di variazione più accurata di quanto non si avrebbe con l'analisi dei casi completi. Tra i due metodi di imputazione proposti, infine, quello basato sui minimi quadrati pesati sembra fornire stime più precise dei valori mancati di quanto non faccia il metodo non pesato.

Va tuttavia rilevato che il metodo scelto introduce, nell'esercizio proposto, una lieve distorsione. Tale distorsione in ogni caso appare destinata a ridursi, fino ad annullarsi, aumentando il numero delle iterazioni. Segue da queste considerazioni la scelta di interpolare i dati mancanti, relativi agli investimenti effettuati e/o programmati, utilizzando i modelli (5) – (10), stimati con il metodo ponderato. I dati pubblicati nei notiziari di aprile e dicembre di ogni anno (cfr. Appendice) sono quindi ottenuti utilizzando il metodo descritto.

5 LE PROCEDURE DI AGGREGAZIONE

Sino all'inchiesta della primavera 2002, le procedure di aggregazione dei microdati raccolti a livello di impresa si basavano, per i dati quantitativi, su una semplice sommatoria dei dati relativi all'ammontare degli investimenti realizzati/programmati dall'impresa; per i dati qualitativi, si effettuavano medie ponderate, usando alternativamente come pesi il numero degli addetti e, per alcune domande, l'ammontare degli investimenti. La tabella 11 sintetizza le

Tab. 11 Metodi di aggregazione dei dati elementari a livello d'impresa

	Sino alla primavera del 2002	Dalla primavera del 2002 in poi	Nuova formula di aggregazione
Investimenti effettuati	Sommatoria semplice	Somma pesata	$\sum_i (w_i I_i^t)$, dove $w_i = N_i/n_i \forall i$ dello strato h
Variazione dei prezzi	Ponderazione con peso interno (numero di addetti dell'impresa)	Ponderazione con peso interno (investimenti dichiarati al tempo t)	$\Delta \pi^t = \sum \Delta \pi_i^t (w_i I_i^t / \sum w_i I_i^t)$
Destinazione della spesa			
- per macro-tipologia	Ponderazione con peso interno (investimenti dichiarati al tempo t)	Ponderazione con peso interno (quota dichiarata degli investimenti destinata alla macro-tipologia in esame)	$D_j^t = \sum \alpha_{ij}^t (w_i I_i^t / \sum (w_i I_i^t))$
- per sotto-tipologia	Ponderazione con peso interno (addetti) ed esterno (valore aggiunto di strato)	Ponderazione con peso interno (quota dichiarata degli investimenti destinata alla macrotipologia di riferimento)	$d_{sj}^t = \sum s_i (\alpha_{ij}^t w_i I_i^t / \sum (\alpha_{ij}^t w_i I_i^t))$
Tendenza investimenti (risposta qualitativa)	Ponderazione con peso interno (addetti) ed esterno (valore aggiunto di strato)	Ponderazione con peso interno (investimenti dichiarati al tempo t)	$T_s^{t+1} = \sum s_i (w_i I_i^{t+1} / \sum (w_i I_i^{t+1}))$
Leasing - ricorso al leasing - incidenza %	Ponderazione con peso interno (addetti) ed esterno (valore aggiunto di strato)	Ponderazione con peso interno (investimenti dichiarati al tempo t-1)	$L_s^{t-1} = \sum s_i (w_i I_i^{t-1} / \sum (w_i I_i^{t-1}))$ $\Phi^{t-1} = \sum \phi_i^s (w_i I_i^{t-1} / \sum (w_i I_i^{t-1}))$
Modifiche ai piani di investimento	Ponderazione con peso interno (investimenti dichiarati)	Ponderazione con peso interno (investimenti dichiarati al tempo t)	$M_s^t = \sum s_i (w_i I_i^t / \sum (w_i I_i^t))$
Fattori influenzanti gli investimenti	Ponderazione con peso interno (investimenti dichiarati)	Ponderazione con peso interno (investimenti dichiarati al tempo t)	$F_{kj}^t = \sum s_{ij}^k (w_i I_i^t / \sum (w_i I_i^t))$

procedure “storicamente” seguite dall’ISAE e le confronta con quelle adottate dall’Istituto dalla rilevazione della primavera del 2002, utilizzate per ricostruire la banca dati ISAEINV relativa al periodo 2000-2004. Nel seguito, il par. 5.1 descrive nel dettaglio le procedure attualmente utilizzate per l’aggregazione dei dati quantitativi, e il par. 5.2 effettua la stessa descrizione per i dati qualitativi.

5.1 L’aggregazione dei dati quantitativi

La procedura in passato utilizzata dall’ISAE per l’aggregazione dei dati elementari quantitativi raccolti a livello d’impresa, basata sulla semplice sommatoria delle informazioni ottenute a livello d’impresa, appariva poco soddisfacente: in effetti, esistono, come ha mostrato il par. 3, alcune differenze tra la struttura del campione effettivamente rispondente all’inchiesta e l’Universo statistico di riferimento. Le distorsioni sono dovute, come accennato in precedenza, da un lato alla necessità di prendere in esame un campione “ragionato”, in grado di fornire rappresentanza anche a strati campionari che, in via strettamente teorica, sarebbero potuti essere scarsamente presenti (ad esempio, alcuni settori in alcune regioni del Mezzogiorno), ma che rivestono invece notevole importanza da un punto di vista economico. D’altro lato, le distorsioni sono anche legate al fatto che il tasso di risposta all’indagine sugli investimenti è notevolmente inferiore a quello dell’indagine mensile sulla situazione congiunturale delle imprese manifatturiere ed estrattive, per le ragioni esposte nel par. 3.2. Inoltre, come osservato precedentemente (cfr. par. 3.2), nel caso dell’Inchiesta sugli Investimenti le potenziali distorsioni potrebbero essere più gravi di quelle eventualmente riscontrabili nell’inchiesta mensile in quanto vi saranno, generalmente, molti più strati vuoti. Per l’aggregazione, si è optato quindi per la fusione di alcuni strati campionari tra loro, cercando di minimizzare il numero di celle vuote e in questo modo ridurre le potenziali distorsioni. Nel dettaglio, si è mantenuta la stratificazione nelle 3 classi di ampiezza e nei 22 settori aggregando, invece, le 19 regioni nelle 4 usuali ripartizioni geografiche (Nord Ovest, Nord Est, Centro e Sud); di conseguenza i pesi di strato sono stati ricalcolati sommando tra loro quelli relativi agli strati originali. Dal nuovo schema di stratificazione risultano, quindi, 4 (ripartizioni geografiche) \times 3 (classi di ampiezza) \times 22 (settori) = 264 strati campionari, con in media 800 interviste.

In analogia alla procedura seguita dalla Banca d’Italia nella stima degli investimenti sulla base della propria indagine periodica sulle imprese della

trasformazione industriale¹⁷, si è adottata quindi per i dati quantitativi dell'inchiesta una procedura di stima in cui l'investimento di ciascuna impresa è ponderato per un fattore di "espansione all'Universo", definito, per ogni impresa i -esima, come il rapporto tra la numerosità di strato nella Popolazione di riferimento e la numerosità di strato effettiva del campione. In pratica si ha che:

$$(11) \quad I^t = \sum_{i=1}^n w_i I_i^t \quad \text{dove } w_i = N_h / n_h \quad \forall i \in \text{strato } h$$

dove I^t è la stima degli investimenti totali del settore manifatturiero ed estrattivo al tempo t , I_i^t sono gli investimenti della singola impresa i -esima, sempre al tempo t , w_i è il fattore di espansione all'Universo, in cui N_h e n_h sono, rispettivamente, la numerosità della Popolazione e quella del campione per ogni h -esimo strato in cui è compresa l'impresa i -esima. Come mostrato in Battipaglia (2000), il sistema di pesi w_i permette di tenere conto, nella stima della quantità complessiva degli investimenti al tempo t , della struttura del disegno campionario. L'ipotesi sottostante la (11) è che ciascuna impresa i -esima sia comunque rappresentativa delle imprese facenti parte dell' h -esimo strato dell'Universo: in pratica, l'utilizzo del fattore di espansione all'Universo consente di attribuire alle restanti imprese di quello strato (non oggetto di rilevazione) lo stesso valore stimato sui dati forniti dalle imprese facenti parte del campione.

A partire dalla stime degli investimenti complessivi per ogni anno t rilevato in ogni singolo questionario, è possibile quindi ricavare una stima del tasso di variazione degli investimenti in due anni contigui, definito come:

$$(12) \quad \Delta I_t = \frac{\sum_i w_i I_i^t}{\sum_i w_i I_i^{t-1}} - 1$$

In pratica, il tasso di variazione degli investimenti è calcolato ponderando i dati relativi agli investimenti di ciascuna impresa con il fattore di espansione all'Universo w_i precedentemente descritto¹⁸.

17 Si veda Banca d'Italia (1996).

18 Il tasso di variazione degli investimenti può anche essere calcolato applicando la winsorizzazione secondo la formula (3).

5.2 L'aggregazione dei dati qualitativi

La procedura storicamente seguita dall'ISAE per l'aggregazione dei dati di natura qualitativa ricalcava da vicino quella seguita per aggregare le risposte al questionario mensile sulle imprese manifatturiere ed estrattive. In pratica, per la maggioranza delle domande qualitative presenti nell'inchiesta sugli investimenti, si effettuava una "doppia ponderazione", utilizzando come peso "interno" il numero degli addetti dell'impresa interpellata e poi, a livello di strato, il peso "esterno" costituito dal valore aggiunto di strato, di fonte ISTAT (si veda ancora la Tab. 11). Facevano eccezione le domande relative alla destinazione degli investimenti per macrotipologia, le domande relative alle modifiche e ai fattori influenzanti i piani di investimento e alla ripartizione degli investimenti su base territoriale e per categorie di beni; per queste domande, veniva usato come peso l'ammontare complessivo degli investimenti dichiarati dall'impresa.

Si è ritenuto opportuno aggiornare anche le procedure di aggregazione dei dati qualitativi, sulla base della considerazione che, nella maggior parte dei casi, il peso da utilizzare dovrebbe essere costituito dalle informazioni fornite dalle stesse imprese circa l'ammontare degli investimenti effettivamente realizzati o programmati nel periodo di riferimento. In particolare, riguardo alle principali variabili qualitative contenute nel questionario, si ha che:

- a. Per la domanda relativa alla destinazione degli investimenti per macrotipologia (rinnovo impianti obsoleti, razionalizzazione, ampliamento della capacità, spese ambientali, altri motivi), per la quale è prevista una modalità di risposta semi-quantitativa¹⁹, si utilizza come peso il rapporto tra gli investimenti effettuati dall'impresa i -esima al tempo t e l'ammontare degli investimenti totali stimati per il tempo t . In pratica si ha che:

$$(13) \quad D_j^t = \sum_{i=1}^n \alpha_{ij}^t \frac{w_i I_i^t}{\sum_{i=1}^n w_i I_i^t}$$

dove D_j^t è la quota di investimenti destinata alla macrotipologia j -esima ($j=1,2,3,4,5$) e α_{ij}^t rappresenta la quota di investimenti destinata a

19 Le imprese indicano direttamente la quota di investimenti che deve essere dedicata a ciascuna macrotipologia.

ciascuna macrotipologia j dichiarata da ogni impresa. La somma in j dei D_j^t risulta per definizione uguale all'unità.

- b. Per la domanda relativa alla destinazione degli investimenti per sottotipologia²⁰, per la quale le risposte sono di tipo strettamente qualitativo (le imprese indicano cioè se effettuano o no tale spesa, senza indicare in alcun modo in quale proporzione gli investimenti si suddividono tra tali modalità), si usa come peso, a livello d'impresa, l'ammontare degli investimenti effettivamente destinato alla relativa modalità dalla singola impresa, normalizzato per la stima di questo ammontare per l'intera popolazione. In pratica si ha che:

$$(14) \quad d_{sj}^t = \sum_{i=1}^n \left(s_i \frac{\alpha_{ij}^t w_i I_i^t}{\sum_{i=1}^n \alpha_{ij}^t w_i I_i^t} \right)$$

dove d_{sj}^t è la frequenza delle risposte che indicano l'effettuazione degli investimenti in una determinata sottotipologia s -esima appartenente alla macrotipologia j , $\alpha_{ij}^t I_i^t$ è la quota di investimenti destinata alla macrotipologia j -esima cui appartiene la sottotipologia s -esima e s_i è l'indicazione (di tipo binario, con valore pari a 0 nel caso in cui non si effettua tale tipo di investimento, e pari ad 1 nel caso in cui invece si effettua) da parte dell'impresa i -esima di effettuare investimenti nella sottotipologia s -esima. E' opportuno sottolineare che in questo caso la sommatoria dei vari d_{sj}^t (tre sottotipologie nel caso degli investimenti di razionalizzazione; due per quelli di ampliamento della capacità, quattro per le spese ambientali) può risultare superiore all'unità, in quanto ciascuna impresa può fornire risposte multiple, cioè indicare più di una sottotipologia all'interno di ciascuna macrotipologia di spesa.

- c. Per la domanda relativa alla tendenza degli investimenti, per la quale è prevista una risposta di tipo qualitativo a tre modalità (gli investimenti

20 Cioè, nel caso della razionalizzazione: per automazione e meccanizzazione, per l'introduzione di nuove tecniche, per economizzare energia; nel caso dell'ampliamento della capacità produttiva: nel quadro dei programmi produttivi esistenti, o per introdurre nuovi prodotti; nel caso delle spese ambientali: per protezione suolo, aria e acqua, per protezione natura e paesaggio, per trattamento rifiuti, per riduzione del rumore.

previsti in t+1 possono essere superiori, uguali o inferiori a quelli in t) si usa come peso il rapporto tra gli investimenti dichiarati/programmati dall'impresa i -esima e l'ammontare degli investimenti dichiarati/programmati dal totale delle imprese per il periodo t:

$$(15) \quad T_s^{t+1} = \sum_{i=1}^n \left(s_i \frac{w_i I_i^{t+1}}{\sum_{i=1}^n w_i I_i^{t+1}} \right)$$

dove s è una variabile binaria (che assume valore 1 nel caso in cui l'impresa abbia selezionato l' s -esima modalità di risposta e 0 altrimenti) e T_s^{t+1} è la frequenza delle risposte relativa alla generica modalità s .

- d. Per la domanda concernente il ricorso al *leasing* da parte dell'impresa nel periodo t-1 per la quale sono previste modalità di risposta qualitative, (le imprese dichiarano se hanno fatto o meno ricorso al *leasing* ed in caso di risposta affermativa per quali motivi), si usa come peso interno il rapporto tra gli investimenti effettuati a livello di impresa al tempo t-1 e l'ammontare degli investimenti effettuati al tempo t-1. La formula è la seguente:

$$(16) \quad L_s^{t-1} = \sum_{i=1}^n \left(s_i \frac{w_i I_i^{t-1}}{\sum_{i=1}^n w_i I_i^{t-1}} \right)$$

dove L_s^{t-1} rappresenta la frequenza di risposte che indicano l'aver fatto o meno ricorso al leasing come forma di investimento, s è l'indicazione di tipo binario con valore pari ad 1 nel caso in cui si ricorre al leasing e 0 nel caso in cui non vi si ricorre.

Anche per la domanda concernente i motivi per cui si è fatto ricorso al *leasing*²¹, per la quale sono previste 6 modalità di risposta qualitative, si usa come peso interno il rapporto tra l'investimento relativo all'impresa i -esima al tempo t-1 e l'ammontare degli investimenti al tempo t-1:

21 Acquisto di Capannoni, Macchine e attrezzature per la produzione, Mezzi di trasporto, Impianti e/o attrezzature per l'elaborazione di dati, *office automation*, Altri motivi.

$$(17) \quad l_s^{t-1} = \sum_{i=1}^n \left(s_i \frac{w_i I_i^{t-1}}{\sum_{i=1}^n w_i I_i^{t-1}} \right)$$

dove l_s^{t-1} è la frequenza delle risposte relativa al motivo s-esimo per cui si è fatto ricorso al *leasing* e s è l'usuale variabile binaria. Anche in questo caso, analogamente a quanto avviene per la domanda relativa alla destinazione degli investimenti per sottotipologia, occorre ricordare che la sommatoria dei vari l_s^{t-1} può risultare superiore all'unità, in quanto ciascuna impresa può fornire risposte multiple, cioè indicare più di un motivo per cui ha fatto ricorso al *leasing* come forma di investimento.

- e. Per la domanda riguardante l'incidenza percentuale del leasing sul totale degli investimenti, per la quale è prevista una risposta di tipo quantitativo, (la percentuale di incidenza del *leasing*), si usa come peso interno il rapporto tra l'investimento relativo all'impresa i -esima al tempo $t-1$ e l'ammontare degli investimenti effettuati al tempo $t-1$.

$$(18) \quad \Phi^{t-1} = \sum_{i=1}^n \left(\phi_i^{t-1} \frac{w_i I_i^{t-1}}{\sum_{i=1}^n w_i I_i^{t-1}} \right)$$

dove ϕ_i^{t-1} indica l'incidenza percentuale dell'impresa i -esima e Φ^{t-1} indica l'incidenza percentuale per il totale delle imprese.

- f. Per la domanda relativa alla variazione dei prezzi per la quale è prevista una risposta quantitativa (le variazioni dei prezzi verificatesi negli anni t e $t-1$), si utilizza come peso, a livello di impresa, il rapporto tra l'investimento dichiarato da ciascuna impresa e il totale degli investimenti effettuati. Con riferimento al generico anno t si ha che:

$$(19) \quad \Delta \pi_t = \sum_{i=1}^n \left(\Delta \pi_i^t \frac{w_i I_i^t}{\sum_{i=1}^n w_i I_i^t} \right)$$

dove $\Delta \pi_t$ è la variazione dei prezzi complessiva e $\Delta \pi_i^t$ è la variazione dei prezzi dell'impresa i -esima nell'anno t .

- g. Per la domanda relativa alle modifiche ai piani di investimento per la quale è prevista una domanda qualitativa a tre modalità di risposta (Sì in più, Sì in meno e NO) si usa come peso interno l'ammontare degli investimenti dichiarati dall'impresa al tempo t diviso la stima dell'ammontare di investimenti al tempo t:

$$(20) \quad M_s^t = \sum_{i=1}^n \left(S_i \frac{w_i I_i^t}{\sum_{i=1}^n w_i I_i^t} \right)$$

dove M_s^t è la frequenza relativa alla modalità di risposta s -esima e s è l'usuale variabile binaria.

Anche per le domande relative ai motivi per cui sono stati modificati i piani di investimento (sia in aumento, sia in diminuzione), per le quali sono previste 6 modalità di risposta qualitative²², si usa come peso interno il rapporto tra l'ammontare degli investimenti dichiarati dall'impresa al tempo t e la stima dell'ammontare di investimenti al tempo t:

$$(21) \quad m_{sj}^t = \sum_{i=1}^n \left(S_{ij} \frac{w_i I_i^t}{\sum_{i=1}^n w_i I_i^t} \right)$$

dove m_{sj}^t è la frequenza delle risposte relative alla motivazione j -esima e s è la consueta variabile binaria.

- h. Per la domanda relativa ai fattori influenzanti gli investimenti (nell'anno appena trascorso o per l'anno successivo), si richiede alle imprese di esprimere valutazioni sull'influenza sulle decisioni di investimento di fattori quali: l'evoluzione della domanda, la disponibilità dei mezzi finanziari o attese sui profitti, fattori tecnici²³ e altri fattori. Per ogni fattore l'impresa indica se questo è considerato molto importante, stimolante, senza

22 Per le modifiche in aumento le modalità previste sono: maggiori possibilità di finanziamento, modifiche nella struttura della domanda, misure di politica economica, variazioni dei prezzi dei beni d'investimento, motivi amministrativi interni, altri motivi; per le modifiche in diminuzioni le modalità sono: minori possibilità di finanziamento, modifiche ai piani di produzione, misure di politica economica, variazioni dei prezzi dei beni d'investimento, motivi amministrativi interni, altri motivi.

23 Quali, ad esempio, sviluppo tecnologico, disponibilità di manodopera e sua capacità di adattamento alle nuove tecnologie.

influenza, limitativo, molto limitativo. Per ogni fattore k ($k=1,\dots,4$) le percentuali relative alle j opzioni ($j=1,\dots,5$) si calcolano usando come peso interno l'ammontare degli investimenti dichiarati dall'impresa al tempo t diviso l'ammontare di investimenti al tempo t :

$$(22) \quad F_{kj}^t = \sum_{i=1}^n \left(s_{ij}^k \frac{w_i I_i^t}{\sum_{i=1}^n w_i I_i^t} \right)$$

dove s_{ij}^k è la variabile binaria che assume valore 1 se l'impresa ha selezionato l'opzione j per il fattore k e F_{kj}^t è la frequenza relativa alla modalità di risposta j -esima del fattore k ²⁴.

6 CONSIDERAZIONI CONCLUSIVE

L'inchiesta semestrale ISAE sugli investimenti si differenzia da quella mensile, rivolta alle imprese manifatturiere ed estrattive, in quanto fornisce anche indicazioni quantitative relative ai piani di investimento e alla loro destinazione, oltre a quelle qualitative quali le modifiche ai piani di investimento, i fattori influenzanti le spese e il ricorso al leasing. Il trattamento di variabili quantitative implica la necessità di un più attento controllo di qualità e coerenza delle risposte, al fine di garantire l'affidabilità dei risultati, soprattutto data la possibilità di elevati tassi di non risposta. Gli elevati tassi di non risposta, infatti, e i problemi di affidabilità legati agli errori campionari, possono ridurre la qualità complessiva dei risultati dell'indagine.

A questo proposito, il presente lavoro mostra che, generalmente, soltanto il 20% del campione effettivo (su un totale di 4,000 unità componenti il campione teorico dell'inchiesta mensile sulle imprese manifatturiere ed estrattive) risponde all'indagine; inoltre, l'analisi *ex-post* dei questionari dimostra che il tasso di non risposta per la domanda sugli investimenti realizzati o pianificati può raggiungere il 6%. D'altra parte, in riferimento all'affidabilità dei

24 Per non perdere le informazioni qualitative fornite dalle imprese aventi peso $I_i^t = 0$, il peso di tali imprese è stato sostituito con una media degli investimenti del triennio.

risultati, l'analisi della distribuzione delle risposte delle imprese alle variabili quantitative mette in evidenza la presenza di valori anomali, i quali persistono nonostante il controllo deterministico effettuato per l'individuazione di errori non campionari.

Per tali ragioni, in questo lavoro si propone l'uso di appropriate tecniche statistiche per trattare, rispettivamente, la presenza di dati mancanti e di dati anomali. Nel dettaglio, per il trattamento delle mancate risposte sugli investimenti realizzati o programmati, si adotta un semplice modello econometrico che stima i dati mancanti sulla base delle informazioni contenute nei questionari compilati dalle imprese nel corso dell'indagine: il numero di addetti, il settore di appartenenza, la ripartizione geografica di appartenenza e, qualora disponibili, i valori degli investimenti effettuati negli anni passati. Sebbene i modelli stimati presentino alcuni problemi di specificazione, essi forniscono stime dei dati mancanti sufficientemente accurate da garantire un valore affidabile per il tasso aggregato di variazione degli investimenti. Un esercizio di simulazione mostra, infatti, che l'adozione del metodo di imputazione permette di ottenere stime del tasso di variazione più accurate di quanto non si abbia con l'analisi dei casi completi (cioè omettendo dal computo del tasso le imprese che non hanno risposto alla domanda sugli investimenti). Sono, ad ogni modo, necessarie ulteriori analisi, tramite le quali formulare un giudizio più accurato sulla bontà del metodo proposto; in particolare, si dovrà in futuro utilizzare più diffusamente le tecniche di simulazione e, eventualmente, modificare i modelli di stima al fine di ottenere una migliore specificazione. Inoltre, l'uso di tecniche di intervista moderne (quale ad esempio la CATI o la somministrazione del questionario, e corrispondente raccolta delle informazioni, tramite Web) potrà certamente incrementare il tasso di partecipazione all'indagine, come già accade per l'inchiesta mensile alle imprese manifatturiere ed estrattive.

Per quanto riguarda il trattamento dei dati anomali, il lavoro mostra che la distribuzione delle risposte alla domanda sugli investimenti effettuati e/o programmati, è fortemente asimmetrica e presenta una coda destra "allungata", con un addensarsi delle osservazioni nella classe aperta finale, anche dopo avere eliminato i dati anomali non rappresentativi tramite le consuete procedure deterministiche di controllo di qualità. L'applicazione delle medie winsorizzate ha confermato la presenza nel data-set di *outlier* rappresentativi, mostrando che la stima del tasso di variazione degli investimenti risente notevolmente della presenza di queste osservazioni estreme. Queste considerazioni suggeriscono l'opportunità di pubblicare le stime winsorizzate. Anche su questi temi, in ogni caso, sono necessari ulteriori approfondimenti, soprattutto per analizzare in dettaglio l'effetto dell'applicazione di questi metodi sulle proprietà degli stimatori

utilizzati. Per concludere, sono state presentate le tecniche di aggregazione per i dati quantitativi e qualitativi utilizzati dall'ISAE a partire dal 2000. I risultati delle indagini condotte da novembre 2000 ad aprile 2004 sono, invece, presentati in appendice. Rispetto ai risultati ottenuti, sarà necessario nel prossimo futuro verificare la bontà degli stimatori utilizzati confrontandoli con le statistiche ufficiali sugli investimenti dell'industria manifatturiera italiana.

APPENDICE:**La Banca dati ISAEINV (novembre 2000 - novembre 2004)**

Tab. A **Tasso di variazione degli investimenti
delle imprese manifatturiere ed estrattive (1)
(valori percentuali)**

	2000/1999	2001/2000	2002/2001	2003/2002	2004/2003	2005/2004
Novembre 2000	3,5	-0,6				
Aprile 2001	5,0	-2,4				
Novembre 2001		-0,2	-3,6			
Aprile 2002		-0,5	-3,0			
Novembre 2002			-2,4	-9,5		
Aprile 2003			7,9	-11,5		
Novembre 2003				-21,3	-3,0	
Aprile 2004				-13,5	-2,3	
Novembre 2004					-5,9	-2,1

(1) Medie winsorizzate ottenute sostituendo ai valori estremi (sia di segno positivo, sia di segno negativo) delle distribuzioni delle variazioni annue degli investimenti i valori del 5° e 95° percentile.

Tab. B **La tendenza degli investimenti
(valori percentuali)**

Data	Previsioni per l'anno	Superiori	Uguali	Inferiori	Saldo
Novembre 2000	2002	12	40	48	-36
Aprile 2001	2002	31	37	32	-1
Novembre 2001	2003	28	31	41	-13
Aprile 2002	2003	28	45	27	1
Novembre 2002	2004	32	37	31	1
Aprile 2003	2004	38	28	34	4
Novembre 2003	2005	13	42	45	-32
Aprile 2004	2005	46	26	28	18
Novembre 2004	2006	14	54	54	-18

**Tab. C Destinazione della spesa per investimenti per macrotipologia
(percentuale sull'ammontare degli investimenti)**

Data	2000				2001				2002						
	SOST.	RAZ.	AMPL.	ALTRE	SOST.	RAZ.	AMPL.	ALTRE	SOST.	RAZ.	AMPL.	ALTRE			
Novembre 2000	30	23	35	12	28	21	39	12							
Aprile 2001	23	24	40	13	28	21	36	15							
Novembre 2001					27	31	31	11	28	31	32	9			
Aprile 2002					29	31	31	9	31	32	28	9			
Novembre 2002									30	24	35	11			
Aprile 2003									34	27	28	11			
Novembre 2003															
Aprile 2004															
Novembre 2004															
Data	2003					2004					2005				
	SOST.	RAZ.	AMPL.	AMBI.	ALTRE	SOST.	RAZ.	AMPL.	AMBI.	ALTRE	SOST.	RAZ.	AMPL.	AMBI.	ALTRE
Novembre 2000															
Aprile 2001															
Novembre 2001															
Aprile 2002															
Novembre 2002	32	22	36	-	10										
Aprile 2003	35	30	26	-	9										
Novembre 2003	34	24	29	-	13	36	25	28	-	11					
Aprile 2004	25	26	37	6	6	28	25	35	7	5					
Novembre 2004						29	22	33	10	6	30	24	31	9	6

Legenda:

SOST.=sostituzione e/ rinnovo impianti obsoleti;

RAZ.=razionalizzazione;

AMPL.=ampliamento capacità produttiva;

AMBI.= spese ambientali;

ALTRE =altre destinazioni (sicurezza, ecc.).

**Tab. D La destinazione della spesa per investimenti per sottotipologia
(valori percentuali)**

a - Razionalizzazione(*)																		
Data	2000			2001			2002			2003			2004			2005		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Novembre 2000	65	50	32	57	41	27												
Aprile 2001	68	54	43	75	57	50												
Novembre 2001				36	32	55	37	59	15									
Aprile 2002				76	26	57	35	64	55									
Novembre 2002							48	45	33	55	44	29						
Aprile 2003							39	56	46	36	58	42						
Novembre 2003										57	50	15	55	38	13			
Aprile 2004										41	62	44	47	57	36			
Novembre 2004													67	46	28	65	43	25
b - Ampliamento(*)																		
Data	2000		2001		2002		2003		2004		2005							
	4	5	4	5	4	5	4	5	4	5	4	5						
Novembre 2000	57	42	50	43														
Aprile 2001	78	72	73	68														
Novembre 2001			72	56	62	47												
Aprile 2002			64	51	61	47												
Novembre 2002					71	55	69	59										
Aprile 2003					62	27	44	35										
Novembre 2003							48	49	41	45								
Aprile 2004							31	64	69	68								
Novembre 2004									66	58	63	48						
c - Spese ambientali (*)																		
Data	2003				2004				2005									
	6	7	8	9	6	7	8	9	6	7	8	9						
Aprile 2004	41	9	32	18	27	11	33	17										
Novembre 2004					24	2	15	13	34	8	23	21						

Legenda: 1 = per automatizzare e meccanizzare i processi produttivi esistenti;
 2 = per introdurre nuove tecniche produttive;
 3 = per economizzare l'energia;
 4 = nel quadro dei programmi produttivi esistenti;
 5 = per introdurre nuovi prodotti;
 6 = protezione suolo, aria, acqua;
 7 = protezione natura e paesaggio;
 8 = trattamento rifiuti;
 9 = riduzione rumore.

(*) la somma delle percentuali può differire da 100 a causa della presenza di mancate risposte.

Tab. E

**Il ricorso al leasing
(valori percentuali)**

	SI	NO	Se SI per						Incidenza leasing (%)
			Capannoni	Macchine produzione	Mezzi di trasporto	Impianti e/o attrezzature per elaborazione dati	Impianti e/o attrezzature per l'automazione d'ufficio	Altri motivi	
Aprile 2001									
TOTALE	53	47	17	22	32	7	21	57	11,1
Piccole imprese	51	49	9	22	32	3	6	63	23,8
Medie imprese	49	51	25	34	30	12	15	55	16,8
Grandi imprese	54	46	22	16	33	9	43	51	2,4
Aprile 2002									
TOTALE	35	65	22	19	28	5	27	44	9,5
Piccole imprese	53	47	11	18	33	7	5	58	29,3
Medie imprese	52	48	36	32	39	6	7	43	17,4
Grandi imprese	26	74	25	14	20	2	49	35	1,6
Aprile 2003									
TOTALE	32	68	25	29	38	7	3	48	11,2
Piccole imprese	47	53	14	18	38	5	2	53	20,7
Medie imprese	52	48	37	45	27	9	4	45	22,6
Grandi imprese	17	83	33	31	47	9	4	42	2,5
Aprile 2004									
TOTALE	52	48	37	11	48	3	4	50	7,3
Piccole imprese	46	54	25	17	35	6	3	54	14,2
Medie imprese	46	54	34	39	52	3	18	43	16,6
Grandi imprese	55	45	42	3	52	2	2	50	2,3

**Tab. F Modifiche dei piani di investimento previste per l'anno in corso
(valori percentuali)**

Data	Si in più	Si in meno	NO	Modifiche in aumento						Modifiche in diminuzione					
				1	2	3	4	5	6	1	2	3	4	5	6
Novembre 2000	34	34	32	26	55	11	9	10	28	3	87	6	1	21	23
Novembre 2001	17	49	34	26	38	13	4	13	39	4	81	58	1	2	38
Novembre 2002	27	22	51	15	24	31	9	15	49	8	74	5	1	12	15
Novembre 2003	24	17	59	14	40	13	9	20	42	13	52	30	1	15	28
Novembre 2004	44	17	39	13	15	0	5	64	20	22	35	25	1	11	17

Legenda: 1 = Maggiori possibilità di finanziamento; 2 = Modifiche nella struttura della domanda; 3 = Misure di politica economica; 4 = Variazione dei prezzi dei beni di investimento; 5 = Motivi amministrativi interni; 6 = Altri motivi.

**Tab. G Fattori influenzanti gli investimenti
(valori percentuali)**

	Anno corrente					Anno successivo					
	MI	ST	NI	LI	ML	MI	ST	NI	LI	ML	
<i>Novembre 2000</i>											
Domanda	36	56	7	1	0	Domanda	38	53	7	2	0
Risorse finanziarie	49	34	12	4	1	Risorse finanziarie	48	34	13	4	1
Fattori tecnici	39	44	12	3	2	Fattori tecnici	23	54	15	4	4
Altri fattori	8	34	42	7	9	Altri fattori	13	18	49	9	11
<i>Novembre 2001</i>											
Domanda	69	20	6	4	1	Domanda	64	22	6	2	6
Risorse finanziarie	22	22	47	7	2	Risorse finanziarie	26	26	40	7	1
Fattori tecnici	16	32	50	1	1	Fattori tecnici	16	36	46	1	1
Altri fattori	5	9	48	33	5	Altri fattori	5	12	38	39	6
<i>Novembre 2002</i>											
Domanda	66	20	9	4	1	Domanda	62	22	11	3	2
Risorse finanziarie	28	28	34	8	2	Risorse finanziarie	28	27	18	25	2
Fattori tecnici	21	59	14	4	2	Fattori tecnici	18	62	16	3	1
Altri fattori	13	36	38	7	6	Altri fattori	8	35	44	7	6
<i>Novembre 2003</i>											
Domanda	47	30	14	7	2	Domanda	54	27	9	4	6
Risorse finanziarie	29	39	17	13	2	Risorse finanziarie	32	33	20	12	3
Fattori tecnici	34	40	21	4	1	Fattori tecnici	32	41	23	3	1
Altri fattori	9	22	45	12	12	Altri fattori	6	21	49	12	12
<i>Novembre 2004</i>											
Domanda	53	32	10	3	2	Domanda	53	35	10	2	0
Risorse finanziarie	36	44	12	6	2	Risorse finanziarie	41	41	10	5	3
Fattori tecnici	32	48	15	2	3	Fattori tecnici	32	46	18	3	1
Altri fattori	6	15	63	9	7	Altri fattori	7	16	64	7	6

Legenda: MI = molto importante; ST = stimolante; NI = senza influenza; LI = limitativo; ML = molto limitativo.

BIBLIOGRAFIA

- Battipaglia, P. (2000), "Robust Estimates of Investment from the Bank of Italy's Business Survey", London School of Economics Statistics Research Report, LSERR56, Aprile 2000.
- Banca d'Italia (1996), "Metodi e risultati dell'indagine sugli investimenti delle imprese industriali", Supplementi al Bollettino Statistico, Note Metodologiche e informazioni statistiche, Nuova Serie, Anno VI, n. 59, 21 Ottobre.
- Banca d'Italia (vari anni), Indagine sugli investimenti delle imprese industriali, Sintesi dei risultati, Roma.
- Chambers, R.L. (1986), "Outlier Robust Final Population Estimation", Journal of the American Statistical Association, Vol. 81, pp. 1063-69.
- Commissione Europea (2002), The Joint Harmonized EU Programme of Business and Consumer Surveys, User Guide 2002.
- Gismondi R. (2002) "Un confronto tra metodi di identificazione di osservazioni outliers in indagini longitudinali finalizzate alla stima di una variazione: proposte teoriche e verifiche empiriche", Statistica, n. 62(2), Università degli Studi di Bologna, Bologna, pag. 323-343.
- Hidiroglou e M. Srinath K. (1981), "Some estimators of a population total from simple random sample containing large units", Journal of the American Statistical Association, Vol.76, pp. 690-695.
- Hulliger B. (1999), "Simple and robust estimators for sampling", American Statistical Association, Proceedings of the survey research methods section ASA (1978-2000), <http://www.amstat.org/sections/srms/Proceedings/allyears.html>.
- Malgarini M., Margani P. e B.M. Martelli (2005), "Re-engineering the ISAE Manufacturing Survey", Documento di lavoro n. 47/2005, ISAE, Roma.
- Martelli B. M. e G. Rocchetti (2004), "The ISAE Market Services Survey: Methodological Upgrading, Survey Reliability, First Empirical Results", 27th CIRET Conference, Varsavia, Settembre 2004.
- OCSE (2003), Business Tendency Surveys: A Handbook, Parigi 2003.
- Smith T.M.F. (1987), "Influential observation in survey sampling", Journal of Applied Statistics, Vol. 14, pp.143-152.

White H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica* vol. 48, pp. 817-838.

Winship C. and L. Radbill (1994), "Sampling Weights and Regression Analysis", *Sociological Methods and Research*, Vol. 23, No. 2, pp 230-257.

Working Papers available:

n. 25/02	M. CAGIANO DE AZEVEDO	Recent Developments in the Environmental Debate Before and after the Kyoto Protocol: A Survey
n. 26/02	M. BOVI	The Nature of the Underground Economy. Some Evidence from OECD Countries
n. 27/02	R. BASILE	Acquisition versus Greenfield Investment: the Location of Foreign Manufacturers in Italy
n. 28/02	G. BRUNO M. MALGARINI	An Indicator of Economic Sentiment for the Italian Economy
n. 29/02	G. ARBIA R. BASILE M. SALVATORE	Regional Convergence in Italy 1951-1999: a Spatial Econometric Perspective
n. 30/03	F. KOSTORIS PADOA SCHIOPPA	Il mercato e le politiche economiche in Italia
n. 31/03	S. DE NARDIS C. VICARELLI	The Impact of Euro on Trade: the (Early) Effect Is not So Large
n. 32/03	S. LEPROUX	L'inchiesta ISAE-UE presso le imprese del commercio al minuto tradizionale e della grande distribuzione: la revisione dell'impianto metodologico
n. 33/03	G. BRUNO C. LUPI	Forecasting Euro-area Industrial Production Using (Mostly)\ Business Surveys Data
n. 34/03	C. DE LUCIA	Wage Setters, Central Bank Conservatism and Economic Performance
n. 35/03	E. D'ELIA B. M. MARTELLI	Estimation of Households Income from Bracketed Income Survey Data
n. 36/03	G. PRINCIPE	Soglie dimensionali e regolazione del rapporto di lavoro in Italia
n. 37/03	M. BOVI	A Nonparametric Analysis of the International Business Cycles
n. 38/03	S. DE NARDIS M. MANCINI C. PAPPALARDO	Regolazione del mercato del lavoro e crescita dimensionale delle imprese: una verifica sull'effetto soglia dei 15 dipendenti

Working Papers available:

n. 39/03	C. MILANA ALESSANDRO ZELI	Productivity Slowdown and the Role of the Ict in Italy: a Firm-level Analysis
n. 40/04	R. BASILE S. DE NARDIS	Non linearità e dinamica della dimensione d'impresa in Italia
n. 41/04	G. BRUNO E. OTRANTO	Dating the Italian Business Cycle: a Comparison of Procedures
n. 42/04	C. PAPPALARDO G. PIRAS	Vector-auto-regression Approach to Forecast Italian Imports
n. 43/04	R. DE SANTIS	Has Trade Structure Any Importance in the Transmission of Currency Shocks? An Empirical Application for Central and Eastern European Acceding Countries to EU
n. 44/04	L. DE BENEDICTIS C. VICARELLI	Trade Potentials in Gravity Panel Data Models
n. 45/04	S. DE NARDIS C. PENSA	How Intense Is Competition in International Markets of Traditional Goods? The Case of Italian Exporters
n. 46/04	M. BOVI	The Dark, and Independent, Side of Italy
n. 47/05	M. MALGARINI P. MARGANI B.M. MARTELLI	Re-engineering the ISAE manufacturing survey
n. 48/05	R. BASILE A. GIUNTA	Things change. Foreign market penetration and firms' behaviour in industrial districts: an empirical analysis
n. 49/05	C. CICONI	Building smooth indicators nearly free of end-of-sample revisions