rivista di statistica ufficiale

n. 1
2007

Temi trattati

Metodologia per l'analisi dei comportamenti
e della soddisfazione degli utilizzatori
di una banca dati statistica su Internet
Natale Renato Fazio, Mauro Politi

Will Italy's Tax Reform Reduce the Corporate Tax Burden?
A Microsimulation Analysis
Filippo Oropallo, Valentino Parisi

Piccole e medie imprese: le innovazioni
nei metodi di calcolo dei principali aggregati economici
di contabilità nazionale
Alessandro Faramondi, Claudio Pascarella, Augusto Puggioni

Istat
Istituto nazionale
di Statistica

rivista di statistica ufficiale

n. 1
2007

Temi trattati

Istat

Direttore responsabile: Patrizia Cacioli

Coordinatore scientifico: Giulio Barcaroli

per contattare la redazione o per inviare lavori scrivere a:
Segreteria del Comitato di redazione delle pubblicazioni scientifiche
c/o Carlo Deli
Istat - Via Cesare Balbo, 16 - 00184 Roma
e-mail: rivista@istat.it

rivista di statistica ufficiale

n. 1/2007

Periodico quadrimestrale
ISSN 1828-1982

Registrazione presso il Tribunale di Roma
n. 339 del 19 luglio 2007

Istituto nazionale di statistica
Servizio Produzione editoriale
Via Cesare Balbo, 16 - Roma

Videoimpaginazione:
Raffaella Rose, Patrizia Balzano

Stampa:
Istat - Produzione libraria e centro stampa
Via Tuscolana 1776 - Roma
Gennaio 2009 - Copie 350

Metodologia per l'analisi dei comportamenti e della soddisfazione degli utilizzatori di una banca dati statistica su Internet[1]

Natale Renato Fazio[2], Mauro Politi [3]

Sommario

Gli Istituti Nazionali di Statistica diffondono sempre di più dati attraverso il Web e tra le forme più evolute di questa diffusione vi sono le banche dati. Tra i nuovi compiti della pubblica amministrazione indotti dall'uso di moderne tecnologie vi è quello di misurare sia la qualità dei servizi offerti on line che la soddisfazione degli utenti. Nel presente lavoro viene proposta e illustrata una metodologia basata anche su tecniche di data mining per rilevare il gradimento degli utilizzatori del servizio della banca dati statistica sul commercio estero offerta sul Web dall'Istat, analizzarne i comportamenti e quindi migliorare l'offerta di servizi.

Abstract

National Statistical Institutes disseminate data more and more on the Web: the data banks represent developed tools of this dissemination. One of the new tasks of the public administration (derived from the usage of modern technologies) is to measure the quality of the services offered on line and the users' satisfaction. In this paper it is proposed and presented a methodology based also on techniques of data mining for detecting the satisfaction of the users of the statistical data bank on external trade offered by Istat on the Web, for analyzing their behaviours and therefore for improving the supply of services.

Parole chiave: banca dati statistica, tecniche di data mining, comportamenti e soddisfazione degli utenti

1. Introduzione

Nel continuo processo di trasformazione e modernizzazione delle amministrazioni pubbliche, hanno assunto particolare importanza il tema della qualità dei servizi pubblici e il ruolo centrale del cittadino, non solo come destinatario dei servizi, ma anche quale risorsa strategica da coinvolgere per valutare la rispondenza dei servizi erogati ai bisogni reali. Contemporaneamente l'attenzione degli statistici si è concentrata sulla crescente richiesta

[1] Sebbene il lavoro sia frutto dell'opera di entrambi gli autori, sono da attribuire: i paragrafi 1, 2, 3 e 6 a Mauro Politi; i paragrafi 4 e 5 a Natale Renato Fazio

[2] Primo Tecnologo (Istat), e-mail: nafazio@istat.it.
[3] Dirigente di Ricerca (Istat), e-mail: politi@istat.it.

degli Enti Pubblici di poter disporre di tecniche statistiche utili per affrontare le tematiche della valutazione dell'efficacia, dell'efficienza e della qualità dei servizi forniti, anche con riferimento alla soddisfazione degli utenti.

La centralità di tali tematiche traspare anche dalla emanazione, da parte del Ministro per l'innovazione e le tecnologie, di concerto con il Ministro per la funzione pubblica, della *"Direttiva per la qualità dei servizi on line e la misurazione della soddisfazione degli utenti"*(G.U. n. 243 del 18 ottobre 2005). In essa viene ribadito il ruolo centrale del cliente, la multicanalità e la qualità dei servizi resi dalle amministrazioni pubbliche.

Con l'avvento del Web l'Istat ha incrementato notevolmente i servizi erogati alla collettività predisponendo varie banche dati e sistemi informativi statistici utilizzabili direttamente tramite Internet. Questo studio si propone di monitorare i servizi offerti on line per il caso particolare della *Banca dati delle statistiche del commercio estero Coeweb* (http://www.coeweb.istat.it): l'obiettivo è quello di analizzare i comportamenti degli utilizzatori di Coeweb per migliorare la qualità del servizio stesso nei confronti della collettività e per aumentare la soddisfazione degli utenti stessi.

Per rilevare la reazione degli utilizzatori del servizio, secondo quanto esposto nella Direttiva suddetta, è stata utilizzata una modalità indiretta fondata sulle informazioni acquisite attraverso le e-mail ricevute ed il contact center, ad essa è stata affiancata una modalità di analisi tecnica basata sull'osservazione dei comportamenti di navigazione, che ha il vantaggio di essere non invasiva per l'utente. L'analisi diretta, ovvero quella scaturita dal contatto diretto con gli utilizzatori era stata effettuata in fase di progettazione e test della banca dati.

Si è fatto ricorso a tecniche di *data mining* per effettuare un'analisi dettagliata dei comportamenti di navigazione, ovvero si è proceduto alla selezione, esplorazione e modellazione di grandi masse di dati, al fine di scoprire regolarità o relazioni non note a priori e allo scopo di ottenere risultati univoci e utili al produttore del database.

2. Definizione degli obiettivi

I dati statistici riguardanti il commercio estero formano una rilevante parte dei dati pubblicati regolarmente dall'Istat. Poiché l'insieme di questi dati, elaborati sulla base delle rilevazioni mensili sul commercio estero, è di una dimensione ragguardevole, le richieste degli utenti sono tra le più varie: la numerosità delle variabili statistiche considerate (merce, paese, movimento ecc.) fa sì che la domanda teorica dell'utenza sia molto ampia. L'Istat ha quindi sviluppato un "Piano della diffusione delle statistiche del commercio estero" che regola quello che è "diffondibile"; nel contesto di questo piano è stata progettata e realizzata la banca dati Coeweb su Internet.

La fase di progettazione ha coinvolto la maggior parte degli utilizzatori "fini" (*stakeholder*) delle statistiche di commercio estero: una serie di incontri ristretti con gruppi di studiosi e cultori della materia hanno permesso di raffinare sempre più la produzione della banca dati che, nella sua versione ufficiale, viene presentata al pubblico il 3 ottobre 2001 riscuotendo un immediato, grande e crescente interesse da parte degli utilizzatori, tra cui le Camere di Commercio, il Ministero degli Affari Esteri, gli uffici dell'Istituto del Commercio Estero, le associazioni di categoria, la Banca d'Italia, gli enti di ricerca, le istituzioni economiche italiane e straniere.

Tavola 1 - Tavole statistiche prodotte da Coeweb - Anni 2001-2005

	2001	2002	2003	2004	2005
Nr.totale	49.795	184.533	286.168	302.456	300.893

I punti di forza dimostrati da Coeweb sono stati la tempestività e l'accuratezza nella fornitura dei dati, la facile accessibilità, la completezza delle informazioni offerte, la leggerezza applicativa a fronte di una complessa struttura di *data warehouse* che permette l'esecuzione on line di complesse elaborazioni con tempi di risposta di pochi secondi.

In questo studio sarà illustrata una metodologia per l'analisi della banca dati statistici Coeweb dell'Istat e verranno utilizzate anche tecniche di *data mining* al fine di migliorarne la qualità e aumentare il soddisfacimento del cliente/utente.

L'obiettivo generale del *web usage mining*, cioè il *data mining* applicato al web, è quello di estrarre informazioni preziose sui comportamenti di visita al fine di migliorare la qualità mediante una valutazione e personalizzazione del sito.

La valutazione dei siti web consiste nel determinare se l'utilizzo effettivo di un sito corrisponde alle intenzioni del suo progettista. Se i percorsi seguiti dalla maggior parte dei visitatori non coincidono con quelli attesi, è probabile che sia difficile navigare all'interno del sito in questione. In questo caso il progettista del sito deve valutare un cambiamento dell'architettura del sito per soddisfare al meglio le esigenze degli utenti. Il *web usage mining* può essere utile per valutare i siti web tramite la determinazione dei *pattern* frequenti e dei percorsi seguiti dalla popolazione degli utenti.

Scopo della personalizzazione di un sito web è offrire ad un utente ciò che più interessa, senza richiederlo in modo esplicito. La personalizzazione può essere implementata manualmente o automaticamente con l'aiuto di metodi di *data mining*. Le tecniche manuali richiedono all'utente di registrarsi a un sito web, di rispondere a qualche domanda e di riempire alcune caselle; esse però non sono in grado di offrire una raffigurazione reale delle azioni attese degli utenti. Utilizzando il *web usage mining* per rendere automatica la personalizzazione, la soggettività delle risposte degli utenti è sostituita dal vero comportamento dell'utente stesso.

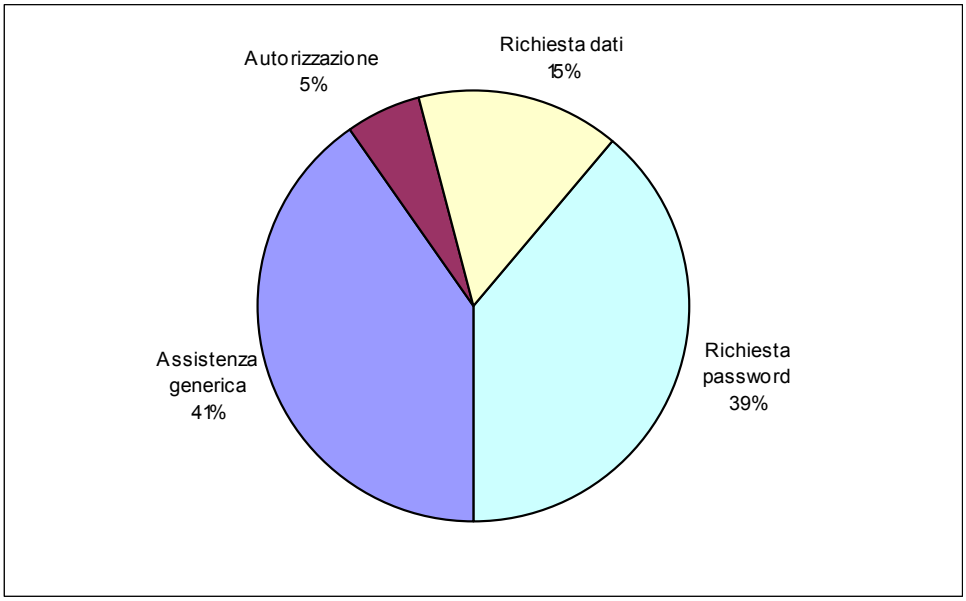
3. Analisi con modalità diretta e indiretta

L'analisi dei bisogni degli utenti di una banca dati sulle statistiche di commercio estero è stata fatta, come già accennato, attraverso un coinvolgimento degli utilizzatori "fini". Durante la fase di progettazione e di costruzione della banca dati sono state effettuate riunioni di *focus group*, presentazioni di versioni provvisorie e sono stati sviluppati contatti bilaterali per cercare di giungere ad un prodotto finale che avesse risposto il più possibile alle esigenze dell'utenza. Questa fase può essere catalogata come *l'analisi diretta,* così come definita nella *Direttiva per la qualità dei servizi on line e la misurazione della soddisfazione degli utenti* del 27 luglio 2005.

Invece successivamente al lancio e all'utilizzo della banca dati è stata effettuata una ulteriore analisi sul comportamento degli utilizzatori di Coeweb attraverso la modalità, definita *indiretta* nella già citata Direttiva. A tale scopo sono state analizzate le e-mail inviate allo staff di Coeweb all'indirizzo coeweb@istat.it. Sono state considerate tutte le 357 e-mail inviate nel 2005 con le quali si poneva una qualsiasi richiesta riguardante Coeweb.

Lo staff di Coeweb opera principalmente su 4 tipologie di interventi (fig.1):

Figura.1 – Distribuzione delle e-mail pervenute a coeweb@istat.it **nell'anno 2005**



Assistenza generica (41%): comprende l'assistenza alla consultazione e alle classificazioni utilizzate.

Assistenza sulla registrazione (39%): una buona percentuale di e-mail si riferisce a utenti di Coeweb che, o non riesce correttamente a registrarsi o fa confusione tra diverse banche dati dell'Istat; non è raro il caso infatti che un utente si registri a Conistat, la banca dati Istat delle serie storiche delle statistiche congiunturali, e che poi provi a collegarsi a Coeweb con la stessa utenza e password.

Richiesta dati (15%): alcuni utenti (9%) trovano difficoltoso estrarre i dati autonomamente e chiedono il supporto allo staff di Coeweb: usualmente la difficoltà consiste nella selezione dei parametri da impostare nelle query dinamiche. Altri (6%) richiedono un dettaglio maggiore rispetto al data base Coeweb su Internet: in questo caso vengono attivate le procedure per verificare se i dati richiesti possono essere forniti tenendo conto del Piano di diffusione delle statistiche di commercio estero. Qualora le verifiche siano positive e l'utente accetti il preventivo dei costi per l'elaborazione e la fornitura dei dati, si procede.

Autorizzazione (5%): una parte di utenti registrati chiede la possibilità di usufruire di maggiori potenzialità nella elaborazione delle tavole statistiche dinamiche mediante l'ampliamento del proprio profilo utente: usualmente viene concesso l'ampliamento del profilo utente per un periodo limitato di tempo.

Per la quasi totalità delle risposte date dallo staff dell'Istat si è avuto un feedback positivo di soddisfazione da parte degli utenti.

4. Analisi tecnica - Trattamento dei dati

I dati disponibili come risultato di una o più sessioni di un navigatore del web sono immagazzinati in log file. Un tipico log file contiene informazioni che descrivono le sequenze di *click* seguite da un utente. I log file molto spesso offrono informazioni nella forma nota come *common log file format*.

Il data set utilizzato per l'analisi contiene dati inerenti alle pagine visitate del sito www.coeweb.istat.it nell'anno 2005. E' composto dai file di log giornalieri del *web server* secondo lo standard *W3C extended log file*, che è quello che fornisce il maggior numero di informazioni ed è composto dai seguenti campi:

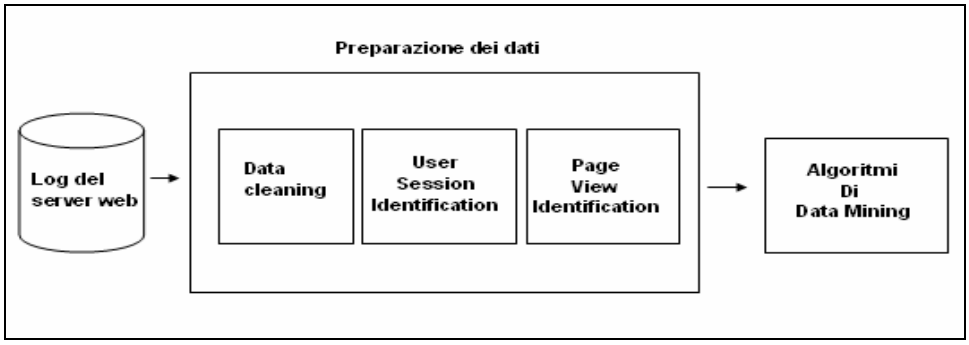
Tavola 2 - W3C Extended Log File Fields

Campo	Appare come	Descrizione
Data	date	Data in cui si è verificata l'attività
Ora	time	L'ora in cui si è verificata l'attività
Indirizzo IP del Client	c-ip	L'indirizzo IP del client che ha effettuato la richiesta
User Name	cs-username	nome dell'utente autenticato che ha acceduto al server. Utenti anonimi sono indicati da un trattino
Indirizzo IP del server	s-ip	L'indirizzo IP del server
Porta del server	s-port	Il numero di porta del server
Metodo	cs-method	L'azione richiesta, per esempio un metodo GET
URI Stem	cs-uri-stem	L'obiettivo dell'azione, per esempio Default.htm
URI Query	cs-uri-query	La richiesta, se esiste, che il cliente sta cercando di operare
HTTP Status	sc-status	Il codice di stato http
Bytes Sent	sc-bytes	Il numero di byte che il server invia
Bytes Received	cs-bytes	Il numero di byte che il server riceve
Time Taken	time-taken	Il tempo, in millisecondi, per la risoluzione dell'azione
User Agent	cs(User-Agent)	Il tipo di browser utilizzato dal cliente
Cookie	cs(Cookie)	Il contenuto del cookie mandato o ricevuto, se presente

Il numero totale di righe dei file di log è 4.324.868, per circa 900 MB di spazio disco occupato. Per comodità di elaborazione i file di log sono stati caricati in tabelle del RDBMS Oracle. Il lavoro di preparazione dei dati, consistente nell'estrarre i dati rilevanti dai log e nel creare i file adatti al processo di *web usage mining*, si divide in tre fasi:

- Data cleaning;
- User/session identification;
- Pageview identification.

Figura. 2 - Processo di web usage mining



4.1 Data cleaning

La fase di *data cleaning* è una fase di trattamento dei file di log per eliminare informazioni irrilevanti o dati anomali rispetto all'analisi condotta. I dati vengono dunque depurati da:

- registrazioni relative a file grafici (.jpg, .gif) che vengono automaticamente catalogati nei log a seguito di una richiesta di una pagina che li comprende;
- registrazioni relative ad *agent* e *spider*[4];
- registrazioni il cui *status code* è diverso da 200 (lo *status code* 200 significa che la richiesta del client è stata ricevuta, compresa ed accettata dal server) ;
- registrazioni effettuate con metodo HEAD ovvero usualmente non associate a richieste fatte dagli utenti;
- registrazioni effettuate da utenti interni all'Istat[5].

4.2 User/Session identification

Esistono vari metodi per l'identificazione dell'utente o delle singole sessioni; quelli maggiormente utilizzati sono i *cookie*, l'identificazione esplicita mediante *userid* e *password*, gli agenti software e gli indirizzi IP.

- I *cookie* sono dei file di testo di piccole dimensioni che vengono trasmessi dal web server del sito Internet e memorizzati nel computer dell'utente. Questi file possono contenere varie informazioni come ad esempio un identificativo dell'utente, una password, e possono essere letti dal web server ad esempio al verificarsi di certi eventi. Con questo metodo il sistema identifica una macchina e non un utente; quindi, l'assunzione che si deve fare è che ad ogni macchina sia associato un utente diverso. Inoltre il metodo non è sempre applicabile poiché l'utente potrebbe impostare il proprio browser per escludere l'utilizzo dei *cookie*.
- Esiste un altro tipo di *cookie* inviato dal web server, che memorizza un identificativo univoco di sessione e che non viene salvato sul disco del client, ma rimane in memoria del computer fintantoché il browser rimane attivo (*cookie* di sessione).
- Il metodo con *userid* e *password* è sicuramente il più utilizzato per identificare un utente. In questo caso il sistema permette di identificare lo specifico utente e non la sua macchina, per contro è applicabile solo se l'utente è disponibile a registrarsi.
- L'agente software è un applicativo da scaricare ed installare sulla propria macchina che usualmente gestisce l'accesso ai servizi del sito che lo ha predisposto. Tale software può fornire automaticamente le informazioni per il riconoscimento dell'utente ma presenta gli stessi svantaggi del metodo dei cookie poiché in realtà riconosce la macchina e non l'utente, inoltre la necessità di scaricare ed installare un software rappresenta un ostacolo per l'utente sia dal punto di vista strettamente tecnico che per la sempre maggiore preoccupazione che questi software possano acquisire informazioni personali su di essi in modo del tutto invisibile.
- L'indirizzo IP identifica univocamente un computer connesso ad Internet. Il metodo

[4] Tipo di software robot che esplora il Web seguendo tutti i link che trova in una pagina, ne legge i contenuti (e altre informazioni) e crea le voci degli indici dei motori di ricerca. Tutti i maggiori motori di ricerca ne hanno uno o più di uno e sono anche noti con i nomi di crawler e bot (per robot).

[5] Le tavole statistiche prodotte da utenti interni dell'Istat costituiscono lo 0,7 per cento del totale.

che sfrutta tale indirizzo potrebbe riconoscere dunque la sola macchina dell'utente, ma solo nel caso in cui l'indirizzo IP sia stato assegnato in modo statico. Con l'assegnazione dinamica (al computer viene assegnato un numero IP diverso tra quelli disponibili ad ogni connessione) non c'è alcuna possibilità di riconoscimento dell'utente e/o della macchina.

Per questo lavoro si è scelto di utilizzare un metodo combinato *cookie* di sessione – *userid-password* per risalire agli utenti e alle singole sessioni. I *cookie* di sessione sono molto meno invasivi dei *cookie* permanenti e hanno l'unico scopo di mantenere lo stato della connessione tra web server e browser.

Il web server IIS (Internet Information Services) di Coeweb e la sua parte applicativa ASP (Active Server Pages) usano i *cookie* per identificare gli utenti all'interno di una sessione di una applicazione inviando un *cookie* denominato "di sessione" al browser dell'utente. Il web server crea automaticamente il *cookie* di sessione quando l'utente accede per la prima volta ad un file del sito elaborato dalla parte applicativa ASP.

Il *cookie* è inviato al browser senza una data di scadenza ed esso viene caricato in memoria dal browser. Ogni volta che l'utente accede ad un'altra pagina nell'applicazione, il web server richiede il *cookie* di sessione e verifica le sessioni aperte su server: se trova una associazione sa che l'utente ha una sessione aperta e permette all'utente stesso di interagire con le variabili di sessione e quindi continuare a lavorare con l'applicazione.

Dopo aver individuato l'utente, per studiare il comportamento durante la navigazione del sito, è necessario riuscire a rilevare tutte le azioni compiute, tracciando e ricostruendo l'intera sessione, ossia l'intera sequenza di pagine richieste dall'utente nello stesso sito in un determinato periodo di tempo. I processi di *tracking* e di ricostruzione delle sessioni devono però tenere conto di tre specifici problemi che ne ostacolano l'attuazione:

- *Tasto "Back" del browser.* L'utente può muoversi all'interno del sito in esame utilizzando il tasto "Back" del browser invalidando in parte la corretta sequenza di visita poiché nessuna richiesta verrà inoltrata in questo caso al web server .
- *Local caching.* I programmi che vengono utilizzati per visualizzare pagine web dispongono di una memoria, detta *cache*, per memorizzare una copia delle pagine che sono state visitate allo scopo di accelerarne la visualizzazione all'accesso successivo, spesso senza neppure inoltrare la richiesta al web server del sito. In questo caso è possibile che alcune pagine appartenenti al percorso di un certo utente vengano fornite direttamente tramite l'uso della copia presente nella cache locale e che non vengano chieste al sito.
- *Proxy server.* Il *proxy server* è un computer che si interpone tra i computer degli utenti ed il web server del sito Internet e che memorizza sul proprio disco le pagine web che gli utenti hanno visitato. Quando un utente richiede una qualunque pagina, il *proxy* ne verifica la presenza sul proprio disco e, in caso affermativo, la fornisce direttamente all'utente senza contattare il sito web.

Mentre il primo problema ha ben poche soluzioni, gli altri due hanno una soluzione nell'uso della tecnica degli URL (*uniform resource locator*) dinamici: in pratica tale sistema consiste nel modificare, ad ogni invio di pagina verso l'utente, gli indirizzi o URL contenuti nelle pagine con l'aggiunta di una stringa finale composta da caratteri casuali. Tale metodo non permette né al browser né al *proxy* di associare l'URL correntemente richiesto con pagine già memorizzate e quindi la richiesta viene indirizzata al server remoto. Questa tecnica, che rallenta la navigabilità del sito, non è stata implementata nel sito Coeweb per cui diventa quasi impossibile poter risolvere i due problemi esposti. D'altronde essi riguardano principalmente la parte statica del sito Coeweb, osservato che le

interrogazioni dinamiche al sito subiscono in misura minore il *local* ed il *proxy caching*. Si analizzeranno quindi i dati con l'assunzione che soprattutto le pagine statiche del sito possono essere state lievemente sottostimate, inoltre si escluderanno dall'analisi le sessioni non identificabili tramite *cookie* poiché non permettono la tracciabilità della sessione (il loro peso è comunque trascurabile, pari allo 0,7% del totale delle tavole statistiche).

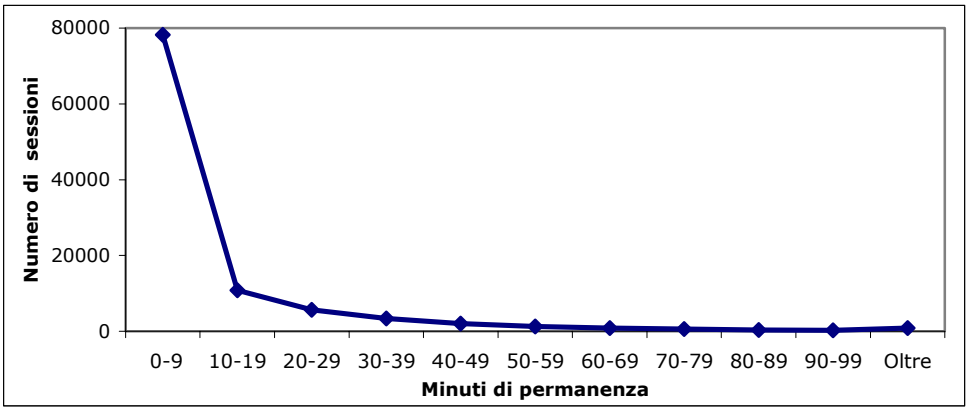
Ultima tematica da affrontare nel processo di *user/session identification* è la suddivisione della sequenza di richieste di ciascun utente in sessioni, secondo il criterio standard di utilizzare un timeout di 30 minuti per individuare la fine della sessione: se il tempo che intercorre tra due richieste supera il timeout si assume che l'utente abbia iniziato una nuova sessione poiché si vedrà in ogni caso costretto a registrarsi nuovamente o ad impostare nuovamente i parametri di ricerca dei dati. Il seguente algoritmo, implementato in PL/SQL, risolve il problema: in pratica, ogni volta che all'interno della stessa sessione il periodo di inattività supera i 30 minuti, viene creata un nuova sessione utente che avrà, come identificativo univoco, la concatenazione della stringa di sessione ad un opportuno contatore di sessioni multiple k.

Figura. 3 - Euristica per l'identificazione delle sessioni

```
1.  Ordina i dati per Session, Time (Data e Ora)
2.  k=0 ; Session0 = NULL ;
3.  Definisci timeout T = 30 min.
4.  Per ogni record i = [1..N]
5.     Se Sessioni <> Session0 allora
6.        (k=0 ;  Session0 = Sessioni )
7.     altrimenti
8.        Se ((Timei – Timei-1 ) > T.) allora incrementa k
9.     Se k>0 allora Sessioni = Concatena(Sessioni ; k )
```

Una prima analisi degli accessi a Coeweb, dopo aver effettuato le elaborazioni della fase di preparazione dei dati, produce l'output mostrato nella figura.4. Si osserva che le sessioni entro i 9 minuti coprono il 74,9% del totale e che il 94% è coperto dalle sessioni entro i 39 minuti. Tenuto conto che ben il 50,3% delle sessioni si conclude entro i due minuti, questa prima serie di informazioni porta ad approfondire l'analisi considerando in particolare la variabile "tempo di permanenza".

Figura. 4 - Numero di sessioni per minuti di permanenza al sito Coeweb - Anno 2005



4.3 Pageview Identification

Per pageview si intende il caricamento, da parte del browser, di una intera pagina web (HTML, ASP, PHP) composta da uno o più page file (ad esempio frames, grafici, documenti). La pageview identification si basa sui risultati delle attività di estrazione del contenuto e della struttura del sito web. Scopo di ambedue le attività è convertire il contenuto e la struttura del sito in una forma utilizzabile dal web usage mining, permettendo di identificare il contenuto associato a ciascun page file e quali page file possono fare parte di una pageview.

Nella figura.5 si nota la struttura principale del sito Coeweb: in essa vengono riportati i page file principali che identificano le singole pageview. Ai fini dell'analisi tali page file vengono classificati in 9 aree tematiche:

- "*Metodologia*" include tutte le note sulla rilevazione statistica, sui numeri indici, sugli operatori, il glossario. E' una sezione statica del sito, aggiornata ogni qualvolta vi sia necessità di diffondere nuove note metodologiche a fronte di modifiche nella metodologia o anche per chiarire particolari cambiamenti nelle classificazioni geografiche o merceologiche, oppure nei regolamenti internazionali che disciplinano le statistiche di commercio estero.
- "*Approfondimenti*" include tutte le tavole statistiche della sezione omonima. Questa è una sezione statica del sito costituita da tavole statistiche che, comprendendo incroci tra variabili al di fuori del piano della diffusione, devono essere elaborate precedentemente al fine di verificarne il rispetto della riservatezza attiva nella diffusione delle stesse.
- "*Manuali*" include tutti i documenti riguardanti Coeweb e di ausilio alla navigazione.
- "*Classificazioni*" include tutte le pagine del sito che trattano le classificazioni adottate, sia geografiche che merceologiche. Include inoltre un motore di ricerca per risalire al codice merceologico di una merce attraverso una descrizione della stessa e il viceversa. A meno di modifiche per eventi particolari tale sezione viene aggiornata annualmente.
- "*Performance*" include tutte le tavole statistiche della sezione omonima. E' una sezione statica che offre all'utente un insieme di tavole statistiche aggiornate mensilmente e riguardanti i maggiori flussi commerciali (top-ten) per una analisi settoriale, territoriale e dei principali mercati di sbocco.
- "*Statistiche tematiche*" include tutte le tavole statistiche create dinamicamente dall'utente utilizzando la sezione del sito "Consultazione tematica".
- "*Statistiche libere*" include tutte le tavole statistiche create dinamicamente dall'utente utilizzando la sezione del sito "Ricerca puntuale", escluso le serie storiche.
- "*Serie storiche*" include tutte le tavole statistiche create dinamicamente dall'utente utilizzando la sezione del sito "Ricerca puntuale", sottosezione "Serie Storiche".
- "*Numeri indici*" include tutti i file .xls .zip di tavole statistiche sui numeri indici del commercio estero. E' una sezione statica del sito che viene aggiornata mensilmente.

Da una prima analisi delle pagine visitate dagli utenti (figura.6) si ricava che le sezioni "statistiche libere" e "statistiche tematiche" coprono insieme il 77,3% del totale considerato (rispettivamente il 35,1% e il 42,2%). L'interesse degli utenti quindi si sposta sull'area "metodologia" con un 7,1%, per poi ritornare alla produzione di tavole statistiche delle sezioni "serie storiche" (4,4%) e "performance" (4,2%). "Indici", una sezione riguardante un argomento per addetti ai lavori, è l'area meno visitata (0,8%).

Figura. 5 - Struttura principale del sito Coeweb

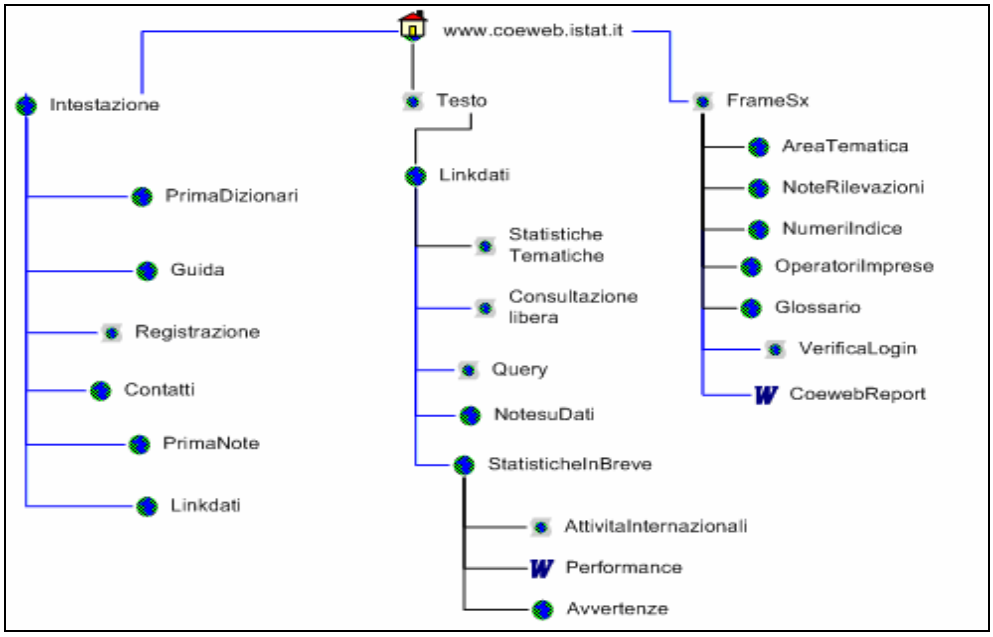
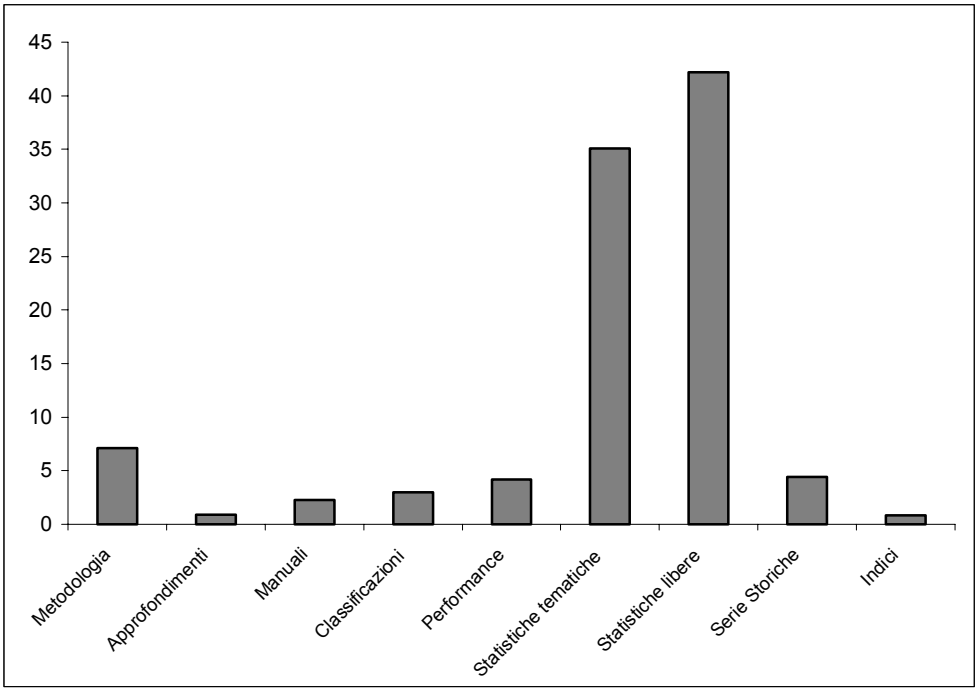


Figura. 6 - Distribuzione delle pagine visitate rispetto alle aree tematiche

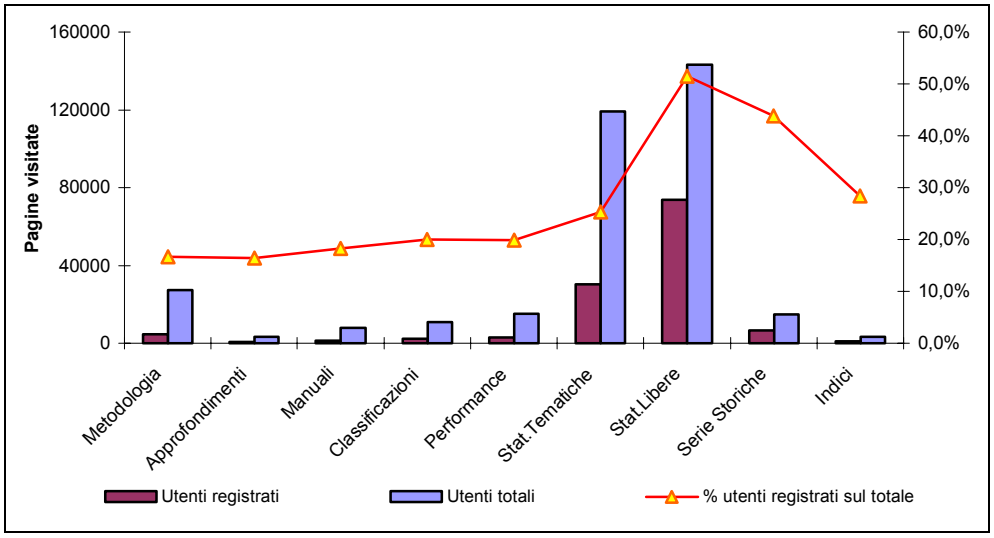


4.4. Sessioni eseguite da utenti registrati

La banca dati Coeweb permette all'utente navigatore del sito di registrarsi al fine di avere un profilo utente più elevato rispetto all'utente "ospite", ad esempio per poter selezionare contemporaneamente più di 10 modalità per variabile nelle sezioni "statistiche libere" e "serie storiche" e poter salvare l'interrogazione per poi eseguirla successivamente con tutti i parametri già impostati. Altra possibilità che viene offerta all'utente registrato è la personalizzazione temporanea del profilo a seguito di una richiesta di autorizzazione all'elaborazione di particolari tavole statistiche molto dettagliate della sezione "statistica tematica". A giugno 2006 le registrazioni utente alla banca dati Coeweb hanno superato le 13.000 unità (sono state 3.072 le registrazioni nell'anno 2005), anche se gli utenti registrati che navigano costantemente il sito sono circa un quinto.

Mediante opportune trasformazioni delle tabelle delle sessioni e procedure di *linkage* con specifiche tabelle del database Oracle di Coeweb, si risale agli utenti registrati ed alle sessioni effettuate da tali utenti. Operando in questo modo si tralascia una parte delle interrogazioni in quanto un utente registrato può accedere al sito anche non identificandosi con la coppia di chiavi *userid/password.* Da rilevare come l'utente registrato copra più della metà (51,5%) delle elaborazioni prodotte dalla sezione "statistiche libere" ed il 43,9% nella sezione "serie storiche": una probabile giustificazione sta nel fatto che, a causa della complessità delle *query* che possono essere prodotte in queste sezioni, Coeweb dà l'opportunità all'utente registrato di salvare le *query*; è chiaro dunque che proprio chi utilizza maggiormente queste sezioni del sito sente la necessità di registrarsi. Al contrario, gli utenti registrati producono solo un quarto delle tavole statistiche elaborate nell'area "statistiche tematiche", ciò è dovuto alla semplicità di creazione delle tavole e dal fatto che non si ha alcun vantaggio o funzionalità aggiuntiva (escludendo gli utenti che richiedono una autorizzazione speciale) ad essere utente registrato per questa sezione del sito.

Figura. 7 — Distribuzione del totale delle pagine visitate per aree tematiche e registrazione



Se si continua con l'analisi esaminando le sessioni che si concludono "entro i due minuti" (da ora LE2m), si nota (figura. 8) come si siano mantenute le percentuali sulle sezioni dinamiche del sito mentre si siano abbassate nelle sezioni statiche. Si nota inoltre un più ampio uso delle sezioni statiche del sito ed un maggior utilizzo delle statistiche tematiche rispetto alle altre statistiche dinamiche.

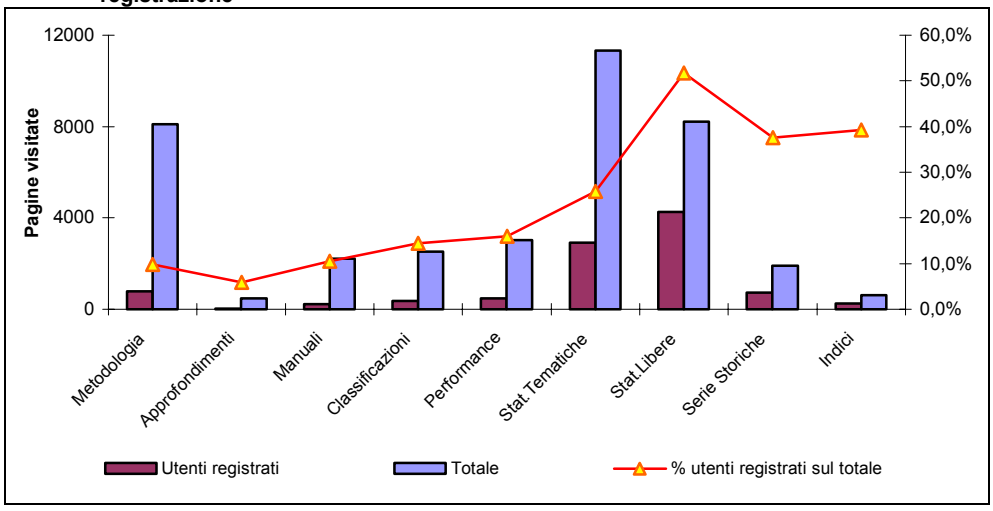
Da ulteriori analisi si evince che:

- Il 23% di sessioni ha iniziato un percorso nell'area "statistiche tematiche" e l'83% di queste ha raggiunto la conclusione producendo tavole statistiche.
- Il 15% di sessioni ha iniziato un percorso nelle aree "statistiche libere" o "serie storiche" ed il 58% di queste ha prodotto tavole statistiche.

Una prima considerazione potrebbe essere che, affinché gli utenti *LE2m* apprezzino in pieno le potenzialità e vastità di informazioni che Coeweb può offrire, occorrerebbe puntare su due obiettivi complementari:

- Rendere più accessibile ed "appetibile" la sezione dinamica del sito al fine di stimolare il 67% di utenti LE2m a conoscerla ed utilizzarla.
- Implementare modalità di aiuto passo-passo nella creazione di tavole statistiche nelle aree "statistiche libere" e "serie storiche".

Figura. 8 – Distribuzione delle pagine visitate nelle sessioni entro 2 minuti per aree tematiche e registrazione



5. Data mining

A supporto delle analisi che scaturiscono dal data mining, si considererà dapprima la fase di "abbandono" (pagine di uscita) del sito da parte degli utenti *LE2m.*

Successivamente verranno utilizzate due diverse tecniche di data mining sui dati elaborati al fine di trovare associazioni tra le pagine del sito Coeweb:

- Analisi degli *odds ratio*;
- Analisi degli indici di supporto e confidenza.

5.1 Analisi delle pagine di uscita

Le pagine di uscita sono le pagine principali in cui un visitatore ha abbandonato un sito. Nel migliore dei casi, le pagine rappresentano la conclusione di un processo naturale, al contrario se la pagina risiede all'inizio o nel mezzo di un flusso di processo logico, si ha la prova diretta che qualcosa impedisce la "naturale" logicità delle operazioni da svolgere per ottenere un risultato, quindi il rischio che il visitatore non si converta in un cliente. Analizzare dove si verificano questi abbandoni può dunque aiutare a comprendere dei punti di debolezza della banca dati.

Sono state analizzate le sessioni utente *LE2m* che non si concludono con una *pageview foglia* dell'albero delle *pageview* delle aree tematiche considerate: rappresentano il 47% del totale e possono essere suddivise nelle seguenti:

- Pagina iniziale (14%)
- Autenticazione (5%)
- Classificazione (8%)
- All'interno di un percorso di statistiche tematiche (24%)
- All'interno di un percorso di consultazione libera (34%)
- In fase di messaggistica (15%).

All'interno della fase di messaggistica abbiamo:

- 25% in fase di autenticazione
- 5% in fase di registrazione
- 70% in fase di impostazione dei parametri.

Una prima lettura di queste informazioni permette di formulare alcune ipotesi:

- Si direbbe che il 14% di utenti si blocca sulla pagina iniziale, ritenendo probabilmente il sito non esattamente ciò che cercava; potrebbe essere utile in questo caso, per stimolare il proseguimento delle navigazione, inserire già nella prima pagina una tavola riassuntiva di statistiche import/export
- Circa il 15% di utenti abbandona il sito dopo aver ricevuto un messaggio di operazione non corretta. Nelle sezione successiva sulle regole di sequenza si cercherà di chiarire il percorso che ha portato l'utente a ricevere una messaggistica di errore e quindi ad abbandonare il sito.

5.2 Analisi degli odds ratio

I dati trattati come descritto nei paragrafi precedenti sono stati sintetizzati in una matrice di 53.872 righe (corrispondenti a tutte le sessioni identificabili in modo univoco) e 9 colonne, una per ciascuna delle variabili dicotomiche indicante l'appartenenza a ciascuna area tematica definita nel paragrafo precedente. L'obiettivo è di ottenere delle variabili dicotomiche descriventi la visita ad almeno una pagina del gruppo.

E' stata effettuata una misura sintetica delle associazioni tra gruppi mediante l'utilizzo dell'odds ratio, un parametro utile nei modelli statistici per l'analisi dei dati qualitativi di tabelle di contingenza 2x2.

Si consideri una tabella 2x2 relativa alle variabili X ed Y, rispettivamente sulle righe e sulle colonne della tabella, il *rischio relativo* (RR) per la variabile Y è definito come il rapporto delle probabilità di successo di Y (modalità 1) nei due livelli della variabile X (modalità 0 ed 1).

$$RR = \frac{\pi_{1/1}}{\pi_{1/0}} = \frac{P(Y=1 \mid X=1)}{P(Y=1 \mid X=0)}$$

Questa quantità può assumere qualsiasi numero reale non negativo, in particolare:

- $RR = 1$ quando Y ed X sono indipendenti
- $RR \in [0,1) se \pi_{1|1} < \pi_{1|0}$, ossia se la probabilità di successo risulta maggiore nella riga 0 rispetto alla riga 1
- $RR \in (1,+\infty)$ se la probabilità di successo è maggiore nella riga 1.

Definiamo l'odds di successo come:

$$odds_1 = \frac{\pi_{1|1}}{\pi_{0|1}} = \frac{P(Y=1 \mid X=1)}{P(Y=0 \mid X=1)}$$

Gli odds risultano non negativi, con valore maggiore di 1, quando un successo (modalità 1) è più probabile di un insuccesso (modalità 0). Quindi se ad esempio risulta odds=3 significa che un successo è tre volte più probabile di un insuccesso; quindi ci si aspetta di osservare tre successi per ogni insuccesso. Analogamente odds = 1/3 = 0,33 significa che un insuccesso è tre volte più probabile di un successo.

Per la riga 0 l'odds di successo risulta:

$$odds_0 = \frac{\pi_{1|0}}{\pi_{0|0}} = \frac{P(Y=1 \mid X=0)}{P(Y=0 \mid X=0)}$$

Definiamo l'odds ratio (θ) come:

$$\theta = \frac{\pi_{1/1} / \pi_{0/1}}{\pi_{1/0} / \pi_{0/0}}$$

Usando la definizione di probabilità congiunta si ha che:

$$\theta = \frac{\theta_1}{\theta_0} = \frac{\pi_{1/1} \cdot \pi_{0/0}}{\pi_{1/0} \cdot \pi_{0/1}}$$

Per il calcolo effettivo dell'odds ratio si sostituiranno le probabilità con le frequenze osservate.

Si verifica che:

- $\theta \in [0,+\infty)$

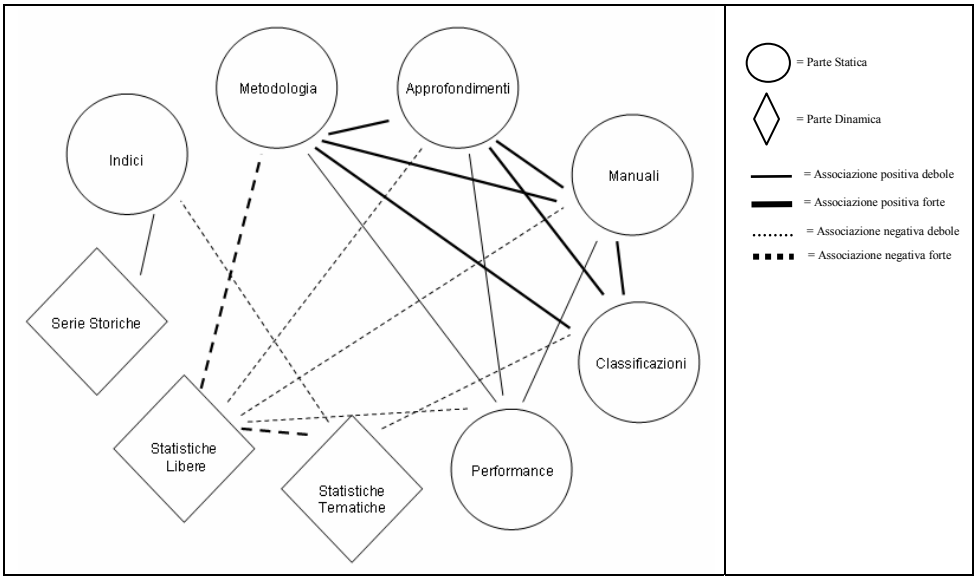
- quando X ed Y sono indipendenti $\theta = 1$; per $1 < \theta < \infty$ si ha associazione positiva, poiché gli odds di successo sono maggiori nella riga 1 che nella riga 0; per $0 < \theta < 1$ si ha associazione negativa, poiché gli odds di successo sono maggiori nella riga 0 che nella riga 1.
- La relazione che permette di calcolare l'odds ratio e usarlo per approssimare il rischio relativo è:

$$\theta = RR(\frac{1-\pi_{1/0}}{1-\pi_{1/1}})$$

- L'odds ratio può essere considerato, dal punto di vista modellistico, l'analogo qualitativo del coefficiente di correlazione lineare. In particolare, per quanto riguarda la costruzione di una regola decisionale che permetta di stabilire se un certo valore osservato dell'odds ratio indichi una associazione significativa tra le corrispondenti variabili, è possibile utilizzare un intervallo di confidenza per campioni: se il valore unitario dell'odds ratio risulta esterno a tale intervallo, l'associazione risulta significativamente diversa da 1.

Il calcolo degli odds ratio ha permesso di individuare la presenza di associazioni tra le visite ai diversi gruppi di pagine. Sono stati calcolati i 36 odds ratio marginali tra le variabili binarie, con i relativi intervalli di confidenza. I risultati sono stati rappresentati in un grafo che ha le variabili di gruppo come nodi e un arco fra di esse se il valore di 1 del relativo odds ratio marginale è esterno all'intervallo di confidenza e l'odds ratio è esterno all'intervallo [0,3; 3,0]. Nel grafo gli archi con linea continua rappresentano le associazioni positive mentre le linee tratteggiate rappresentano le associazioni negative. Il tratto più spesso indica una associazione maggiore.

Figura. 9 - Grafo marginale esplorativo



Il grafo evidenzia in modo netto la differenza di utilizzo del sito da parte degli utenti: un primo gruppo naviga principalmente nella parte statica ed usualmente esplora tutte le sue parti; inoltre le sezioni che hanno dei collegamenti sulla pagina molto vicini tra loro hanno anche una associazione positiva maggiore. Un secondo gruppo utilizza il sito come "banca dati" e non esplora quasi per nulla la parte statica. Rilevante è la massima associazione negativa che si verifica tra le sezioni "Statistiche libere" e "Statistiche tematiche": un utente che utilizza una delle due modalità di estrazione dinamica di tavole statistiche si affeziona a tale modalità e non prende in considerazione la possibilità di estrarre i dati con un diverso procedimento.

Tavola.3 - Odds ratio per variabili con associazione positiva e negativa

Coppia di variabili	Odds ratio	95% Limite di confidenza inferiore	95% Limite di confidenza superiore
Approfondimenti * Metodologia	16,783	14,795	19,038
Manuali * Metodologia	14,593	13,706	15,537
Approfondimenti * Manuali	9,098	8,211	10,081
Metodologia * Classificazioni	5,904	5,578	6,249
Approfondimenti *Classificazioni	4,368	3,919	4,868
Manuali * Classificazioni	4,173	3,910	4,455
Approfondimenti * Performance	3,802	3,415	4,232
Manuali * Performance	3,769	3,539	4,014
Metodologia * Performance	3,710	3,517	3,913
Serie Storiche * Indici	3,048	2,572	3,612

Coppia di variabili	Odds ratio	95% Limite di confidenza inferiore	95% Limite di confidenza superiore
Stat.Libere * Approfondimenti	0,333	0,291	0,381
Stat.Libere * Performance	0,328	0,307	0,350
Stat.Tematiche * Classificazioni	0,302	0,281	0,325
Manuali * Stat.Libere	0,292	0,272	0,314
Stat.Tematiche * Indici	0,287	0,237	0,346
Stat.Libere * Metodologia	0,244	0,231	0,258
Stat.Tematiche * Stat.Libere	0,106	0,101	0,112

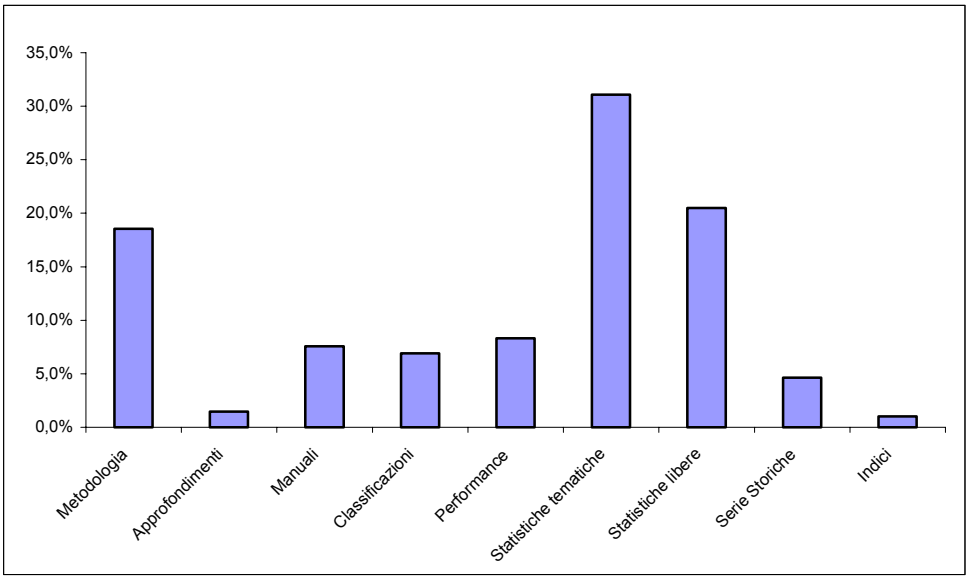
E' opportuno quindi approfondire l'analisi alle sole sessioni *LE2m*, cioè quelle che hanno avuto un tempo di permanenza minore o uguale a due minuti, al fine di studiare il comportamento di utenti a rischio "fedeltà", cioè utenti che navigano apparentemente nel sito per curiosità, ma lo abbandonano abbastanza rapidamente.

Nonostante il breve tempo di permanenza, l'osservazione rivela che *l'utente entro i due minuti* ha un profilo simile ad un utente generico.

L'analisi degli odds ratio per le sessioni utente *LE2m* dimostra, così come osservato nell'analisi di tutte le sessioni, una suddivisione in tre principali tipi di utenti:

- l'utente "curioso" che naviga nella parte statica del sito, tra le sezioni "Approfondimenti", "Metodologia", "Manuali" e "Classificazioni" (circa il 34,5%),
- l'utente "pratico" che elabora tavole statistiche dinamicamente dalla sezione "Statistiche tematiche" (31,1%),
- l'utente "esperto" che elabora tavole statistiche dinamicamente dalla sezione "Statistiche libere" o riprendendo query già salvate (20,5%).

Figura. 10 – Distribuzione delle sessioni utente "*LE2m*" per le aree del sito considerate



5.3. Analisi degli indici di supporto e confidenza

Si introducono gli indici comunemente utilizzati nel *web mining* per lo studio delle sequenze di visita (*clickstream analysis*) . Si consideri la sequenza indiretta A→B (in cui A e B sono pagine Web) e si indichi con $N_{A\to B}$ il numero di sessioni utente in cui tale sequenza compare almeno una volta. Sia *N* il numero complessivo delle sessioni utente. Il supporto per la regola A→B si ottiene dividendo il numero di sessioni utente che soddisfano tale regola per il numero totale di sessioni utente:

$$\text{supporto}(A\to B) = \frac{N_{A\to B}}{N}$$

Si tratta quindi di una frequenza relativa che indica la percentuale degli utenti che hanno visitato in successione le due pagine. In presenza di un numero elevato di sessioni utente si può affermare che il supporto per la regola A→B esprime la probabilità che una sessione utente contenga le due pagine, in sequenza:

$$\text{supporto}(A\to B) = \Pr(A\to B)$$

La confidenza per la regola A→B si ottiene invece dividendo il numero di sessioni utente che soddisfano la regola per il numero di sessioni utente che contengono la pagina A

$$\text{confidenza}(A\to B) = \frac{N_{A\to B}}{N_A} = \frac{\frac{N_{A\to B}}{N}}{\frac{N_A}{N}} = \frac{\text{supporto}(A\to B)}{\text{supporto}(A)}$$

Quindi l'indice di confidenza esprime la frequenza (e al limite la probabilità) che in una sessione utente in cui è stata visualizzata la pagina A possa essere successivamente richiesta la pagina B.

Si useranno due diverse tecniche che permettono di ottenere in un caso regole di associazione tra pagine in cui non viene considerata la dimensione temporale (Market Basket Analysis), nell'altro regole di sequenza (Clickstream Analysis).

Per effettuare tali analisi di pagine del sito Coeweb si farà riferimento alle sole sessioni che si concludono entro i due minuti.

5.3.1 Regole di associazione

E' stato utilizzato Artool, un software gratuito di pubblico dominio[6]. Artool ricerca regole associative in database binari. I database binari sono quelli i cui attributi possono avere solo due valori, presente o assente. Ogni riga del database rappresenta una sessione utente e viene chiamata transazione, gli attributi binari sono gli *items* i cui insiemi sono gli *itemset*.

Artool è composto da vari moduli, se ne utilizzeranno fondamentalmente due:

- asc2db: è un modulo che trasforma un file ascii già formattato secondo precise regole definite nel progetto Artool in un file con estensione db (file database), input per il software di analisi Artool;
- Artool: è un software grafico realizzato in Java che ha come input un particolare file (denominato database). L'utilizzo di questo strumento si sintetizza in tre passaggi:
    - selezione e controllo integrità del database;
    - analisi esplorativa degli itemset frequenti del database considerato;
    - analisi esplorativa delle regole associative del database considerato.

La prima fase è consistita nello sviluppo di un programma PL/SQL che preparasse i dati dei log file secondo il formato richiesto dal modulo "asc2db" del software Artool, come da esempio seguente.

[6] Artool, distribuito con licenza GNU General Public Licence, rappresenta una collezione di algoritmi e strumenti software per l'analisi di regole associative in database binari. Tali tecniche vengono utilizzate principalmente nel ramo del data mining definito Market Basket Analysis.

Supponiamo di volere analizzare le pagine di Coeweb visitate dagli utilizzatori: ogni pagina di interesse sarà codificata con un codice numerico ordinato sequenzialmente; ogni riga della sezione compresa tra BEGIN_DATA e END_DATA corrisponde alle singole sessioni utente ed è composta dalle singole pagine visitate, ad esempio la prima sessione utente avrà visitato le pagine 1,3,5, la seconda le pagine 1,6,7,9 ecc)

1 pagina iniziale
2 pagina introduzione dati
3 pagina di consultazione tematica
4 pagina di ricerca puntuale
.
.
.
609 pagina classificazioni ateco
BEGIN_DATA
1 3 5
1 6 7 9
3 5
.
.
END_DATA

Per far questo è stata dapprima predisposta una tabella di tutte le pagine web del sito Coeweb di interesse per una analisi associativa del tipo *market basket*, quindi, partendo dalle tabelle delle sessioni all'interno dei due minuti, si sono estratte le informazioni secondo lo standard illustrato nel precedente riquadro in un file di 609 modalità o colonne e 22.924 righe. Dopo aver eseguito il modulo "asc2db" di Artool, si è creato il database di input per il software grafico Artool. Effettuato un controllo di integrità sul database (figura. 11), la fase successiva è consistita nel trovare tutti gli itemset all'interno di un valore di supporto definito (figura. 12): è stato scelto un valore abbastanza basso (0,1) al fine di individuare anche le associazioni tra pagine poco visitate.

Il passaggio successivo è consistito nel definire un valore minimo di confidenza e generare tutte le regole associative a partire dagli itemset trovati al passo precedente (figura. 13). Questo passo richiede meno tempo di elaborazione rispetto al passo precedente, ma il risultato può dar luogo ad un grande numero di regole, nell'ordine delle decine di migliaia. Navigare tra queste regole e trovare quelle utili è esso stesso un problema di *data mining*. Ci sono tuttavia algoritmi che possono generare regole che sono interessanti o non ridondanti secondo alcune definizioni date a priori di "ridondanza" e "interesse".

Figura.11 - Artool - Passo 1: selezione database e controllo di integrità

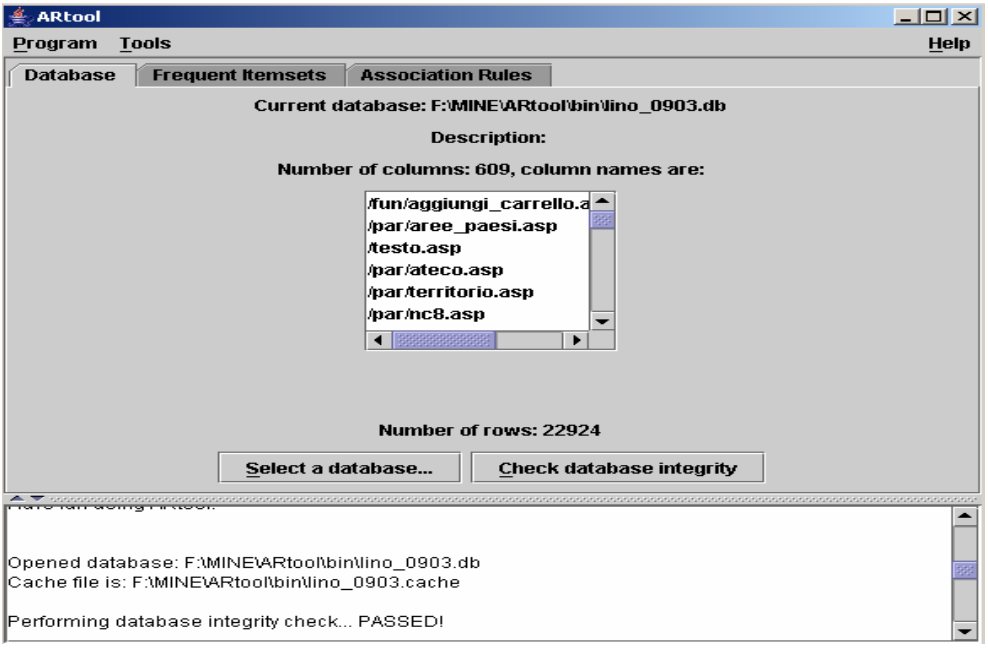


Figura.12 - Artool - Passo 2: definizione di un valore minimo di supporto e creazione itemset

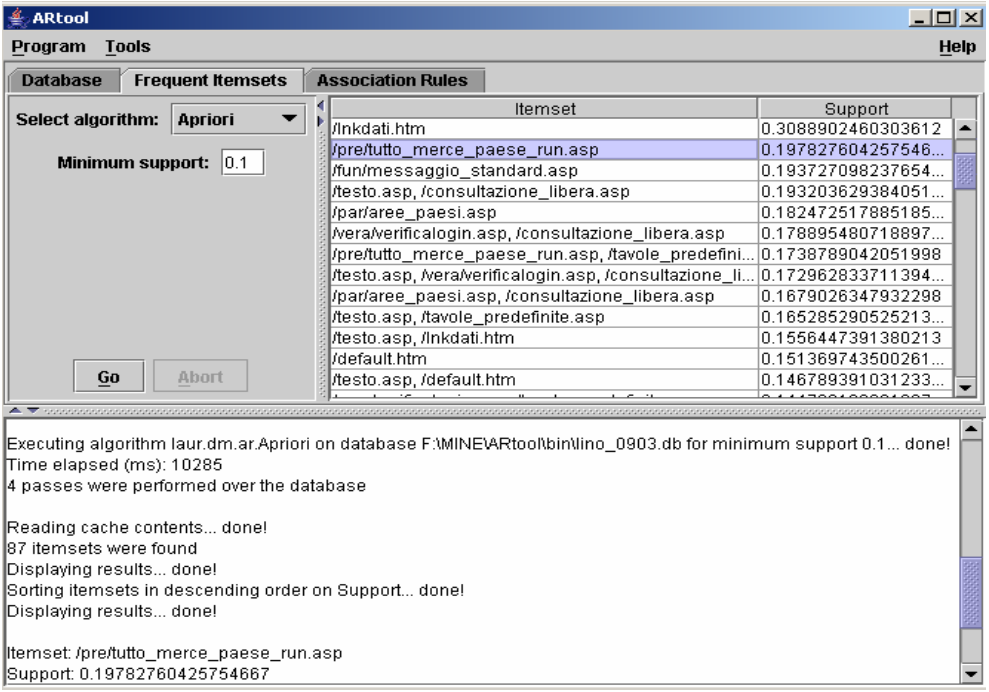
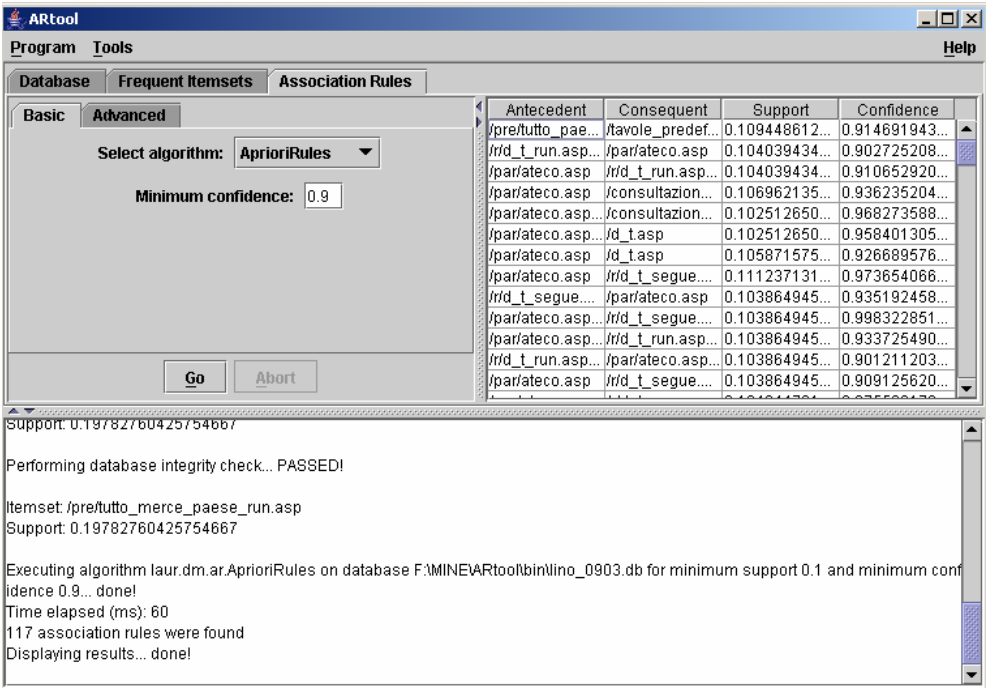


Figura.13 - Artool - Passo 3: generazione regole associative assegnato un valore minimo di confidenza



Gli indicatori utilizzati da Artool sono i seguenti:

- La misura Piatetsky-Shapiro di una regola A→C è definita come:
  *support(AC) – support(A) * support(C)*
  Se si considera il supporto di un itemset come approssimativamente la probabilità che quell'itemset appaia in una riga della tabella, allora la misura Piatetsky-Shapiro esprime quanto la reale occorrenza di AC differisce dal valore aspettato. Si noti che questa misura è simmetrica, cioè ha lo stesso valore per A→C e per C→A
- La misura *lift* è definita come:*support(AC) / (support(A) * support(C))*
- Il lift è una misura che esprime quanto la presenza di un antecedente influenza la comparsa del conseguente. E' una misura simmetrica.
- La misura *influence* è definita come: *support(AC) / support(A) – support(C)* Essa deriva dalla misura lift ma è asimmetrica.

Qui di seguito sono riportati gli estremi di alcune regole associative individuate con Artool.

Antecedent A [pagina impostazione parametri sezione "NC8 per tutti i paesi" in Statistiche tematiche]
Consequent C: [pagina "risultato" dopo elaborazione secondo i parametri richiesti]
Support: 0.10239520958083832
Confidence: 0.9437086092715231
Piatetsky-Shapiro: 0.0808505432249274
Lift: 4.752694141988672
Influence: 0.7451457350200261

L'output prodotto evidenzia che la combinazione delle pagine A e C è presente nel 10,2% delle sessioni utente e la probabilità che come conseguenza di avere scelto A vi sia C è del 94,4%: in pratica un 10,2% di sessioni utente accedono alla pagina di impostazione della nomenclatura combinata (NC8) nella sezione di statistiche tematiche e di questa ben il 94,4% continuerà l'elaborazione ottenendo la tabella richiesta.

Antecedent: [pagina impostazione parametri ricerca codice o descrizione merceologica]
Consequent: [pagina "risultato" secondo i parametri impostati]
Support: 0.012868609317745593
Confidence: 0.793010752688172
Piatetsky-Shapiro: 0.012656244096208783
Lift: 60.596594982078855
Influence: 0.7799240313480917

In questo caso si nota come chi accede alla prima pagina del motore di ricerca codici e descrizioni merceologiche della banca dati Coeweb (1% sessioni utente), per il 79% dei casi arriva alla risposta richiesta, mentre per il 21% dei casi abbandona la ricerca.

Antecedent: [pagina impostazioni parametri classificazione CPATECO]
Consequent: [pagina "risultato" nella sezione ricerca puntuale tavole territoriali per CPATECO]
Support: 0.1040394346536381
Confidence: 0.9106529209621993
Piatetsky-Shapiro: 0.09087241414134986
Lift: 7.901516866062625
Influence: 0.7954025283605591

Avere selezionato alcuni codici merceologici CPAteco nella sezione di ricerca puntuale delle tavole territoriali (10% sessioni utente) conduce per il 91% dei casi alla reale esecuzione della tavola richiesta.

5.3.2 Regole di sequenza

Le regole di sequenza sono molto simili alle regole di associazione, ma si concentrano non tanto sul fatto che due o più pagine facciano parte di una stessa sessione, quanto sulla dimensione temporale delle pagine all'interno della sessione stessa.

E' stato predisposto un programma apposito per estrarre tutte le regole di sequenza di lunghezza massima 4 (cioè del tipo A→B,C,D A,B,C→D, etc.) dalle sessioni *LE2m*. Per ogni regola si è calcolato l'indice di supporto, confidenza ed il lift.

Le regole di sequenza del tipo A,B,C→D con il più alto supporto e con lift maggiore di 1 sono state le seguenti :

REGOLA	SUPPORTO	CONFIDENZA	LIFT
dati, statistiche tematiche, selezione settore paese per merce (NC8) → Tavola statistica x merce e paese	0,041643128	0,840390879	7,540393474
Consultazione libera, dati territoriali, dati territoriali segue → parametri Ateco	0,029577919	0,66879562	16,69092957

Questo conferma l'utilizzo della banca dati per informazioni al massimo livello di disaggregazione a livello nazionale (NC8 per paese) ed anche il grande interesse sulle statistiche a livello territoriale provinciale (CPAteco per paese).

Si è quindi focalizzata l'attenzione su alcune pagine particolari poste o come antecedente o come conseguente delle regole di sequenza estratte.

Tavola 4 – Analisi dell'antecedente: "Dati, statistiche tematiche"

REGOLA	SUPPORTO	CONFIDENZA	LIFT
Dati, statistiche tematiche → selezione NC8 x tutti i paesi	0,049552094	0,288601645	4,252156
Dati, statistiche tematiche→selezione SH6 x tutti i paesi	0,007384392	0,043008226	3,918492
Dati, statistiche tematiche→selezione SH4 x tutti i paesi	0,006617706	0,038542891	4,152913
Dati, statistiche tematiche→selezione paese x tutti NC2	0,006617706	0,038542891	3,631825
Dati, statistiche tematiche → selezione paese x tutti L1_T	0,004922928	0,02867215	2,973026
Dati, statistiche tematiche→selezione paese x tutti AT3_T	0,004922928	0,02867215	2,612328
Dati, statistiche tematiche→consultazione libera	0,004600113	0,026792009	0,137808
Dati, statistiche tematiche → selezione paese x tutti L1	0,004035187	0,023501763	3,033441
Dati, statistiche tematiche→selezione paese x tutti L2_T	0,003510613	0,020446533	2,585235
Dati, statistiche tematiche → selezione NC2 x tutti i paesi	0,003349205	0,019506463	3,428434
Dati, statistiche tematiche→selezione paese x tutti AT3	0,003228149	0,01880141	3,503282
Dati, statistiche tematiche→selezione AT3_T x tutti i paesi	0,00258252	0,015041128	2,374199
Dati, statistiche tematiche → selezione paese x tutti L2	0,002542168	0,01480611	3,190652
Dati, statistiche tematiche → selezione AT3 x tutti i paesi	0,002421112	0,014101058	3,392742
Dati, statistiche tematiche → selezione AT4 x tutti i paesi	0,002259705	0,013160987	4,235787
Dati, statistiche tematiche → selezione paese x tutti AT2_T	0,002057945	0,011985899	2,106628
Dati, statistiche tematiche → selezione AT5 x tutti i paesi	0,002017593	0,011750881	3,46679
Dati, statistiche tematiche → statistiche in breve	0,001654427	0,009635723	0,149995
Dati, statistiche tematiche → prima dizionari	0,001331612	0,007755582	0,14174
Dati, statistiche tematiche → selezione paese x tutti AT2	0,00129126	0,007520564	2,740803

Legenda:
NC8 = nomenclatura combinata a 8 posizioni, SH6 = sistema armonizzato a 6 posizioni, SH4 = sottocapitoli del sistema armonizzato, NC2 = capitoli del sistema armonizzato, L1_T = Sezioni della classificazione CPAteco (dati territoriali), CPateco = Classificazione dei prodotti associata alle attività economiche, L2_T = Sottosezioni della classificazione CPAteco (dati territoriali), AT2_T = Divisioni della classificazione CPAteco (dati territoriali)
AT3_T = Gruppi della classificazione CPAteco (dati territoriali), L1 = Sezioni della classificazione CPAteco, L2 = Sottosezioni della classificazione CPAteco, AT2 = Divisioni della classificazione CPAteco, AT3 = Gruppi della classificazione CPAteco, AT4 = Classi della classificazione CPAteco, AT5 = Categorie della classificazione CPAteco

La tavola 4 permette di verificare il percorso seguito dagli utenti e la scelta operata nel caso di statistiche tematiche. Sono state scelte le 20 regole con supporto maggiore con antecedente la sequenza "Consultazione dati → Statistiche tematiche". Si ricava un'attenzione dell'utente concentrata principalmente su due classificazioni (la Nomenclatura Combinata e la CPAteco), mentre non viene considerata la classificazione CTCI (Classificazione tipo del commercio internazionale). In effetti anche le frequenze di utilizzo delle statistiche per CTCI dimostrano un interesse molto limitato per questa classificazione che è invece molto utile per il confronto di dati internazionali. Infine il lift minore di uno nella regola "Consultazione dati, statistiche tematiche → Consultazione libera" rivela ancora una volta la scarsa associazione tra i due tipi di consultazione dinamica del sito.

Tavola.5 - Analisi del successivo: "Messaggistica"

REGOLA	SUPPORTO	CONFIDENZA	LIFT
Verifica Login → messaggistica	0,033653458	0,132696897	1,153453
Sezione interrogazioni salvate → messaggistica	0,003954483	0,135546335	1,178221
Statistiche libere (dati territoriali) → messaggistica	0,002622871	0,570175439	4,956187
Statistiche libere (dati nazionali) → messaggistica	0,001614075	0,519480519	4,515527
Registrazione → messaggistica	0,001573723	0,75	6,519291
Modifica Login → messaggistica	0,000887741	0,611111111	5,312015

Dalla tavola 5 si ottengono informazioni sugli antecedenti la pagina di messaggistica, con lift maggiore di uno. Anche se queste regole non si presentano frequentemente, è stato opportuno evidenziarle poiché, come analizzato nel paragrafo sulla pagine di uscita, il 15% di utenti *LE2m* abbandona il sito proprio dopo aver ricevuto un messaggio di errore. Si nota in generale una non facile comprensione delle funzioni legate alla login dell'utente, soprattutto l'utilizzo improprio di utenze non riconosciute dal sistema; inoltre la fase di registrazione e modifica dei dati personali porta molto frequentemente a messaggistica di errore. La messaggistica seguente a interrogazioni impostate nella sezione Statistiche libere deriva fondamentalmente dall'assenza di selezione di parametri obbligatori per l'elaborazione dell'interrogazione dati statistici richiesta.

Conclusioni

In questo lavoro è stata illustrata una metodologia e varie tecniche di data mining per analizzare il comportamento degli utilizzatori del servizio della banca dati statistici Coeweb dell'Istat e quindi valutare l'efficacia e l'efficienza del sito web nel quale è presentata la banca dati. E'stata utilizzata una modalità "indiretta" fondata sulle informazioni acquisite attraverso le e-mail ricevute ed le richieste al contact center e una modalità di analisi "tecnica" basata sull'osservazione dei comportamenti di navigazione degli utenti.

L'applicazione della metodologia proposta consente di fare alcune valutazioni sul sito Coeweb.

In generale il sito è visitato in tutte le sue sezioni con una netta predominanza delle aree di elaborazione dinamica delle tavole statistiche, quindi la banca dati Coeweb è fondamentalmente utilizzata in tutta la sua interezza dagli utenti alla ricerca di dati statistici sui flussi di commercio estero. Gli utenti possono essere divisi in tre gruppi principali:

- Utente "curioso" che si documenta sulle statistiche del commercio estero ed estrae tavole statistiche predefinite.
- Utente "pratico" che elabora tavole statistiche dinamiche pre-formattate dalla sezione "statistiche tematiche"
- Utente "esperto" che elabora tavole statistiche dinamiche con possibilità di selezionare le variabili e le modalità di interesse e formattazione della tavola.

Al fine di fidelizzare maggiormente gli utenti, occorre affrontare le seguenti tematiche:

- *Home page*. Si è evidenziato come una parte di utenti abbandoni il sito già dalla pagina iniziale: può essere utile inserire in essa una tavola statistica d'effetto che costituisca, in termini di comunicazione, uno "strillo", come ad esempio una top-ten per le importazioni e le esportazioni, che faccia da ponte (mediante link sulla tavola) alle sezioni statistiche del sito. Per aumentare l'interesse la tavola proposta dovrebbe

cambiare frequentemente, ogni settimana o, meglio, ogni giorno.

- *Impostazione dei parametri nelle query nelle sezioni "statistiche libere"*. E' un punto critico evidenziato sia attraverso l'analisi con modalità indiretta che con quella tecnica. Effettivamente il processo di selezione dei parametri per l'estrazione delle tavole statistiche non brilla per semplicità, quindi un intervento per semplificare o chiarire i passi che l'utente deve seguire si rende necessario.
- *Registrazione e/o autenticazione utente.* Sia attraverso la modalità tecnica che attraverso la modalità indiretta si è osservato che l'utente trova difficoltà nella procedura di registrazione e/o autenticazione. Occorre dunque agire secondo due linee principali, una a breve ed una a medio-lungo termine. Dapprima si deve cercare di semplificare le procedure di registrazione all'interno del sito Coeweb. Quindi si deve migliorare la procedura di registrazione utente implementando, a livello di Istituto, il *single sign-on* che permette a qualunque utente registrato in una qualunque banca dati Istat di poter accedere con la stessa utenza/password alle altre banche dati di Istituto.
- *Associazione tra le sezioni del sito.*Come le analisi hanno evidenziato, alcune parti del sito (sezioni statiche) si dimostrano essere scollegate dal nucleo principale (sezioni dinamiche). Generalmente questa scarsa associazione può dirsi legata alla diversa tipologia di utenti, pur tuttavia, in alcuni casi (ad esempio la ricerca di codici di classificazione merceologica al fine di impostare i parametri merceologici nella *query)* l'ampliamento dei collegamenti tra le sezioni favorirebbe un uso completo del sito.

Un'ulteriore tema di ricerca deriva dalla possibilità di seguire la traccia dei percorsi effettuati dagli utenti per consentire l'adozione di meccanismi automatici di personalizzazione del sito. La personalizzazione adattativa del sito, basata sulla *clusterizzazione* degli utenti, prevede un sistema che non possiede alcuna indicazione esplicita riguardo ai contenuti da fornire all'utente, ma è in grado di costruire automaticamente il profilo in base alle informazioni sui bisogni dell'utente ricavate autonomamente dall'osservazione del comportamento dello stesso durante la navigazione del sito. In base a queste informazioni il sistema è in grado di selezionare automaticamente i contenuti che più si avvicinano ai bisogni utente.

Si ritiene, in conclusione, necessario ottenere il massimo di integrazione tra l'analisi diretta, indiretta e tecnica per consentire di monitorare costantemente la qualità dei servizi offerti on line e quindi operare per il loro miglioramento.

Riferimenti bibliografici

Berthon P., Pitt L.F., Watson R.T. (1996), *The world wide web as an advertising medium*, Journal of Advertising Research, 36: 43-54

Camillo F., Tassinari G. (2002), *Data mining, web mining e CRM: metodologie, soluzioni e prospettive*, Franco Angeli, Milano

Chang G. (2001), *Mining the world wide web: an information search approach,* Kluwer Academic, Boston

Del Ciello N., Dulli S., Saccardi A. (2000), *Metodi di data mining per il Customer Relationship Management,* Franco Angeli, Milano

Gazzetta Ufficiale Serie Generale n 243 del 18/10/2005 (2005),*"Direttiva per la qualità dei servizi on line e la misurazione della soddisfazione degli utenti"*, Poligrafico di Stato

Fazio N.R. (2002), *COEWEB: la banca dati per la diffusione delle statistiche di commercio estero su Internet* , mimeo

Giudici P.(2001), *Data mining*, McGraw-Hill, Milano

Roiger R.J.,Geatz M.W. (2003), *Introduzione al data mining*, McGraw-Hill, Milano

Scanu A. (2007), *L'applicazione della customer satisfaction nel portale nazionale del cittadino*, Innovazione, n.4, luglio 2007, CNIPA

Spiliopoulou M. (2000), *Web usage mining for Web site evaluation*, Communications of the ACM, 8: 127-134

Srivastava J., Cooley R., Deshpande M., Tan P. (2000), *Web usage mining: discovery and applications of usage patterns from Web data,* ACM SIGKDD Explorations, 2: 12-23

Tasso C., Omero P. (2002), *La personalizzazione dei contenuti web*, Franco Angeli, Milano

Will Italy's Tax Reform Reduce the Corporate Tax Burden? A Microsimulation Analysis[1]

Filippo Oropallo[2], Valentino Parisi [3]

Abstract

This paper analyses the impact of the corporate tax reform introduced in Italy in early 2004 on firms' tax burden with respect to 2001 tax legislation. For this purpose we build a microsimulation model reproducing in detail the Italian corporate tax system under the two regimes. The model is based on an integrated dataset combining ISTAT (National Institute for Statistics) survey data on firms and company accounts for the year 2000. The results show that the mean ex-post implicit tax rate increases by 0.26 percentage points, although for firms belonging to groups and opting for tax consolidation the implicit tax rate falls by 1.18 percentage points, demonstrating that groups are favoured by the new system. We also examine the features of both regimes concerning neutrality over company funding decisions. To this end, we develop a sensitivity analysis in which we consider two scenarios in terms of company financial policy (debt, internal sources) and, using the microsimulation tool, compute implicit tax rates in each regime. We find that the new regime widens the distortion in favour of debt an can thus be regarded as less efficient than the previous system.

Keywords. Corporate tax, Data matching, Microsimulation, Effective tax rates, Performance indicators, Tax efficiency; JEL classification: H25, H32

1. Introduction

From its inception in the early 1970s, the Italian business income tax regime changed only marginally for over twenty years.[4] Then, in 1997 and again in early 2004 it was

[1] This paper stems from the authors' participation in the DIECOFIS (Development of a system of Indicators on Economic COmpetitiveness and FIScal impact on enterprise performance) project financed by the Information Society Technologies Programme (IST-2000-31125) of the European Commission. The model presented here relies on micro data from the SCI (Sistema dei Conti delle Imprese) and PMI (Piccole e Medie Imprese) ISTAT surveys and company accounts data from the Italian Chamber of Commerce, both available at ISTAT within the DIECOFIS project. Data were used at ISTAT to run the model and produce results analysed here. ISTAT bears no responsibility for analysis or interpretation of the data. An earlier version of this paper was presented to the 60th Congress of the International Institute of Public Finance (Fiscal and Regulatory Competition), Milan, August 2004. The authors wish to thank the participants at the IIPF Congress and also Laura Castellucci, Maria Grazia Pazienza, Paolo Roberti (scientific coordinator of the project DIECOFIS) for their comments and suggestions. The authors are also grateful to an anonymous referee for helpful comments. The usual disclaimer applies.

[2] Ricercatore (Istat), e-mail: oropallo@istat.it.

[3] Ricercatore (Università di Cassino), e-mail: valentino.parisi@eco.unicas.it.

[4] Until the mid-1990s, while other countries adopted reforms of the base-broadening/statutory rate cut type (Messere et al., 2003), Italy moved in the opposite direction, actually increasing the corporate tax rate. In 1997, on the eve of the first reform discussed here, the system contemplated a corporate income tax (IRPEG) with a rate of 37%, the so-called local income tax (ILOR) with a rate of 16.2%, basically an additional tax on profits, and a 0.72% tax on companies' net assets. The combined rate amounted to 53.95%.

overhauled with the declared objective of simplifying the system and reducing the tax burden on firms. However, a closer look at the rationale behind these two reforms reveals important differences (Maurizi and Monacelli, 2003, Giannini, 2002).

In 1997 the primary policy objective was a selective reduction in the burden of taxation, to reduce the tax distortion between equity and debt financing. The main change introduced to this end was the so-called Dual Income Tax (DIT) system, basically an allowance for corporate equity, with a lower statutory rate applied to the portion of profits representing the opportunity cost of new equity financing compared with other forms of capital investment. This system structurally reduced the corporate tax burden depending on the amount of the capital increase (new capital subscription and retained earnings) carried out by the company.

By contrast, the policy design envisaged by the 2004 reform posits that tax measures aimed at modifying firms' financial decision tend to introduce distortions in firms' behaviour and should therefore be eliminated. Consequently, the reform abolishes the DIT system and reinstates a uniform tax rate. Furthermore, it modifies the corporate tax base by introducing a participation-exemption regime and eliminating the full imputation of dividends, and brings in an optional consolidated tax treatment for corporate groups, with a view to simplifying computation of the tax base.[5]

In this paper we review the key elements of the two regimes and offer an assessment of the 2004 reform by analysing its impact on firms' tax burden. In the present context of European Monetary Union, where competitive devaluations of the domestic currency are ruled out, this is clearly seen as a key factor in driving firms' competitiveness and, in general, in fostering economic growth. Furthermore, given the international trend of increasing fiscal competition (Devereux and Sørensen, 2005), reducing the corporate tax burden is deemed to be desirable in order to attract multinational companies and to deter domestic enterprises from locating abroad.

To explore the effects of the reform we develop a microsimulation model that reproduces in detail the corporation tax system under the two regimes. The model is based on an integrated dataset built by the authors of this paper by integrating company accounts data with survey data on firms for the year 2000. The microsimulation model is static in the sense that it does not include firms' behavioural responses, and so the empirical analysis only examines the first-round impact of the tax policy changes on firms.

In evaluating the impact of corporate taxation on enterprise activity, the empirical literature offers two type of effective tax rates, ex-post implicit tax rates and ex-ante marginal tax rates. The first relate taxes paid by the company to some aggregate item of the company accounts, such as gross profit or gross operating profits. As they use ex-post real-life data, they are often described as backward-looking indicators reflecting the fact that measures of effective taxation imply past investment decisions. By contrast, ex-ante marginal tax rates follow a forward-looking approach focussing on the enterprise's marginal decisions and are based on computations of the impact of taxes on the cost of capital. Specifically, ex-ante tax rates measure the theoretical tax burden on a hypothetical marginal investment (giving no extra-profits) that produces cash-flow subject to tax and, therefore, are calculated to analyse how the tax system affects a marginal investment undertaken by the company, using alternative financial sources (equity, debt, retained

[5] It is noteworthy that the new system actually mirrors some features of the reform introduced in 2000 in Germany (Keen, 2002).

earnings).[6] The methodology to derive ex-ante marginal tax rates was first developed by King and Fullerton (1984) and then extended by Devereux and Griffith (1998) to infra-marginal investments, i.e. investments with different rates of profitability. In the latter case the literature refers to ex-ante average tax rates.

Being simplified measures, forward-looking indicators do not take into account the complexity and the interaction of all elements of the tax system (definition of profits for tax purposes, carry-forward losses provisions, allowances, tax credits and so on) that crucially alter effective company taxation. By contrast, implicit tax rates can be derived considering the various features of the tax system and therefore give a precise measure of the effective tax burden supported by the firm. Such rates are especially appropriate if the objective is to study the effects of the tax system on enterprise cash flows and to focus on distributional burdens (for instance, at sectoral level or on firms of different size).

In this paper we estimate ex-post implicit tax rates calculated as ratios of tax paid on companies' operating surplus to study the impact of the 2004 tax reform on corporate cash-flow.

Clearly one central issue regards the efficiency effects of the new regime on firms' funding choices, given that the dual-rate system was meant to correct the distortion in favour of debt financing present in the pre-1997 system (Bordignon et al., 2001). To explore these aspects in detail we perform a sensitivity analysis where we posit two different scenarios in terms of firms' financing decisions (debt funding, capital) and employ the microsimulation model to compute the implicit tax rates under these hypothetical scenarios. The purpose of this analysis is twofold: first, to provide a measure of the distortion associated with the pre and the post-reform corporate tax system in using debt rather than equity as a financing source, second, to obtain indications on the possible variations of the implicit tax rates if one moves away from the 'static' scenario and assumes changes in firms' behaviour regarding financial policy.

The paper is organised as follows. Section 2 is devoted to the theoretical issues underlying the aspects discussed in this paper. The main features of the two reforms are then discussed in sections 3 and 4. Section 5 describes the dataset used in the empirical analysis, and section 6 presents and discusses the simulation results. In section 7 we explore the efficiency aspects of the reform on firms' funding decisions. Finally, in section 8 we offer some concluding remarks. The methodology used to build the microsimulation model is explained in the Appendix.

2. Distortionary effects of corporate taxation: the theoretical background

Economic efficiency, usually identified with neutrality, is by far the most important consideration when designing a corporate tax system. Generally speaking, a tax on firms is efficient when it leaves the firm behaviour unchanged after taxation, that is when decisions undertaken by the firm is unaffected by the presence of the tax system. In this sense the efficiency features of a corporate tax system can be studied regarding the investment decisions of the firm as well the as company's financing decisions. The latter aspect has

[6] It must be noted that the theoretical model on which this approach is based implies restrictive assumptions, such as perfect information, perfect competition, and no risk.

received great attention both in the theoretical and empirical literature.

As well known from the Modigliani and Miller theorem, in the absence of taxation, and in the absence of imperfections in the capital markets and information, firms are indifferent whether they finance their investments through debt or equity capital. This result changes when taxes are introduced. As the firm has three main financing sources, i.e. debt, retained earnings, and new shares issues, it can the be demonstrated that the corporate tax system is efficient (neutral) over the company financing decisions if the flow of before-tax profits remains unchanged after taxes for the marginal investors, whether the return of the investors takes the form of interest, dividends, capital gains. The tax system changes this picture as interest payments are usually deductible from the corporate tax base while dividends are not. In this sense a corporate tax system is distortionary, that is it is not neutral over the company funding sources in that it favours debt over equity capital financing. The magnitude of these distortions then depends on the specific features of the corporate tax and the personal tax systems, and the quantification of these distortions has been at the centre of the discussions on the optimal design of corporate tax systems.

Specific systems can be proposed to address this issue. In 1991, the Institute for Fiscal Studies (1991) suggested introducing an Allowance for Corporate Equity (ACE) on which as we will see below the Italian DIT system is inspired. The basic idea is to provide a deduction of a notional return on the company equity from taxable profits so to address the difference in the tax treatment of debt and equity. Such systems have been in operation in several countries, though with differences in their practical application (Klemm, 2007). A number of interesting properties of the ACE systems can be listed and two specific aspects are worth mentioning here. Assuming that parameters are chosen correctly, the first obvious feature of this system is that it ensures neutrality between debt and equity financing. The second important feature is that the corporate tax is not levied on marginal investments. Indeed, under the ACE system the tax is charged only on business extra-profits (economic rents) while normal profits are exempt from taxation. An ACE system is therefore neutral to firm investment decisions.

3. The corporate tax reform of 1997 and the DIT system: an overview

The DIT scheme was implemented in 1997 with the general aim of reducing the discrimination against equity finance and lowering the effective tax rate.[7] It remained in place until its repeal at the beginning of 2004, although some modifications were introduced in July 2001 in order to rein in its effects.

Table 1 summarises the main changes to the corporate tax system introduced in the period 1997-2004.

[7] The Dual Income Tax systems of some northern European countries, as well as the ACE system, were clearly taken into consideration when designing the 1997 tax reform. On these aspects see Bordignon, Giannini and Panteghini (2001). It is important to stress that the DIT allowance, like the ACE system, applied to both the corporate and non-corporate sector.

Table 1 - Changes to the corporate tax system enacted in the period 1997-2004

1997	Introduction of a dual rate (DIT) system *Main features*: ▪ profits are divided into two component: normal profits (imputed return on capital increases) are taxed at the preferential rate of 19%, extra-profits at the statutory rate of 37% ▪ in computing the DIT allowance, the effective corporate tax rate must not fall below 27%
2000	Introduction of the so-called multiplier: in computing normal profits, capital increases are multiplied by 20%
2001 (before July)	1. The multiplier is increased to 40% 3. The floor of 27% for the effective rate is removed 4. The statutory rate is lowered to 36%
2001 (after July)	The DIT system is frozen (introduction of changes in the computation of ordinary income, abolition of the multiplier)
2003	The statutory rate is cut to 34%
2004	Introduction of the corporate tax reform. The DIT system is definitely repealed. *Main features:* ▪ the statutory rate is 33% ▪ participation-exemption regime for both capital gains and dividends; repeal of dividend tax relief ▪ thin capitalisation rules ▪ consolidated group taxation (optional)

The DIT system works as a dual-rate schedule in which overall profits are divided into two components. The first approximates normal profits or ordinary income, i.e. the opportunity cost of new financing with equity capital (in the form of new capital subscriptions and retained earnings) compared with other forms of capital investments, and is taxed at the preferential rate of 19%. Ordinary income is calculated by applying an assigned nominal rate of return to equity capital injected after 30/09/1996 (when the reform was actually presented) net of the increases (again after 30/09/1996) in loans to subsidiaries, loans to parent companies, or other investments held as fixed assets by the firm. The nominal rate, set yearly by the government, was 7% from 1997 to 2000 and 6% in 2001.

The second component of overall profits is computed residually from total profits after ordinary income and represents business extra-profits. It was taxed at the prevailing statutory rate of 37% up to 2000, cut to 36% in 2001. In order to limit revenue losses resulting from the introduction of the dual-rate schedule, the law fixed a floor of 27% for the average effective corporate tax rate. Furthermore, it permitted firms to bring allowable DIT profits forward up to five years whenever they could not benefit from the reduced rate, i.e. when they incurred losses and when ordinary profits exceeded total taxable income.

In the first years of application, the dual-rate system mainly benefited new and less-well capitalised enterprises rather than strongly capitalised companies (Bordignon et al., 2001). In order to accelerate the impact of the reform, in 2000 some adjustments were made to the original mechanism.[8] Specifically, when computing ordinary income capital increases were

[8] In addition, in the years 1999-2001 a temporary measure was introduced for both corporations and unincorporated firms that worked basically as an incentive scheme for investments. This allowance could be cumulated with the DIT system, strengthening its effects and its general purposes. The share of profits corresponding to the amount of investments in new producer goods financed out of the company's own capital was taxed at a reduced rate of 19% rather than the statutory tax rate. In this way, profits corresponding to the amount of new investments were taxed at a lower rate when investments were made, while ordinary income resulting from the same capital increases could benefit from the reduced rate in the following periods.

to be multiplied (up to the enterprise net wealth threshold) by a conventional parameter set first at 20% in 2000 and then at 40% in 2001. Obviously, the idea the policy maker had in mind was to make the system a regime in which normal profits would be computed on the enterprise's entire capital stock rather than on capital increases. Moreover, in 2001 the constraint under which the average statutory rate resulting from the application of the DIT could not fall below 27% was removed.

Formally, under the DIT regime in place in July 2001, the total amount of corporate tax (T_C) can be written as follows:

$$T_C = t\left(\Pi - 1.4r\Delta K_{96}\right) + t'1.4r\Delta K_{96} \tag{3.1}$$

where Π represents total taxable profits, r is the imputed nominal rate (6%), t the statutory corporate tax rate (36%), t' the preferential tax rate (19%), and ΔK_{96} net capital increases evaluated with reference to 1996, as explained above. Therefore, under the DIT scheme the effective statutory rate ranges between t and t', depending on the amount of profits qualifying for the allowance (ΔK_{96}).

In July 2001, when the new government took office, some changes were made to the DIT scheme in order to curb its effects. These changes anticipated the intention of the (new) policy maker to repeal the dual-rate allowance (it was in fact repealed at the beginning of 2004 when the tax reform came into effect). The measures in question froze the capital increases to be taken into account when computing ordinary income at those carried out until July 2001, lowered the imputed nominal rate from 6% to 3%, and abolished the 'multiplier'.[9]

Lastly, in 2003 the statutory corporate tax rate was reduced by 2 percentage points, to 34% .

4. The Corporate tax reform of 2004

As summarised in Table 1, the main characteristics of the 2004 corporate tax reform and the new corporate income tax (IRES, *Imposta sul Reddito delle Società*) are:

i) the abolition of the DIT scheme and the introduction of a single rate of 33%;

ii) the introduction of a participation-exemption regime;

iii) the exemption of corporate dividends along with the abolition of the dividend tax credit;

iv) the introduction of thin capitalisation rules;

iv) the introduction of an optional consolidated tax declaration for groups that can be extended also to foreign subsidiaries.

Another feature of the full reform project, initially presented at the end of 2001, is the abolition of IRAP. This is a regional tax paid by corporations and unincorporated firms on their value added net of depreciation and amortisations, i.e. with no deduction of interest expense and labour costs from the tax base. The statutory tax rate is 4.25%, although since 2000 regions may vary the rate within specific limits. IRAP was introduced in 1998 as a replacement for other taxes[10] and heath insurance contributions, for reasons of simplification. As IRAP is the basic source of revenue for the National Health System, its abolition will necessarily be gradual.

[9] An optional regime contemplating the application of the multiplier could be used but under the constraint of a minimum average rate of 30%. In July 2001 a new temporary (for the second half of 2001 and for the year 2002) investment tax incentive replaced the previous one (see note 5).

[10] ILOR and the tax on firms' net assets.

The declared policy aim of the 2004 tax reform is to simplify the tax treatment of firms through standardisation of capital income taxation, the abolition of the dividend tax credit and group taxation, as well as to foster firms' competitiveness. Concerning the neutrality issue, as we will see in greater detail in section 6, the idea behind the reform is that the tax system should not interfere with firms' financing decisions. Consequently, the combined system of incentives for equity capital (provided by the DIT allowance) and taxation of interest (by IRAP), designed in the previous tax regime to balance fiscal discrimination, is eliminated.[11]

As mentioned, the corporate tax reform abolishes the dual-rate system and establishes a uniform corporate tax rate of 33%.

Among its most important innovations is the introduction of a consolidated tax regime for groups. Companies belonging to the same group can opt for tax consolidation, making it possible to offset profits and losses between group companies. The control requirement for consolidation is met when a company holds, directly or indirectly, more than 50% of the share capital of another company and the parent company can select which subsidiaries will be included in tax consolidation. Group taxation can also be extended to non-resident subsidiaries, although in this case consolidation must include all foreign subsidiaries[12] and their income can be attributed to the parent company only in proportion to the percentage of ownership, while in the domestic case there is no such restriction. The Italian system appears to be favourable compared with the group tax regimes of other EU countries, where eligibility rules are generally more restrictive (European Commission, 2001).

A second important feature of the reform is the introduction of a participation-exemption regime, where inter-corporate capital gains are exempt from taxation, and the exemption of dividends along with the abolition of the full imputation of dividend tax relief. These rules aim in general at avoiding double taxation of inter-corporate incomes (capital gains as well as dividends) and, as far as dividend taxation is concerned, respond to international issues, as the imputation system tends to favour resident tax payers over non-residents (Giannini, 2003, Keen, 2002). Capital gains on shareholdings in other companies (resident or non-resident) are exempt from taxation provided that: (i) the equity interest is recorded as a long-term asset and has been owned for at least one year; (ii) the subsidiary carries out a business activity; (iii) the subsidiary is not resident in a tax haven. Symmetrically, capital losses are not tax deductible if the above conditions are met. Dividends paid by an investee company (resident or non-resident) are 95% excluded from the corporate tax base; in the case of consolidated taxation, the exemption is 100%. Again, these exemptions do not apply if the investee is resident in a tax haven.

The post-reform regime also establishes rules against thin capitalisation, mainly for anti-avoidance purposes. Accordingly, it introduces a debt-to-equity ratio,[13] to prevent thin capitalisation. When the financial debts owed to or secured by holders of an equity interest of 10% or more in the company exceed this threshold, interest expense is treated as dividends paid and cannot be deducted from the tax base. Should the debt-to-equity ratio be disallowed, the company must demonstrate that the excess amount of the financial debt is based on the company's (rather than the shareholder's) credit capacity.

[11] The announced abolition of IRAP also reflects the necessity of eliminating a tax that has no counterpart in the tax systems prevailing in most EU countries.

[12] The option remains in effect for at least three years in the case of domestic consolidation, five years in the regime for foreign subsidiaries.

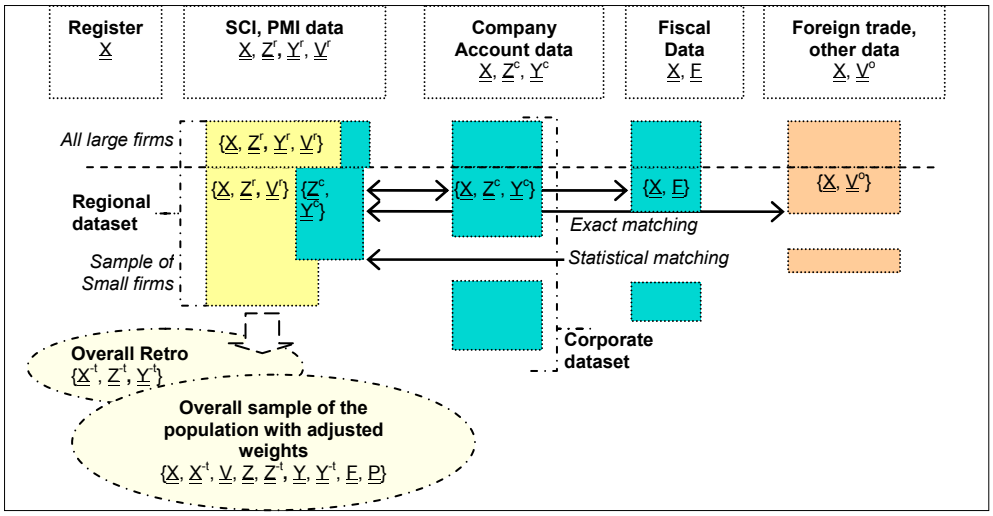
[13] The law sets a ratio of 5:1 for the first year, 4:1 for the subsequent year. In the simulations we use this ratio.

5. Data description

The Corporate Tax microsimulation Model[14] (CTM) used in this paper is based on a specific dataset obtained by integrating survey data on firms with company accounts data which makes it possible to have a complete representation of the corporate tax system. In the analysis we use data of the year 2000.

Figure 1 illustrates the features of the data sources and the steps we followed in order to obtain the integrated final dataset.

Figure 1 - Integration scheme: sources, units and variables - Year 2000



Legend:
↔ Exact matching (one to one)
← Statistical matching (similar to similar)
$\underline{X}$ = Matrix **Register** (around 4 million of firms)
$\underline{Z}^r$ = Matrix Profit & Loss of SCI and PMI surveys combined in the **Regional** dataset (62.900 firms)
$\underline{Y}^r$ = Matrix Assets & Liabilities of SCI survey (roughly 9,300 large firms)
$\underline{V}^r$ = Matrix Employment and other variables (SCI and PMI surveys combined in the **Regional** dataset)
$\underline{Z}^c$ = Matrix Profit & Loss of **Corporate** dataset (around 489 thousand corporate firms)
$\underline{Y}^c$ = Matrix Assets & Liabilities of Corporate dataset (around 489 thousand corporate firms)
$\underline{F}$ = Matrix of the sections RF RN RJ RU RS of **Fiscal** returns declaration (tax receipts fiscal datasets)
$\underline{V}^o$ = Matrix of other datasets
$\{\underline{X}^{-t}, \underline{Z}^{-t}, \underline{Y}^{-t}\}$ = Matrices with retrospective information (t= 1996, 1997, 1998, 1999)

The sources involved in the integration process are:

- The statistical Register (Asia);
- The Business Structural Surveys (SCI and PMI);
- Administrative data (Company accounts and Fiscal data);
- Other statistical sources (Foreign trade survey etc.).

[14] The authors developed the microsimulation model and the dataset as part of the DIECOFIS project carried out by a consortium made up of: ISTAT, the Board of Inland Revenue (UK), the Joint Research Centre of the European Commission (Applied Statistics Sector), Informer S.A., the London School of Economics, the University of Cambridge, the University of Economics and Business Administration of Vienna (Wirtschaftsuniversitaet), the University of Rome Tor Vergata, the University of Florence, and the Centre of Economic and Social Research (CERES, Italy).

The (spine) information used as a basis for the integration process is represented by the statistical register (matrix $\underline{\underline{X}}$) of Italian active enterprises (acronym ASIA), which covers all economic activities except agricultural, public and non-profit sectors. As of 2000 the business register covers roughly 4,222,657 firms, counting 555,621 corporations and 3,656,509 unincorporated enterprises (sole proprietorships and partnerships). The register includes basic information on the firm as well as variables (geographical reference, activity sector, legal status, size, turnover) that can be used as auxiliary variables in the imputation process when integrating the various data sources.

The main statistical sources are two surveys conducted yearly by ISTAT on both incorporated and unincorporated firms: the survey of small and medium-sized enterprises (acronym PMI) regarding firms with fewer than 100 workers, and the survey of large enterprises (acronym SCI) concerning firms with more than 99 workers.

Table 2 - Number of statistical units in survey sources by legal types - Year 2000

Sources and legal types	Unincorporated	Corporations	Total
SCI survey	1,175	8,043	9,218
PMI survey	35,220	18,374	53,594
Total sample	**36,395**	**26,417**	**62,812**

Source: ISTAT

The SCI survey is exhaustive, embracing the universe of large firms (of which 8 thousand corporations), whereas the PMI survey is carried out on a sample of firms (of which 18 thousand corporations).

The integrated dataset compounds two main administrative sources, the company accounts database containing information about Assets and economic accounts of about 489 thousand corporate firms, and tax returns data containing information about differences between the balance sheets profits and the corporate tax base. Fiscal data are available for all large corporate firms and for a sample of small and medium sized firms (PMI survey sample).

As shown in Figure 1, surveys contain variables of the company accounts (matrices $\underline{\underline{Z}}^s$ and $\underline{\underline{Y}}^s$) and variables pertaining to the firm's employment, investments and other information on the activity of the firm (matrix $\underline{\underline{V}}^s$). As the PMI survey includes only the profit and loss account and because both for PMI and SCI surveys specific items in the administrative archive (box Company Accounts of the chart) are reported at a more disaggregated level, survey data are matched against the administrative data. The integration process[15] allows us to reconstruct the balance sheet of firms covered by the PMI survey, as well as to impute specific variables that are needed for tax modelling purposes for companies of both the PMI and SCI surveys. In the data reconstruction process two main issues have been faced: (i) incoherent values across survey and administrative sources; (ii) mismatches between survey and administrative data. To overcome the first problem, we calculate a discrepancy variable in order to identify the incoherent units that must be deleted.

[15] For a detailed description see Oropallo (2005).

For the second problem, a statistical matching procedure is used in order to impute data of similar units. Imputation of missing information uses the deck imputation technique based on the nearest neighbour search (Little and Rubin, 1987), in which similar units are found by means of a mixed distance function (Abbate, 1998). At the end of the process the sample weighs are recalculated to comply with the corporate sector population.

Furthermore, for tax modelling purposes, the final database also includes data of previous years (1996-1999) for specific variables[16], as shown in Figure 1 (*Overall-retro* data).

Table 3 displays the total number of companies present in the final dataset by business sector, comprising 18,187 small and medium-sized companies and about 8,000 large corporations; overall, the dataset includes 26,196 companies out of a population of about 556,000.

Table 3 - Number of companies present in the database by sector of activity - Year 2000

Sector of activity	Small and medium- sized firms	Large firms	Total	Not part of groups	Part of groups	
					Parents	Subsidiaries
Products of mining and quarrying	218	13	231	170	13	48
Manufacturing	6,978	4,443	11,421	6,719	807	3,895
Electrical, energy, gas, steam and water	245	74	319	188	23	108
Construction	705	299	1,004	626	125	253
Wholesale and retail trade	3,243	711	3,954	2,680	255	1,019
Hotel and restaurant services	326	197	523	322	36	165
Transport, storage and communication services	1,248	673	1,921	1,302	132	487
Real estate renting and business services	3,634	1,037	4,671	3,126	287	1,258
Education services	250	11	261	229	5	27
Health and social work services	373	387	760	635	47	78
Other social and personal services	967	164	1,131	848	46	237
Total	**18,187**	**8,009**	**26,196**	**16,845**	**1,776**	**7,575**

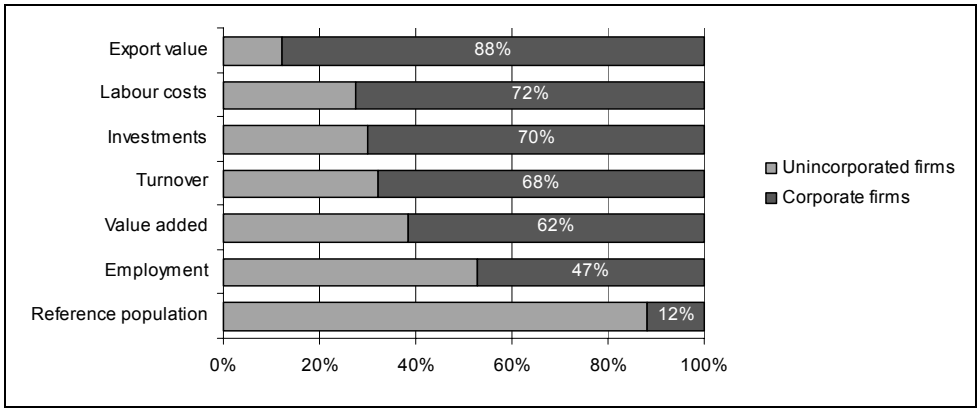
Source: ISTAT

[16]Integration between survey data and company accounts is also applied to 1999 data and, for the SCI survey alone, to the year 1998. Therefore the model simulates the corporate tax for the year 1998 (large companies) and for the years 1999 and 2000 (both large and small and medium sized firms).

The data also contain information on group structures,[17] which proves to be of crucial importance to analyse the impact of 2004 corporate tax reform. The ASIA register contains 17,968 corporate groups, comprising about 103,000 companies.[18] The dataset enumerates 1,776 parent companies and 7,575 subsidiaries, roughly 10% of the reference population. Table 3 also reports the number of group companies (parent companies and subsidiaries) and non-group companies in the dataset, by business sector.

In order to give a picture about the importance of the corporate sector over the population of Italian firms, Figure 2 shows their weight in terms of number of firms, employment and economic performance.

Figure 2 - Weight of corporate firms in terms of firms, employment and economic performance - Year 2000



Source: ISTAT

Only 12% of the total number of enterprises are limited liability companies. They employ 47% of total workers and produce almost two thirds of the total value added, while the share of total investments and export over the total is, respectively, 72% and 88%.

To complete the descriptive analysis of the dataset, table 4 offers some economic indicators for the population of Italian firms and for the corporate ones involved in the tax simulation analysis.

[17] The reconstruction of groups structures performed at ISTAT uses other data sources, namely: (i) the Shareholders Database available from the Italian Chambers of Commerce; (ii) the Ownership Transparency Database available from the Italian securities regulator (Consob); (iii) the Chambers of Commerce Consolidated Financial Statement data. The procedure adheres to the Italian Civil Code's definition of a controlling company as one that holds, directly or indirectly, more than 50% of the share capital of another company. This is the same requirement established by the Italian tax code after the 2004 reform for companies electing to consolidated taxation. The algorithm developed at ISTAT makes it possible to reconstruct for each group the 'chains of control' and to identify the company with no companies controlling it that is at the top of the group (parent company).

[18] As of 2000 ISTAT estimates on the whole 48,331 groups (Garofalo, Morganti, 2000 and Cerroni, Morganti, 2003), including unincorporated and corporate parent enterprises residing in Italy as well as abroad. It is noteworthy that in Italy a substantial share (28.8% in 2000) of parent firms are represented by individuals which are subject to the personal income tax.

Table 4 - Main economic indicators by legal type, economic activity and size - Year 2000

Legal type	Economic Activity	Size	Average size	Turnover per worker (€)	Value added per worker (€)	Labour cost per employee (€)	Investment per worker (€)	Share of export on sales (%)
Unincorporated	Industry	*Small-Medium*	*3.0*	*69,598*	*24,411*	*19,805*	*4,040*	*7.3*
		Large	*320.2*	*326,829*	*85,321*	*38,530*	*17,088*	*15.8*
		Total	3.2	86,118	28,323	21,849	4,878	9.4
	Service activities	*Small-Medium*	*1.8*	*84,112*	*23,886*	*20,296*	*3,163*	*1.3*
		Large	*425.3*	*192,274*	*56,213*	*30,053*	*15,765*	*1.2*
		Total	1.9	89,408	25,469	21,732	3,780	1.3
	Total		**2.2**	**88,214**	**26,505**	**21,791**	**4,178**	**4.1**
Corporate firms	Industry	*Small-Medium*	*12.2*	*162,604*	*39,723*	*26,318*	*7,555*	*18.0*
		Large	*345.5*	*280,942*	*68,407*	*36,849*	*16,090*	*26.2*
		Total	21.3	215,096	52,447	31,269	11,341	22.7
	Service activities	*Small-Medium*	*6.1*	*227,919*	*37,246*	*25,399*	*7,279*	*4.4*
		Large	*468.6*	*166,402*	*48,355*	*30,627*	*14,862*	*2.5*
		Total	10.8	201,006	42,106	27,982	10,596	3.7
	Total		**14.6**	**208,400**	**47,533**	**29,763**	**10,987**	**14.0**
Total			**3.7**	**144,879**	**36,419**	**27,030**	**7,388**	**10.8**

Source: ISTAT

In particular, the table above shows the difference in terms of employment structure and economic performance between corporate firms and unincorporated firms (single ownership and partnerships). Corporations employ a greater number of workers and have higher productivity in terms of value added per worker. Also, if we look at the investment expenditure per worker and at the share of export on sales, we see that their performance is sharply above the unincorporated ones'.

6. Simulation results

The empirical analysis considers two policy scenarios. The base-case reproduces the corporate tax structure existing at July 2001, just before the practical abolition of the DIT mechanism,[19] while the second scenario considers the 2004 corporate tax reform. Simulations of both scenarios are run on the 2000 dataset.

The impact of the tax reform depends both on the modifications of the corporate tax base under the new regime and on the introduction of the unified rate of 33%, as opposed to the effective rate prevailing in the 2001 scenario under the DIT system. As explained in section 2, the effective rate ranges between the preferential rate of 19% and the statutory rate of 36%, depending on the amount of profits eligible for the allowance. To estimate the effects of the DIT system on

[19] The idea behind the empirical analysis carried out in this paper is to compare the structure of the DIT system before this was 'frozen' with the new regime. Therefore, in the base-case we do not consider the temporary incentive on investments introduced in 1999 and then repealed in 2001.

companies as of 2001 we therefore first compute the effective statutory tax rates,[20] reported in Tables 5 and 6 respectively by firms' sector of activity and size.[21]

Table 5 - Effective statutory corporate tax rates (ESTR) resulting from the DIT system in 2001. Breakdown by activity sector (percentages)

Sector of activity	ESTR
Mining and quarrying	29.74
Manufacturing	33.14
Electricity, gas, steam and water	29.10
Construction	32.08
Wholesale and retail trade	32.31
Hotels and restaurants	31.05
Transport, storage and communication	32.53
Real estate renting and business activities	32.82
Education	33.22
Health and social work	34.27
Other services	31.65
Total	**32.59**

Source: Authors' estimates.

Table 6 - Effective statutory corporate tax rates (ESTR) resulting from the DIT system in 2001. Breakdown by firm size (number of workers, percentages)

Size	ESTR
From 1 to 2	31.90
From 3 to 9	32.85
From 10 to 19	33.13
From 20 to 49	33.43
From 50 to 99	33.25
From 100 to 249	33.35
From 250 to 499	33.07
From 500 to 999	33.09
More than 999	33.19
Total	32.59

Source: Authors' estimates.

In the reform scenario, 2004, we simulate the effects of:

- the abolition of the DIT scheme and the introduction of a single rate of taxation of 33%;
- the exemption of capital gains on shares owned for at least one year and recorded as long- term

[20] These effective statutory rates measure the average statutory rates actually applied to the tax base because of the dual-rate schedule existing in 2001. They are calculated as ratios of the gross corporate tax (before tax reliefs) to taxable profits (computed from reported profits adjusted for tax purposes after losses from previous years carried forward and before the dividend tax relief). For details on how these aggregates are defined see the Appendix.

[21] To estimate fully the effects of the DIT system, ideally the 2001 scenario simulation should be run using data of July 2001, before this system was frozen. This is generally true for all simulations referring to tax laws of different years, which should use data for the same years, and therefore also in the 2004 regime. The other possibility could be to update the main company accounts variables, but this procedure would inevitably be imprecise and present strong biases. We perform analyses using 2000 accounts both in the base-case and in the reformed scenario. As regards the effects of the DIT system, therefore, we might expect the effective statutory rates to be lower than the estimated ones owing to larger capital increases carried out by companies between January and July 2001.

assets, and the symmetric non-deductibility of capital losses if the same conditions occur;[22]

- the introduction of thin capitalisation rules limiting the amount of interest expense that can be deducted from the tax base;
- the exemption of 95% of dividends and the abolition of the dividend tax relief;
- the introduction of the optional group taxation regime for domestic companies,[23] in which case dividends from companies of the same group are fully exempted from taxation.

Tables 7 and 8 present the estimated ex-post implicit tax rates both in the 2001 regime and in the reformed scenario, along with the absolute differences, by sector of activity and firm size. The implicit rates are computed as ratios of the corporate tax owed to the operating surplus recorded in the 2000 company account.[24] The tables also display the percentage number of firms by sector and size class.

To complement the analysis of the changes in the tax burden under the reform with a descriptive statistic summarising differences between companies of the various sectors/size in terms of economic performance, we construct a specific indicator by considering three dimensions of firm performance – value added, export and investments – for the years 1996, 1998 and 2000. The methodology is briefly described below.

Using the decomposition of the Gini index,[25] total inequality can be broken into three components, respectively within and between inequality and an overlapping term due to the fact that the Gini index is not perfectly decomposable, as follows:

$$G = \sum_{k=1}^{K} G_k p_k \pi_k + \frac{1}{\mu} \sum_{k}^{K} \sum_{k>i}^{K} (y_k - y_i) p_k p_i + L \qquad (6.1)$$

where p_k represents the weight of class k (k=1,2,…,K), π_k the share of variable y in class k, μ the mean of variable y. After disaggregating firms by classes k (sector, size) and ordering enterprises by values of y, the between component gives a weighted distance measure between each class k, with classes i exhibiting lower mean values of y (k>i). The indicator ranges between 0, when the mean of y is the same in each class and total inequality is thus due only to differences within classes and to the overlapping component, and 100, when only inequality between classes is present and there are no within-class differences and no overlapping effects.

Generally, we can derive a composite indicator (BC) by aggregating several dimensions (d) of firm performance, as follows:

$$BC_k = \frac{1}{D} \sum_{d=1}^{D} b_k^d \quad \forall k \qquad (6.2)$$

[22] The information available in the dataset is not detailed enough to compute the amount of capital gains/losses potentially eligible for the exemption/non deductibility rule or to identify interest expense subject to the thin capitalisation rule (as described in point iii) from the aggregate variables. Accordingly, in the analysis we follow the same procedure developed in the Government's technical report to Parliament on the tax reform (Ministero dell'Economia e delle Finanze, 2003).

[23] As the data do not cover foreign subsidiaries, we can only simulate the impact of consolidated taxation for resident firms. In simulating the optional regime for each group we assume all subsidiaries are included in group consolidation. In addition, excess tax credits that firms can carry forward up to five years can be transferred to the parent company. Any pre-consolidation losses can be set against future profits of the company that incurred the losses but cannot be deducted from the group tax base.

[24] Formally, the law lays down that for firms electing consolidation the (aggregate) corporate tax is to be paid by the parent company, and so for firms belonging to groups the results refer only to the parent company. In this case the implicit rate is calculated as the ratio of group tax to group operating surplus.

[25] This analysis builds on that proposed by Milanovic (2002) for poverty studies.

where b_k^d is the performance indicator for each class *k* and dimension *d*, as defined by the second term of equation 2. As mentioned, in this study the composite indicator is computed using three dimensions: value added, export, investments ($BC=b^{va}+b^{exp}+b^{inv}$). The final idea is to rank firms belonging to different classes from the most competitive (best performing) to the least. The upper part of Figures 3 and 4 plots the percentage values of the performance indicator with reference to business sector and firm size, respectively. For the sake of exposition, these figures also reports, in the lower part, the absolute change in ex-post corporate tax rates for each sector and size class. The classes are shown in descending order on the basis of the performance statistic.

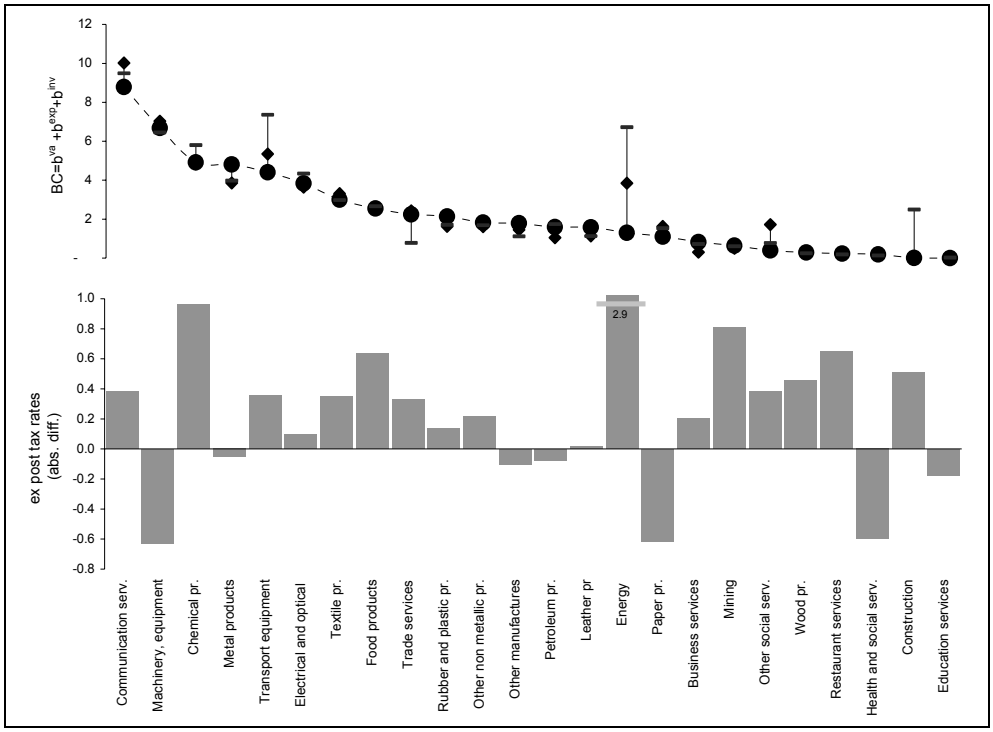
As a total effect, the reform reduces the corporate tax base by about 3 percentage points, while total tax revenue decreases by 0.9 percentage points. Table 7 shows that in the new regime the mean implicit tax rate rises by 0.26 percentage points, from 18.01% to 18.27%. The figures highlight some differences in the effective company tax burden due to firm-specific characteristics (production function) and features of the corporate tax system (depreciation rates, allowances, tax reliefs and so on). In the 2001 regime, the implicit rate of taxation ranges from about 10% for 'Other social and personal services' to about 20% for firms in the 'Electricity' sector.

Table 7 - Ex-post implicit tax rates: breakdown by sector of activity *(percentages)*

Sector	Companies	2001 regime	2004 regime	Diff. (percentage points)
Mining and quarrying	0.9	13.69	14.50	0.81
Manufacturing industry	43.4	19.36	19.37	0.01
Food products	*4.3*	*10.13*	*10.76*	*0.64*
Textile products	*4.6*	*19.80*	*20.15*	*0.35*
Leather products	*1.4*	*14.73*	*14.75*	*0.02*
Products of wood	*1.0*	*14.99*	*15.45*	*0.46*
Paper products	*3.2*	*24.64*	*24.02*	*-0.61*
Petroleum products	*0.3*	*16.39*	*16.31*	*-0.07*
Chemical products	*3.2*	*18.60*	*19.57*	*0.96*
Rubber and plastic products	*1.9*	*18.50*	*18.63*	*0.13*
Other non metallic mineral products	*2.8*	*18.99*	*19.21*	*0.22*
Basic metals and fabricated metal products	*6.2*	*18.51*	*18.46*	*-0.05*
Machinery and equipment	*5.3*	*24.48*	*23.85*	*-0.63*
Electrical and optical equipment	*4.4*	*20.57*	*20.67*	*0.10*
Transport equipment	*2.0*	*20.99*	*21.35*	*0.36*
Other manufactured goods	*3.0*	*15.78*	*15.68*	*-0.10*
Electricity, energy, gas, steam and water	1.2	20.32	23.23	2.91
Construction	3.8	17.64	18.14	0.51
Wholesale and retail trade services	15.1	19.08	19.41	0.33
Hotel and restaurant services	2.0	11.44	12.09	0.65
Transport, storage and communication services	7.3	17.62	18.01	0.39
Real estate, renting and business services	17.9	18.73	18.93	0.20
Education services	1.0	17.18	17.01	-0.18
Health and social work services	2.9	14.15	13.55	-0.60
Other social and personal services	4.3	10.38	10.77	0.39
Total	**100.0**	**18.01**	**18.27**	**0.26**

Source: Authors' estimates.

Figure 3 - Enterprise performance and simulated ex-post tax rates changes by business sector - *(Years: — 1996 ♦ 1998 ● 2000)*



Source: Authors' estimates.

The effects of the reform are not homogeneous across sectors, both in their magnitude and in their sign. Firms in 'Education' and 'Health and social services' exhibit a reduction in the implicit tax rate, while those in the remaining sectors show ex-post tax rate increases.

For 'Manufacturing' industry as a whole, we see that the new regime leaves the implicit tax rate basically unchanged. We then note that some sub-sectors actually benefit from the new system (e.g. 'Machinery and equipment' and 'Paper products'), while others bear higher tax rates (for instance, 'Chemical products'). The largest tax rate decrease is recorded in 'Machinery and equipment' (0.63 points). Turning to the magnitude of the tax rate increases for the main sector classification, the largest rises occur in 'Electricity' (2.9 points) and 'Mining' (0.8 points). This finding is somewhat expected, as companies of these sectors record the lowest effective statutory rates in the base line, because of the dual-rate system.[26]

Lastly, companies in 'Health and social work' enjoy a substantial reduction (0.6 points) in the tax rate after the reform. This result too can be explained in light of the fact that companies of this sector did not benefit largely from the DIT allowance, as in the base-line they show the highest effective statutory rate.

The effects of the 2004 reform as regards firm size are shown in Table 8 and Figure 4.

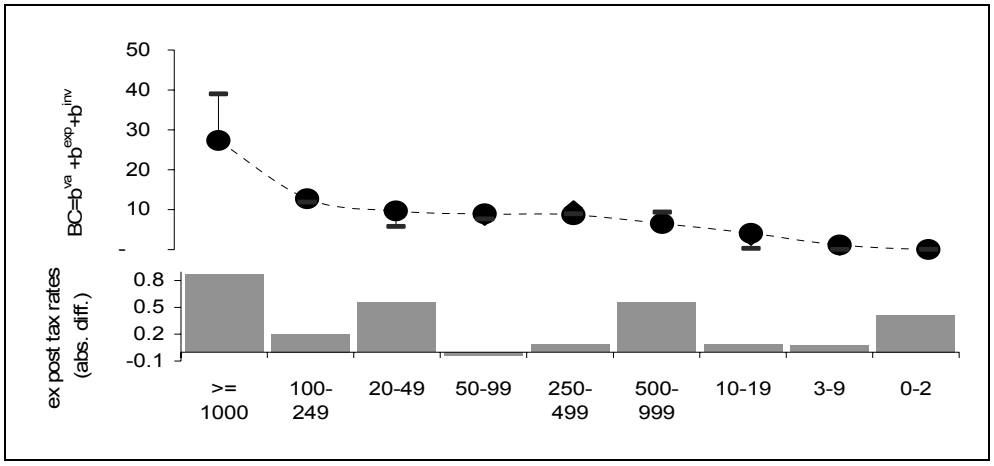
[26] It must be noted, however, that while in 'Mining' the tax base rises by almost 1 percentage point, in 'Electricity' it falls by 4 points, partially offsetting the increase in the rate of taxation for such companies under the reform.

Table 8 - Ex-post implicit tax rates: breakdown by firm size (number of workers, percentages)

Size	*Number of companies (%)*	*2001 regime*	*2004 regime*	*Difference (percentage points)*
From 1 to 2	15.1	16.64	17.06	0.42
From 3 to 9	17.4	17.75	17.83	0.08
From 10 to 19	17.2	19.67	19.75	0.09
From 20 to 49	13.8	22.03	22.58	0.56
From 50 to 99	6.3	23.22	23.18	-0.04
From 100 to 249	21.1	19.42	19.62	0.20
From 250 to 499	5.5	19.65	19.74	0.09
From 500 to 999	2.1	19.74	20.30	0.56
More than 999	1.5	22.43	23.30	0.87
Total	**100.0**	**18.01**	**18.27**	**0.26**

Source: Authors' estimates.

Figure 4 - Enterprise performance and simulated ex-post tax rates changes by classes of employed persons - *(Years: — 1996 ♦ 1998 ● 2000)*



Source: Authors' estimates.

In both scenarios the implicit tax rates vary across firm size,[27] ranging from almost 17% (firms with fewer than 3 workers) to 23% (firms with between 50 and 99 workers). As one might have expected, Figure 4 gives evidence that large companies (with at least 1,000 workers) are the most competitive while small firms (up to 20 workers) perform less well than companies of larger size.

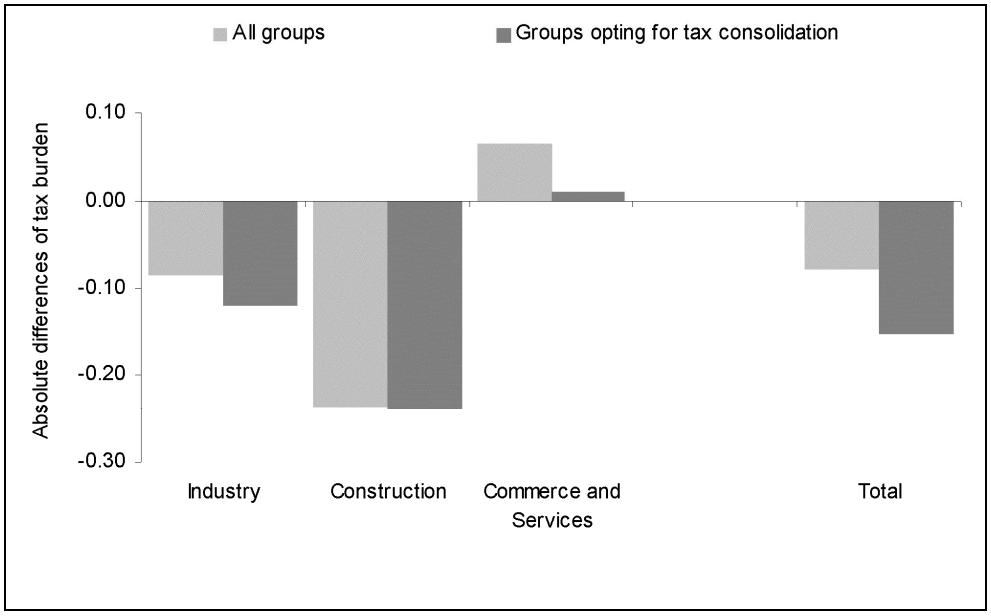
Firms benefiting from the reform are concentrated in the size class of companies with between 50 and 99 workers, while for all the remaining size classes tax burden increases after the reform. The magnitude of the increases differs across size classes. The biggest rise (0.87 points) is recorded for large firms, those with at least 1,000 workers, although very small firms, with fewer than 3 workers, also experience a significant increase in the tax

[27] For firms opting for group taxation, the number of workers refer to the aggregate number of employees of the firms electing tax consolidation.

burden (0.42 points). For very small firms, again, the result can be partially explained considering that in the base-case scenario these firms enjoy greater benefits than larger firms from the DIT system (in terms of a lower effective statutory tax rate) and therefore actually experience an increase in the statutory rate of taxation after the reform.

The results discussed so far consider the overall effects of the corporate tax reform. To analyse the reform's impact on companies belonging to groups in greater depth,[28] Figure 5 depicts the absolute changes in the estimated ex-post implicit tax rates both for all firms belonging to groups and for groups opting for tax consolidation, by sector of activity.[29]

Figure 5 - Effects of the 2004 corporate tax reform for firms belonging to groups and for firms opting for tax consolidation: absolute variations of ex-post implicit tax rates by business sector *(percentage values)*



Source: Authors' estimates.

Although the overall result examined above shows an increase in the mean ex-post implicit corporate tax rate, for firms belonging to groups we obtain an opposite finding: the tax reform lowers the mean tax burden on corporate groups by 0.29 percentage points and 1.18 points on groups opting for tax consolidation. Figure 5 shows that after the reform the implicit tax rate declines for groups in 'Industry' by 2.4 points and in 'Construction' by 0.3 points, while in 'Commerce and services' it increases by about 0.6 points. Restricting

[28] In the simulations we assume that companies opt for tax consolidation when the gross group tax liability under the new regime is lower than the tax due in the base-case. Results show that out of 1,776 parent companies (groups) present in the dataset, 276 opted for tax consolidation, corresponding to a grossed-up figure of 4,273 enterprises when set in relation to the population.

[29] As implicit tax rates show in this case high variability across sectors as defined by the NACE classification, we consider the three main sector classification. The group sector of activity refers to that of the parent company.

the analysis to groups opting for tax consolidation we find that the tax rate drops by 3.08 points in 'Industry' and by 0.43 points in 'Construction', while the reduction for groups in 'Commerce and services' is rather modest, amounting to 0.07 points.

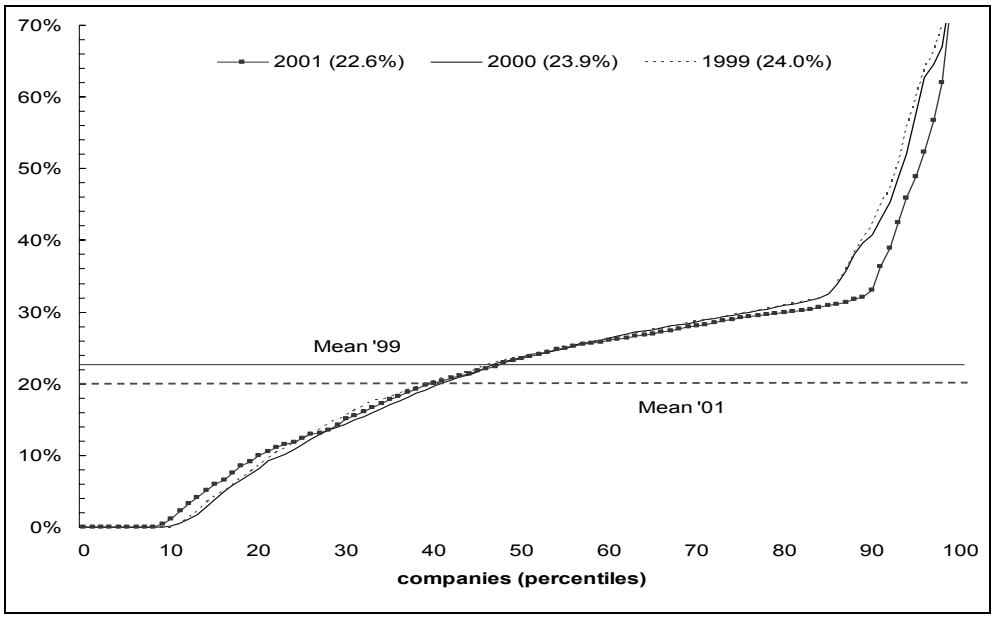
This result suggests that the reform might give companies an incentive to change their strategies and organisational behaviour in order to take advantage of the consolidated tax system.

7. Efficiency issues

In the early 1990s Italy's system of corporate taxation was pointed to as one of the main reasons for companies' over-reliance on debt financing, which policy makers and critics viewed as a potential threat to the financial stability of the corporate sector and an obstacle to the development of capital markets. The financial structure of Italian companies in the 1990s was also weak by international standards; the debt-equity ratio of non-financial firms was the highest among the main European countries (De Bondt, 1998).

The DIT system was introduced to address these issues and to encourage a gradual increase in firms' capitalisation. Figure 6 illustrates the change in the debt ratio (calculated as financial debts over net assets) for the companies of our dataset in the years 1999-2001. Firms are sorted into percentiles of the debt ratio.

Figure 6 - Debt ratio: Financial debts on net assets - Years 1999-2001



Source: Authors' computations based on ISTAT data.

The average debt ratio fell by 1.4 percentage points in the period 1999-2001, and this might suggest a significant effect of the DIT mechanism in driving the desired incentives on firms' capitalisation.

Plainly, an important question raised by the 2004 reform is how far the abolition of the

DIT allowance will alter the bias towards debt financing. To discuss this aspect we perform a sensitivity analysis by considering two alternative behavioural scenarios in terms of firms' financing choices. Specifically, we assume that companies increase their assets by 10% of the total value recorded in the company account through debt (scenario *A*) or through equity capital (scenario *B*).[30] We then run the microsimulation model to estimate the implicit tax rates in each scenario under the pre-reform regime, with the DIT system, and after the 2004 reform. This exercise yields indications on the incentive provided by each regime in using debt finance as opposed to internal sources, in the year 2000, as well as to explore the variation of the implicit tax rates with respect to the static case if one assumes different financing behaviour on the part of the company.

Table 9 - Sensitivity analysis: variation of ex-post implicit tax rates in alternative scenarios in terms of company financial choices, *(percentage values and variation in percentage points)*

Scenarios	Ex-post implicit tax rates		Debt ratio
	2001 regime	2004 regime	
Static case	18.01	18.27	23.8
Scenario A - debt funding	-1.19	-1.24	+7.8
Scenario B - equity funding	-0.30	0.00	-2.2

Source: Authors' estimates.

The results by sector of activity and firm size are presented in Table 10. The table also shows the implicit corporate tax rates estimated in the static case and discussed in the previous section.

In both scenarios the corporate tax burden increases after the reform. Roughly speaking, if we move away from the static case and assume that a company changes its investments by an amount equivalent to 10% of its net assets, under the new regime the implicit tax rate goes up for both equity and debt-financed investments (by 0.56 and 0.20 points respectively). As one would have expected, the magnitude of this increase is greater in the case of equity financing. This result can be explained if we consider that while both regimes offer a tax subsidy to debt, thereby lowering the tax burden, in the case of equity finance the 2001 system reduces the average implicit tax rate because of the DIT allowance, whereas the 2004 system leaves the rate basically unchanged.[31]

[30] The 10% variation in the company assets used in the simulations is in line with the one recorded for firms of the dataset in the period 1999-2000. In the simulations, we assume that the new investments do not alter the structure of the income statement (revenue, production costs), as any other assumption would have strong biases. Of course, debt financing changes interest expense and in this way affects profits, whereas equity financing leaves company profits unchanged.

[31] In the 2004 regime the only direct effect is given by changes in the debt-to-equity ratio when defining the amount of deductible interest expense under the thin capitalisation rules. This effect is very modest, as is confirmed by the distribution of the implicit tax rates in the static case and in scenario *B* under the reformed regime.

Table 10 - Sensitivity analysis: variation of ex-post implicit tax rates in alternative scenarios in terms of company financial choices

Type of firm and performance	Ex-post implicit tax rates								
	Static case			Scenario A Debt funding			Scenario B Equity funding		
	2001 regime	2004 regime	Diff.	2001 regime	2004 regime	Diff.	2001 regime	2004 regime	Diff.
Mining and quarrying	13.69	14.50	0.81	12.97	13.64	0.68	13.37	14.50	1.13
Manufacturing	19.36	19.37	0.01	18.14	18.20	0.06	19.09	19.33	0.24
Electricity	20.32	23.23	2.91	18.97	21.80	2.83	20.01	23.44	3.43
Construction	17.64	18.14	0.51	16.01	16.32	0.31	17.26	18.12	0.86
Wholesale and retail trade	19.08	19.41	0.33	18.00	18.39	0.39	18.76	19.50	0.74
Hotels Restaurants	11.44	12.09	0.65	10.79	11.44	0.64	11.19	12.09	0.90
Transport- comm..	17.62	18.01	0.39	16.28	16.91	0.62	17.30	17.98	0.68
Real estate. bus. services	18.73	18.93	0.20	17.48	17.45	-0.03	18.44	18.92	0.48
Education	17.18	17.01	-0.18	16.67	16.56	-0.10	16.63	16.85	0.22
Health	14.15	13.55	-0.60	13.53	13.00	-0.52	14.04	13.60	-0.44
Other services	10.38	10.77	0.39	9.82	10.24	0.41	10.12	10.72	0.60
from 1 to 2	16.64	17.06	0.42	15.42	15.71	0.29	16.35	17.05	0.71
from 3 to 9	17.75	17.83	0.08	16.69	16.72	0.02	17.47	17.90	0.43
from 10 to 19	19.67	19.75	0.09	18.38	18.68	0.30	19.32	19.72	0.40
from 20 to 49	22.03	22.58	0.56	20.64	21.12	0.48	21.60	22.41	0.81
from 50 to 99	23.22	23.18	-0.04	22.05	21.96	-0.09	23.14	23.22	0.08
from 100 to 249	19.42	19.62	0.20	18.06	18.35	0.29	19.30	19.59	0.30
from 250 to 499	19.65	19.74	0.09	18.20	18.28	0.08	19.62	19.77	0.15
from 500 to 999	19.74	20.30	0.56	18.85	19.32	0.47	19.67	20.30	0.63
more than 999	22.43	23.30	0.87	21.07	21.98	0.91	22.33	23.30	0.96
1st productivity quintile	11.86	12.17	0.31	10.99	11.12	0.13	11.63	12.17	0.55
2nd productivity quintile	15.60	15.79	0.19	14.92	15.08	0.16	15.31	15.88	0.57
3rd productivity quintile	20.89	21.15	0.26	18.95	19.34	0.39	20.46	21.11	0.65
4th productivity quintile	20.33	20.74	0.41	19.32	19.66	0.34	20.01	20.76	0.75
5th productivity quintile	21.39	21.50	0.11	19.97	19.97	0.00	21.14	21.43	0.29
1st investment quintile	15.90	16.47	0.57	14.73	15.06	0.34	15.60	16.47	0.87
2nd investment quintile	21.67	21.59	-0.08	21.00	21.06	0.06	21.58	21.94	0.36
3rd investment quintile	19.57	19.71	0.14	18.54	18.71	0.17	19.17	19.65	0.49
4th investment quintile	18.59	18.47	-0.13	17.34	17.30	-0.05	18.33	18.47	0.14
5th investment quintile	18.41	18.72	0.31	17.00	17.31	0.30	18.10	18.68	0.59
Non exporting firms	17.44	17.67	0.23	16.24	16.42	0.18	17.14	17.68	0.53
Exporting firms	20.69	21.09	0.40	19.59	19.90	0.31	20.37	21.06	0.69
Total	**18.01**	**18.27**	**0.26**	**16.83**	**17.03**	**0.20**	**17.71**	**18.27**	**0.56**

Source: Authors' estimates.

Although there is evidence (Panteghini et al. 2001) that the introduction of the DIT system significantly reduced the tax advantage in favour of debt in the previous regime, the distortion is still considerable in the tax system in force in 2001. The simulation exercise for the 2001 regime shows that when a company finances its investments out of equity capital, the implicit corporate tax rate decreases by 0.30 points,[32] while in the case of debt-financed investments the reduction is almost fourfold, amounting to 1.19 points. This result can be traced to the technical features of the DIT mechanism (and especially to the fact that capital increases are taxed although at a preferential rate rather than being tax exempt), as well as to the fact that the statutory corporate tax rate was still very high, thus making the tax subsidy in favour of debt substantial. Regarding this point, it must be emphasized that the system envisaged with the 1997 reform was still underway when the DIT allowance was frozen in 2001. Indeed, the provisions enacted in the period 2000-2001 revealed the policy maker's intention to extend the initial incremental system to a final one in which the allowance would be computed on the company entire capital stock. In the short term, this aim, as well as that of further reductions in the statutory corporate tax rate, had to be treated with great care, if Italy was to meet its public finance obligations within the European Monetary Union process. In other words, the objective of rapidly increasing the neutrality of the tax system with respect to firms' financial policy had to be sacrificed owing to the tight constraints imposed by the budget.

One clear conclusion that we draw from the analysis is that the 2004 reform widens the distortion in favour of debt, given that the quantitative difference of the tax subsidy between debt and equity-financed investments increases after the reform. If we then compare the magnitude of the tax subsidy to debt offered by the two systems, we find that the tax burden falls by 1.24 points after the reform, compared with 1.19 points in the pre-reform system. As a general result we might expect the reduction to be greater under the 2001 system than in the new one, given the higher statutory tax rate of the former and the smaller tax base changes in the 2004 regime as a consequence of the rules against thin capitalisation. However, it must be emphasized that the simulated tax base changes interact in a very complex way with the various elements of the tax system, for instance with the DIT scheme or with tax consolidation (in the way companies offset their tax base between members of the group) in the reformed regime. This might explain the larger decrease in the tax burden under the 2004 regime in the case of debt-financed investments.

Moreover, in Table 10 we also report the company implicit tax rates according to the firm's performance evaluated in terms of labour productivity (value added on employment), investments and export, again in the static scenario as well as in scenarios *A* and *B*. Looking at the results for the static case, we see that the tax burden increases as we move from the first to the fifth quintile of productivity; the lowest increase is recorded for firms of the fifth quintile ('best performers'). If we then look at the investment intensity, we find that the tax burden decreases for the second and the fourth quintiles. Lastly, the new tax regime seems to penalize exporting firms.

[32] As the purpose of the sensitivity exercise is to analyse the likely impact of the 2004 reform on financing choices' neutrality, we do not consider the effects of IRAP, which remains unchanged in the pre and post-reform scenarios. However, it must be noted that interest expense is included in the IRAP tax base and this reduces the overall tax advantage of debt with respect to equity-financed investments.

8. Conclusions

At the beginning of 2004 Italy undertook in a comprehensive reform of the corporate tax system with the aim of simplifying the tax treatment of firms and reducing their tax burden. In the present context of European Monetary Union, this is seen as an important means of stimulating the competitiveness of domestic firms and attracting inward investment.

In this paper we have assessed the effects of the 2004 reform on firms' tax burden by comparing the new system with the pre-existing one. For this purpose we have built a microsimulation model reproducing in detail the corporate income tax system. The model is based on an integrated dataset combining survey data on firms and company accounts for the year 2000, both collected by the Italy's National Institute for Statistics (ISTAT). The data do not cover firms of the 'Agriculture, forestry and fishing' sector, the Public sector and financial companies, which are therefore excluded from the analysis.

In the empirical analysis we have considered two policy scenarios. The base-line scenario replicates the structure of the corporate tax system in place in July 2001, when a dual-rate scheme (the so-called Dual Income Tax) offering a reduced rate (19% rather than 36%) on the portion of profits deemed to be derived from capital increases was present, just before some changes were made to this system to reduce its effects. In fact, the new corporate tax system moves back to a single-rate (33%), changes the definition of the tax base by exempting corporate dividends and symmetrically eliminating dividend tax relief, exempting capital gains on long-term assets owned for at least one year, and limiting the tax deductibility of interest expense under thin capitalisation rules. The reform also introduces an optional consolidated group tax regime that can be extended to foreign subsidiaries.

To analyse the impact of the reform we have estimated ex-post implicit tax rates computed as ratios of the simulated tax liabilities to the operating surplus. The results show an increase of 0.26 percentage points in the mean tax burden, although the effects of the reform, both in the sign and in the magnitude of the implicit tax rate variations, are not homogeneous across sectors. Focussing on corporate groups, we find evidence that the new system favours firms belonging to groups and opting for tax consolidation, whose average ex-post tax rate decreases by almost 1.2 points.

One important issue regards the impact of the abolition of the DIT system on neutrality with respect to company funding choices. To investigate this aspect, we perform a sensitivity analysis in which we assume two hypothetical scenarios in terms of firms' financing choices (debt, internal sources) and compute implicit tax rates in the pre and post-reform regime. Specifically, we assume that companies finance new investments amounting to 10% of the total value recorded in the company account either through debt or through equity capital. The results show that the 2004 reform annuls the encouragement given to equity capital funding by the DIT allowance and therefore widens the distortion in favour of debt-funding.

The analysis carried out in this paper is static in that it only considers the first-round impact of the tax changes on firms. Obviously firm's behaviour is also endogenous to the tax changes and in this sense the long-term impact of the new regime will also depend on how firms will react to the new system. However, on the basis of the results obtained so far, a few aspects can be pointed out. Given the increasing international competitive pressure to reduce the tax burden on firms, in the present context of European Monetary Union reducing the company tax burden is seen as a key factor in driving firms' competitiveness,

and at the same time to deter firms from locating abroad. The new system will not help the competitive position of domestic companies.

Companies belonging to groups seem to be favoured by the new regime. This suggests that the new system might provide a strong incentive to companies to change their strategic behaviour so as to take advantage of the consolidated tax system.

Finally, the abolition of the DIT system 2004 reform may actually reverse the impact that the DIT mechanism had in deterring firms from over-reliance on debt as a source of finance and, in general, in stimulating firms to redress their undercapitalisation, which still represent two specific weaknesses of the Italy's corporate sector.

References

Abbate C. (1997), "Completeness of Information and Imputation from Donor with Minimum Mixed Distance", *Quaderni di Ricerca ISTAT*, n. 4.

Bardazzi, R., Parisi, V., Pazienza, M.G. (2004), "Modelling direct and indirect taxes on firms: a policy simulation", *Austrian Journal of Statistics*, Volume 33 2004, Number 1&2.

De Bondt, G.J. (1998), "Financial Structure: Theories and stylized facts for six EU Countries", *De Economist*, n. 2, 146.

Bordignon, M., Giannini, S., Panteghini, P. (2001), "Reforming Business Taxation: Lessons from Italy?", in *International Tax and Public Finance*, vol. 8, n. 2.

Cerroni F., Morganti E. (2003), "La metodologia e il potenziale informativo dell'archivio sui gruppi di impresa: primi risultati", *Contributi Istat*, 2003

Devereux M., Griffith R. (1998), "The Taxation of Discrete Investment Choices", *The Institute for Fiscal Studies, Working Paper series* N. W98/16, London.

Devereux M., Sørensen P. (2005), "The Corporate Income Tax: International Trends and Options for Fundamental Reform", mimeo.

European Commission (2001), "Company Taxation in the Internal Market". COM(2001)5822001.

Garofalo G.,Morganti E. (2000), "Relazione finale Gruppo di lavoro per la Progettazione di un archivio statistico sui gruppi d'impresa", mimeo.

Giannini, S. (2003), "La nuova tassazione dei redditi di impresa: verso un sistema più efficiente e competitivo?", mimeo

Institute for Fiscal Studies (1991), Equity for Companies: A Corporation Tax for the 1990s", London

ISTAT (2002), "CONCORD (Generalized Data Editing Software) CONtrollo e CORrezione dei Dati", mimeo

Little R. J. A, Rubin D. B. (1987), *Statistical Analysis with Missing Data*, Wiley & Sons, New York.

Keen, M. (2002), "The German Tax Reform of 2000", *International Tax and Public Finance*, Vol. 9 n.5.

Klemm, A. (2007), "Allowances for Corporate Equity in Practice", *CESifo Economic Studies*, June 12.

King M., Fullerton D. (1984), *The taxation of income from capital*, University of Chicago Press.

Maurizi G., Monacelli, D. (2003), "Corporate Tax Reform in Italy in the late 1990s and beyond", paper presented at the *Conference Public Finance and Financial Markets, 59th International Institute of Public Finance Congress*, Prague, August 2003, mimeo.

Messere K., de Kam F., Heady C. (2003), *Tax Policy. Theory and Practice in OECD Countries*, Oxford University Press.

Milanovic, B. (2002) "True world income distribution, 1988 and 1993: First calculation based on household surveys alone", *The Economic Journal*, January, 51-99.

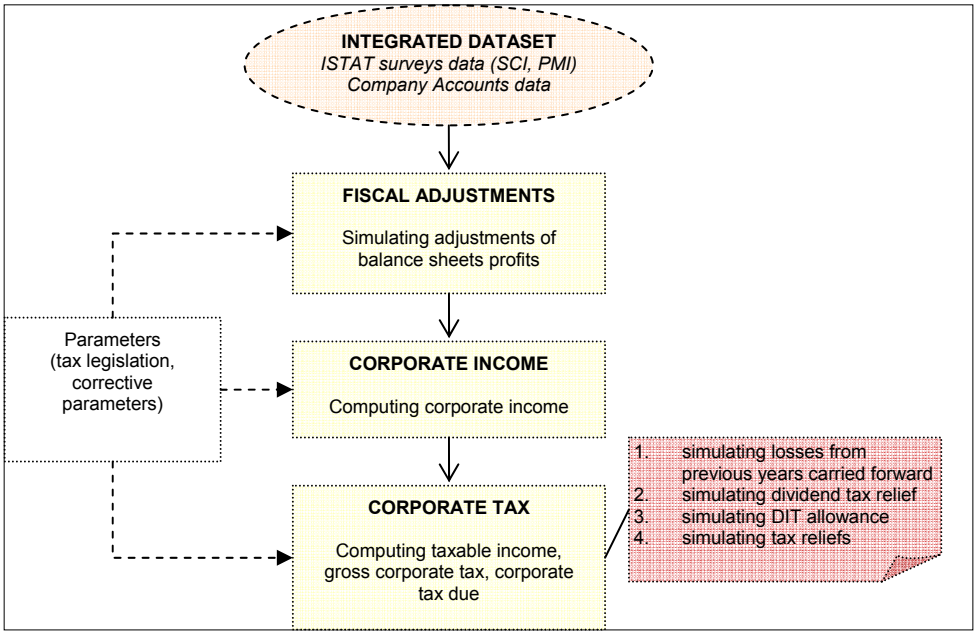
Ministero dell'Economia e delle Finanze (2003), "Decreto legislativo recante riforma dell'imposizione sul reddito delle società in attuazione dell'art. 4, comma 1, lettere da a) ad o) delle legge 7 aprile 2003, n. 80, Relazione Tecnica", mimeo.

Oropallo F. (2005) "Enterprise Microsimulation Models and Data Challenges: Preliminary Results from the Diecofis Project" – *Contributi Istat 09/2005* – pagg. 159-188 in "*L'integrazione di dati di fonti diverse: tecniche e applicazioni del record linkage e metodi di stima basati sull'uso congiunto di fonti statistiche e ammnistrative*" by P.D. Falorsi, A. Pallara e A. Russo, Franco Angeli editor.

Appendix. The microsimulation model

Figure 1.A shows the basic structure of the Corporate Tax Model (CTM). This is part of an integrated model that is currently also simulating social insurance contributions paid by enterprises, IRAP, and excises from 1998 to 2000.[33]

Figure 1.A - The structure of the microsimulation model



In Italy the (gross) corporate tax is a proportional tax applying the statutory rate (or the dual-rate schedule provided up to 2004 by the DIT allowance) on corporate taxable income. The CTM is built following a modular structure and the order in which these sub-modules are implemented obviously reflects the corporate income tax rules described below. As shown in Figure A.1, the main building blocks of the CTM are the routines Fiscal Adjustments, Corporate Income, Corporate Tax, which run sequentially

The first two modules compute corporate income for tax purposes. In Italy, as in many other countries, corporate income is obtained from total business profits (losses) shown in company accounts, adjusted according to specific tax rules. These adjustments reflect the difference between the conventional accounting rules and business accounting for tax purposes. Information available in the dataset, and more generally in company accounts, is not detailed enough to simulate tax adjustments of reported profits. The model reproduces tax adjustments of write-down to receivables, amortisation of tangible/intangible assets, and

[33] The IRAP, social insurance contributions and excises modules were built at the University of Florence within the DIECOFIS project. In this case the model runs on both incorporated and unincorporated enterprises. For a description of the methodology used in constructing the integrated model and the possible interactions of the single modules, see Bardazzi, Parisi and Pazienza (2004).

certain expenses, while fiscal adjustments that cannot be modelled on the basis of the available data are 'imputed' using parameters computed from the corporate tax returns micro data collected by ISTAT for a sample of firms.[34]

Once corporate income for tax purposes has been computed, taxable income is obtained by adding the dividend tax credit (given the imputation system subsequently abolished with the 2004 reform) to corporate income and by deducting losses from previous periods that can be brought forward up to five years.[35] The gross tax is computed by applying the prevailing tax rates, and the corporate tax due, or the tax actually paid by the company, is obtained by subtracting the dividend tax credit and the main tax reliefs (specifically for innovative investments, for research expenses, for job creation, the tax relief for small enterprises of the 'Commerce and tourism' sector) from the gross tax. Tax reliefs that cannot be simulated because of lack of information in the data, which are however of modest importance, are again imputed on the basis of corrective parameters computed form the fiscal microdata.

The final output of the module contains the main variables generated within the corporate tax module, i.e. taxable income, allowable DIT income, tax reliefs, gross tax, tax due. At intermediate levels, the model also generates variables reflecting eligible amounts of specific allowances that companies can bring forward to subsequent years, whenever companies do not benefit for the full amount. This is the case of the tax loss for the year, income eligible for the reduced rate under the DIT system, and tax reliefs.

In order to get a precise picture of the performance of the model in reproducing the corporate tax system, model outputs are validated against tax returns micro data. Table 1.A displays the (un-weighted) mean amounts of taxable income, gross corporate tax and corporate tax due estimated by the model and the amounts calculated from the tax returns, along with the percentage differences.

Table 1.A - Comparison between model estimates and tax returns data; mean amounts (euros) **and percentage differences - Year 2000**

	Model output	Tax returns data	% differences
Taxable income	1,174,286	1,121,892	4.46
Gross corporate tax	425,294	404,906	4.79
Corporate tax due	383,671	368,913	3.85

Source: Authors' estimates and computations from the 2000 corporate tax returns micro data

As can be seen, the model overestimates the corporate tax due by only 3.9% and this shows that the microsimulation model's fit is very good.

[34] This sample includes about 5279 corporations and is representative of the population covered by the SCI and PMI surveys. The parameters reflect the incidence of the specific adjustments/provisions on some variables (usually reported profits). To improve the accuracy of these corrections, coefficients are computed on a sectoral and dimensional basis.

[35] Given that PMI survey data are collected for a sample of firms, losses from previous years carried forward can be simulated only for firms covered by the SCI survey, which is exhaustive as discussed in section 4. For companies of the PMI survey previous losses are imputed using parameters computed from the tax returns microdata.

Piccole e medie imprese: le innovazioni nei metodi di calcolo dei principali aggregati economici di Contabilità Nazionale[1]

Alessandro Faramondi[2], Claudio Pascarella[3], Augusto Puggioni[4]

Sommario

Nel 2005 i Conti Economici Nazionali annuali sono stati oggetto di una revisione generale per gli anni 1970-2004, sia per la disponibilità di nuove fonti, che per l'introduzione di nuove metodologie. Tra le principali innovazioni è stata effettuata un'approfondita revisione delle metodologie di stima, con l'introduzione di uno stimatore indiretto, e del metodo di rettifica del valore aggiunto, dei ricavi e dei costi intermedi, dichiarati dalle piccole e medie imprese. Nel paper si dà ampio spazio sia alla descrizione dei metodi precedentemente applicati, sia alle innovazioni introdotte, effettuando un confronto in termini di efficienza e adeguatezza.

Abstract

A major National Accounts general revision was completed by 2005 with a complete application of SEC95 and implementation of NACE rev.1.1 classification of economic activities. In particular, the last revision has needed more detailed estimates – NACE rev.2 four digits, five size classes and legal form, than published "Small and Medium-size Enterprises Survey" figures. So a new estimator, based on an indirect approach, has been defined. At the same time the method to estimate the underreporting on value added was revised.In the paper both the old estimator, based on a direct approach, and the new estimator are described and an empirical analysis, based on Monte-Carlo simulation, is presented in order to evaluate their efficiency.

Parole chiave: contabilità nazionale, rettifica della dichiarazione mendace dell'impresa, stimatori indiretti

1. Introduzione

Nel corso del 2005 i Conti Economici Nazionali annuali sono stati oggetto di una revisione generale per gli anni 1970-2004. Tale processo è stato avviato per molteplici

[1] Sebbene il lavoro sia frutto di tutti gli autori, sono da attribuire: il paragrafo 3.1.1, 3.2, il capitolo 4 e 5 a Alessandro Faramondi; il capitolo 2, il paragrafo 3.1, 3.1.2, 3.1.3, 3.1.4, 3.1.5 a Augusto Puggioni; il capitolo 1 e 6 a Claudio Pascarella.

[2] Ricercatore (Istat), e-mail:faramond@istat.it

[3] Dirigente di ricerca (Istat), e-mail: pascarel@istat.it

[4] Primo ricercatore (Istat), e-mail: puggioni@istat.it

necessità delle quali le principali erano: introdurre alcune modifiche del sistema contabile imposte da nuovi regolamenti europei intervenuti a completare o modificare il SEC95 (Sistema Europeo dei Conti[5]), tener conto delle raccomandazioni fatte dall'Eurostat nell'ambito del Comitato Gross National Income (GNI)[6], adottare la nuova classificazione delle attività economiche NACE-Rev.1.1[7], implementare le risultanze dei censimenti del 2001 e della nuova Indagine sulle Forze di Lavoro, adottare un sistema di bilanciamento basato sulle tavole *supply and use* anziché sulla *input-output*[8]. In questo contesto è stata effettuata un'approfondita revisione sia delle metodologie di stima che del metodo di rivalutazione del valore aggiunto dichiarato dalle imprese, al fine di meglio perseguire l'esaustività delle stime degli aggregati macroeconomici, secondo le esigenze di completezza imposte dagli standard europei[9].

In particolare è stata rilevante la revisione che ha interessato le stime dei valori economici per addetto delle imprese, che congiuntamente alla stima delle unità di lavoro (ULA), rappresentano le variabili principali del modello di stima degli aggregati economici interni dal lato dell'offerta per quanto riguarda il settore *market* dell'economia, prima del bilanciamento con gli aggregati della domanda.

Il riporto all'universo attraverso le ULA è uno degli strumenti che garantisce l'esaustività delle stime degli aggregati economici (produzione e valore aggiunto) in quanto le ULA, come è noto, comprendono una valutazione dell'input di lavoro non regolare. Gli altri due strumenti rilevanti sono: la correzione del valore aggiunto e della produzione per addetto per ovviare alla sottodichiarazione del fatturato da parte delle imprese interessate a dissimulare gli introiti per fini fiscali; la riconciliazione delle stime degli aggregati dell'offerta con quelle degli aggregati della domanda (bilanciamento delle risorse e degli impieghi di beni e servizi), meno affetti dal fenomeno della sottodichiarazione e, viceversa, più soggetti alla sovradichiarazione, per l'interesse delle imprese a far figurare costi intermedi più elevati.

---

[5] Il regolamento per la costruzione dei conti nazionali è il SEC – Sistema dei Conti Economici, nella versione SEC95. Il SEC costituisce il punto di riferimento per la costruzione delle stime effettuate dai diversi istituti nazionali di statistica, nell'ambito dell'Unione europea, garantendo una migliore comparabilità internazionale degli aggregati stimati.

[6] Il Comitato *Gross National Income* dell'Eurostat vigila sulla qualità dei dati della contabilità nazionale ed il rispetto delle regole contabili presenti nel SEC95 da parte degli Istituti Nazionali di Statistica della Comunità Europea, ai fini della contribuzione al bilancio dell'UE da parte degli Stati Membri, essendo il reddito nazionale lordo (RNL) un parametro per il dimensionamento di tale contribuzione.

[7] Dal 2002 è stata introdotta dall'Istat una nuova classificazione, l'ATECO 2002 (Istat, 2003) che costituisce una versione, adattata alla struttura dell'economia italiana, della NACE Rev.1.1. La precedente classificazione adottata dall'Istat, l'ATECO 91, era a sua volta una versione nazionale della NACE Rev.1 (ISTAT, 1991).

8 Per una trattazione completa delle motivazioni della revisione in argomento vedasi: Caricchia A. "Perché la revisione dei conti nazionali?" in www.Istat.it documenti del Convegno su "La revisione generale dei conti nazionali del 2005", Roma 21-22 giugno 2006.

9 Sulla base delle definizioni dello SNA93 e del SEC95, i conti nazionali forniscono una misura esaustiva degli aggregati economici quando coprono la produzione, il reddito primario e la spesa osservati direttamente e non direttamente attraverso le indagini statistiche e gli archivi amministrativi. Le stesse definizioni stabiliscono che l'economia non (direttamente) osservata include le seguenti principali aree: illegale, sommersa, informale. L'economia sommersa è costituita dalla produzione legale di cui la pubblica amministrazione non ha conoscenza per diverse ragioni (evasione fiscale, evasione di contributi sociali, non osservanza di regole dettate dalla legge relativamente a salario minimo, numero di ore di lavoro, sicurezza sul lavoro, eccetera, e infine mancato rispetto di norme amministrative come nel caso della mancata compilazione dei questionari statistici o di altri moduli amministrativi). In particolare è definito "sommerso economico" l'insieme delle attività produttive dissimulate ai fini dell'evasine fiscale e contributiva così da ridurre i costi di produzione. Si rimanda per una completa trattazione del fenomeno a: Calzaroni M., L'occupazione come strumento per la stima esaustiva del PIL e la misura del sommerso, Atti del seminario "La nuova contabilità nazionale" ISTAT 12-13 gennaio 2000.

Nel presente lavoro, l'interesse è rivolto alle innovazioni introdotte nella stima degli aggregati economici delle piccole e medie imprese (fino a 99 addetti). Le fonti su cui si basa la contabilità nazionale presentano infatti le problematicità delle stime campionarie ed, inoltre, è in relazione alle imprese di tale fascia dimensionale che viene sostanzialmente operata la correzione per sottodichiarazione.

E' opportuno sottolineare che, per le caratteristiche strutturali dell'economia italiana, le piccole e medie imprese rivestono un ruolo rilevante, sia in termini di numerosità che di contributo al PIL. Infatti, il numero di imprese fino a 99 addetti è circa il 99,8% del totale delle imprese dell'industria e dei servizi e il 57% in termini di fatturato (in base all'Archivio Statistico delle Imprese Attive (ASIA), anno 2000, del quale di dirà più avanti).

La *Rilevazione sulle Piccole e Medie imprese e sull'esercizio di arti e professioni* (indagine di tipo campionario sulle imprese fino a 99 addetti) costituisce la principale fonte statistica per detta fascia dimensionale, insieme all'archivio ASIA e all'archivio che contiene i bilanci civilistici delle società di capitale (d'ora in poi BILANCI). Nel capitolo che segue (capitolo 2), sono descritte le principali caratteristiche delle fonti considerate.

Nel capitolo 3, viene effettuata un'ampia descrizione delle problematiche e delle soluzioni adottate in merito al problema della dichiarazione mendace da parte delle imprese, che comporta una sottodichiarazione del valore aggiunto. Tale aspetto, come evidenziato nell'inventario della contabilità nazionale italiana, rappresenta uno dei principali fattori del sommerso economico[10].

Fra i numerosi obiettivi dell'attività di revisione generale dei conti nazionali è stato posto quello della verifica della soluzione metodologica per la correzione della sottodichiarazione del valore aggiunto, adottata dall'Istat a partire dalla grande revisione del 1987 con la quale l'Istituto iniziò ad affrontare sistematicamente i problemi di stima dell'economia sommersa[11]. In vista della revisione generale del 2005, è parso opportuno sviluppare una riflessione sul metodo in uso, per verificare se, a circa 17 anni dalla sua introduzione, fosse ancora soddisfacente per la soluzione dei problemi per i quali fu inventato, ciò anche alla luce degli approfondimenti suggeriti dall'Eurostat, nell'ambito del Comitato GNI[12]. Su questo argomento è stato costituito un gruppo di lavoro fra Istat ed Ufficio Studi dell'Agenzia delle Entrate e il capitolo 3 del presente lavoro, riporta sostanzialmente le problematiche affrontate nell'ambito di tale gruppo, nonché la descrizione della nuova metodologia adottata nella revisione del 2005.

Nel capitolo 4 è descritta la metodologia di stima degli aggregati economici delle piccole e medie imprese, confrontando il nuovo metodo con il metodo adottato fino all'ultima revisione.

Nel capitolo 5 è presentata la stima dell'errore, valutato a livello di branca di attività economica. La disponibilità di tali stime rappresenta un elemento di estrema utilità, in quanto consente di tenere conto della precisione delle stime degli aggregati dal lato dell'offerta di beni e servizi nella fase di bilanciamento con gli aggregati della domanda.

Le conclusioni e le prospettive future di ricerca sono presentate nel capitolo 6.

---

[10] Cfr Istat 2004

[11] Trattasi del metodo proposto da Alfred Franz (1985), che fu applicato dall'ISTAT a partire dal 1987, secondo le modalità descritte più avanti nel paragrafo 3.1.2. Dopo il 1987 la contabilità nazionale italiana ha visto un'altra importante revisione nel 1999 in occasione dell'adozione del SEC95, ma in quell'occasione il metodo di correzione della sottodichiarazione non subì modifiche.

[12] Gross National Income (GNI) Assessment – Italy, Eurostat, 2006

2. Descrizione e qualità delle fonti utilizzate

La fonte principale per la stima degli aggregati economici relativi alle imprese fino a 99 addetti è l'indagine sulle Piccole e Medie Imprese (PMI), indagine di tipo strutturale condotta annualmente dall'Istat su base campionaria con universo di riferimento dato dall'archivio ASIA. Al fine di ridurre l'entità dell'errore campionario sui domini di analisi per le stime di Contabilità Nazionale (CN), i dati dell'indagine sono utilizzati a livello micro (per impresa) e le stime sono ottenute dall'applicazione dello stimatore descritto più avanti.

In realtà, per le società di capitale si effettua anche un'integrazione con i dati di bilancio (per una descrizione della quale rimandiamo a Puggioni, Sassaroli, 2004): dopo una fase di raccordo delle definizioni delle voci e degli aggregati economici a quelle dell'indagine, e una di individuazione e rimozione dei dati anomali, la parte di universo delle imprese relativa alle società di capitale è coperta in modo virtualmente completo dalla fonte BILANCI.

A livello generale, va rimarcato come già da diverso tempo (dalla revisione generale del 1988) i dati delle indagini sulle imprese siano utilizzati a livello micro dalla CN. L'esigenza di disporre di dati micro, come prima ricordato, nasce in particolare dal fatto che i domini di stima sono differenti rispetto a quelli fissati dalle indagini, comportando la necessità di effettuare ulteriori processi di editing, oltre che a ricorrere a stimatori differenti.

I controlli sui microdati sono sia preliminari che integrati nel processo di costruzione ed analisi dell'accuratezza delle stime degli aggregati di CN, in quanto generalmente forniscono indicazioni sull'entità degli errori che incidono sui parametri stimati. È opportuno sottolineare come detti controlli siano specifici del processo di costruzione dei conti nazionali e non vadano a sostituire, ma solo a completare ed integrare, i processi di analisi propri di ogni indagine statistica.

Si consideri che motivi di tempestività possono indurre la CN ad utilizzare un set informativo composto anche da dati provvisori, sia dal punto di vista formale (es. dati di bilancio provvisori rilasciati dall'impresa e suscettibili di rettifica entro una certa data), sia per numero di unità disponibili rispetto a quelle finali (es. l'indagine delle piccole e medie imprese al tempo t-2, viene utilizzata anche se non completa).

L'integrazione della mancata risposta rappresenta, in questo caso, un modo per minimizzare la distorsione delle stime provvisorie, rispetto a quelle definitive, che si hanno con la disponibilità del set informativo completo dell'indagine.

L'utilizzo di uno specifico metodo di controllo e correzione dei dati è dovuto alla differenza che esiste tra il concetto di qualità dell'indagine e quello delle stime di CN: il processo utilizzato per massimizzare la qualità delle stime di CN a partire dai dati dell'indagine può allora essere diverso da quello dell'indagine. Ovviamente, i due aspetti tendono, anche se parzialmente, a sovrapporsi (Puggioni, 2000).

L'integrazione delle fonti da parte della CN consente di individuare eventuali altri errori nei microdati, oltre che di valutare l'ordine di grandezza di componenti di errore di natura non campionaria (coverage errors, measurement errors, non-response errors).

2.1. Indagine sulle Piccole e Medie Imprese

L'indagine sulle piccole e medie imprese rileva il conto economico delle imprese fino a 99 addetti (fino al 1997, l'universo di riferimento era quello delle imprese fino a 19 addetti). L'indagine è condotta estraendo dall'archivio ASIA un campione di imprese stratificato per attività economica (prime 4 cifre ATECO), Regione (NUTS 2) e Classe dimensionale.

La somministrazione del questionario avviene mediante invio postale (De Gregorio, Monducci, 2002, Istat, 2005).

Sono rilevate dall'indagine tutte le attività economiche, ad esclusione dell'agricoltura-zootecnia-caccia e pesca, delle attività finanziarie (eccetto le attività ausiliarie dell'intermediazione finanziaria e delle assicurazioni), della amministrazione pubblica e delle attività di organizzazioni associative e svolte da famiglie e convivenze.

È opportuno ricordare che il regolamento sulle statistiche strutturali richiede stime per Classe di attività economica (ATECO 4 digit) senza limiti di fascia dimensionale, per Gruppo di attività economica (ATECO 3 digit) e Classi di addetto e infine per Divisione di attività economica (ATECO 2 digit) e Regione.

2.2. Archivio Statistico delle Imprese Attive (ASIA)

L'archivio ASIA è nato nel 1995, a partire dall'esigenza di disporre di uno strumento che consentisse di seguire, a cadenza annuale, l'apparato produttivo del Paese, attraverso la conoscenza di tutte le unità economiche operanti sul territorio. Tale universo era normalmente noto a cadenza decennale, in occasione dei censimenti generali e con alcuni anni di ritardo rispetto alla data di riferimento, a causa dei tempi tecnici necessari all'elaborazione dei modelli di rilevazione. L'archivio ASIA è invece disponibile a t+16 mesi dalla conclusione dell'anno di riferimento.

L'archivio è stato realizzato a partire dall'Anagrafe tributaria, dal Registro delle imprese, dagli archivi dell'INPS e dell'INAIL, dall'archivio delle utenze elettriche "non domestiche" dell'ENEL. Le informazioni di carattere amministrativo sono state poi arricchite da quelle provenienti dalle rilevazioni dell'ISTAT, in modo tale da evitare di richiedere alle imprese informazioni già fornite in precedenti occasioni (Istat, 1998). Per ogni impresa, oltre a varie informazioni di natura anagrafica, sono disponibili informazioni sul numero di addetti e dipendenti, sull'attività economica e sul volume di affari.

Le unità statistiche oggetto di registrazione nell'archivio ASIA sono le imprese e le istituzioni attive in senso economico. Nel 1996 e a partire dal 2004 con cadenza annuale sono presenti nell'archivio ASIA anche le unità locali.

La disponibilità dell'archivio ASIA ha contribuito in modo rilevante al miglioramento della qualità delle indagini sulle imprese, che rappresentano, come già sottolineato, la principale fonte per le stime di contabilità nazionale, dal lato dell'offerta.

2.3. Bilanci aziendali civilistici delle società di capitali

L'uso dei dati di bilancio delle società di capitale rientra nel grande tema dell'utilizzo a fini statistici degli archivi amministrativi. L'archivio, una volta acquisito dall'Istat, è sottoposto ad una attenta analisi per un corretto uso a fini statistici. Ciò comporta l'individuazione del quadro di riferimento concettuale relativo alle informazioni oggetto di trattamento, l'individuazione dell'universo di riferimento, delle unità di rilevazione e di analisi, dei caratteri, delle classificazioni, della tempistica e delle modalità di aggiornamento e l'identificazione delle regole di conversione del dato amministrativo in informazione statistica (Vaccari, 2002).

Le società di capitale sono tenute a depositare i bilanci presso le Camere di Commercio, le quali devono renderli disponibili e consultabili all'interno della rete camerale (Infocamere). I bilanci, si sono dimostrati adeguati a rappresentare molte delle variabili contenute nei questionari Istat e in particolare si evidenzia:

- l'elevato allineamento tra informazioni desumibili da bilancio e informazioni rilevate dalle indagini; il confronto è stato effettuato sia con SCI - Sistema dei Conti delle Imprese (cfr. Dabbicco, De Gregorio - 2002) sia con PMI (cfr. Faramondi - 2005);
- il grado di precisione e controllo delle variabili registrate;
- il grado di tempestività della fonte;
- la consistente riduzione degli oneri di risposta;
- l'adeguatezza (per un set delimitato di variabili) rispetto alle definizioni Eurostat contenute nel Regolamento UE 58/97.

I soggetti obbligati al deposito del bilancio presso le Camere di Commercio sono i seguenti: società a responsabilità limitata, società per azioni, società accomandita per azioni, società cooperativa a responsabilità limitata, società cooperativa a responsabilità illimitata, piccola società cooperativa., consorzio con attività esterna, società estera avente sede secondaria in Italia, gruppo di interesse economico, società consortili per azioni o a responsabilità limitata. Sono esclusi i bilanci presentati da imprese che svolgono attività di intermediazione monetaria e finanziaria (settore J che comprende i codici ATECO a due cifre da 65 a 67) e che hanno presentato il bilancio secondo lo schema previsto per le società finanziarie.

In Italia le società di capitale ammontano a circa 554.000 unità, per un numero di addetti pari a circa 7.969.000, equivalente al 52% del totale del numero di addetti "regolari". Il peso delle società di capitali cresce inoltre all'aumentare della dimensione delle imprese. Infatti, nella classe 1-10 addetti le società di capitale incidono, in termini di unità, per l'11%, nella classe 10-19 per il 51%, mentre nella classe 20-99 addetti la percentuale di società di capitali è dell'81% (ASIA 2000).

3. Rivalutazione dei principali aggregati economici delle imprese per dichiarazione mendace dei ricavi o dei costi

3.1 Rivalutazione del valore aggiunto

In Contabilità Nazionale i dati dell'indagine sulle piccole e medie imprese, prima di essere utilizzati per la stima della produzione e del valore aggiunto, sono sottoposti ad una procedura di correzione del valore aggiunto, spesso caratterizzato da problemi di distorsione per la tendenza che le imprese hanno a sottodichiarare i ricavi o a indicare un valore dei costi superiore al reale.

Il problema è statistico per le questioni inerenti la stima dei dati, ma è anche di natura fiscale, tanto che nel 1998 l'Agenzia delle Entrate ha deciso di introdurre gli Studi di Settore con l'obiettivo di ridurre l'evasione fiscale da parte delle imprese regolari. Gli Studi di Settore sono stati istituiti, a partire dai risultati d'esercizio del 1998, dall'Agenzia delle Entrate, per valutare la capacità di produrre ricavi dalle singole attività economiche. Detti studi sono realizzati attraverso la raccolta sistematica di dati di carattere fiscale e di numerosi altri elementi "strutturali" che caratterizzano l'attività economica delle imprese (con fatturato non superiore a 10 miliardi delle vecchie Lire).

La logica che sottostà agli Studi di Settore, è quella di definire delle soglie sotto le quali viene posta in essere una procedura di accertamento da parte dell'Agenzia delle Entrate. L'algoritmo utilizzato dagli Studi di Settore si basa in buona parte su

tutta una serie di indicatori non rilevati dalle indagini dell'Istat sulle imprese.

L'esperienza maturata in questi anni nell'analisi dei dati di indagine PMI per la costruzione degli aggregati economici, le osservazioni espresse da Eurostat sul metodo di rivalutazione adottato fino alla revisione del 2000 avvenuta nel 2005, suggerite anche dal confronto con le modalità seguite da Statistik Austria - che è l'altro paese dell'UE ad applicare il metodo Franz, hanno tuttavia suggerito di sottoporre il metodo ad una nuova formulazione in occasione della nuova revisione generale.

3.1.1 Modello teorico di rivalutazione

L'ipotesi alla base del modello di rivalutazione è che il reddito netto d'impresa debba garantire agli indipendenti una remunerazione non inferiore al reddito percepito da un lavoratore dipendente che opera nella stessa attività economica con analoghe competenze ed analogo orario di lavoro (Franz, 1985). Qualora l'indipendente dovesse trovarsi nella condizione di guadagnare meno di un lavoratore dipendente con queste caratteristiche, allora l'ipotesi è che preferisca modificare il proprio status occupazionale, da indipendente a dipendente, pur di aumentare il proprio reddito da lavoro. Se, in base ai dati di conto economico dichiarati dall'impresa, si presenta una situazione incoerente con l'ipotesi formulata, allora si assume che l'indipendente sia stato reticente nel dichiarare i ricavi, oppure abbia gonfiato i costi intermedi. Le imprese che si trovano in tale condizione sono identificate come sottodichiaranti e quindi sottoposte a rivalutazione.

Di seguito si dà una formulazione analitica del modello, secondo uno sviluppo per fasi, che consente di separare i diversi momenti del processo, dalla definizione dei parametri, all'identificazione delle imprese sottodichiaranti e per finire all'eventuale rivalutazione:

Fase 1 - si calcola il reddito da lavoro per dipendente;

Fase 2 - si effettua la rettifica del pro-capite della Fase 1, per tenere conto del diverso orario di lavoro tra i lavoratori dipendenti ed indipendenti; infatti gli indipendenti lavorano mediamente un numero di ore superiore (cfr. dati dell'indagine sulle Forze di Lavoro);

Fase 3 - per ciascuna ditta si calcola il reddito netto d'impresa, come differenza tra il valore aggiunto al costo dei fattori e la somma delle seguenti voci:

- spese per il personale
- interessi passivi e spese bancarie
- ammortamenti dei capitali fissi;

Fase 4 - si calcola il reddito netto per indipendente, come rapporto tra il risultato della Fase 3 ed il numero degli indipendenti dell'impresa considerata;

Fase 5 - qualora il risultato della fase 4 fosse negativo o inferiore a quello della Fase 2 si ricalcola il reddito netto di impresa attribuendo a ciascun imprenditore il pro-capite risultante dalla Fase 2;

Fase 6 - si somma al valore aggiunto dell'impresa, la differenza tra il risultato della Fase 5 ed il risultato della Fase 3.

A partire dal modello generale sono state proposte diverse varianti, a seconda della tipologia di imprese sottoposte a rivalutazione e dei metodi per realizzare le diverse fasi. Di seguito, viene presentata la soluzione adottata dall'Istat a partire dalla revisione del 1987 ed il nuovo metodo adottato a partire dalla revisione del 2005.

3.1.2 Il metodo ISTAT adottato fino alla revisione del 2005

Il metodo definito del "reddito pro-capite da lavoro dipendente della singola impresa" è

stato applicato in CN fino alla revisione generale del 2005, alle imprese da 1 a 19 addetti. Il pro-capite della fase 1 (reddito da lavoro per dipendente), era calcolato per ogni impresa sulla base dei suoi dipendenti, considerando la media tra dirigenti, quadri, impiegati, operai e commessi.

La rettifica per tenere conto del diverso orario di lavoro dei dipendenti e degli indipendenti (fase 2), era ottenuta moltiplicando il pro-capite della fase 1 per un coefficiente orario, dato dal rapporto tra la media delle ore lavorate dagli indipendenti e dai dipendenti, in base ai dati dell'indagine sulle Forze di Lavoro per ramo di attività economica.

Qualora nell'impresa non fossero presenti lavoratori dipendenti, il reddito teorico dell'indipendente veniva confrontato con il reddito medio da lavoro dipendente rettificato, calcolato per dominio di appartenenza dell'impresa a livello di attività economica (ATECO 3 digit), ripartizione territoriale e classe di addetti (1-5, 6-9, 10-19 addetti) (cfr. Istat, 2004).

3.1.3 Aspetti critici del metodo del "reddito pro-capite da lavoro dipendente della singola impresa"

Gli aspetti analizzati[13] nel corso della revisione generale del 2005 hanno riguardato:

1. universo da sottoporre a verifica ed eventuale rivalutazione;
2. problematiche nell'applicazione del metodo connesse alla anzianità dell'impresa ;
3. analisi della veridicità delle voci di costo dichiarate;
4. analisi della correttezza dei dati riguardanti il numero degli indipendenti;
5. ipotesi di stessa retribuzione oraria per indipendenti e dipendenti.

1. L'universo considerato era quello delle imprese fino a 19 addetti, in accordo con i principi generali proposti in Franz (1985), dove si proponeva di applicare il metodo alle sole imprese di piccole e piccolissime dimensioni, dove è plausibile ritenere che gli indipendenti dedichino per intero il loro tempo lavorativo all'attività imprenditoriale. Solo se accade ciò è infatti corretto confrontare il reddito dell'indipendente con il reddito teorico del lavoratore dipendente.

A tale proposito, nelle fasi di studio e analisi effettuate per la revisione del 2005, si è visto che nella realtà economica italiana il modello di rivalutazione poteva essere applicato con successo anche alle imprese di medie dimensioni (fino a 99 addetti). Infatti si è potuto constatare che in molte imprese con più di 19 addetti gli indipendenti prestano per intero il loro tempo lavorativo all'attività dell'impresa e ciò è particolarmente vero se vengono considerate le "società non di capitali" (imprese individuali e familiari, liberi professionisti, società semplici o di fatto, società in nome collettivo, società in accomandita semplice, studio associato).

2. Dalle analisi empiriche è emerso un andamento non costante del tasso di rivalutazione del valore aggiunto, per anni di attività dell'impresa. Da un'analisi basata sui dati rilevati dall'indagine PMI nel 2000, si evince un fattore di rivalutazione decrescente. In particolare, dal quarto anno di vita si nota un netto calo

[13] Parte delle analisi e considerazioni riportate nel documento sono state presentate e discusse nell'ambito del *Gruppo di lavoro avente l'obiettivo di verificare l'esaustività del PIL, attraverso l'utilizzo di informazioni disponibili presso l'Anagrafe tributaria*, presieduto dal dott. Claudio Pascarella.

del fattore di rivalutazione. Dato che i primi tre anni di vita dell'impresa sono in genere caratterizzati da una fase di avviamento, nella quale sono plausibili livelli di reddito per gli indipendenti anche inferiori a quelli dei loro dipendenti, è credibile ritenere che la rivalutazione induca una sovrastima nei primi anni di vita dell'impresa (0-3 anni).

Di conseguenza, nella revisione del metodo andava senz'altro considerato l'aspetto legato agli anni di attività dell'impresa. In un'economia di mercato, infatti, le imprese sostengono costi prima di conseguire ricavi, circostanza generatrice del fabbisogno finanziario d'impresa. Durante la fase di avviamento è fisiologico attendersi che gli esborsi generati dal sostenimento dei costi possano superare anche ampiamente i ricavi conseguenti al collocamento sul mercato dei beni e servizi prodotti.

3. Il metodo agiva sul valore aggiunto e sulla produzione, mantenendo invariati i costi. In realtà, è verosimile che le imprese "occultino" parte del loro valore aggiunto agendo anche sui costi, oltre che sul fatturato. Si è ritenuto pertanto utile approfondire anche gli aspetti riguardanti la veridicità dei costi dichiarati dall'impresa, con l'obiettivo di individuare eventuali dichiarazioni mendaci e correggere i dati a partire dalla rivalutazione del valore aggiunto[14].

4. Il tempo dedicato al lavoro da parte dell'indipendente è uno dei fattori più importanti del modello di rivalutazione (paragrafo 3.1.1) In tale quadro teorico, sono presi in esame solamente gli indipendenti che dedicano interamente il proprio tempo lavorativo all'attività imprenditoriale. Vengono esclusi quegli indipendenti che intervengono nell'attività solo attraverso un apporto di capitale o per un tempo limitato. Tale scelta è coerente con la filosofia del modello, che considera la convenienza dell'indipendente a cambiare status, da indipendente a dipendente, qualora, impegnando lo stesso tempo per la medesima attività economica e con analoghe competenze ed analogo orario di lavoro, percepisse un reddito maggiore. Una tale impostazione prevedeva, per ogni impresa, di conoscere l'equivalente a tempo pieno del numero di indipendenti che lavorano nell'ambito dell'impresa.

Dalle analisi effettuate sono emersi alcuni elementi di criticità nei dati dell'indagine PMI, che portano a supporre, in alcuni casi, un'erronea indicazione del numero di addetti che lavorano a tempo pieno nell'impresa in posizione indipendente. In particolare si è osservata una relazione di concordanza tra numero di indipendenti per impresa e percentuale di imprese rivalutate. Tale relazione non avendo alcun fondamento teorico ha rappresentato un campanello di allarme, rispetto all'affidabilità delle variabili coinvolte. Quindi è stato effettuato un supplemento di analisi, approfondendo gli aspetti di natura economica.

Le imprese sono state classificate in base al numero di indipendenti ed è stata calcolata la probabilità di un risultato incoerente, cioè che il rapporto fatturato/costi sia maggiore tra le imprese sottodichiaranti. Come si può osservare dalla tavola 3.1, il punto di svolta è rappresentato dalle imprese con 4 indipendenti: la probabilità di registrare un'incoerenza, passa da 0,125 per le imprese con un numero di indipendenti compreso tra 3 e 4, a 0,625 per le imprese con un numero di indipendenti compreso tra 4 e 5 (Tav. 3.1).

[14] Nell'Inventario relativo alla precedente revisione generale (Istat, 2004. pag.68) era infatti evidenziato come il metodo di rivalutazione utilizzato, rivalutando sempre il fatturato e non intervenendo sui costi, producesse una sovrastima dei ricavi. La distorsione era corretta in fase di bilanciamento dei conti.

Le risultanze basate sulla probabilità di un risultato incoerente del rapporto fatturato/costi avvalorano l'ipotesi che la concordanza tra il numero di imprese rivalutate e il numero di indipendenti per impresa rilevato dall'indagine PMI non sia sempre corretta. Un'erronea rilevazione del numero di indipendenti a tempo pieno può comportare la falsa identificazione di un impresa quale sottodichiarante. Dalla nostra analisi dei dati dell'indagine PMI deriva quanto segue:

a. le imprese tendono a dichiarare il numero di indipendenti senza tenere conto di coloro che effettivamente lavorano a tempo pieno nell'impresa;

il dato relativo agli indipendenti è affidabile fino a 4, massimo 5 indipendenti;

b. l'informazione relativa al numero di indipendenti risulta fortemente compromessa nel caso delle società cooperative, dove il grado di sovrapposizione tra dipendenti e indipendenti è particolarmente elevato.

Tav. 3.1 - Probabilità di incoerenza del rapportofatturato/costi

Numero imprenditori	Probabilità
1	0,125
2	0,000
3	0,000
4	0,125
5	0,625
6	0,750
7	0,857
8	0,333
9	0,333
Oltre 9	0,667

5. Il confronto della retribuzione oraria degli indipendenti e dei dipendenti della stessa impresa, alle base del metodo del "reddito pro-capite da lavoro dipendente della singola impresa", presenta diverse criticità. In particolare, è poco plausibile ritenere che l'indipendente valuti la redditività della sua attività tenendo conto solo della propria realtà imprenditoriale e non dell'intero sistema.

La variante descritta nel paragrafo seguente, mettendo a confronto il reddito dell'indipendente della singola impresa con un reddito da lavoro dipendente di strato, supera la logica del rapporto 1:1 tra retribuzione oraria indipendente/dipendente, in quanto i valori ottenuti non sono più paragonabili a quelli osservati per i dipendenti della singola impresa.

3.1.4 Il nuovo metodo ISTAT adottato a partire dalla revisione del 2005

Il nuovo metodo, definito del "reddito pro capite da lavoro per strato" è stato applicato per la prima volta nella revisione generale del 2005, a partire dai dati d'indagine dell'anno 2000.

Al fine di superare le problematiche emerse e descritte nel paragrafo precedente, è stato definito un nuovo metodo di rivalutazione del valore aggiunto, con la medesima struttura teorica del precedente (vedere par. 3.1.1), ma più coerente con l'universo delle imprese sottoposte a rivalutazione e con le implicazioni statistiche che un tale metodo comporta.

Uno dei principali aspetti rivisti è stata la ridefinizione dell'universo delle imprese sottoposte a rivalutazione. Il nuovo universo è rappresentato dalle società non di capitale, indipendentemente dalla dimensione, e dalle società di capitale fino a 50 addetti (con oltre

50 addetti non sono sottoposte ad analisi in quanto non presentano lavoratori indipendenti secondo l'accezione prevista dal metodo). Sia le società non di capitale che di capitale sono comunque escluse dall'applicazione del metodo di rivalutazione se presentano un numero di indipendenti superiore a 5 (i motivi di tale scelta sono illustrati nel paragrafo 3.1.3). Le cooperative, a causa delle difficoltà di identificazione dei dipendenti e degli indipendenti, sono state escluse dall'universo delle imprese da rivalutare.

Il pro-capite della fase 2 del modello base di Franz (1985) (reddito da lavoro dipendente pro-capite rivalutato) è calcolato per strato e non per singola impresa come nel metodo utilizzato fino alla revisione del 2005. Inoltre, un ulteriore elemento di differenza con il metodo precedente è che viene considerato il valore massimo tra il reddito pro-capite della qualifica "dirigenti/quadri/impiegati" e della qualifica "operai/commessi" e non un generico valore medio relativo ai lavoratori dipendenti nel loro insieme.

Lo strato è definito dall'incrocio dell'attività economica (ATECO 3 digit), della classe di fatturato (<0.5, 0.5-5, >5 milioni di euro), della forma giuridica (società di capitale e società non di capitale), degli anni di vita dell'impresa (0-3, 4-6, 7-9, 10-19, 20 e oltre anni), e dalla ripartizione territoriale. La scelta degli strati è stata effettuata sulla base delle seguenti considerazioni:

- la scelta dei primi 3 digit per ATECO garantisce sia un sufficiente grado di efficienza delle stime sia un adeguato livello di omogeneità delle imprese dal punto di vista dell'attività economica. Al riguardo si tenga presente che i dati sono pubblicati per 101 branche di attività economica, risultato dell'aggregazione delle stime per ATECO a 3 digit;
- la classe di fatturato è stata scelta come variabile di stratificazione dato l'elevato grado di correlazione tra i livelli di fatturato e quelli di produttività; le classi scelte (<0.5, 0.5-5, >5 milioni di euro) sono coerenti con gli Studi di Settore, per tener conto anche del diverso comportamento fiscale delle imprese;
- la forma giuridica, insieme alla classe dimensionale, è discriminate del tipo di rapporto che esiste tra impresa ed imprenditore, dal punto di vista della sua partecipazione alla gestione diretta dell'impresa; infatti nel modello di rivalutazione è determinante la quantità di tempo che l'imprenditore dedica all'attività;
- la variabile "anni di vita" tiene conto del diverso stadio di sviluppo delle imprese. Infatti, le imprese in avviamento possono presentare dei costi e dei ricavi sensibilmente diversi dalle imprese che da diversi anni operano sul mercato nello stesso settore, potendo contare queste ultime su un livello di esperienza e rapporti commerciali ormai consolidati, oltre ad avere assorbito in parte o completamente i costi d' avviamento;
- infine, è stata considerata come ulteriore fattore di stratificazione una variabile geografica, per tener conto dei diversi sistemi economici che esistono sul territorio nazionale. Non potendo considerare la Regione, per l'elevato grado di variabilità delle stime, è stata scelta una soluzione intermedia tra la stratificazione regionale e quella a tre ripartizioni ("Nord", "Centro" e "Mezzogiorno"), rappresentata dalla suddivisione del territorio in cinque aree ("Nord-Ovest", "Nord-Est", "Centro", "Sud", "Isole"). In tal modo è stato assicurato un adeguato livello di accuratezza delle stime dei parametri richiesti dal modello.

3.1.5 Analisi empirica

Nella tavola 3.2 sono riportati i valori, per l'anno 2000, dei coefficienti di rivalutazione del valore aggiunto a seguito dell'applicazione dei due metodi. I coefficienti sono stati calcolati sia con riferimento alla popolazione campionaria, sia con riferimento all'universo

delle imprese. Il dato campionario è stato ottenuto applicando il metodo di rivalutazione ai dati rilevati dall'indagine, mentre il dato relativo all'universo è stato ottenuto applicando i coefficienti di riporto all'universo ai dati d'indagine.

I metodi applicati sono rispettivamente il metodo del reddito pro-capite da lavoro dipendente della singola impresa (nella tavola è indicato per brevità come il metodo A) ed il metodo del reddito pro capite da lavoro per strato (nella tavola indicato come metodo B). In particolare il metodo B è stato applicato sia nella versione adottata per la revisione (numero di indipendenti inferiore a 5), sia nella versione che non tiene conto del vincolo sul numero di indipendenti.

Come si evince dai dati, il nuovo metodo (metodo B) ha portato un significativo incremento della rivalutazione del valore aggiunto. Inoltre tale incremento risulta maggiore in corrispondenza dei dati campionari, mentre si attenua leggermente a livello dell'universo.

Tav. 3.2 - Coefficiente di rivalutazione a seguito dell'applicazione dei metodi di rivalutazione del valore aggiunto (PMI 2000)

	Metodo A	Metodo B	
		Indipendenti < 5	Indipendenti qualsiasi
Campione			
Coefficiente di rivalutazione	8,9%	14,6%	14,9%
Valore aggiunto per addetto (migliaia di euro)	38,0	39,9	40,1
Dato riportato all'universo			
Coefficiente di rivalutazione	23,8%	27,2%	27,5%
Valore aggiunto per addetto (migliaia di euro)	33,4	34,3	34,4

Per quanto concerne il metodo B, le differenze tra il metodo "con meno di 5 indipendenti" e quello "esteso a tutti gli indipendenti" sono minime, sia a livello dei dati campionari che di quelli riportati all'universo.

3.2 Nuovo metodo di rivalutazione della produzione, del fatturato e dei costi

Effettuata la rivalutazione del valore aggiunto (VA), per mantenere la coerenza di tutti gli elementi contabili si pone il problema di rettificare anche gli altri aggregati economici: fatturato, valore della produzione e costi. Per quanto concerne fatturato e costi, la rivalutazione si basa sulla seguente ipotesi: l'indipendente che evade, è stato reticente nel dichiarare i ricavi, oppure ha gonfiato i costi (Pisani, 2000). A partire da questa ipotesi, vengono definiti due indicatori, rispettivamente del fatturato e dei costi intermedi:

$$(F)ij = (K^*)j / (K)ij \qquad (3.1)$$

$$(C)ij = (H)ij / (H^*)j \qquad (3.2)$$

dove

Kij = fatturato per addetto dell'impresa i-esima sottodichiarante appartenente allo strato j

K*j = fatturato per addetto medio delle imprese non sottodichiaranti appartenenti allo strato j

Hij = costi intermedi per addetto dell'impresa i-esima sottodichiarante appartenente allo strato j

H*j= costi intermedi per addetto medio delle imprese non sottodichiaranti appartenenti allo strato j

i = indice dell'impresa nell'ambito delle imprese sottodichiaranti

j = indice dello strato: ATECO 2 digit, classe di addetti, ripartizione geografica, classe di avviamento, forma giuridica, classe del valore aggiunto rivalutato.

I due indicatori mettono in relazione la situazione, in termini di fatturato e costi, dell'impresa sottodichiarante con i valori medi delle imprese non sottodichiaranti. Nel caso di sottodichiarazione del fatturato è ipotizzabile che (F)ij sia in misura rilevante maggiore di 1, mentre qualora le dichiarazioni mendaci riguardino i costi è ipotizzabile che sia (C)ij ad essere fortemente maggiore di 1.

Il modello di rettifica proposto si basa sulle seguenti assunzioni, coerenti sia con l'andamento empirico dei due indicatori osservato sulle imprese sottodichiaranti, sia con la maggiore facilità, che hanno le imprese che intendano sottodichiarare il valore aggiunto, di occultare parte dei ricavi piuttosto che "gonfiare" i costi, date le ridotte dimensioni delle imprese e la clientela composta prevalentemente da consumatori finali:

a. la probabilità che le imprese sottodichiaranti agiscano sul fatturato è circa 4 volte superiore alla probabilità che le imprese agiscano sui costi;

b. è verosimile attribuire la rivalutazione del valore aggiunto al fatturato, a meno di non avere evidenze sufficientemente significative che l'impresa abbia agito sui costi.

La traduzione operativa delle due assunzioni è:

se vale il sistema di disequazioni che segue si rettificano i costi intermedi, diminuendoli di un ammontare uguale all'entità di rivalutazione del valore aggiunto :

$(I\_cost)ij > (I\_fatt)ij$
$(I\_cost)ij > 2$

in tutti gli altri casi si rettifica il fatturato, aumentandolo di un ammontare uguale all'entità di rivalutazione del valore aggiunto.

Per quanto concerne la rettifica del valore della produzione (Prod), vale la seguente relazione:

$$riv(Prod) = riv(VA) - aggiustamento(Costi) \qquad (3.3)$$

dove

riv(VA)>= 0 è la differenza tra il valore aggiunto eventualmente rivalutato ed il valore aggiunto prima dell'eventuale rivalutazione;

aggiustamento(Costi) >=0 è l'ammontare dell'eventuale rettifica effettuata sui costi intermedi;

riv(Prod)>=0 è l'ammontare attribuito al valore della produzione nel caso di rettifica del fatturato.

La relazione 3.3 si dimostra nel seguente modo:

data la relazione contabile

$$VA = Prod. - Costi \qquad (3.4)$$

dopo la rivalutazione del valore aggiunto, essa diventa

$$(VA + riv(VA)) = (Prod + riv(Prod)) - (Costi - aggiustamento\ (Costi)), \qquad (3.5)$$

e quindi

$$VA + riv(VA) = Prod + riv(Prod) – Costi + aggiustamento\ (Costi). \quad (3.6)$$

Dopo aver semplificato (in base alla (3.4)) si ha:

$$riv(VA) = riv(Prod) + aggiustamento\ (Costi), \quad (3.7)$$

e quindi, cvd:

$$riv(Prod) = riv(VA) - aggiustamento\ (Costi) \quad (3.8)$$

Tale modello prevede la rettifica, in termini di fatturato e costi, di tutte le imprese che vengono identificate come sottodichiaranti.

Nelle tavole che seguono, relative ai dati dell'indagine PMI dell'anno 2000, sono riportati i principali indicatori dell'impatto del metodo di rettifica del fatturato e dei costi. Nella tavola 3.3 sono state considerati i dati campionari senza il riporto all'universo. La variazione percentuale del fatturato, è stata del 2,28%, mentre la variazione dei costi è stata di -0,5%. Dette variazioni sono state calcolate anche sui dati di indagine riportati all'universo (Tavola 3.4). In questo caso si ha un incremento delle variazioni percentuali assolute, infatti il fatturato aumenta del 5,23%, mentre i costi si riducono dello 0,93%. Nella tavola 3.5 è stato considerato l'impatto delle rettifiche effettuate sulle piccole e medie imprese sull'intero universo, considerando anche le imprese con oltre 99 addetti. In questo caso la variazione del fatturato (3,2), della produzione (4,4) e dei costi in valore assoluto (0,53) sono ovviamente inferiori alle variazioni registrate in riferimento all'universo delle imprese fino a 99 addetti (Tavola 3.4).

Tav. 3.3 - Impatto del metodo sui dati dell'indagine PMI interessati dalla rivalutazione

Variazione % del fatturato	2,28
Variazione % dei costi	-0,50
Variazione % della produzione	3,10
Fatturato/costi	
- NON SOTTODICHIARANTI	200,34
- SOTTODICHIARANTI	
prima della rivalutazione	191,34
dopo la rivalutazione	220,06
- SULL'INTERO CAMPIONE	
prima della rivalutazione	198,58
dopo la rivalutazione	204,12
Valore aggiunto/produzione	33,74

Tav. 3.4 - Impatto rispetto all'universo delle imprese fino a 99 addetti

Variazione % del fatturato	5,23
Variazione % dei costi	-0,93
Variazione % della produzione	7,41
FATTURATO/COSTI	
prima della rivalutazione	224,22
dopo la rivalutazione	238,17
Valore aggiunto/produzione	40,62

Tav. 3.5 - Impatto rispetto all'universo delle imprese

Variazione % dei costi	-0,53
Variazione % della produzione	4,42
FATTURATO/COSTI	
prima della rivalutazione	206,19
dopo la rivalutazione	213,95
Valore aggiunto/produzione	35,55

4. Stima dei parametri e domini di studio dell'indagine sulle piccole e medie imprese

Effettuata la rettifica del valore della produzione, del fatturato, del valore aggiunto e dei costi intermedi, si passa alla fase propriamente di stima.

Il dominio di interesse è rappresentato dalla combinazione delle modalità delle seguenti variabili:

- classe di attività economica (ATECO 4 digit),
- classe di addetti (1-2, 3-5, 6-9, 10-19, 20-99),
- forma giuridica (società non di capitale, società di capitale).

I domini considerati rappresentano un dominio di studio non pianificato - in quanto il disegno di campionamento dell'indagine sulle piccole e medie imprese prevede, come indicato nel Paragrafo 2.1, una stratificazione delle imprese che non contempla la forma giuridica. Inoltre, la classificazione per addetti adottata nel disegno di campionamento dell'indagine PMI (Tav. 4.1) è diversa da quella adottata in Contabilità Nazionale. Da ciò deriva che alcuni domini di piccole dimensioni possono non essere rappresentati nel campione, oppure possono essere rappresentati ma solo con pochissime unità campionarie, con la conseguenza che l'eventuale stima diretta o non esiste oppure presenta un livello di errore molto elevato.

Tab. 4.1 - Classi di addetti adottate nel disegno di campionamento dell'indagine sulle piccole e medie imprese

Divisioni di attività economica (ATECO 2)	Classi di addetti
Industria: 10-45	1-9, 10-19, 20-49, 50-99
Servizi: 50, 51, 52	1, 2-4, 5-9, 10-19, 20-49, 50-99
Servizi: 55, 60, 61, 62, 63, 64, 70, 71, 72, 73, 74	1-4, 5-9, 10-19, 20-49, 50-99
Servizi: 67, 80, 85, 90, 92, 93	1-9, 10-19, 20-49, 50-99

4.1 La metodologia di stima fino all'attuale fase di benchmark

Fino all'attuale fase di benchmark le stime della CN erano ottenute adottando uno stimatore diretto.

La procedura prevedeva l'utilizzo del seguente stimatore:

$$_{\text{dir}}\hat{\text{Y}}_{\text{ijk}} = \frac{N_{ijk}}{n_{ijk}} \sum_{l \in ijk} \text{Y}_l$$

Dove

i = indice di classe di attività economica (i=1, ….., 461),
j = indice di classe dimensionale di addetti (j=1, ….., 5),
k = indice della forma giuridica (k=1, 2),
l = indice delle imprese campionarie contenute nel dominio "ijk"
n= addetti campionari
N= addetti riferiti all'universo (archivio ASIA)

4.2 La nuova metodologia di stima

Nell'attuale fase di benchmark si è cercato di adottare una metodologia coerente, sia con gli obiettivi delineati nell'ambito delle attività di Contabilità Nazionale, sia con il disegno di campionamento. In particolare, trattandosi di domini non pianificati si è cercato di considerare dei metodi di stima indiretti, al fine di sfruttare al meglio l'informazione campionaria, non solo del dominio oggetto di stima ma dell'indagine nel suo complesso. A tale proposito, nell'ambito dell'attività di ricerca propedeutica alla revisione del 2005, è stato realizzato uno stimatore ad-hoc, definito Sample Error Dependent Estimator (SEDE).

Tale metodo è stato ottenuto dallo sviluppo di un altro stimatore, noto in letteratura con il nome di Sample Size Dependent (SSD) (Ghosh e Rao, 1994; Russo, 1995). Quest'ultimo stimatore si basa sulla considerazione che la precisione del diretto dipende dal numero di unità del campione presenti nel dominio d'interesse. Se tale numero è sufficientemente ampio, è ragionevole avere "fiducia" nello stimatore diretto; altrimenti, si ritiene più ragionevole introdurre un'informazione esterna, da una macro-area che contiene il dominio considerato, assumendo quindi una media ponderata tra lo stimatore relativo alle unità campione del dominio considerato (stimatore diretto), e lo stimatore delle unità campione della macro-area (stimatore sintetico). Con l'introduzione di tale stimatore si vuole raggiungere l'obiettivo di ridurre il Mean Square Error (MSE) in quelle situazioni dove la numerosità campionaria non è adeguata, anche se ciò potrebbe comportare un aumento del bias.

Inoltre, al fine di rispettare la coerenza tra domini gerarchicamente ordinati, è stata definita una procedura di stima articolata in più fasi. Nella prima fase si effettua la stima a livello di ATECO a 2 digit. Nella seconda fase si calcola la stima a livello di ATECO a 3 digit, utilizzando come informazione indiretta la stimatore sintetico a livello di ATECO a 2 digit. Nella terza fase si passa alla stima a livello di ATECO a 4 digit, considerando come informazione indiretta la stima che proviene dall'ATECO a 3 digit.

Prima di descrivere la procedura in dettaglio, si da una breve presentazione della simbologia adottata:

d = indice di divisione di attività economica (ATECO a 2 cifre),
g = indice di gruppo di attività economica (ATECO a 3 cifre),
i = indice della classe di attività economica (ATECO a 4 cifre),
j = indice di classe dimensionale di addetti (j=1, ….., 5),
k = indice della forma giuridica (k=1, 2),
n = numerosità nel campione
N = numerosità nell'universo
λ = parametro maggiore di zero, che consente di modulare il peso dello stimatore diretto e quindi indirettamente esprime il grado di fiducia nella stima ottenuta con il metodo diretto.

fase 1 - stima a livello di ATECO a 2 digit

Lo stimatore del parametro d'interesse, della d-esima divisione di attività economica, j-esima classe dimensionale di addetti e k-esima forma giuridica, è dato da:

$$_{SSD}\hat{Y}_{djk} = w_{dljk} \;\; {}_{\text{dir}}\hat{Y}_{djk} + (1 - w_{djk}) \;\; {}_{\text{S}}\hat{Y}_{djk} \quad (4.1)$$

dove:

$${}_{\text{dir}}\hat{Y}_{djk} = \frac{N_{djk}}{n_{djk}} \sum_{l \in djk} Y_l$$ è lo stimatore diretto del dominio "djk",

l = indice delle imprese campionarie contenute nel dominio "djk",

$${}_{\text{S}}\hat{Y}_{\text{djk}} = \frac{N_{dj}}{n_{dj}} \sum_{l \in dj.} Y_l$$ è lo stimatore sintetico del dominio "djk", calcolato nello stesso modo del diretto, ma per macro-area: "divisione economica e classe di addetti. Quindi, aggregando i domini rispetto alla forma giuridica.

$$w_{ijk} = \begin{cases} 1 & \text{se } \frac{n_{ijk}}{n} \geq \lambda \frac{N_{ijk}}{N} \\ \frac{n_{ijk}}{n} \Big/ \lambda \frac{N_{ijk}}{N} & \text{altrimenti} \end{cases}$$

$$\lambda = 1$$

fase 2 - stima a livello di ATECO a 3 cifre

In questa fase, si procede alla stima a livello di ATECO a 3 cifre, adottando uno stimatore simple size dependent, dove lo stimatore sintetico è dato dalla stima della fase 1. Il dominio d'interesse è lo stesso della fase 1, con la sola eccezione del gruppo di attività economica (ATECO 3) al posto della divisione di attività economica (ATECO 2). Anche in questo caso lo stimatore è dato da:

$$_{SSD}\hat{Y}_{gjk} = w_{gjk} \;\; {}_{\text{dir}}\hat{Y}_{gjk} + \; (1\text{-}w_{gjk}) \;\; {}_{\text{S}}\hat{Y}_{gjk} \quad (4.2)$$

dove:

$${}_{\text{dir}}\hat{Y}_{\text{gjk}} = \frac{N_{gjk}}{\hat{n}_{gjk}} \sum_{l \in gjk} Y_l$$ è lo stimatore diretto del dominio "gjk"

l = indice delle imprese campionarie contenute nel dominio "gjk"

$_{S}\hat{Y}_{gjk} = {}_{SSD}\hat{Y}_{djk}$ è lo stimatore sintetico, per (g ∈ d), dove d è l'indice dell'ATECO a 2 cifre

$$w_{gjk} = \begin{cases} 1 & \text{se } \dfrac{n_{gjk}}{n} \geq \lambda \dfrac{N_{gjk}}{N} \\ \dfrac{n_{gjk}}{n} \Big/ \lambda \dfrac{N_{gjk}}{N} & \text{altrimenti} \end{cases}$$

$$\lambda = 9/2$$

fase 3 - stima a livello di ATECO a 4 cifre

In questa fase, si procede alla stima a livello di ATECO a 4 cifre, adottando uno stimatore simple size dependent, dove lo stimatore sintetico è dato dalla stima della fase 2. Il dominio d'interesse è lo stesso della fase 2, con la sola eccezione della classe di attività economica (ATECO 4) al posto del gruppo di attività economica (ATECO 3). Lo stimatore è dato da:

$$_{SSD}\hat{Y}_{cjk} = w_{cjk} \;\; {}_{\text{dir}}\hat{Y}_{cjk} + (1 - w_{cjk}) \;\; {}_{S}\hat{Y}_{cjk} \qquad (4.3)$$

dove:

$$_{\text{dir}}\hat{Y}_{cjk} = \frac{N_{ijk}}{n_{ijk}} \sum_{l \in ijk} \hat{Y}_{l}$$ è lo stimatore diretto del dominio "cjk"

l = indice delle imprese campionarie contenute nel dominio "cjk"

$_{S}\hat{Y}_{cjk} = {}_{SSD}\hat{Y}_{gjk}$ è lo stimatore sintetico, per (c ∈ g), dove g è l'indice dell'ATECO a 3 cifre

$$w_{cjk} = \begin{cases} 1 & \text{se } \dfrac{n_{cjk}}{n} \geq \lambda \dfrac{N_{cjk}}{N} \\ \dfrac{n_{cjk}}{n} \Big/ \lambda \dfrac{N_{cjk}}{N} & \text{altrimenti} \end{cases}$$

$\lambda = 11/2$.

È opportuno sottolineare che il valore w, nelle tre versioni dello stimatore, assume valori diversi: è uguale a 1 nel caso dell'ATECO 2; è uguale a 9/2 nel caso dell'ATECO 3; è uguale a 11/2 nel caso dell'ATECO 4. Il motivo di tale diversità risiede nel ruolo di λ nel definire il peso ω dello stimatore SSD. Come si può vedere, ω è funzione decrescente di λ, con massimo uguale a 1. In particolare, ω rappresenta il peso della "nostra fiducia" nello stimatore diretto. Quindi λ ci consente di modulare tale "fiducia" a seconda del livello di precisione dello stimatore diretto.

Nel caso specifico, la precisione del diretto diminuisce passando dall'ATECO 2 all'ATECO 3 e quindi all'ATECO 4, per questo motivo il valore di λ aumenta ai diversi livelli. In particolare la scelta dei diversi λ è stata il frutto di un'attenta analisi che ha preso in considerazione sia la precisione delle stime sul singolo anno sia la precisione delle variazioni delle stime tra diversi anni.

Dalle analisi effettuate sull'efficienza della procedura gerarchica basata sul metodo di stima SSD, si è potuto evincere un miglioramento rispetto allo stimatore diretto, nonostante rimanessero ancora domini con errore elevato, soprattutto dove la variabilità era particolarmente elevata. Infatti uno dei limiti del SSD è proprio quello di non tener conto della variabilità all'interno dei domini, ma di considerare solamente la frazione campionaria.

Per tali motivi, si è cercato di introdurre delle migliorie che consentissero di implementare nel metodo elementi legati alla variabilità dei parametri nei domini di interesse. E stato così definito un nuovo stimatore, denominato "Sample Error Dependent Estimator", che si configura come una via di mezzo tra lo stimatore composto[15] e lo stimatore SSD, in quanto tiene conto sia della dimensione campionaria, sia di una misura della variabilità all'interno dei rispettivi domini. L'idea è di considerare lo stimatore diretto solo in quei casi in cui la dimensione campionaria consente di ottenere, con un certo grado di fiducia, un errore inferiore ad un livello prefissato. Per contro nei casi in cui tale condizione non si verifica, si considera una media ponderata tra lo stimatore diretto e lo stimatore sintetico, dove il peso è funzione della dimensione campionaria e di una "soglia" prefissata.

Anche in questo caso è stato adottato il medesimo schema gerarchico. Infatti il modello di stima prevede una prima fase a livello di ATECO 2 cifre, quindi a livello di ATECO a 3 cifre e per finire a livello di ATECO a 4 cifre.

Una formulazione generale dello stimatore adottato nelle 3 diverse fasi è la seguente:

$$ {}_{new}\hat{Y}_{mjk} = \alpha_{mjk} \ {}_{\text{dir}}\hat{Y}_{mjk} + (1 - \alpha_{mjk}) \ {}_{\text{S}}\hat{Y}_{mjk} \qquad (4.4) $$

dove:

m = indice dell'ATECO (equivale all'indice d nel caso dell'ATECO a 2 cifre (fase 1); equivale all'indice g nel caso dell'ATECO a 3 cifre (fase 2); equivale all'indice c nel caso dell'ATECO a 4 cifre (fase 3)),

[15] Tale stimatore è una media ponderata tra lo stimatore diretto e lo stimatore sintetico, dove i pesi sono funzione della varianza del diretto e del MSE del sintetico. Per ulteriori approfondimenti si può consultare "Il campionamento da popolazioni finite" (1999) di Frosiini-Montinaro-Nicolini.

j = indice di classe dimensionale di addetti (j=1, ….., 5),
k = indice della forma giuridica (k=1, 2),

$$_{\text{dir}}\hat{Y}_{mjk} = \frac{N_{mjk}}{\hat{n}_{mjk}} \sum_{l \in ijk} Y_l$$ è lo stimatore diretto del dominio "mjk"

l = indice delle imprese campionarie contenute nel dominio "mjk"

$_{\text{S}}\hat{Y}_{mjk}$ è lo stimatore sintetico del dominio "mjk" (vedere stimatore sintetico della 4.1 nel caso di ATECO a 2 cifre, stimatore sintetico della 4.2 nel caso di ATECO a 3 cifre e stimatore sintetico della 4.3 nel caso di ATECO a 4 cifre)

$$\alpha_{ijk} = \begin{cases} 1 & \text{se } n_{\text{ijk}} \geq n'_{ijk} \\ n_{\text{ijk}} / n'_{ijk} & \text{altrimenti} \end{cases}$$

dove

n_{ijk} = dimensione campionaria osservata del dominio oggetto d'interesse;

n'_{ijk} = è la dimensione campionaria teorica, determinata in funzione dell'efficienza dello stimatore diretto (errore massimo del 20% a livello di ATECO a 2 cifre; errore massimo al livello del 30% a livello di ATECO a 3 cifre; errore massimo al livello del 40% a livello di ATECO a 4 cifre 40), ad un livello di confidenza dello 0,95.

Le soglie di errore sono state calcolate utilizzando l'archivio ASIA. In particolare è stata presa in considerazione la variabile economica del volume d'affari, in quanto fortemente correlata con i principali aggregati economici della CN.

4.3 - Analisi empirica

Al fine di valutare l'efficienza dei diversi stimatori (Diretto[16], SSD e SEDE) è stata effettuata un'analisi empirica sulle imprese fino a 99 addetti.

Gli stimatori sono stati applicati a 1000 simulazioni campionarie. Tale sperimentazione è stata effettuata considerando un disegno di campionamento identico a quello adottato nell'anno 2000 nell'indagine sulle PMI:

- stratificazione ad uno stadio, con selezione delle unità con probabilità uguali;
- gli strati sono definiti dalla concatenazione di regione, classi di attività economica (ATECO a 4 cifre) e classi di addetti, questa ultima secondo lo schema della Tab. 4.1.

L'universo di riferimento, dal quale sono stati estratti i campioni, è l'archivio ASIA dell'anno 2000. Il parametro d'interesse è il volume d'affari (per il quale è noto, da ASIA, l'ammontare per

[16] Ai fini della simulazione, per rendere comparabili i risultati, nei domini dove non esisteva il diretto è stata utilizzata una stima sintetica, calcolata come diretto a livello di ATECO a 3 cifre.

l'universo delle imprese oggetto di rilevazione dell'indagine PMI).

Il dominio d'interesse è rappresentato dalla combinazione delle modalità delle seguenti variabili[17]:

- classe di attività economica (ATECO a 4 cifre),
- classi di addetti (1-2, 3-5, 6-9, 10-19, 2 0-99).

Le misure per valutare le performances degli stimatori sono state l'Absolute Relative Bias (ARB) e il Relative Root Mean Square Errors (RRMSE):

$$ARB\ (\hat{Y}_d) = \left| \frac{1}{1000} \sum_{r=1}^{1000} \frac{\hat{Y}_d(r) - Y_d}{Y_d} \right| \tag{4.5}$$

$$RRMSE \quad (\hat{Y}_d) = \frac{1}{Y_d} \sqrt{\frac{\sum_{r=1}^{1000} (\hat{Y}_d(r) - Y_d)^2}{1000}} \tag{4.6}$$

dove

$\hat{Y}_d(r)$ = stima della replicazione r-esima;

Yd = valore vero del volume d'affari, relativo alle imprese di ASIA;

r = indice delle simulazioni (r = 1……..1000);

d = generico dominio.

Di seguito (Tab. 4.2) sono presentati i dati relativi alla sperimentazione, considerando la media dei domini per classe di addetti, dell'Absolute Relative Bias (AARB) e del Relative Root Mean Square Errors (ARRMSE).

Tab. 4.2 - AARB e ARRMSE degli stimatori ("Diretto", "SSD" e "SEDE") del Volume d'affari, per classe di addetti (valori percentuali)

	ARRMSE			AARB		
Classe di	DIR	SSD	SEDE	DIR	SSD	SEDE
1-2	37,5	35,5	31,2	0,6	3,6	14,8
3-5	43,1	39,7	36,2	2,0	7,2	17,3
6-9	48,6	43,3	40,9	7,0	12,6	20,3
10-19	30,6	29,7	28,2	1,9	2,7	14,0
20-99	32,8	32,0	29,8	2,5	3,4	18,1
Tutte	38,8	36,2	33,4	3,1	6,3	17,2

Come si evince dai dati, lo stimatore migliore in termini di errore relativo è il SEDE (cfr. i dati del AARMSE), infatti il livello medio di errore è sempre inferiore agli altri due stimatori. In particolare il guadagno in termini di precisione è particolarmente evidente

[17] I risultati dell'analisi prendono in considerazione solamente il sottodominio delle società non di capitale, in quanto per le società di capitale l'informazione è desunta dai dati di bilancio, che per loro natura risultano censuari e quindi non affetti da errori di campionamento.

nelle classi con un numero minore di addetti (la "1-2", la "3-5" e la "6-9"). È interessante osservare che tale riduzione si deve esclusivamente alla forte riduzione della varianza, in quanto l'altra componente del MSE e cioè la distorsione (cfr. AARB) è significativamente più alta nel SEDE rispetto al SSD e al Diretto. Inoltre, come era lecito attendersi, la distorsione di questo ultimo è comunque la più bassa, coerentemente con la proprietà di correttezza di cui gode.

Visti i risultati della sperimentazione, il SEDE è stato scelto come nuovo stimatore dei parametri d'interesse delle piccole e medie imprese nell'ambito delle attività di Contabilità Nazionale. Tale scelta è stata ulteriormente validata dalle analisi effettuate sulle stime di più anni. Infatti, come si è potuto notare, in particolar modo sul valore aggiunto, tale metodo consente una significativa riduzione dei casi di variazioni "anomale".

5. L'errore relativo nelle stime di Contabilità Nazionale a livello di branca per le imprese fino a 99 addetti

Nel capitolo 4 è stato presentato il livello medio dell'errore di stima del volume di affari a livello di ATECO 4 cifre, 5 classi dimensionali e forma giuridica (società non di capitale, società di capitale). Tale dominio, utilizzato nelle stime di Contabilità Nazionale, non rappresenta il dominio finale di stima, bensì il dominio iniziale su cui vengono effettuate le stime che sono sottoposte successivamente ad aggregazione, fino ad arrivare al dominio d'interesse, che è rappresentato dalle 101 branche di attività economica[18]. In tal senso si è ritenuto opportuno dare una misura dell'errore che viene commesso a livello di branca per le imprese fino a 99 addetti.

Inoltre è anche opportuno tener conto dell'integrazione che viene effettuata con i dati dell'archivio BILANCI, per la parte relativa alle società di capitale. Infatti tale archivio, contenendo tutte le società di capitale rappresenta di fatto un universo esaustivo e quindi non affetto da errore campionario.

Partendo da tali premesse si è cercato di fornire una misura dell'errore della stima del valore della produzione livello di branca, tenendo conto anche dell'integrazione con l'archivio BILANCI:

$$RRMSE\ (\hat{Y}_b) = \frac{1}{Y_b}\sqrt{\frac{\sum_{r=1}^{1000}({}_{SNC}\hat{Y}_b(r) - {}_{SNC}Y_b)^2}{1000}} \qquad (5.1)$$

dove

${}_{SNC}\hat{Y}_b(r)$ = stima volume d'affari delle società non di capitale per la branca b-esima nella replicazione r-esima;

[18] I valori per addetto dei domini di analisi vengono riportati all'universo di CN utilizzando le ULA. Per ogni branca poi si aggiungono le componenti non calcolate tramite i dati di output rilevati presso le imprese ed i segmenti produttivi *non market*. Sono poi applicate le correzioni necessarie per adeguare le definizione proprie delle rilevazioni statistiche ed alcune particolari definizione dettate dal SEC95. Infine, si aggiungono le stime delle imposte e dei contributi sulla produzione e sui prodotti per elaborare le versioni dell'offerta ai prezzi base e al costo dei fattori.

$_{SNC}Y_b$ = valore vero del volume d'affari delle società non di capitale della branca b-esima;

Y_b = valore vero del volume d'affari delle società sia di capitale che non di capitale della branca b-esima;

r = indice delle simulazioni (r = 1……..1000);

b = generica branca.

La 5.1 può essere dimostrata nel seguente modo:
sia

$$RRMSE\ (\hat{Y}_b) = \frac{1}{Y_b}\sqrt{\frac{\sum_{r=1}^{1000}(\ \hat{Y}_b(r) - Y_b)^2}{1000}} \tag{5.2}$$

il Relative Root Mean square error a livello di branca, indipendentemente dalla forma giuridica;
introducendo la scomposizione per forma giuridica, la 5.2 diventa

$$RRMSE\ (\hat{Y}_b) = \frac{1}{(_{SC}Y_b + _{SNC}Y_b)}\sqrt{\frac{\sum_{r=1}^{1000}((_{SC}\hat{Y}_b(r) + \ _{SNC}\hat{Y}_b(r)) - (_{SC}Y_b + _{SNC}Y_b))^2}{1000}},$$

poiché i dati relativi alle società di capitale sono desunti dall'archivio BILANCI, che come detto è esaustivo, si può supporre, che $_{SNC}\hat{Y}_b(r) \cong _{SNC}Y_b$ e quindi

$$RRMSE\ (\hat{Y}_b) = \frac{1}{(_{SC}Y_b + _{SNC}Y_b)}\sqrt{\frac{\sum_{r=1}^{1000}(_{SC}\hat{Y}_b(r)\ - _{SC}Y_b)^2}{1000}} \tag{5.3}$$

che coincide proprio con la 5.1, cvd.

A livello di 101 branche, l'errore è risultato mediamente del 2,9%. In particolare, solo otto branche hanno fanno registrare un errore superiore al 5%, mentre nel 55% dei casi l'errore è inferiore al 2%. I casi di errori elevati, superiori al 10% si registrano solo su tre branche.

6. Conclusioni

A partire dalla revisione generale del 1987, i dati delle indagini sulle imprese vengono acquisiti per microdato. Tale approccio permette di utilizzare in modo più efficiente le fonti statistiche, consentendone un'elaborazione ad hoc finalizzata alle specificità della contabilità nazionale. L'utilizzo dei dati micro è infatti una delle caratteristiche salienti del metodo di costruzione dei conti nazionali da parte dell'Istat. Tale approccio è stato mantenuto anche con la revisione generale del 2005, riguardante gli anni 1970-2004, ma ad esso sono state apportate alcune considerevoli innovazioni delle quali, in questo lavoro, sono state presentate le principali, relative al trattamento dei dati di base sulle piccole e medie imprese (fino a 99 addetti). Le innovazioni qui descritte hanno una particolare rilevanza dato il peso delle imprese di tale fascia dimensionale nel tessuto produttivo italiano e nello sviluppo del fenomeno dell'economia sommersa al suo interno.

Nel contesto della revisione del 2005 si è inteso migliorare il metodo di rivalutazione del valore aggiunto per dichiarazione mendace da parte delle imprese adottato nel 1987, avendone verificato i limiti e la ristrettezza di alcune ipotesi di base; inoltre, si è operato sullo stimatore dei valori economici per addetto, mirando a tener conto contemporaneamente delle esigenze della contabilità nazionale e della coerenza con i dati pubblicati della rilevazione sulle piccole e medie imprese.

Da questo punto di vista si è dimostrato che l'introduzione del nuovo stimatore indiretto consente una maggiore precisione e, allo stesso tempo, garantisce una coerenza generale con i dati dell'indagine PMI, nonostante che i domini di analisi utilizzati nella contabilità nazionale per le stime della produzione e del valore aggiunto, siano più "fini" rispetto alla stratificazione che è alla base del dimensionamento del campione dell'indagine stessa. Il miglioramento dell'accuratezza, infatti, non comporta un allontanamento significativo dai livelli "originali" degli aggregati economici dell'indagine, pubblicati. Nel lavoro si sono altresì evidenziate le differenze con lo stimatore precedente, sia in termini analitici che in termini di efficienza.

Rilevante è stato anche l'apporto dell'integrazione dei dati di indagine con i dati dei bilanci delle società di capitale, nel favorire una riduzione dell'errore totale.

I risultati conseguiti non hanno solo una rilevanza in sé, in quanto portatori di una metodologia statistica tale da non generare incoerenze fra dati di base rilevati, dati di input della contabilità nazionale e sue stime finali, ma hanno anche rilevanza nel contesto, per così dire, "politico-amministrativo" dell'UE.

A partire da dicembre 2006, infatti, gli Istituti Nazionali di Statistica dell'Unione hanno l'obbligo di compilare, in appendice agli "Inventari sulle fonti ed i metodi di calcolo del Pil e del Rnl" da consegnare ad Eurostat, le cosiddette "process tables", nelle quali siano esplicitate tutte le trasformazioni che subiscono le cifre derivanti dalle fonti statistiche di base, che si dichiara di utilizzare, per dar luogo alle stime della contabilità nazionale (integrazioni per carenze statistiche, correzioni per economia sommersa, coerenza con le definizioni Sec, ecc.). Tale documentazione, approntata ai fini dei controlli da parte del Comitato GNI, costituisce la base per un giudizio sulla qualità delle stime della contabilità nazionale, sia in termini di rispondenza alle definizioni del Sec, sia sul loro grado di affidabilità e trasparenza. Sulla scorta di questa documentazione possono essere avanzate delle riserve su tali stime da parte della Commissione Europea.

In questo contesto, i conti nazionali non hanno solo una rilevanza ai fini della conoscenza della realtà economica, e come tali interessa che siano fatti "al meglio", ma anche una rilevanza "amministrativa", essendo la base per il calcolo di una parte della contribuzione che gli stati membri devono versare all'Unione. Come tali sono soggetti all'attenzione e all'indagine non solo da parte della Commissione, tramite l'Eurostat e il

Comitato GNI, ma anche da parte della Corte dei Conti Europea. La trasparenza e la coerenza del "processo di formazione dei numeri", dunque, non ha solo una valenza scientifica, ma anche una valenza politica, potendo, su tali numeri, innescarsi un contenzioso fra UE e singolo Stato. Sotto l'aspetto tecnico-scientifico, garantire la coerenza fra i dati risultanti dalle rilevazioni statistiche di base e i dati, da esse derivati, utilizzati come input della contabilità nazionale, significa non immettere elementi distorsivi nella formazione delle stime finali degli aggregati macroeconomici. Dal particolare punto di vista politico, significa non esporre tali stime a critiche e riserve per una ingiustificata lontananza dalle evidenze statistiche di base.

L'efficacia delle soluzioni adottate, per il futuro, non può che spingere verso il mantenimento dell'approccio intrapreso in un contesto di una sua evoluzione nella ricerca di nuove metodologie basate su stimatori che ottimizzino l'eventuale informazione ausiliaria e nell'utilizzo mirato di nuove fonti che si rendessero disponibili.

Un altro campo d'azione che si ritiene aprirsi in seguito alle ricerche condotte, è quello della messa in coerenza della rilevazione sulle piccole e medie imprese con le esigenze di stratificazione della contabilità nazionale: come si è detto, i domini di studio di questa sono più fini di quelli sulla base dei quali è calibrato il campione dell'indagine.

In questo lavoro si è evidenziato che le stime di contabilità nazionale relative alle imprese con meno di 100 addetti sono affette da un errore statistico relativamente contenuto, sia nel complesso sia per singola branca, ma, all'interno di questo quadro rassicurante, c'è un numero limitato di branche nelle quali l'errore è di una certa consistenza. Da questo dovrebbe trarsi una buona indicazione per calibrare diversamente la stratificazione del campione della rilevazione sulle piccole e medie imprese, così da riportare in limiti più "tranquillizzanti" l'intervallo di confidenza delle stime in argomento.

Riferimenti bibliografici

Calzaroni M. (2000), "L'occupazione come strumento per la stima esaustiva del PIL e la misura del sommerso" , *Atti del seminario "La nuova contabilità nazionale"*, ISTAT 12-13 gennaio 2000.

Caricchia A. (2006), "Perché la revisione dei conti nazionali?", *Convegno Istat "La revisione generale dei conti nazionali del 2005"*, Roma 21-22 giugno 2006.

Dabbicco G., De Gregorio C. (2002), L'utilizzo dei dati dei bilanci civilistici per l'integrazione delle mancate risposte totali alla rilevazione sul Sistema dei Conti delle Imprese (SCI), documento del *Gruppo di lavoro su"Utilizzo di dati amministrativi a fini statistici per la produzione di statistiche strutturali sui risultati economici delle imprese*, Istat.

De Gregorio C., Monducci R. (2002), "SBS data for Italy: integrated use of administrative sources and survey data. New experiences, and medium and long term strategies", *Working paper*, Istat, Department of Economic Statistics – Structural Business Statistics.

Di Consiglio L., Discenza A., Faramondi A. (2003), "Comparison of different macro-areas for the definition of the composite estimators in the estimation of employment and unemployment at sub-provincial level", *Atti Convegno CLADAG-SIS 2003*, Università di Bologna, 22-24 Settembre.

Eurostat (1996), *European System of Accounts 1995*, Lussemburgo.

Eurostat (1997), European *Regulation on Structural Business Statistics (SBS) n. 58/97*, Lussemburgo.

Faramondi A.(2005), "Analisi di coerenza dell'archivio bilanci e dell'indagine PMI", *Documento interno DCCN*, Istat.

Faramondi A., Piras M.G. (2003), "Le nuove stime di aggregati socio-economici per i sistemi locali del lavoro" in *Sviluppo locale*, n. 20 (2002), Rosenberg & Selier, Torino

Faramondi A., Foschi F., Puggioni A. (2004), "Alcune evidenze empiriche sul metodo Franz applicato sull'indagine PMI", nota presentata al *Gruppo di lavoro avente l'obiettivo di verificare l'esaustività del PIL, attraverso l'utilizzo di informazioni disponibili presso l'Anagrafe tributaria*.

Faramondi A., Foschi F., Puggioni A. (2006) "Le innovazioni introdotte nel trattamento dei dati di impresa per le stime di contabilità nazionale", *Convegno Istat "La revisione generale dei conti nazionali del 2005"*, Roma 21-22 giugno 2006.

Franz A. (1985), *Estimates of the hidden economy in Austria on the basis of official statistics*, The Review of Income and Wealth, 4, 1985.

Frosini B.V., Montinaro M., Nicolini G. (1999), *Il campionamento da popolazioni finite*, UTET, Torino.

Ghosh M., Rao J.N.K. (1994) "Small area Estimation: an Appraisal", in *Statistical Science*, n. 9, 55-93.

Istat (1991), *Classificazione delle attività economiche*, Metodi e Norme, serie C – n.11

Istat (1998), *L'impianto normativo, metodologico e organizzativo – Censimento intermedio dell'industria e dei servizi 31 dicembre 1996*.

Istat (2003), *Classificazione delle attività economiche*, Metodi e Norme, n.18

Istat (2004), *Metodologie di stima degli aggregati di contabilità nazionale a prezzi correnti – Italia – Inventario SEC95*, Istat, Metodi e Norme n.21.

Istat (2005), *Conti economici delle imprese*, Informazioni, n.6.

Pisani S., (2000) "L'identificazione delle imprese sottodichiaranti", Nota interna, Istat

Puggioni A. (2000), "L'analisi di qualità delle stime di contabilità nazionale ", *Seminario "La nuova contabilità nazionale"*, ISTAT, Roma, 12-13 gennaio 2000.

Puggioni A., Sassaroli A. (2004), Procedura di integrazione dati PMI e BILANCI per la revisione dei conti nazionali dal 2000, *Documento interno DCCN*, Istat.

Russo A. (1995) "Stimatori per piccole aree: problemi aperti", in *Società Italiana di Statistica, 100 anni di indagini campionarie,* Atti del convegno, CISU, Roma, 287-311.

Vaccari C. (2002), Caricamento e analisi dei dati di bilancio forniti dalla soc. Pitagora. Accoppiamento con ASIA, *Documento interno DCCN* - Gruppo di lavoro per l'acquisizione e gestione dei dati amministrativi.

Norme redazionali

La Rivista di Statistica Ufficiale pubblica contributi originali nella sezione "Temi trattati" ed eventuali discussioni a largo spettro nella sezione "Interventi". Possono essere pubblicati articoli oggetto di comunicazioni a convegni, riportandone il riferimento specifico. Gli articoli devono essere fatti pervenire al Comitato di redazione delle pubblicazioni scientifiche Istat corredati da una nota informativa dell'Autore contenente: appartenenza ad istituzioni, attività prevalente, qualifica, indirizzo, casella di posta elettronica, recapito telefonico e l'autorizzazione alla pubblicazione firmata dagli Autori. Ogni articolo prima della pubblicazione dovrà ricevere il parere favorevole di un referente scelto tra gli esperti dei diversi temi affrontati. Gli originali, anche se non pubblicati, non si restituiscono.

Per l'impaginazione dei lavori gli autori sono tenuti a conformarsi rigorosamente agli standard editoriali fissati dal Comitato di redazione e contenuti nel file Template.doc disponibile on line o su richiesta. In base a tali standard la lunghezza dei contributi originali per entrambe le sezioni dovrà essere limitata entro le 30–35 pagine.

Tutti i lavori devono essere corredati di un sommario nella lingua in cui sono redatti (non più di 12 righe); quelli in italiano dovranno prevedere anche un *Abstract* in inglese. La bibliografia, in ordine alfabetico per autore, deve essere riportata in elenco a parte alla fine dell'articolo. Quando nel testo si fa riferimento ad una pubblicazione citata nell'elenco, si metta in parentesi tonda il nome dell'autore e l'anno di pubblicazione. Ad esempio (Bianchi, 1987, Rossi, 1988). Quando l'autore compare più volte nello stesso anno l'ordine verrà dato dall'aggiunta di una lettera minuscola accanto all'anno di pubblicazione. Ad esempio (Bianchi, 1987a, 1987b).

Nella bibliografia le citazioni di libri e articoli vanno indicate nel seguente modo. Per i libri: cognome dell'autore seguito dall'iniziale in maiuscolo del nome, il titolo in corsivo dell'opera, l'editore, il luogo di edizione e l'anno di pubblicazione. Per gli articoli: dopo l'indicazione dell'autore si riporta il titolo tra virgolette, il titolo completo in corsivo della rivista, il numero del fascicolo e l'anno di pubblicazione. Nei riferimenti bibliografici non si devono usare abbreviazioni.

Nel testo dovrà essere di norma utilizzato il corsivo per le parole in lingua straniera e il corsivo o grassetto per quei termini o locuzioni che si vogliano porre in particolare evidenza (non vanno adoperati, per tali scopi, il maiuscolo, la sottolineatura o altro).

Gli articoli pubblicati impegnano esclusivamente gli Autori, le opinioni espresse non implicano alcuna responsabilità da parte dell'Istat.

Per contattare la redazione delle pubblicazioni scientifiche Istat e per inviare lavori: rivista@istat.it. Oppure scrivere a:

Comitato di redazione delle pubblicazioni scientifiche
C/O Carlo Deli (cadeli@istat.it)
Via Cesare Balbo, 16
00184 Roma

La Rivista di Statistica Ufficiale accoglie lavori che hanno come oggetto la misurazione e la comprensione dei fenomeni sociali, demografici, economici ed ambientali, la costruzione di sistemi informativi e di indicatori come supporto per le decisioni pubbliche e private, nonché le questioni di natura metodologica, tecnologica e istituzionale connesse ai processi di produzione delle informazioni statistiche e rilevanti ai fini del perseguimento dei fini della statistica ufficiale.

La Rivista di Statistica Ufficiale si propone di promuovere la collaborazione tra il mondo della ricerca scientifica, gli utilizzatori dell'informazione statistica e la statistica ufficiale, al fine di migliorare la qualità e l'analisi dei dati.

La pubblicazione nasce nel 1992 come collana di monografie "Quaderni di Ricerca ISTAT". Nel 1999 la collana viene affidata ad un editore esterno e diviene quadrimestrale con la denominazione "Quaderni di Ricerca - Rivista di Statistica Ufficiale". L'attuale denominazione, "Rivista di Statistica Ufficiale", viene assunta a partire dal n. 1/2006 e l'Istat torna ad essere editore in proprio della pubblicazione.

1B0120080010000000

ISSN 1828-1982

€ 10,00