

istat working papers

N.5
2011

Metodologia per l'attribuzione del codice Ateco 2007 - Registro Asia

Simone Ambroselli

istat working papers

N. 5
2011

Metodologia per l'attribuzione del codice Ateco 2007 - Registro Asia

Simone Ambroselli

Comitato di redazione

Coordinatore: Giulio Barcaroli

Componenti:

Rossana Balestrino	Francesca Di Palma	Luisa Picozzi
Marco Ballin	Alessandra Ferrara	Mauro Politi
Riccardo Carbini	Angela Ferruzza	Alessandra Righi
Claudio Ceccarelli	Danila Filipponi	Luca Salvati
Giuliana Coccia	Cristina Freguja	Giovanni Seri
Fabio Crescenzi	Aurea Micali	Leonello Tronti
Carla De Angelis	Nadia Mignolli	Sonia Vittozzi

Segreteria:

Lorella Appolloni, Maria Silvia Cardacino, Laura Peci, Gilda Sonetti, Antonio Trobia

Istat Working Papers

Metodologia per l'attribuzione del codice Ateco 2007 – Registro Asia

N. 5/2011

ISBN 978-88-458-1685-0

Istituto nazionale di statistica
Servizio Editoria
Via Cesare Balbo, 16 – Roma

Metodologia per l'attribuzione del codice Ateco 2007 – Registro Asia

A cura di Simone Ambroselli¹

Sommario

L'implementazione della nuova versione della classificazione delle attività economiche (Ateco 2007 derivata dalla nomenclatura comunitaria Nace Rev. 2) nel Registro Asia (Archivio statistico delle imprese attive) ha rappresentato per l'Istat un compito affrontato con la predisposizione di un piano di azione con il duplice obiettivo di sviluppare una nuova metodologia per l'attribuzione del codice Ateco 2007 per tutte le unità del registro (approccio micro), garantendo contestualmente agli utilizzatori la disponibilità di dati in doppia codifica (vecchio e nuovo codice) per un certo numero di anni. I principali obiettivi del presente documento sono: illustrare la metodologia utilizzata per la stima del carattere "attività economica" in classificazione Ateco 2007, principalmente nel processo di realizzazione del Registro Asia 2008; analizzare nel dettaglio le fonti disponibili. Un ampio spazio è dedicato alla fonte Studi di Settore che costituisce l'elemento centrale del nuovo processo di stima.

Parole chiave: Ateco 2007, Registro Asia, Fonti amministrative, Studi di Settore.

Abstract

The implementation of the new version of the statistical classification of economic activities (Ateco 2007 derived from the European community nomenclature Nace Rev. 2) in the Business Register Asia was a task that Istat faced setting up a project plan with two main objectives: i) to develop a new methodology to assign the Ateco 2007 codes to each unit of the register (micro-approach); ii) guaranteeing the users with a "double coding" system (both the old and the new classification) for a certain number of years. The main aims of this paper are: i) to show the methodology developed to assign the Ateco codes according to the new version of the national classification (2007) especially for the implementing of the BR Asia in the year 2008; ii) to analyse the available sources. A broad examination is devoted to the administrative source "Studi di Settore" (Statistics-based tax assessment, a surveys managed by the Italian Tax Authority) that played the key role in the new process.

Keywords: Nace Rev. 2, Business Register, Administrative Sources, Studi di Settore (Statistics-based tax assessment).

¹ Il lavoro è stato curato da Simone Ambroselli ma i seguenti paragrafi sono da attribuire a: il paragrafo 3.1.1 a Roberto Antonello Palumbo, i paragrafi 4.1.2 e 4.1.3 a Salvatore Cusimano, i paragrafi 4.1.1 e 4.1.4 a Stefano Campea, il paragrafo 3.3 e l'Appendice 1 a Alessandra Fiori.

1. Introduzione

La nuova classificazione delle attività economiche Ateco 2007 costituisce la versione nazionale della classificazione europea Nace Rev. 2, entrata in vigore il 1° gennaio 2008 con Regolamento della Commissione n. 1893/2006 (pubblicato su Official Journal del 20 dicembre 2006). Nel Regolamento sono evidenziati gli ambiti e i tempi di applicazione. Eurostat² ha individuato ben 36 differenti domini statistici coinvolti nella *migrazione* alla nuova versione e riguardo le statistiche economiche, sono interessati al cambio di classificazione di riferimento i registri delle unità economiche, le statistiche strutturali, le statistiche congiunturali e la contabilità nazionale. I tempi di adozione si differenziano per ambito di riferimento, infatti le Statistiche strutturali adottano la nuova classificazione dal 2008, mentre le Statistiche Congiunturali dal 2009 e la Contabilità Nazionale a partire dal 2010.

La classificazione Ateco 2007 risulta profondamente diversa dalla precedente versione Ateco 2002, non solo in termini strettamente numerici, per la presenza di più voci di dettaglio, ma anche riguardo i contenuti. Certamente il processo di convergenza internazionale delle classificazioni adottate a livello mondiale e la volontà di disporre di uno strumento più adeguato, in grado di cogliere le attuali dinamiche economiche, hanno determinato un deciso salto di qualità della nomenclatura di riferimento. A livello nazionale è, invece, da sottolineare come con l'Ateco 2007, per la prima volta, il mondo della statistica ufficiale, il mondo fiscale e quello camerale abbiano adottato la stessa classificazione delle attività economiche semplificando notevolmente il sistema delle dichiarazioni per le imprese.

L'Istituto Nazionale di Statistica (Istat) è stato impegnato nella predisposizione della nuova versione nazionale della classificazione e nell'implementazione delle necessarie strategie per produrre e diffondere i dati utilizzando Ateco 2007. Da considerare, inoltre, che per un certo numero di anni di riferimento è stato necessario produrre i dati in doppia codifica (Ateco 2002 e Ateco 2007) al fine di garantire il necessario supporto informativo ad utilizzatori ed utenti durante le fasi di estrazione dei campioni, di riporto all'universo o nelle analisi longitudinali, e che l'adozione della nuova classificazione Ateco 2007 per gli enti amministrativi interessati non è avvenuta in maniera contemporanea.

Tutti questi elementi hanno determinato la necessità di implementare un piano d'azione in grado di garantire la doppia codifica per un certo numero di anni e di sviluppare una nuova metodologia per l'assegnazione del codice di attività economica prevalente. Infatti l'impossibilità di utilizzare la vecchia metodologia (per la mancanza, negli anni di riferimento, di informazioni congrue fra le varie fonti amministrative), l'esistenza di fonti aggiuntive rispetto a quelle utilizzate in precedenza, la necessità – come suggerito dagli organismi internazionali – di migliorare la qualità dell'informazione prodotta, l'opportunità di inserire eventuali cambiamenti metodologici proprio in questa occasione, evitando così ulteriori *break* nei dati diffusi, hanno reso necessario lo sviluppo di una nuova metodologia di stima della variabile “attività economica” nel registro Asia.

L'anno ufficiale di pubblicazione dei dati in doppia codifica è il 2007. Riguardo la diffusione dei dati per gli utenti, il Registro Asia 2007 contiene le variabili Ateco 2002 e Ateco 2007 e rappresenta il punto di partenza per eventuali ricostruzioni in serie storiche. L'attribuzione del codice Ateco 2007 per gli anni 2005 e 2006, con un insieme di fonti a disposizione ridotto, ha avuto solo finalità di supporto per la produzione delle statistiche economiche. Nel 2008, come da Regolamento, i dati del Registro sono stati pubblicati nella sola classificazione Ateco 2007, avendo a disposizione, comunque, anche informazioni in termini di Ateco 2002 utili agli utilizzatori interni per il rispetto dei relativi obblighi comunitari.

Obiettivo del presente documento è quello di illustrare la nuova metodologia adottata, le fonti utilizzate e il complesso delle azioni implementate dall'Istituto riguardo l'adozione dell'Ateco 2007 nel Registro.

² Nace Rev. 2 - Statistical classification of economic activities in the European Community - Eurostat

Dopo una prima breve descrizione delle principali novità della nuova versione della classificazione (§2) saranno illustrate le fondamentali scelte strategiche del piano di implementazione (§3) e analizzate nel dettaglio le fonti disponibili (§4). Un ampio spazio è dedicato alla fonte Studi di Settore che costituisce l'elemento centrale del nuovo processo di stima. Successivamente è presentata la metodologia utilizzata per la stima della variabile Ateco 2007 nel processo di realizzazione del Registro Asia 2008, che sarà in seguito utilizzata a regime (§5). Infine, con riferimento alle modalità di assegnazione del carattere "attività economica", sono illustrati i principali risultati ottenuti per il Registro Asia nel quadriennio 2005-2008 (§6).

2. La nuova classificazione Ateco 2007

La nuova classificazione delle attività economiche Ateco 2007³ costituisce la versione nazionale della classificazione europea Nace Rev. 2. A sua volta, la Nace Rev. 2 è la versione europea della nuova Isic Rev. 4 definita e adottata in ambito Onu la quale rappresenta il risultato di un processo di aggiornamento e convergenza di classificazioni diverse utilizzate a livello mondiale. La nuova versione nazionale della classificazione ha in comune con la Nace le prime quattro cifre e con la Isic Rev. 4 le prime due.

La revisione ha determinato una classificazione profondamente diversa rispetto alla precedente e certamente più aderente all'attuale realtà economica. Il dettaglio è aumentato per tutti i livelli della classificazione, basti considerare che l'Ateco 2002 comprendeva 17 sezioni e 62 divisioni mentre l'attuale versione ha 21 sezioni e 88 divisioni.

In aggiunta ad un primo obiettivo di convergenza internazionale, garantito dalla comparabilità a livello di divisione, si è ottenuta, per la prima volta, una classificazione unica a livello nazionale, utilizzata sia dall'Istat sia dagli enti titolari degli archivi amministrativi in cui sono classificate le unità economiche.

Una novità sostanziale della nuova classificazione è l'introduzione della sesta cifra (livello di sotto-categoria), non presente nelle precedenti versioni. In questo modo è stato possibile garantire la comparabilità tra la nuova classificazione e le precedenti classificazioni delle fonti amministrative. Da sottolineare che, attualmente, le informazioni statistiche diffuse dall'Istat e dagli altri enti del Sistan, fanno riferimento solo al livello di categoria (cinque cifre), mentre per fini amministrativi e fiscali viene utilizzata anche la sotto-categoria.

In termini di contenuti, i cambiamenti e le novità, rispetto alla versione Ateco 2002, riguardano ciascun livello della classificazione. In particolare a livello di sezione, le modifiche più rilevanti sono:

- accorpamento in un'unica sezione (A) delle attività "Agricoltura, silvicoltura e pesca", precedentemente codificate in due sezioni distinte;
- scomposizione in due sezioni delle attività "Fornitura di energia elettrica, gas, vapore e aria condizionata" (D) e "Fornitura di acqua; reti fognarie, attività di gestione dei rifiuti e risanamento" (E); le due nuove sezioni derivano, principalmente, dalla vecchia sezione "Produzione e distribuzione di energia elettrica, gas e acqua" (E) e dalle attività inserite nelle divisioni "37 - Recupero e preparazione per il riciclaggio" e "90 - Smaltimento dei rifiuti solidi, delle acque di scarico e simili";
- creazione di una nuova sezione per le attività di "Servizi di informazione e comunicazione" (J), in cui confluiscono principalmente i vecchi gruppi "22.1 - Editoria", "64.2 - Telecomunicazioni", "92.1 - Produzioni e distribuzioni cinematografiche e di video", "92.2 - Attività radiotelevisive" e la divisione "72 - Informatica e attività connesse";
- suddivisione della vecchia sezione "Attività immobiliari, noleggio, informatica, ricerca, ser-

³ Istat Collana Metodi e norme n. 40 - Classificazione delle attività economiche - Ateco 2007, diffuso il 15 giugno 2009 (http://www.istat.it/dati/catalogo/20090615_00/)

vizi alle imprese” (K) in tre sezioni distinte: “Attività immobiliari” (L); “Attività professionali, scientifiche e tecniche” (M); “Noleggio, agenzie di viaggio, servizi di supporto alle imprese” (N);

- suddivisione della vecchia sezione “Altri servizi pubblici, sociali e personali” (O) in due sezioni distinte: “Attività artistiche, sportive, di intrattenimento e divertimento” (R) e “Altre attività di servizi” (S).

E’ da segnalare, in quanto di notevole portata innovativa, la creazione, all’interno del settore manifatturiero, di una divisione autonoma che accorpa tutte le attività di “Riparazione, manutenzione ed installazione di macchine ed apparecchiature” (divisione 33), che precedentemente erano inserite in parti diverse della classificazione. Per contro la riparazione di computer e di beni per uso personale e per la casa è stata inserita nella sezione S dedicata alle altre attività di servizi.

Nel Regolamento n. 1893/2006 sono indicati anche i tempi e gli ambiti di applicazione e, in particolare, nell’articolo 8 è stabilito che:

- le unità statistiche cui è fatto riferimento nei registri di imprese (Reg. CE n. 2186/93 successivamente modificato nel Reg. 177/2008) sono classificate secondo la Nace Rev. 2;
- gli Stati membri elaborano le statistiche relative alle attività economiche svolte a partire dal 1° gennaio 2008 in base alla Nace Rev. 2 o a una classificazione nazionale ad essa conforme;
- in deroga al paragrafo 2, dal 1° gennaio 2009 le statistiche congiunturali, elaborate a norma del reg. (CE) n. 1165/98, e le statistiche sul costo del lavoro, elaborate a norma del reg. (CE) n. 450/2003, sono prodotte in base alla Nace Rev. 2.

I semplici riferimenti agli anni 2008 e 2009 presenti nel regolamento, nascondono, in realtà, un complesso sistema di relazioni tra registri statistici, statistiche strutturali e statistiche congiunturali tali da richiedere la disponibilità di serie storiche nelle due classificazioni di attività economica per più anni, rispetto ai due considerati (Prospetto 1). Basti pensare, ad esempio, che fornire indici congiunturali nell’anno 2009 in Ateco 2007 comporta la necessità di avere dati nella stessa classificazione anche per l’anno 2005, considerato come anno-base per l’elaborazione delle statistiche. La base al 2005 è costruita su dati strutturali e questo determina un obbligo di estensione temporale anche per queste statistiche, che, a loro volta, necessitano di dati in doppia classificazione dal Registro delle imprese. Se, a questo schema, si aggiungono anche i flussi verso la Contabilità Nazionale, da considerarsi come utilizzatore finale dei diversi processi di produzione, si ha un’idea della complessità insita nel regolamento.

Prospetto 1 - Principali statistiche economiche per anno di diffusione e di riferimento dei relativi dati

Anno di diffusione	Anno di riferimento		
	-	2008	2009
2008	BR(a)		
2009			STS(b)
2010		SBS(c)	

Fonte: Regolamento n. 1893/2006 Art. 8

(a) BR: Business register

(b) STS: Short-term business statistics

(c) SBS: Structural business statistics

Sulla base di queste considerazioni, in seno all’Istituto è stato deciso di garantire la doppia classificazione (Ateco 2002 e Ateco 2007) per tutte le imprese del Registro Asia per un periodo di quattro anni, dal 2005 al 2008, fornendo, dunque, informazioni utili per gestire al meglio il complesso passaggio fra le due classificazioni per tutto il sistema delle indagini economiche. Tale circostanza ha richiesto l’implementazione di una nuova metodologia per l’assegnazione del codice di attività economica prevalente in classificazione 2007, da affiancare a quella già in uso per la precedente versione. Bisogna considerare, infatti, che la vecchia metodologia di scelta del codice Ateco basata sui dati amministrativi (in prevalenza, Anagrafe Tributaria e Camere di Commercio) non poteva essere usata per Ateco 2007, almeno per gli anni 2005, 2006 e 2007, poiché solo successiva-

mente, nel 2008, gli enti che utilizzano la classificazione hanno cominciato ad adeguarsi alla nuova versione, come prescritto dal regolamento.

Dovendo scegliere tra un approccio micro, che prevede una riclassificazione unità per unità, e uno macro, che fa uso di matrici di conversione probabilistiche, la scelta è ricaduta sul primo che assicura un'accuratezza maggiore. La possibilità di utilizzare questo tipo di approccio è emersa fin dalle prime analisi ed è assicurata dalla disponibilità di una serie di strumenti utili a garantire un processo di codifica puntuale in Ateco 2007 per tutte le unità del Registro e per tutto il periodo di riferimento. Il vantaggio principale della codifica per microdati è quello di garantire risultati maggiormente affidabili, in grado di cogliere sia il peso relativo delle novità della nuova versione della classificazione, sia la reale evoluzione strutturale dell'economia. L'uso delle matrici di conversione è stato residuale e utilizzato per completare, in alcuni casi particolari, il processo di assegnazione del codice per tutte le unità.

La necessità di predisporre una nuova metodologia è stata, dunque, trasformata in opportunità, con l'introduzione di nuove fonti informative e il miglioramento della qualità del processo di stima dell'attività economica principale, in modo tale da poter evitare ulteriori *break* nelle serie storiche dei dati diffusi, così come esplicitamente richiesto da Eurostat.

Per quattro anni di riferimento 2005-2008, sono stati prodotti i dati del Registro Asia nelle due versioni della classificazione ognuna generata utilizzando una metodologia diversa. La scelta di fondo è stata quella di lasciare separati i due processi di stima, in modo da avere un trattamento omogeneo e coerente delle diverse informazioni a disposizione. L'attribuzione di codici mediante l'uso di matrici probabilistiche è stato limitato ai soli casi di completa assenza di fonti utili per la classificazione in Ateco 2007⁴. Di conseguenza, anche nei casi in cui sarebbe stata possibile la trascodifica di tipo 1 a 1 (ad un codice 2002 corrisponde un unico codice 2007), l'attribuzione del codice Ateco 2007 è stata effettuata secondo la nuova metodologia, pertanto la coerenza tra i codici assegnati nelle due versioni della classificazione per una stessa unità statistica potrebbe non essere garantita. Fanno eccezione le imprese di maggiori dimensioni, che nel corso dei quattro anni di riferimento sono state classificate coerentemente nelle due classificazioni.

3. Il piano di implementazione

La struttura imprenditoriale italiana è caratterizzata dalla presenza di circa 4,5 milioni di imprese attive, delle quali solo una minima parte, circa 10.000 unità, ha più di 100 addetti. Una struttura simile induce alla ricerca di una metodologia di attribuzione del codice Ateco 2007 in grado di integrare i diversi strumenti a disposizione per il trattamento differenziato delle imprese di grandi dimensioni e operanti in alcuni particolari settori economici e di quelle di medio-piccola dimensione (PMI). Inoltre, la necessità di sviluppare il processo di doppia codifica per il Registro Asia per quattro anni ha influito notevolmente nella definizione delle strategie da adottare.

Prima di tutto gli sforzi dell'Istituto sono stati indirizzati alla predisposizione della versione nazionale della classificazione delle attività economiche. La pubblicazione del volume, con relativa versione *on line* corredata da tutti i necessari strumenti di supporto, ha rappresentato il momento conclusivo di un lungo processo in cui l'attività di formazione del personale ha certamente avuto un peso rilevante. In particolare, a partire dal mese di marzo 2007, il personale dell'Istituto ha potuto partecipare a specifici corsi di formazione sulla nuova classificazione Ateco. Gli eventi sono stati organizzati entrando molto nel dettaglio perché funzionali alle attività lavorative dei colleghi operanti in ambiti interessati dal cambio di classificazione. Dalla fine del 2008, invece, sono stati organizzati incontri con un taglio maggiormente divulgativo nei confronti di enti, istituzioni e associazioni di categoria.

⁴ A partire dai dati già acquisiti, sono stati calcolati i pesi per tutti i flussi previsti dalle tabelle di raccordo Ateco 2002-2007 (Cfr. <http://www.istat.it/strumenti/definizioni/ateco/>) sulla base dei quali, nei casi previsti, sono stati assegnati i nuovi codici alle imprese interessate.

Parallelamente alle attività divulgative e formative, è stato realizzato il primo piano di implementazione della variabile Ateco 2007 del Registro Asia per gli anni di riferimento 2005 e 2006. Per queste due prime versioni, il processo di trascodifica è basato principalmente sull'utilizzo della nuova fonte amministrativa Studi di Settore (SDS) e dall'applicazione del software ACTR per l'attribuzione di un codice della classificazione Ateco 2007 a partire dalle descrizioni fornite dalle imprese al momento della loro iscrizione al Registro delle Imprese delle Camere di Commercio. Sulla base delle analisi svolte è subito emersa anche la necessità di predisporre delle indagini statistiche *ad hoc* per quei settori di attività economica dove sono registrate le novità maggiormente rilevanti con il passaggio alla nuova versione della classificazione (Ambroselli e Vicari, 2007).

Per il Registro Asia anno di riferimento 2007, è stato possibile utilizzare per la prima volta una fonte amministrativa che fornisce dati direttamente in Ateco 2007, l'Archivio dichiarazioni annuali delle imposte indirette (IVA annuale). Si tratta di una fonte amministrativa già utilizzata per il processo di aggiornamento del Registro ed in particolare per la stima dello stato di attività delle imprese e per l'attribuzione del loro fatturato. A partire dal 2007 è stata impiegata anche ai fini della stima della variabile Ateco 2007.

Infine, per il Registro del 2008, le tradizionali forniture amministrative di tipo dichiarativo, Registro delle Imprese delle Camere di Commercio (CCIAA) e Anagrafe Tributaria (AT), hanno fornito informazioni sull'attività economica delle imprese sulla base della classificazione Ateco 2007. Da sottolineare, comunque, come la fornitura dell'Anagrafe Tributaria fornisca dati sull'attività economica utilizzando la nuova classificazione solo per il sottoinsieme delle nuove iscrizioni a partire dal 1° gennaio 2008 e per quelle unità che dopo questa data hanno comunicato una qualsiasi variazione nei caratteri registrati.

In conclusione, ciò che emerge è che più che parlare di una nuova metodologia per l'attribuzione del codice Ateco 2007, per i quattro anni considerati, se ne possono identificare tre distinte che fanno uso delle differenti fonti a disposizione (Prospetto 2): una per gli anni 2005 e 2006, una per il 2007 e una per il 2008.

Prospetto 2 - Modalità di utilizzo delle fonti informative a disposizione per anno di riferimento e dimensione delle unità – Anni 2005-2008

Anni	Imprese						GRANDI (a)
	PMI						
	SDS	ACTR	Indagini Speciali	Anagrafe Tributaria	CCIAA	IVA	
2005	X	X	X	-	-	-	X
2006	X	X	X	-	-	-	X
2007	X	-	-	-	-	X	X
2008	X	-	-	X(b)	X	X	X

Fonte: Istat, Archivio Statistico delle Imprese Attive

(a) si intendono le imprese di grandi dimensioni e quelle operanti in particolari settori economici

(b) fornitura parziale

L'implementazione di nuovi processi di stima per la variabile Ateco 2007 ha riguardato le piccole e medie imprese (PMI). Per quelle di grandi dimensioni e per quelle più rilevanti (GRANDI) che operano in particolari settori economici, annualmente, gli aggiornamenti sono stati effettuati dai revisori garantendo, così, la massima coerenza possibile tra le due classificazioni.

3.1 Piccole e medie imprese

Riguardo il sottoinsieme delle PMI, i Registri 2005 e 2006 sono stati realizzati adottando la stessa metodologia volta, in particolare, al soddisfacimento dell'approccio micro (classificazione unità per unità) adottato. Gli strumenti a disposizione (SDS, ACTR e Indagini Speciali) hanno permesso di raccogliere dati sufficienti per risolvere gran parte dei problemi legati alla migrazione

verso la nuova Ateco 2007. Per il Registro 2007, i processi sviluppati hanno riguardato in particolare l'integrazione dei codici ottenuti dalle procedure sviluppate per analizzare le fonti Studi di Settore e Iva, quest'ultima già disponibile in Ateco 2007. Infine, nel 2008, a SDS e IVA si sono aggiunte le "tradizionali" forniture delle Camere di Commercio (CCIAA) e dell'Anagrafe Tributaria arrivando a definire quella che sarà in futuro la metodologia a regime.

Nel seguito è descritta la metodologia utilizzata per la realizzazione dei Registri Asia anni di riferimento 2005 e 2006 (§3.1.1), mentre, riguardo gli anni 2007 e 2008 sono solo sintetizzate le scelte di fondo del nuovo processo di stima per la variabile Ateco 2007 (§3.1.2). Un'analisi dettagliata delle fonti e delle procedure sviluppate per il 2008 è ampiamente illustrata nei successivi paragrafi 3 e 4. Da sottolineare come la metodologia del 2008 rappresenti il risultato finale del complesso piano di implementazione predisposto dall'Istituto per garantire la doppia codifica per i quattro anni di riferimento. Le procedure sviluppate per il Registro Asia 2007 non sono descritte nel dettaglio perché sono da considerarsi inglobate nel più complesso sistema realizzato per il 2008.

3.1.1 Metodologia e fonti per i Registri Asia 2005 e 2006

Come già evidenziato (Prospetto 2), per i Registri Asia 2005 e 2006 è stata utilizzata una metodologia comune. Il lavoro di aggiornamento puntuale delle imprese più grandi è stato svolto dopo un'intensa fase di aggiornamento professionale dei revisori sui principi e le novità della nuova classificazione.

Oltre gli aggiornamenti puntuali e i risultati del piano di controllo del Registro Asia, le fonti principali utilizzate per l'attribuzione del codice Ateco 2007 per gli anni di riferimento 2005 e 2006 sono state: gli Studi di Settore (fonte principale per oltre il 65% del totale nel 2005 e per il 63,5% nel 2006); il software ACTR; i risultati delle Rilevazioni sulla nuova classificazione delle attività economiche. Notevole anche il ruolo rivestito da una serie di forniture speciali di dati da parte di Enti e Associazioni di categoria, utilizzate per i primi due anni di doppia codifica in modo da ridurre al minimo il peso delle matrici di transizione.

Degli Studi di Settore, che rappresentano la principale fonte di aggiornamento, si parlerà più diffusamente in seguito (cfr. §§3 e 4). Da sottolineare, comunque, che, in generale, la fonte SDS è stata sempre considerata come prioritaria nella scelta del codice da attribuire quando proveniente da più fonti.

Lo strumento ACTR, sviluppato originariamente dall'US Census Bureau e successivamente da ricercatori di Statistics Canada, è un software generalizzato (indipendente dalla lingua e dalla classificazione) che rientra tra i cosiddetti "*weighting algorithms*". Il software è stato utilizzato per attribuire un codice della classificazione Ateco 2007 a partire dalle descrizioni fornite dalle imprese alle Camere di Commercio. Fondamentale, per ottenere buoni risultati, è la costruzione di un apposito *dizionario*. La costruzione di tale dizionario è stata effettuata sulla base delle seguenti fonti: i titoli, le declaratorie e le note esplicative della classificazione considerata; tutte le descrizioni presenti nelle nomenclature connesse, come la Prodcom; le "descrizioni" già a disposizione da precedenti indagini Istat. In sintesi, il software si basa sul confronto, preceduto da una fase di standardizzazione dei testi detta "*parsing*", tra la descrizione da codificare e le voci contenute nel dizionario. Il risultato può essere un abbinamento esatto (*direct match*), che dà luogo inequivocabilmente all'assegnazione di un codice unico, o uno parziale (*indirect match*), con individuazione del codice o dei codici del dizionario con descrizione più simile a quella analizzata.

I principali problemi nell'utilizzo di ACTR sono legati alla presenza di elementi (caratteri speciali e parole) non collegati direttamente all'attività economica svolta e alla lunghezza delle descrizioni che spesso sono riferite a più di un'attività. ACTR, infatti, opera al meglio con lunghezze di testo inferiori ai 200 caratteri e con descrizioni il più possibile vicine al "linguaggio" presente nel dizionario.

Pur avendo a disposizione un insieme variegato di strumenti, è apparso fin da subito evidente che, almeno per i primi due anni, gli strumenti a disposizione non sarebbero stati sufficientemente in grado di cogliere le novità introdotte nella nuova classificazione. Per questo motivo sono state predisposte delle indagini *ad hoc* per individuare sia le attività economiche "nuove", emerse in ma-

niera rilevante negli ultimi anni, sia quelle con particolari problemi nel passaggio tra le due versioni (trascodifiche 1 a n). I settori indagati, in classificazione 2007, sono stati i seguenti:

- servizi di informazione e comunicazione (sezione J);
- attività degli studi di architettura e d'ingegneria; collaudi ed analisi tecniche (divisione 71);
- ricerca e sviluppo sperimentale nel campo delle scienze naturali e dell'ingegneria (gruppo 72.1);
- pubblicità (gruppo 73.1);
- attività di design specializzate (gruppo 74.1);
- attività di servizi per edifici e paesaggio (divisione 81);
- altre attività professionali, scientifiche e tecniche n.c.a. (non classificato altrove); attività di supporto per le funzioni d'ufficio; servizi di supporto alle imprese n.c.a. (74.9, 82.1, 82.9).

Nel complesso sono state coinvolte circa 45 mila imprese: tutte le unità con oltre 10 addetti e un campione di quelle più piccole, con l'obiettivo di verificare la qualità della codifica in Ateco 2002 e di attribuire un nuovo codice in Ateco 2007.

Infine, la soluzione di complessi casi di trascodifica 1 a n è stata ottenuta utilizzando nuove forniture (liste, elenchi speciali) sia come fonte principale per la riclassificazione, sia per completare le informazioni già a disposizione (Prospetto 3).

Prospetto 3 - Liste utilizzate per l'attribuzione del codice Ateco 2007 come fonte principale

Ateco 2007	Descrizione attività	Ente Fornitore	Tipo fonte
64.30.2	Sicav	Banca d'Italia	Principale
64.91.0	Leasing finanziario	Banca d'Italia	Principale
64.92.0	Consorzi garanzia fidi	Banca d'Italia	Principale
64.99.1	Intermediazione mobiliare	Banca d'Italia	Principale
64.99.2	Factoring	Banca d'Italia	Principale
64.99.4	Società veicolo	Banca d'Italia	Principale
66.19.2	Promotori e mediatori finanziari	Banca d'Italia	Principale
66.19.5	Money transfer	Banca d'Italia	Principale
66.21.0	Periti e liquidatori assicurazioni	Isvap	Principale
66.22.0	Agenti e broker assicurazioni	Isvap	Principale
69.10.2	Notai	Enti di previdenza	Principale
71.11.0	Architetti	Enti di previdenza	Principale
71.12.1	Ingegneri	Enti di previdenza	Principale
71.12.3	Geometri	Enti di previdenza	Principale
74.90.9	Periti industriali	Enti di previdenza	Principale
47.73.1	Farmacie	Enti di previdenza	Ad integrazione
69.10.1	Studi legali	Enti di previdenza	Ad integrazione
69.20.1	Studi commerciali, tributari e di revisione contabile	Enti di previdenza	Ad integrazione
69.20.3	Consulenti del lavoro	Enti di previdenza	Ad integrazione
75.00.0	Veterinari	Enti di previdenza	Ad integrazione
86.21.0	Studi medici generale	Enti di previdenza	Ad integrazione
86.22.0	Studi medici specialistici	Enti di previdenza	Ad integrazione
86.23.0	Studi odontoiatrici	Enti di previdenza	Ad integrazione
86.90.2	Attività paramediche	Enti di previdenza	Ad integrazione
86.90.3	Psicologi	Enti di previdenza	Ad integrazione
90.01.0	Rappresentazioni artistiche	Enti di previdenza	Ad integrazione

Fonte: Istat, Archivio Statistico delle Imprese Attive

Per alcuni codici Ateco, inoltre, è stata implementata una procedura di analisi testuale della ragione sociale (non ACTR) con cui individuare delle parole chiave direttamente riconducibili a particolari attività svolte. Può accadere, infatti, che per alcune attività sia richiesta espressamente la presenza di particolari dizioni all'interno della ragione sociale (ad esempio, *SIM – società di inter-*

mediazione mobiliare, codificata in Ateco 2007 nella 64.99.1). In altri casi, invece, risulta frequente l'uso di alcune parole in grado di cogliere specificità che l'impresa vuole rendere immediatamente manifeste (ad esempio, *trasloco* per rilevare, il nuovo codice 49.42.0). In questo modo, sono stati attribuiti codici ad oltre 20 mila imprese per circa trenta categorie Ateco coinvolte in trascodifiche di tipo complesso (1 a n).

3.1.2 Linee guida per i Registri Asia 2007 e 2008

Per il sottoinsieme delle piccole/medie imprese, la nuova metodologia a regime (Asia 2008) è basata su diversi elementi: l'utilizzo di nuove fonti amministrative (SDS), lo sfruttamento, in un'altra ottica, di informazioni già a disposizione (IVA) e l'uso delle tradizionali forniture delle Camere di Commercio e dell'Anagrafe Tributaria.

In estrema sintesi, i processi sui dati di provenienza SDS e IVA hanno come fine l'ottenimento di codici di attività economica a cui sono associati valori di fatturato e/o volume d'affari (in termini assoluti o percentuali).

In particolare, il cuore della metodologia è costituito dal trattamento a fini statistici degli oltre 200 questionari che costituiscono gli Studi di Settore. Questi ultimi rappresentano un fondamentale strumento fiscale da cui è possibile ricavare una serie di indicazioni utili alla codifica per attività economica. Sono stati utilizzati per tutti e quattro gli anni di riferimento ed hanno consentito, annualmente, l'aggiornamento della variabile Ateco 2007 per circa i due terzi delle imprese del Registro. Ogni questionario è stato analizzato singolarmente, in modo da individuare il set di informazioni rilevanti ai fini della codifica. La metodologia implementata ha come obiettivo l'individuazione di un insieme di regole in grado di collegare variabili, anche eterogenee, compilate in diverse sezioni del modello ma che riconducono ad una particolare attività economica. In pratica, si tratta di un processo di costruzione di regole deterministiche che lega, anche in modo complesso, le singole risposte date a specifiche domande. In questo modo, dal processo derivano sia i codici sia le percentuali di fatturato relativi ad ogni singola attività. Avendo a disposizione dati numerici è possibile applicare, ove necessario, il metodo dell'ordine decrescente, basato però sul fatturato e non sul valore aggiunto, per attribuire il codice di attività prevalente e quelli secondari.

Meno complesso ma con finalità simili è il trattamento delle informazioni dell'archivio IVA annuale. In questo caso, si dispone di informazioni distinte per le singole attività svolte da ciascuna impresa e, dopo un'accurata fase di trattamento volta ad evitare soprattutto duplicazioni di valori, è possibile ordinare sulla base del volume d'affari i codici di attività economica.

Poiché il miglioramento della qualità nell'attribuzione del codice di attività prevalente è un obiettivo da perseguire costantemente, l'uso di fonti con questa caratteristica, codici più valori ordinabili, rappresenta una sostanziale novità e un indubbio vantaggio da sfruttare nel processo di codifica. Per questo motivo, le fonti tradizionali, non essendo corredate da dati quantitativi, occupano le posizioni più basse della scala gerarchica all'interno della metodologia di scelta.

Come visto, gli SDS sono presenti in tutti e quattro gli anni considerati e sono sempre stati posti al vertice della gerarchia delle fonti, mentre la fornitura parziale dell'Anagrafe Tributaria è stata usata ad integrazione, per l'anno 2008, per le sole imprese non coperte dalle altre fonti. La prima operazione di trattamento è consistita nel distinguere le imprese potenzialmente aggiornabili con altre fonti (almeno un codice Ateco valido, non necessariamente completo, nelle fonti SDS, IVA e CCIAA) da quelle presenti solo nell'Anagrafe Tributaria.

Per circa 7,5 milioni di unità è presente almeno un codice Ateco valido di provenienza SDS, IVA o CCIAA mentre sono oltre 4,2 milioni i record non coperti dalle principali fonti di aggiornamento. Queste ultime possono essere ulteriormente suddivise in tre sottoinsiemi:

- a) imprese con Ateco 2007 fornito solo dall'Anagrafe Tributaria (609.767 imprese);
- b) imprese con il solo dato in classificazione 2002 con trascodifica 1 a 1 (1.923.571 imprese);
- c) imprese con il solo dato in classificazione 2002 con trascodifica 1 a n (1.217.380 imprese).

Per il sottoinsieme b) è stata applicata la conversione automatica del codice 2002 in classificazione 2007. Per il sottoinsieme c) sono state predisposte delle regole di conversione in grado di garantire almeno le due cifre Ateco in classificazione 2007.

In conclusione, è da sottolineare come rispetto all'universo di partenza per la creazione del Registro Asia 2008, costituito dalle circa 11,3 milioni di unità giuridiche presenti nell'archivio dell'Anagrafe Tributaria, sono risultate senza alcuna fonte circa 400 mila unità per le quali non si dispone di codici neanche nella precedente versione della classificazione.

3.2 Le imprese di grandi dimensioni

Per l'aggiornamento di tutti i caratteri delle imprese di grandi dimensioni nel registro Asia è prevista un'attività continua di monitoraggio che fa uso di tutte le fonti amministrative e statistiche a disposizione. In particolare, per l'attività economica è garantita la coerenza tra i codici espressi nelle due classificazioni Ateco. L'attività è svolta da revisori esperti, personale qualificato, integrando dati statistici provenienti dalle principali indagini economiche dell'Istituto (Indagine sulle unità locali delle imprese, Sistema dei conti delle imprese, Rilevazioni della produzione industriale, Indagine Congiunturale Grandi Imprese e Indagine Congiunturale fatturato e ordinativi) e dati amministrativi disponibili *on line* aggiornati in tempo reale (Inps – DM10, CCIAA - visura storica, bilanci d'esercizio e consolidati).

Il controllo sulle imprese rilevanti per dimensione (grandi imprese con oltre 100 addetti oppure con elevato peso specifico all'interno di particolari *cluster*) è condotto in maniera continuativa per ogni anno di riferimento, indipendentemente dalla coerenza o meno dei dati con l'anno precedente. In questo modo è possibile seguire l'evoluzione delle principali imprese italiane, tenendo conto anche di possibili eventi demografici, quali fusioni, scorpori, acquisizioni o cessioni di rami d'azienda. L'esame analitico dei dati di bilancio (ricavi e costi del Conto Economico con gli ulteriori chiarimenti presenti in Nota Integrativa) e della Relazione di gestione, consente anche l'individuazione delle attività secondarie e l'applicazione del metodo dell'ordine decrescente⁵.

3.3 Le imprese operanti in settori rilevanti

Per particolari settori economici sono disponibili elenchi speciali, in alcuni casi corredati da dati di tipo economico, provenienti da Autorità di Vigilanza, Ministeri e Associazioni di categoria. Per le imprese che operano in questi settori, al controllo puntuale delle imprese più grandi se ne affianca un altro relativo all'intero sottoinsieme considerato. Riguardo l'Ateco, un primo obiettivo da raggiungere è quello di attribuire un determinato codice alle sole imprese che hanno i necessari attributi per svolgere particolari attività economiche, evitando, contestualmente, che nel settore siano inserite unità operanti in altri ambiti. Un classico esempio è relativo al settore finanziario dove, grazie alle forniture annuali delle Autorità di Vigilanza, Banca d'Italia e Isvap, è possibile disporre dei sottoinsiemi delle imprese autorizzate allo svolgimento delle attività bancarie e assicurative. A partire da questi elenchi speciali (liste satellite) è possibile selezionare sia le imprese a cui attribuire i codici Ateco relativi all'attività svolta, sia quelle per le quali le fonti amministrative forniscono indicazioni errate, da valutare e eventualmente correggere automaticamente.

Questa attività è da inquadrare in un contesto più ampio, quello del *piano di check*, redatto annualmente per il controllo e la correzione dei dati, per tutte le variabili del Registro. In generale sono predisposte regole di correzione deterministica (del tipo *edit-imputation*) e di accertamento su dati anomali con cui validare i risultati, caso per caso, dei settori considerati particolarmente rilevanti (Appendice 1).⁶

⁵ Regole di base di classificazione, tratte dal volume Ateco 2007 – “Ad ogni unità registrata nei registri statistici delle imprese si associa un codice Ateco in base all'attività economica principale, cioè quella che contribuisce maggiormente al valore aggiunto dell'unità”. “Nel caso complesso in cui un'unità svolge più attività classificate in più di due posizioni Ateco, e nessuna delle attività incide per oltre il 50 per cento del valore aggiunto, la classificazione per attività dell'unità dovrà essere stabilita secondo il metodo top-down” che “segue il principio gerarchico per cui la classificazione di un'unità al livello più basso della classificazione deve essere coerente con la classificazione dell'unità al livello più alto della struttura”. Il processo “identifica la sezione con la percentuale di valore aggiunto più alta; identifica, all'interno della sezione, la divisione con la percentuale di valore aggiunto più alta; identifica, all'interno della divisione, il gruppo con la percentuale di valore aggiunto più alta; identifica, all'interno del gruppo, la classe con la percentuale di valore aggiunto più alta”.

⁶ L'appendice 1 dal titolo “Piano di *check* per la variabile codice di attività economica” è stata curata da Alessandra Fiori ed è disponibile su richiesta al curatore.

I settori per i quali l'attribuzione del codice Ateco è basata sulla disponibilità di liste satellite ed elenchi speciali sono riepilogati nel Prospetto 4. Nel caso in cui una lista sia considerata esaustiva, il corrispondente codice di attività economica è assegnato esclusivamente alle imprese ad essa appartenenti. Per le liste non esaustive, invece, sono solo tenute sotto controllo tutte le "entrate" e le "uscite" dalla lista dovute all'assegnazione di codici di attività economica che la caratterizzano; tuttavia la disponibilità di liste parziali e i controlli originati dall'attivazione delle regole consentono, in genere, di determinare correttamente il sottoinsieme di riferimento.

Anche per queste unità, così come per quelle di grandi dimensione, l'attribuzione dei codici nelle due classificazioni è stata fatta contemporaneamente garantendo la coerenza tra le due versioni.

Prospetto 4 - Liste utilizzate nel piano di controllo del Registro Asia con riferimento al codice di attività economica - Anni 2005-2008

Ateco 2002	Ateco 2007	Descrizione attività	Ente Fornitore	Tipologia lista
24.20.0	20.20.0	Prodotti fitosanitari	Agrofarma	Esaustiva
52.11.1	47.11.1	Ipermercati	Nielsen	Esaustiva
52.11.2	47.11.2	Supermercati	Nielsen	Non esaustiva
52.11.3	47.11.3	Discount	Nielsen	Non esaustiva
52.11.5	47.11.5	Commercio al dettaglio di surgelati	Nielsen	Non esaustiva
52.12.1	47.19.1	Grandi magazzini	Nielsen	Non esaustiva
52.12.2	47.19.9	Bazar	Nielsen	Non esaustiva
60.10.0	49.10.0	Trasporto ferroviario passeggeri (interurbano)	Ministero trasporti	Esaustiva
	49.20.0	Trasporto ferroviario di merci	Ministero trasporti	Esaustiva
	49.31.0	Trasporto terrestre di passeggeri in aree urbane e suburbane	Ministero trasporti	Non esaustiva
62.10.0	51.10.1(a)	Trasporto aereo di linea	Enac	Esaustiva
62.20.0	51.10.2(a)	Trasporto aereo non di linea	Enac	Non esaustiva
	51.21.0	Trasporto aereo di merci	Enac	Non esaustiva
65.11.0	64.11.0	Banche centrali	Banca d'Italia	Esaustiva
65.12.1	64.19.1	Banche	Banca d'Italia	Esaustiva
66.01.0	65.11.0	Assicurazioni vita	Isvap	Esaustiva
66.03.0	65.12.0	Assicurazioni diverse da quelle sulla vita	Isvap	Esaustiva
	65.20.0	Riassicurazioni	Isvap	Esaustiva
67.11	66.11.0	Amministrazione di mercati finanziari	-	Non Esaustiva(b)
74.12.2	69.20.2	Revisione bilanci	Consob	Esaustiva
74.50.2	78.20.0	Interinali	Ministero del lavoro	Esaustiva
	78.30.0	Staff leasing	Ministero del lavoro	Esaustiva
64.2	61	Telecomunicazioni	Archivio Roc	Non esaustiva

Fonte: Istat, Archivio Statistico delle Imprese Attive

(a) In Ateco 2007 si tratta esclusivamente di trasporto di passeggeri

(b) Elenco Borse Merci non completo

4. Fonti a disposizione

4.1 Studi di Settore

Gli Studi di Settore sono stati introdotti nell'ordinamento normativo italiano nel 1993 con l'obiettivo di valutare la capacità di produrre ricavi o conseguire compensi dalle singole attività economiche svolte dai soggetti interessati. Sono costituiti da un insieme di modelli, collegati a uno o più codici Ateco (in classificazione 2007 a partire dai dati fiscali del 2007) che, inviati ai contribuenti, permettono la raccolta sistematica di dati di carattere economico e di elementi di tipo strutturale, che caratterizzano l'attività e il contesto economico di riferimento. Gli studi di settore sono, dunque, uno strumento fiscale utilizzato sia dal contribuente, per verificare la congruità e la coerenza della propria dichiarazione, sia dall'Amministrazione finanziaria quale ausilio all'attività di controllo.

I singoli questionari sono tutti diversi tra loro, tranne poche eccezioni, perché tengono conto delle particolari tipologie di attività esercitate e, inoltre, la loro redazione, per essere il più possibile aderente alla realtà economica di riferimento, deriva da accordi di reciproca collaborazione tra Amministrazione finanziaria, associazioni di categoria e ordini professionali.

L'universo di riferimento non è lo stesso del Registro Asia. In termini di categorie Ateco 2007, rispetto all'universo di riferimento del Registro Asia,⁷ il grado di copertura della fonte è pari al 71,5%. Oltre ad alcune divisioni del tutto fuori campo, non sono assoggettate agli SDS altre attività particolarmente rilevanti e, in genere, caratterizzate da ricavi superiori alla soglia di riferimento di 7,5 milioni di euro prevista, tra le quali le imprese della divisione 64 – attività di servizi finanziari e gli ipermercati (codice Ateco 47.11.1) (Prospetto 5).

Prospetto 5 - divisioni Ateco 2007 non assoggettate agli Studi di Settore

Divisione	Descrizione
05	Estrazione di carbone
06	Estrazione di petrolio greggio e di gas naturale
07	Estrazione di minerali metalliferi
11	Industria delle bevande
12	Industria del tabacco
35	Fornitura di energia elettrica, gas, vapore e aria condizionata
36	Raccolta, trattamento e fornitura d'acqua
37	Gestione delle reti fognarie
51	Trasporto aereo
61	Telecomunicazioni
65	Assicurazioni, riassicurazioni e fondi pensione
78	Attività di ricerca, selezione, fornitura di personale
80	Servizi di vigilanza e investigazione
87	Servizi di assistenza sociale residenziale
88	Assistenza sociale non residenziale
91	Attività di biblioteche, archivi, musei ed altre attività culturali

Altre importanti differenze rispetto al Registro Asia si hanno a causa di una serie di norme di esclusione di applicazione degli studi, di cui le principali sono le seguenti:

- ricavi, ovvero compensi, di ammontare superiore a euro 7,5 milioni di euro (soglia in vigore dal 2007, in precedenza il limite era di 5.164.569 euro);
- periodo di imposta di durata diversa da 12 mesi, indipendentemente dalla circostanza che tale arco temporale sia o meno a cavallo di due esercizi;
- attività iniziata o cessata nel corso del periodo d'imposta;
- periodo di non normale svolgimento dell'attività.

Per l'anno di imposta 2008 (dati relativi al 2007) gli Studi analizzati sono stati 206, relativi in totale ad oltre 4 milioni di soggetti: 61 nei servizi; 24 per i professionisti; 70 nel commercio; 51 nel manifatturiero.

Da sottolineare come i quattro macro-settori identificati negli Studi di Settore non coincidano perfettamente con i concetti propri della classificazione economica Ateco 2007. Ad esempio, tra i

⁷ Per l'Archivio statistico delle imprese attive (Asia) (ovvero le imprese che hanno svolto un'attività produttiva per almeno sei mesi nell'anno di riferimento) sono escluse dal campo di osservazione le attività economiche relative a: agricoltura, silvicoltura e pesca (sezione A della classificazione NACE Rev. 2); amministrazione pubblica e difesa; assicurazione sociale obbligatoria (sezione O); attività di organizzazioni associative (divisione 94); attività di famiglie e convivenze come datori di lavoro per personale domestico; produzione di beni e servizi indifferenziati per uso proprio da parte di famiglie e convivenze (sezione T); organizzazioni ed organismi extraterritoriali (sezione U); le unità classificate come istituzioni pubbliche e istituzioni private non profit.

“servizi” sono inseriti i questionari rivolti ai settori costruzioni, trasporti e magazzinaggio o anche relativi al settore della pesca (categorie Ateco 03.11.0 e 03.12.0), mentre, nel settore “manifatturiero”, all’interno dei singoli modelli, sono presenti unità codificate nel commercio (tre nel dettaglio e tre all’ingrosso). Inoltre, gli agenti di commercio (divisione Ateco 46) non sono inseriti nel “commercio” ma tra i “servizi”.

In genere, ogni anno circa un terzo dei questionari risulta modificato rispetto all’anno precedente. In questi casi, la sigla identificativa del modello passa prima da S, lettera che identifica un modello iniziale, a T, in caso di una prima evoluzione e successivamente a U, se eventualmente si verifica una seconda evoluzione. Per gli studi che non si sono modificati da un anno all’altro, il trattamento è senz’altro più veloce in quanto è quasi sempre possibile utilizzare le regole applicate nell’anno precedente. Per gli studi modificati, invece, è necessario ripetere nuovamente un processo di analisi e di creazione delle regole, almeno quando le modifiche riguardano i campi realmente utilizzati.

In termini di imprese, la copertura “potenziale” della fonte è intorno al 90%. A posteriori, per il 2008, considerando lo stato di attività, la completezza delle informazioni e la reale possibilità di trattamento in termini di attività economica, sono stati ottenuti i codici Ateco su circa 3,6 milioni di unità (il 78% del totale imprese di Asia).

In media ogni questionario si rivolge a 3,3 attività economiche diverse esercitate (a livello di 5 cifre Ateco), con una punta massima di 54 codici per lo studio TD32U. In 84 questionari il settore esaminato è rappresentato da un’unica attività, mentre per altri 49 sono due i codici esaminati.

Ogni studio è composto da un insieme di quadri con diverso contenuto informativo. La struttura tipo di ogni modello è, in genere, la seguente:

- Quadro A – personale addetto all’attività
- Quadro B – unità locali destinate all’esercizio dell’attività
- Quadro C – modalità di svolgimento dell’attività
- Quadro D – elementi specifici dell’attività
- Quadro E – beni strumentali
- Quadro F – elementi contabili (manifatturiero, servizi, commercio)
- Quadro G – elementi contabili (professionisti)
- Quadro X – altre informazioni rilevanti
- Quadro Z – dati complementari.

In sintesi, le informazioni ottenute dalla fonte possono essere raggruppate in due gruppi:

- gli elementi comuni a tutti i tipi di impresa (occupazione, localizzazione, valori di bilancio);
- un insieme di dati legati alla specifica attività svolta (input, output, processi produttivi, organizzazione e beni materiali).

I dati di questo secondo sottoinsieme, che possiamo definire di tipo strutturale, sono in genere contenuti nei quadri C, D ed E e costituiscono il punto di partenza per la definizione di una nuova metodologia di assegnazione dei codici Ateco in classificazione 2007.

L’obiettivo della nuova procedura è stato quello di individuare un insieme di regole, in grado di collegare variabili, anche eterogenee, compilate in diverse sezioni del modello ma che riconducono ad una particolare attività economica.

L’idea di base della metodologia sviluppata, è stata quella di legare l’attribuzione del codice Ateco all’analisi dei singoli *blocchi* che compongono la definizione di attività economica stessa. Per questo motivo occorre valutare la congruenza di una metodologia così costruita con i seguenti principi tratti dalle linee guida della Nace Rev. 2:

- un’attività economica ha luogo quando risorse come capitali, lavoro, conoscenze tecnologiche o prodotti intermedi sono combinati per produrre specifici beni o servizi; così, un’attività economica è caratterizzata da input di risorse, un processo produttivo e un output di prodotti (beni o servizi);
- l’attività principale di un’unità statistica è quella che contribuisce in misura maggiore alla creazione del valore aggiunto dell’unità stessa;

- l'attività principale è attribuita in base al criterio dell'ordine decrescente e non necessariamente copre il 50% o più del valore aggiunto totale dell'unità.

La nuova metodologia rispetta il primo principio. In generale, infatti, le regole sono state costruite sulla base di “domande chiave” da cui ricavare informazioni su *input*, *output* e *processi* che sono i tre elementi caratterizzanti la definizione di attività economica, “risorse combinate per portare alla creazione di un bene o di un servizio specifico”. Non avendo a disposizione i dati in termini di valore aggiunto, il fatturato (valori percentuali) è stato utilizzato come sua *proxy* per garantire l'individuazione dell'attività prevalente con l'applicazione del metodo dell'ordine decrescente ottenendo anche i codici per le attività secondarie svolte (rispetto del secondo e terzo principio).

Le principali domande utili ai fini dell'attribuzione dei codici di attività economica sono le seguenti:

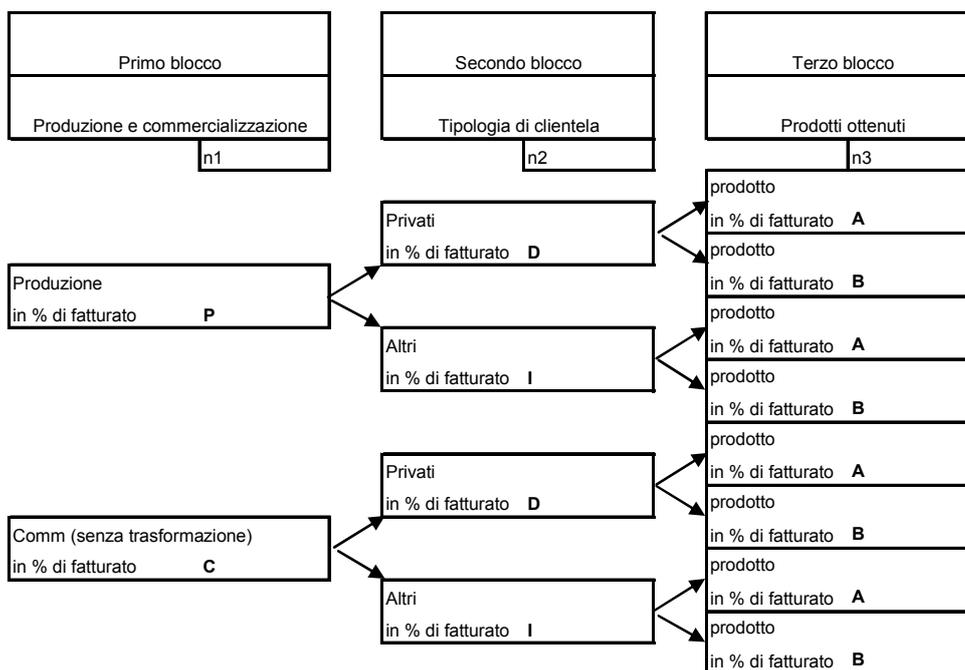
- produzione o commercializzazione (vendita senza ulteriori trasformazioni) in termini di fatturato (valori percentuali);
- tipologia clientela in termini di fatturato (valori percentuali) – analisi dei canali distributivi;
- materie prime (input) utilizzate;
- processi produttivi (fasi della produzione);
- tipo di attività svolta in termini di fatturato (valori percentuali) - (output);
- prodotti in termini di fatturato (valori percentuali) - (output).

La metodologia parte da queste singole componenti, legandole in modo sempre più complesso, fino ad ottenere regole di tipo deterministico da cui derivano sia i codici sia le percentuali di fatturato relativi ad ogni singola attività.

In genere, le variabili chiave utilizzate sono quelle riferite a: tipologia attività (produzione o commercio); tipologia clientela (ingrosso o dettaglio); elenco prodotti/attività. Nel Prospetto 6 è sintetizzato il processo di costruzione delle regole deterministiche tipico del settore manifatturiero. Nell'esempio, sono considerati i tre blocchi di domande chiave: due modalità per il primo set di domande (produzione o commercio); due modalità per la tipologia di clientela (orientamento all'ingrosso o al dettaglio); due soli prodotti considerati. In uno schema così semplificato, le combinazioni tra le differenti modalità da esaminare sono 8, mentre, nella realtà l'applicazione della metodologia dà luogo ad un numero decisamente maggiore di regole di tipo deterministico a cui associare i codici Ateco di riferimento.

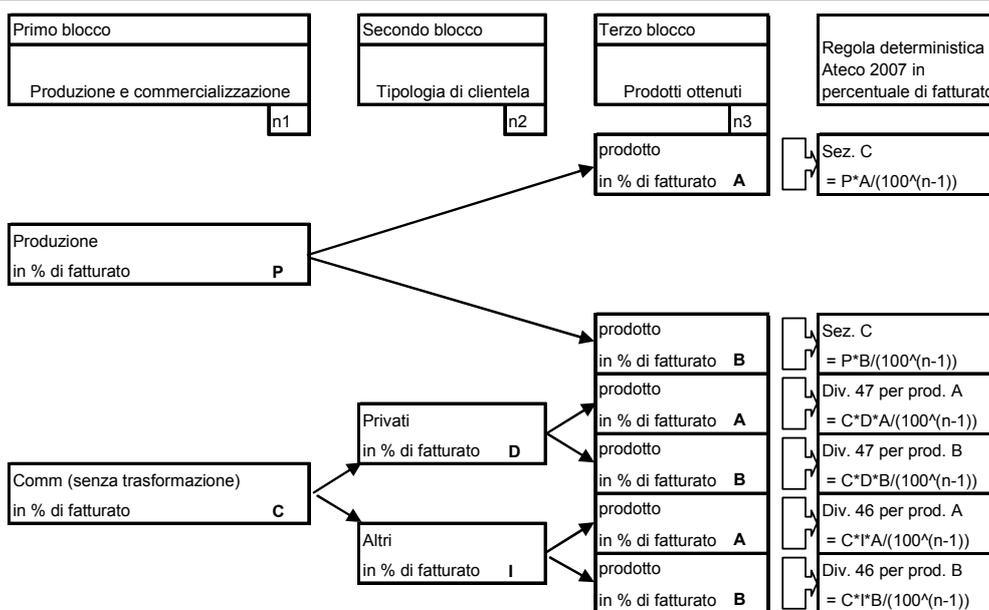
In alcuni casi non è necessario trasformare in regole tutte le combinazioni. Infatti, l'ottica della metodologia è quella di ottenere risultati utilizzando al meglio le informazioni disponibili, evitando procedure inutilmente laboriose. Nel semplice esempio presentato, si deve considerare che l'integrazione dei primi due blocchi è necessaria solo nei casi in cui ci sia l'indicazione di svolgere, almeno in parte, attività di commercializzazione senza rivendita. Solo in questo caso, infatti, è utile distinguere tra commercio all'ingrosso e dettaglio, mentre risulta inutile associare le risposte sulla tipologia di clientela all'attività di produzione. Quindi, per le attività potenzialmente codificabili nella sezione C (Manifatturiero), basta esaminare le combinazioni tra percentuali di produzione e quelle di output, mentre per quelle del commercio è necessario distinguere tra dettaglio e ingrosso per tutti i prodotti interessati.

Prospetto 6 - Il processo di costruzione delle regole deterministiche - un esempio per il settore manifatturiero (schema teorico)



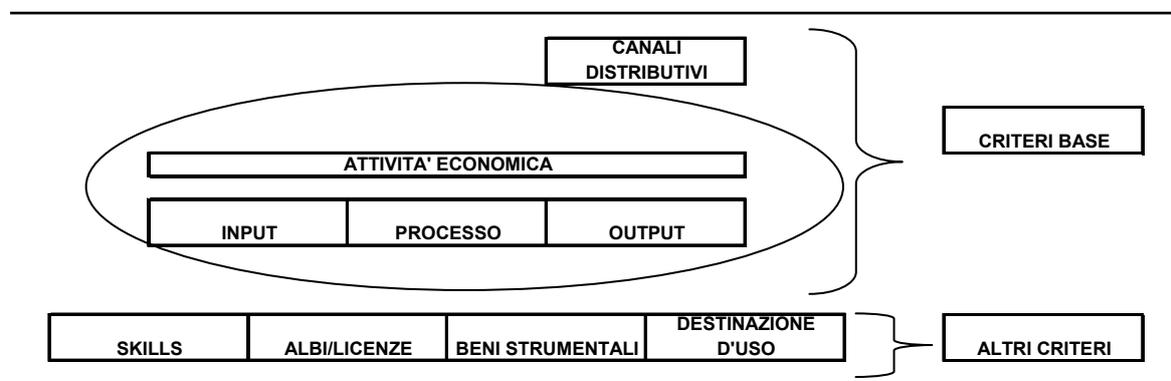
Proseguendo nell'esempio, si può notare come ci sia una semplificazione dello schema con l'individuazione di sei regole deterministiche per l'attribuzione di un codice Ateco sulla base delle modalità di volta in volta considerate (Prospetto 7). Inoltre, ad ogni regola viene automaticamente associata una quota percentuale di fatturato, ottenuta come risultato di una produttrice delle singole variabili utilizzate per la creazione della regola.

Prospetto 7 - Il processo di costruzione delle regole deterministiche - un esempio per il settore manifatturiero (applicazione pratica)



A questo schema base sono stati aggiunti, di volta in volta, altri gruppi di variabili in grado di affinare ulteriormente la codifica come, ad esempio, le materie prime utilizzate (input) o le fasi di produzione svolte (processo) (Prospetto 8).

Prospetto 8 - I criteri di attribuzione



In concreto, l'attribuzione dei codici di attività economica è stata realizzata principalmente basandosi sull'analisi degli *output* (attività svolta o prodotto ottenuto/venduto), allineandosi in questo modo al criterio prevalente seguito nella determinazione teorica delle classi all'interno delle classificazioni Nace e Ateco. Per le stesse situazioni in cui la classificazione deroga a questo principio, la nuova fonte permette di utilizzare dati, ad esempio su materie prime, processi produttivi e mercati di sbocco in modo da poter attribuire un codice aderente alla realtà imprenditoriale in esame:

- analisi degli *input*: necessaria nel caso di prodotti classificati in codici diversi a seconda della materia prima utilizzata (ad esempio, "parti di scarpe");
- analisi dei *processi*: necessaria quando specifiche fasi sono strettamente collegate ad una singola attività economica (ad esempio, "finissaggio");
- analisi dei *mercati/canali distributivi*: per distinguere tra ingrosso e dettaglio e per individuare specifiche forme di commercio (per esempio, intermediari; commercio al dettaglio via mail o via Internet);
- analisi delle *competenze specifiche degli addetti*: ad esempio, figure professionali necessarie allo svolgimento delle attività per distinguere diversi tipi tra i diversi tipi di consulenza (generale, informatica, scolastica);
- analisi dei *beni strumentali*: ad esempio, gondole per individuare trasporti via d'acqua interni;
- analisi di *albi o licenze*: ad esempio, distinguere tassisti (con licenza) da N.C.C. (noleggio con conducente);
- *principio della destinazione d'uso*: ad esempio, per i prodotti in plastica analizzando gli sbocchi produttivi.

In conclusione, il processo è stato condotto seguendo due fondamentali linee guida:

- centralità attribuita alla domanda sui prodotti ottenuti/attività svolta: per ogni SDS è stata costruita prima di tutto una tabella di corrispondenza tra prodotti/attività e codice Ateco;
- scelta delle variabili da usare in una logica di semplificazione: usare, dove possibile, poche variabili in grado comunque di risolvere la maggior parte delle situazioni.

In questo modo si è ottenuto un buon grado di oggettività nel processo di scelta del codice da collegare ad ogni regola e sono stati ottimizzati i tempi di lavoro.

I codici così ottenuti sono indipendenti dal codice di partenza del modello (in classificazione 2002 fino al 2006) e possono avere diversi gradi di dettaglio, da 1 a 5 cifre. Infatti, la descrizione delle attività economiche svolte e quella relativa ai prodotti ottenuti non sono sempre strettamente collegate ad una sola categoria, così, in alcuni casi, il processo di codifica si limita ai livelli più alti della struttura gerarchica della classificazione.

Per alcuni casi limitati, in media dieci ogni anno, la codifica non passa attraverso la costruzione delle regole deterministiche ma per l'utilizzo dei gruppi omogenei (*cluster*) elaborati dall'Agenzia delle Entrate applicando tecniche statistiche quali l'*Analisi in componenti Principali* e la *Cluster Analysis*. Per ogni studio di settore la creazione dei gruppi omogenei si basa sugli aspetti strutturali, cioè gli stessi quadri del modello che sono utilizzati nella procedura di attribuzione del codice Ateco.⁸

Alla fine del processo ad ogni unità rispondente con dati congrui è stato possibile attribuire un insieme di codici di attività economica collegati ad una percentuale di fatturato. Tra questi la scelta del codice di attività principale è stata effettuata applicando il metodo dell'ordine decrescente.

Per il 2008 il 34,5% delle imprese ha fornito dati relativi ad almeno due codici di Attività economica; inoltre, per una quota maggiore del 95% dei casi il processo è riuscito ad arrivare al livello utile di codifica (categoria, cinque cifre della classificazione) (Tavole 1 e 2).

Tavola 1 - Imprese della fonte SDS per numero di codici Ateco attribuiti (valori percentuali)

Codici Ateco attribuiti	Imprese
1	65,5%
2	18,5%
3	6,8%
4	3,8%
5-10	4,7%
11-38	0,5%

Fonte: SDS rielaborazione per la realizzazione del Registro ASIA

Tavola 2 - Imprese della fonte SDS per numero di cifre del codice Ateco attribuito (valori percentuali)

Cifre attribuite all'Ateco principale	Imprese
5 cifre	95,3%
4 cifre	1,6%
3 cifre	1,3%
2 cifre	1,7%
1 cifra	0,1%

Fonte: SDS rielaborazione per la realizzazione del Registro ASIA

Volendo analizzare nel dettaglio la metodologia di individuazione delle regole di codifica, del numero e del tipo di variabili effettivamente considerate, occorre sottolineare come questa sia fortemente dipendente dal macrosettore di riferimento, pertanto nel seguito sarà analizzata separatamente per i quattro macrosettori Manifatturiero (§4.1.1), Servizi (§4.1.2), Commercio (§4.1.3) e Professionisti (§4.1.4).⁹

⁸ Note tecniche Studi di Settore – Agenzia delle Entrate: “La finalità perseguita è di determinare un ‘ricavo/compenso potenziale’ attribuibile ai contribuenti cui si applica lo Studio di Settore tenendo conto non solo di variabili contabili, ma anche di variabili strutturali in grado di influenzare il risultato di un’impresa o di un professionista. A tale scopo, nell’ambito dello studio, vanno individuate le relazioni tra le variabili contabili e le variabili strutturali, per analizzare i diversi modelli organizzativi impiegati nell’espletamento dell’attività... Per segmentare i soggetti oggetto dell’analisi in gruppi omogenei sulla base degli aspetti strutturali, si è ritenuta appropriata una strategia di analisi che combina due tecniche statistiche: una tecnica basata su un approccio di tipo multivariato, che si è configurata come un’analisi fattoriale del tipo *Analyse des données* e nella fattispecie come un’Analisi in Componenti Principali e un procedimento di Cluster Analysis.... La Regressione Multipla è una tecnica statistica che permette di interpolare i dati con un modello statistico-matematico che descrive l’andamento della variabile dipendente in funzione di una serie di variabili indipendenti. La stima della “funzione di ricavo/compenso” è stata effettuata individuando la relazione tra il ricavo/compenso (variabile dipendente) e i dati contabili e strutturali dei contribuenti (variabili indipendenti)”.

⁹ L’elenco degli Studi di Settore con i relativi codici e le descrizioni Ateco con riferimento all’anno 2007 sono consultabili, su richiesta al curatore, nell’Appendice 2. L’appendice è tratta dal sito dell’Agenzia delle Entrate (<http://www.agenziaentrate.gov.it/wps/portal/entrate/home>).

4.1.1 Manifatturiero

I questionari del settore manifatturiero analizzati (anno di riferimento 2008) sono stati 51. I modelli in esame sono certamente i più completi da un punto di vista informativo, perché sono poste domande sulle modalità di svolgimento dell'attività, sul tipo di clientela, sulle fasi svolte, sul tipo di materie prime impiegate e sui beni strumentali utilizzati. Il processo di costruzione delle regole deterministiche per questo settore è certamente quello che più si avvicina al modello teorico presentato in precedenza (§4.1). Il risultato ottenuto in termini di tipo di codifica, distinguendo il caso di codici con valori percentuali del fatturato dalla sola indicazione di tipo qualitativo, insieme con l'indicazione dei blocchi di domande utilizzati per la costruzione delle regole o comunque necessari per completare il processo di codifica, sono sintetizzati, per i questionari del settore manifatturiero, in Prospetto 9. Con “prod/comm” si identifica il gruppo di domande in cui si richiede l'indicazione della parte di fatturato che deriva dall'attività di produzione separatamente da quella derivante dalla commercializzazione di beni che non hanno subito ulteriori lavorazioni.

Prospetto 9 - Questionari degli Studi di Settore per tipo di codifica e domande utilizzate - settore manifatturiero

	STUDIO	Tipo codifica	Numero blocchi utilizzati	blocco1	blocco2	blocco3	blocco4	blocco5
1	TD03U	% fatturato	3	prod/comm	tipo clientela	prodotti		
2	TD04A	% fatturato	4	prod/comm	tipo clientela	prodotti	tipo attività	
3	TD04B	% fatturato	4	prod/comm	tipo clientela	prodotti	tipo attività	
4	TD05U	% fatturato	3	prod/comm	tipo clientela	prodotti		
5	TD09A	cluster	-					
6	TD09B	% fatturato	3	prod/comm	tipo clientela	prodotti		
7	TD11U	% fatturato	4	prod/comm	tipo clientela	prodotti	materie prime	
8	TD15U	% fatturato	4	prod/comm	tipo clientela	prodotti	materie prime	
9	TD16U	% fatturato	3		tipo clientela		modalità lavorazione	specializzazione
10	TD17U	% fatturato	4	prod/comm	tipo clientela	prodotti	materie prime	
11	TD19U	% fatturato	4	prod/comm	tipo clientela	prodotti	fasi	
12	TD20U	% fatturato	5	prod/comm	tipo clientela	prodotti	tipo attività	fasi
13	TD21U	% fatturato	2	prod/comm		prodotti		
14	TD22U	% fatturato	3	prod/comm	tipo clientela	prodotti		
15	TD23U	% fatturato	4	prod/comm	tipo clientela		tipo attività	beni strumentali
16	TD24U	% fatturato	3	prod/comm	tipo clientela	prodotti		
17	TD25U	% fatturato	2	prod/comm	tipo clientela			
18	TD26U	% fatturato	3	prod/comm	tipo clientela	prodotti		
19	TD27U	% fatturato	3	prod/comm	tipo clientela	prodotti		
20	TD28U	% fatturato	3	prod/comm	tipo clientela	prodotti		
21	TD29U	% fatturato	3	prod/comm	tipo clientela	prodotti		
22	TD30U	non ottenuta	-					
23	TD31U	% fatturato	3	prod/comm	tipo clientela	prodotti		
24	TD32U	% fatturato	4	prod/comm	tipo clientela	prodotti	tipo attività	
25	TD33U	non ottenuta	-					
26	TD35U	% fatturato	3	prod/comm	tipo clientela	prodotti		
27	TD36U	% fatturato	4	prod/comm	tipo clientela		materie prime	fasi
28	TD37U	% fatturato	2			prodotti	tipo attività	
29	TD38U	% fatturato	3	prod/comm	tipo clientela	prodotti		
30	TD39U	% fatturato	3	prod/comm	tipo clientela	prodotti		
31	TD40U	% fatturato	4	prod/comm	tipo clientela	prodotti	tipo attività	
32	TD41U	% fatturato	4	prod/comm	tipo clientela	prodotti	tipo attività	

Prospetto 9 segue - Questionari degli Studi di Settore per tipo di codifica e domande utilizzate - settore manifatturiero

STUDIO	Tipo codifica	Numero blocchi utilizzati	blocco1	blocco2	blocco3	blocco4	blocco5	
33	TD42U	% fatturato	4	prod/comm	tipo clientela	prodotti	tipo attività	
34	TD43U	% fatturato	5	prod/comm	tipo clientela	prodotti	tipo attività	fasi
35	TD44U	% fatturato	4	prod/comm	tipo clientela	prodotti	mercato di riferimento	
36	TD45U	% fatturato	3	prod/comm	tipo clientela	prodotti		
37	TD46U	% fatturato	4	prod/comm	tipo clientela	prodotti	mercato di riferimento	
38	TD47U	% fatturato	3	prod/comm	tipo clientela	prodotti		
39	TD49U	% fatturato	3	prod/comm	tipo clientela	prodotti		
40	UD01U	% fatturato	3	prod/comm	tipo clientela	prodotti		
41	UD02U	% fatturato	3	prod/comm	tipo clientela	prodotti		
42	UD06U	% fatturato	2	prod/comm	tipo clientela			
43	UD07A	% fatturato	4	prod/comm	tipo clientela	prodotti	fasi	
44	UD07B	% fatturato	4	prod/comm	tipo clientela	prodotti	fasi	
45	UD08U	% fatturato	7	prod/comm	tipo clientela	prodotti	tipo attività	materie prime
46	UD10B	% fatturato	4	prod/comm	tipo clientela	prodotti	fasi	
47	UD12U	% fatturato	3	prod/comm	tipo clientela	prodotti		
48	UD13U	% fatturato	4	prod/comm	tipo clientela	prodotti	fasi	
49	UD14U	% fatturato	4	prod/comm	tipo clientela	prodotti	fasi	
50	UD18U	% fatturato	4	prod/comm	tipo clientela	prodotti	specializzazione	
51	UD34U	% fatturato	1			prodotti		

Per la quasi totalità dei 51 questionari (48 modelli) è stato possibile ottenere i risultati in termini di codici di attività economica accompagnati dal relativo valore percentuale del fatturato. Solo per due questionari, invece, non è stato possibile applicare la procedura perché le informazioni presenti non sono state ritenute sufficienti per lo sviluppo di regole di codifica indipendenti dai codici Ateco di partenza. Si tratta dei questionari TD30U – riciclaggio e commercio all'ingrosso di rottami e materiali di recupero e TD33U – produzione di metalli preziosi e oggetti di gioielleria e oreficeria. Infine, per lo studio TD09A (prodotti in legno e mobili) sono stati ottenuti risultati solo in termini di codici Ateco principali senza l'indicazione di valori percentuali. In quest'ultimo caso sono stati utilizzati i *cluster* (gruppi omogenei) individuati per fini fiscali dall'Agenzia delle Entrate all'interno di ogni studio di settore. Dell'utilizzo dei gruppi omogenei a supporto dell'individuazione dei codici Ateco si parlerà più diffusamente nel successivo paragrafo dedicato ai servizi.

Per 48 modelli, dunque, è stato possibile utilizzare almeno un gruppo di domande (blocco) con risposte espresse in percentuali di fatturato. In media per ogni questionario sono stati utilizzati circa 3,5 blocchi di domande (con un massimo di 7 per il modello UD08U) e in 40 casi sono presenti le informazioni di base del modello teorico (produzione e/commercializzazione, tipo clientela e prodotti) con cui costruire le regole di codifica.

Nello specifico possono essere segnalate ulteriori particolarità:

- per TD17U (prodotti in gomma e plastica) e UD08U (calzature e parti di calzature) l'uso combinato di informazioni su materie prime utilizzate (input), prodotti e tipo di attività ha consentito anche di applicare il criterio della destinazione d'uso laddove previsto dalla classificazione Ateco; ad esempio, è stato possibile distinguere i prodotti ottenuti in base al materiale utilizzato e per destinazione del bene finale (parti o prodotti finiti);
- per TD03U e TD09B oltre all'uso delle variabili di base è stato necessario integrare l'analisi con i dati sulle materie prime nel primo caso (molitura e lavorazione dei cereali) e sulle fasi della lavorazione per il secondo (taglio e fabbricazione pannelli e imballaggi in legno).

Infine, è da sottolineare come per i modelli più complessi del settore meccanico (TD20U e TD32U che

coinvolgono rispettivamente 26 e 54 codici Ateco) si è in parte tenuto conto della codifica di partenza delle imprese soprattutto in relazione ad attività particolari svolte da un numero ridotto di unità.

4.1.2 Servizi

I questionari analizzati per il settore “servizi” sono 61 ma, come detto in precedenza, in termini di Ateco, sono coinvolte anche imprese classificate nelle costruzioni, nei trasporti e gli agenti di commercio. Questo determina una situazione piuttosto eterogenea pur in presenza comunque di questionari più semplici, almeno come struttura, rispetto al settore manifatturiero (Prospetto 10).

Prospetto 10 - Questionari degli Studi di Settore per tipo di codifica e domande utilizzate - servizi

STUDIO	Tipo codifica	Numero blocchi utilizzati	blocco1	blocco2	blocco3	blocco4
1	SG57U	% fatturato	1	tipo attività		
2	SG96U	% fatturato	2	tipo attività	tipo clientela	
3	SG98U	% fatturato	3	tipo attività	tipo clientela	altre attività
4	SG99U	% fatturato	1	tipo attività		
5	TG31U	% fatturato	2	tipo attività		settore
6	TG33U	% fatturato	1	tipo attività		
7	TG34U	% fatturato	1	tipo attività		
8	TG36U	% fatturato	3	tipo attività	tipo esercizio	beni strumentali
9	TG37U	% fatturato	1	tipo attività		
10	TG38U	% fatturato	1	tipo attività		
11	TG40U	cluster	-			
12	TG41U	% fatturato	1	tipo attività		
13	TG42U	% fatturato	2	tipo attività	tipo clientela	
14	TG46U	% fatturato	1	tipo attività		
15	TG48U	% fatturato	2	tipo attività	tipo clientela	
16	TG50U	% fatturato	1	tipo attività		
17	TG51U	% fatturato	2	tipo attività	tipo input	
18	TG52U	% fatturato	3	tipo attività	tipo input	tipo input
19	TG53U	% fatturato	1	tipo attività		
20	TG54U	% fatturato	3	tipo attività	tipo esercizio	tipo attrezzature
21	TG55U	% fatturato	1	tipo attività		
22	TG58U	% fatturato	1	tipo attività		
23	TG60U	% fatturato	1	tipo attività		
24	TG68U	% fatturato	2	tipo attività	tipo input	
25	TG69U	% fatturato	1	tipo attività		
26	TG70U	% fatturato	1	tipo attività		
27	TG72A	modalità	-			
28	TG72B	% fatturato	2	tipo attività	specializzazione	
29	TG73A	% fatturato	2	tipo attività	tipo input	
30	TG73B	% fatturato	1	tipo attività		
31	TG75U	% fatturato	1	tipo attività		
32	TG76U	% fatturato	3	tipo attività	tipo clientela	specializzazione
33	TG77U	misto	-			
34	TG78U	modalità	-			
35	TG79U	% fatturato	2	tipo attività	beni strumentali	
36	TG81U	% fatturato	1	tipo attività		
37	TG82U	% fatturato	1	tipo attività		
38	TG83U	cluster	-			

Prospetto 10 segue - Questionari degli Studi di Settore per tipo di codifica e domande utilizzate - servizi

STUDIO	Tipo codifica	Numero blocchi utilizzati	blocco1	blocco2	blocco3	blocco4	
39	TG85U	% fatturato	2	tipo attività		tipo servizi	
40	TG87U	% fatturato	2	tipo attività		skills	
41	TG88U	% fatturato	1	tipo attività			
42	TG89U	% fatturato	2	tipo attività		tipo servizi	
43	TG90U	% fatturato	2	tipo attività		specializzazione	
44	TG91U	% fatturato	2	tipo attività		modalità organizzative	
45	TG92U	% fatturato	1	tipo attività			
46	TG93U	% fatturato	2	tipo attività		settore	
47	TG94U	% fatturato	2	tipo attività		settore	
48	TG95U	% fatturato	1	tipo attività			
49	UG39U	% fatturato	2	tipo attività		specializzazione	
50	UG44U	modalità	-				
51	UG61A	% fatturato	3		tipo clientela	prodotti	modalità organizzative
52	UG61B	% fatturato	3		tipo clientela	prodotti	modalità organizzative
53	UG61C	% fatturato	3		tipo clientela	prodotti	modalità organizzative
54	UG61D	% fatturato	3		tipo clientela	prodotti	modalità organizzative
55	UG61E	% fatturato	3		tipo clientela	prodotti	modalità organizzative
56	UG61F	% fatturato	3		tipo clientela	prodotti	modalità organizzative
57	UG61G	% fatturato	3		tipo clientela	prodotti	modalità organizzative
58	UG61H	% fatturato	3		tipo clientela	prodotti	modalità organizzative
59	UG66U	% fatturato	2	tipo attività	tipo clientela		
60	UG67U	% fatturato	2	tipo attività	tipo clientela		
61	UG74U	% fatturato	1	tipo attività			

In questo caso sono 55 i modelli su cui è stato possibile ottenere codici Ateco ordinabili per percentuale di fatturato, con l'uso in media di 1,8 blocchi di domande per SDS. In questi, la domanda maggiormente utilizzata nel processo di codifica è quella sul tipo di attività svolta, espressa in percentuale del fatturato, presente in 47 modelli. A parte questa domanda comune, ogni SDS segue specificità del proprio ambito e di conseguenza i blocchi di domande utilizzate per la costruzione delle regole sono piuttosto eterogenei.

Tra questi 55 studi di settore possono essere segnalati alcuni casi particolari in cui è stato necessario introdurre alcune particolarità nella costruzione delle regole.

Lo studio SG99U (altre attività di servizi, altre attività di supporto alle imprese, altre attività di servizi alla persona), introdotto nel 2006, è rivolto alle imprese che svolgono attività di servizi codificate principalmente nelle categorie residuali (codici Ateco 2007 74.90.9, 82.99.9 e 96.09.0). Nel questionario è presente un elenco di circa 100 attività a cui è associato un codice numerico non riconducibile però a nessuna classificazione adottata dall'Istituto. In questo caso, prima di tutto, è stato necessario assegnare a ciascuna attività indicata il corrispondente codice Ateco. A questo punto sulla base della tabella di corrispondenza creata è stato legato il codice Ateco al relativo valore del fatturato.

Per lo studio TG73A (magazzini di custodia e movimento merci), non riuscendo ad individuare l'ambito di riferimento è stato necessario utilizzare anche i codici di partenza per arrivare ad una codifica a cinque cifre Ateco.

Le regole costruite per il modello TG87U (attività di consulenza) si basano in parte sulle competenze dei lavoratori delle imprese coinvolte (skills) e costituiscono dunque uno degli esempi tipici dell'utilizzo di criteri alternativi a quello dell'output. A questo proposito occorre sottolineare come l'evoluzione dello studio TG81U (noleggio di macchinari edili con o senza operatore) abbia determinato anche un cambio di criterio utilizzato per la codifica. In passato, nella versione SG81U

era presente una domanda (di tipo qualitativo) con cui distinguere il noleggio con e senza operatore (skills), mentre ad oggi è possibile utilizzare la domanda sul tipo di attività svolta con le relative percentuali di fatturato.

Infine, per gli otto studi identificati con la sigla UG61A-H (agenti, rappresentanti e mediatori) è stato necessario costruire prima di tutto una tabella di corrispondenza tra le categorie merceologiche elencate (circa 360) e i rispettivi codici Ateco. Successivamente sono state create le regole deterministiche utilizzando sia i prodotti, ora espressi in termini di codici Ateco, sia le altre informazioni utili (in questo caso, tipo clientela e modalità organizzative).

Per i restanti sei modelli (TG40U, TG72A, TG77U, TG78U, TG83U, UG44U) sono stati ottenuti risultati solo in termini di codici, senza poter usufruire di variabili quantitative di supporto.

In particolare, per TG40U (attività immobiliari) e TG83U (gestione impianti sportivi, palestre e piscine) sono stati utilizzati i cluster utili a fini fiscali per la determinazione del ricavo/compenso potenziale. Come anticipato, l'Agenzia delle Entrate elabora per ogni studio di settore i dati disponibili in base agli aspetti strutturali, cioè gli stessi quadri del modello che sono utilizzati nella procedura di attribuzione del codice Ateco, per la creazione di gruppi omogenei.

Per ogni studio, dunque, all'analisi del contenuto informativo del questionario si affianca lo studio delle Note Tecniche in cui sono descritti i gruppi omogenei ottenuti. L'obiettivo di questa seconda analisi è quello di provare ad assegnare a ciascun cluster un codice Ateco univoco, anche se non sempre completo a 5 cifre, in modo da sfruttare le informazioni già disponibili senza dover creare nuove regole di codifica. In concreto, difficilmente i gruppi resi omogenei per fini fiscali lo sono anche in termini di attività economica. Comunque, anche nei casi in cui lo fossero sarebbe sempre da preferire, se semplice, la creazione di regole basate sulle percentuali di fatturato in modo da applicare il metodo dell'ordine decrescente ottenendo anche informazioni sui codici Ateco secondari. Di conseguenza l'uso dei cluster è stato limitato ai soli casi in cui la costruzione delle regole deterministiche era impossibile o comunque troppo complicata rispetto al dettaglio della codifica ottenibile.

Per i servizi, dunque, i cluster sono stati utilizzati come unico strumento di codifica in due casi. Per TG83U sono state create anche regole ad hoc ad integrazione dei risultati ottenuti con i cluster. Nel secondo caso lo studio di settore TG40U (in precedenza SG40U) ha subito un'evoluzione che ha comportato anche la modifica dei cluster di riferimento mettendo in evidenza, ancora una volta, che gli aggiornamenti dei questionari possono influenzare notevolmente i confronti temporali.

Per gli studi TG72A (trasporto con taxi o tramite noleggio con conducente), TG78U (agenzie di viaggio e tour operator) e UG44U (alberghi, affittacamere e alloggi per studenti) è stato possibile utilizzare solo variabili di tipo qualitativo (modalità). In particolare, in TG72A è stato possibile distinguere sulla base della domanda sul tipo di licenza tra trasporto con taxi e noleggio con conducente, mentre, il tipo di veicolo usato (beni strumentali) ha permesso, in alcuni casi, l'individuazione di più attività economiche senza però dati quantitativi di supporto. Per il questionario TG78U sono disponibili alcuni dati relativi all'ammontare totale dei ricavi (non la percentuale) per tipologia di attività e ai volumi intermediati. I dati sono stati dapprima elaborati per ottenere delle modalità con cui individuare le imprese che svolgono un'unica attività. Per quelle pluri-attività i dati numerici sono serviti anche per ottenere l'ordinamento dei codici Ateco in termini di prevalenza, ma i valori sono comunque da considerare meno attendibili rispetto a quelli ottenuti, per gli altri SDS, sulla base delle percentuali di fatturato. In questo caso, le domande poste nel questionario non sono del tutto coerenti e si riferiscono a valori non sempre pienamente comparabili. Infine, per UG44U è stata utilizzata la domanda di tipo qualitativo (6 modalità) sul tipo di struttura ricettiva gestita.

Per lo studio di settore TG77U (trasporti marittimi, costieri e per vie d'acqua interne) è stato adottato un approccio misto: prima di tutto l'analisi dei beni strumentali (ad esempio, la presenza di gondole) ha permesso l'identificazione di un primo gruppo di unità su cui ottenere una codifica sulla base di modalità (di tipo qualitativo); successivamente, per il resto delle unità, le percentuali di fatturato presenti per tipologia di attività hanno condotto a risultati anche ordinabili.

4.1.3 Commercio

Per i 70 modelli del settore Commercio i dati relativi al tipo di codifica ottenuta e al numero di variabili utilizzate sono sintetizzati nel Prospetto 11.

Prospetto 11 - Questionari degli Studi di Settore per tipo di codifica e domande utilizzate - commercio

STUDIO	Tipo codifica	Numero blocchi utilizzati	blocco1	blocco2	blocco3	blocco4
1	SM47U	% fatturato	1	prodotti		
2	SM81U	% fatturato	2	prodotti	tipo clientela	
3	SM87U	cluster	-			
4	SM88U	cluster	-			
5	TM06A	% fatturato	1	prodotti		
6	TM06B	% fatturato	1	prodotti		
7	TM08U	% fatturato	1	prodotti		
8	TM09A	% fatturato	2	tipo attività		settori
9	TM09B	% fatturato	2	tipo attività		settori
10	TM10U	% fatturato	2	prodotti	tipo clientela	
11	TM11U	% fatturato	2	prodotti	tipo clientela	
12	TM12U	% fatturato	2	prodotti	tipo clientela	
13	TM13U	% fatturato	1	prodotti		
14	TM15B	% fatturato	2	prodotti	tipo clientela	
15	TM16U	% fatturato	2	prodotti	tipo clientela	
16	TM17U	% fatturato	2	prodotti	tipo clientela	
17	TM18A	% fatturato	2	prodotti	tipo clientela	
18	TM18B	% fatturato	2	prodotti	tipo clientela	
19	TM19U	% fatturato	2	prodotti		tipologia vendita
20	TM20U	% fatturato	1	prodotti		
21	TM21A	% fatturato	2	prodotti	tipo clientela	
22	TM21B	% fatturato	2	prodotti	tipo clientela	
23	TM21C	% fatturato	2	prodotti	tipo clientela	
24	TM21D	% fatturato	2	prodotti	tipo clientela	
25	TM21E	% fatturato	2	prodotti	tipo clientela	
26	TM22A	% fatturato	2	prodotti	tipo clientela	
27	TM22B	% fatturato	2	prodotti	tipo clientela	
28	TM22C	% fatturato	2	prodotti	tipo clientela	
29	TM23U	% fatturato	2	prodotti	tipo clientela	
30	TM24U	% fatturato	-	prodotti	tipo clientela	tipologia vendita
31	TM25A	% fatturato	2	prodotti	tipo clientela	
32	TM25B	% fatturato	2	prodotti	tipo clientela	
33	TM29U	% fatturato	1	prodotti		
34	TM30U	% fatturato	1	prodotti		
35	TM31U	% fatturato	2	prodotti	tipo clientela	
36	TM32U	% fatturato	2	prodotti	tipo clientela	
37	TM33U	% fatturato	2	prodotti	tipo clientela	
38	TM34U	% fatturato	2	prodotti	tipo clientela	
39	TM35U	% fatturato	2	prodotti	tipo clientela	
40	TM36U	% fatturato	2	prodotti		tipologia vendita
41	TM37U	% fatturato	2	prodotti		tipologia vendita
42	TM39U	% fatturato	2	prodotti	tipo clientela	
43	TM40B	% fatturato	2	prodotti	tipo clientela	

Prospetto 11 segue - Questionari degli Studi di Settore per tipo di codifica e domande utilizzate - commercio

STUDIO	Tipo codifica	Numero blocchi utilizzati	blocco1	blocco2	blocco3	blocco4
44	TM41U	% fatturato	2	prodotti	tipo clientela	
45	TM42U	% fatturato	2	prodotti	tipo clientela	
46	TM43U	% fatturato	3	prodotti	tipo clientela	servizi offerti
47	TM44U	% fatturato	2	prodotti	tipo clientela	
48	TM45U	% fatturato	2	prodotti	tipo clientela	
49	TM46U	% fatturato	2	prodotti	tipo clientela	
50	TM48U	% fatturato	1	prodotti		
51	TM80U	cluster	-			
52	TM82U	% fatturato	2	prodotti	tipo clientela	
53	TM83U	% fatturato	2	prodotti	tipo clientela	
54	TM84U	% fatturato	3	prodotti	tipo clientela	settori
55	TM85U	% fatturato	1	prodotti		
56	TM86U	% fatturato	3	prodotti	tipo attività	beni strumentali
57	UM01U	misto*	1	prodotti		
58	UM02U	% fatturato	2	prodotti	tipo clientela	
59	UM03A	% fatturato	1			settori
60	UM03B	% fatturato	1			settori
61	UM03C	% fatturato	1			settori
62	UM03D	% fatturato	1			settori
63	UM04U	% fatturato	1	prodotti		
64	UM05U	% fatturato	1	prodotti		
65	UM07U	% fatturato	1	prodotti		
66	UM15A	% fatturato	2	prodotti	tipo clientela	
67	UM27A	% fatturato	1	prodotti		
68	UM27B	% fatturato	1	prodotti		
69	UM28U	% fatturato	2	prodotti		servizi offerti
70	UM40A	% fatturato	2	prodotti	tipo clientela	

Per 66 modelli è stato possibile ottenere dati sui codici Ateco principali e secondari ordinati sulla base delle percentuali di fatturato, mentre in tre casi sono stati utilizzati i cluster. Per un unico studio di settore (UM01U – commercio al dettaglio di prodotti alimentari), invece, è stato necessario seguire un approccio misto combinando variabili qualitative e quantitative.

In media sono stati utilizzati 1,7 blocchi di domande per la costruzione delle regole deterministiche con la presenza congiunta di informazioni sui prodotti venduti e sul tipo di clientela in 40 casi. Solo in 15 casi è stato necessario utilizzare informazioni di diverso tipo (ad esempio, settori e tipologie di vendita).

Da segnalare la necessità di creare tabelle di corrispondenza a partire dalle categorie merceologiche presenti nei cinque modelli TM21A-E (commercio all'ingrosso di prodotti alimentari) e la disponibilità di informazioni per individuare le attività di *e-commerce* per lo studio TM35U (erboristerie).

Discorso a parte merita il modello UM01U particolarmente importante perché interessa il settore della distribuzione al dettaglio (supermercati, discount, minimercati ed esercizi specializzati). La difficoltà maggiore nel trattamento dei dati ha riguardato la necessità di distinguere gli esercizi che operano vendendo prodotti di diverso tipo (la classe Ateco 47.11 – commercio al dettaglio in esercizi non specializzati con prevalenza di prodotti alimentari e bevande) da quelli specializzati (gruppo 47.2 – commercio al dettaglio di prodotti alimentari, bevande e tabacco in esercizi specializzati). Di conseguenza, accanto all'ordinamento dei codici Ateco sulla base delle attività svolte è stato necessario analizzare la variabile "tipologia di esercizio", che presenta quattro modalità: negozio tradizionale, minimercato, supermercato e discount per ciascuna unità locale.

4.1.4 Professionisti

I 24 modelli del settore professionisti sono stati tutti risolti ottenendo codici Ateco ordinati in base alla percentuale sui compensi, fornita principalmente dal blocco di domande sulla tipologia di attività svolta (Prospetto 12). La combinazione dei dati sull'attività svolta e la tipologia di clientela servita è presente in 18 casi, mentre un terzo set di domande (principalmente le aree di specializzazione) è stato analizzato per 16 questionari. In media i blocchi utilizzati sono stati 2,5 per SDS.

La principale particolarità evidenziata dal trattamento di questo tipo di questionari è data dalla necessità di prevedere delle soglie oltre le quali poter assegnare con certezza un codice a 5 cifre. L'esperienza accumulata in questi anni di analisi sul settore professionisti ha portato alla seguente decisione operativa: nel caso di attività svolte con una percentuale superiore all'80% del valore dei compensi si assegna un codice Ateco a cinque cifre; per valori inferiori, in presenza di attività che rientrano comunque nell'ambito di diverse categorie (a parità di classe) ci si limita ad una codifica a quattro cifre (10 modelli).

Prospetto 12 - Questionari degli Studi di Settore per tipo di codifica e domande utilizzate - professionisti

	STUDIO	Tipo codifica	Numero blocchi utilizzati	blocco1	blocco2	blocco3	blocco4
1	SK29U	% compensi	3	tipo attività	tipo clientela	aree specialistiche	
2	TK01U	% compensi	2	tipo attività	tipo clientela		
3	TK08U	% compensi	3	tipo attività	tipo clientela	aree specialistiche	
4	TK16U	% compensi	2	tipo attività	tipo clientela		
5	TK19U	% compensi	3	tipo attività	tipo clientela	tipo prestazioni	
6	TK20U	% compensi	1	tipo attività			
7	TK22U	% compensi	3	tipo attività	tipo clientela	aree intervento	
8	TK56U	% compensi	2	tipo attività	tipo clientela		
9	SK30U	% compensi	3	tipo attività	tipo clientela	aree specialistiche	
10	TK23U	% compensi	3	tipo attività	tipo clientela	aree specialistiche	
11	TK24U	% compensi	2	tipo attività	tipo clientela		
12	TK25U	% compensi	3	tipo attività	tipo clientela	aree specialistiche	
13	UK03U	% compensi	3	tipo attività	tipo clientela	aree specialistiche	
14	UK04U	% compensi	4	tipo attività	tipo clientela	aree specialistiche	modalità organizzativa
15	UK18U	% compensi	3	tipo attività	tipo clientela	aree specialistiche	
16	UK21U	% compensi	2	tipo attività		tipo pazienti	
17	TK26U	% compensi	2	tipo attività		tipo offerta	
18	TK27U	% compensi	3	tipo attività	tipo clientela	profilo professionale	
19	TK28U	% compensi	1	tipo attività			
20	UK02U	% compensi	3	tipo attività	tipo clientela	aree specialistiche	
21	UK06U	% compensi	1	tipo attività			
22	UK17U	% compensi	3	tipo attività	tipo clientela	aree specialistiche	
23	TK10U	% compensi	3	tipo attività	tipo clientela	aree specialistiche	
24	UK05U	% compensi	2	tipo attività	tipo clientela		

Per cinque SDS sono state costruite variabili di tipo qualitativo in grado di orientare l'attribuzione del codice Ateco sulla base di alcune caratteristiche specifiche:

- TK16U (amministrazione condomini e servizi integrati agli edifici) – presenza/assenza di edifici amministrati/gestiti;
- TK25U (consulenza agraria) – presenza/assenza di particolari aree specialistiche;
- UK04U (attività degli studi legali) – presenza/assenza di particolari modalità organizzative;
- TK27U (informatica) – presenza/assenza di particolari profili professionali;
- TK28U (recitazione e regia) – attività svolte da persone fisiche o giuridiche.

4.2 Archivio dichiarazioni annuali delle imposte indirette (IVA Annuale)

In questo archivio sono presenti, in linea generale, i titolari di partita IVA tenuti alla presentazione della dichiarazione annuale. Ai fini dell'attribuzione del codice Ateco, dopo una fase di standardizzazione delle informazioni disponibili, si ottengono i dati sulle singole attività economiche svolte dai dichiaranti in termini di volume d'affari. A partire dal 2007, i modelli sono predisposti utilizzando la nuova classificazione delle attività economiche Ateco 2007.

L'Archivio IVA è, dunque, una fonte corredata da valori numerici, il volume d'affari, utilizzabile come *proxy* del valore aggiunto. Per questo motivo, nella nuova metodologia per l'assegnazione del codice Ateco 2007, la fonte è posta, gerarchicamente, al di sotto degli Studi di Settore ma preferita alle classiche fonti in cui il codice di attività è ricavato da dichiarazioni rese in sede di iscrizione dell'unità.

I codici ottenuti sono stati ordinati in termini di prevalenza ottenendo, per ogni impresa, informazioni sull'attività principale e su quelle secondarie.

In concreto, le prime operazioni svolte nel trattamento dei dati della fonte amministrativa hanno riguardato il controllo di validità e completezza dei codici Ateco 2007 presenti. Per i dati del 2008, il controllo ha riguardato oltre 5,1 milioni di record: tutti i codici sono risultati appartenenti alla classificazione ufficiale e completi a 5 cifre.

L'analisi è proseguita per il sottoinsieme di unità con informazioni presenti in più record (circa 450 mila), potenzialmente riferite a codici di attività economica differenti. I dati di questo sottoinsieme sono relativi a quasi 210 mila imprese (circa il 4% del totale imprese della fornitura). A questo punto, i dati sul volume d'affari riferiti ai singoli codici Ateco sono stati aggregati. Pur in presenza di più record, circa 74 mila unità svolgono comunque un'unica attività economica, mentre per il resto delle imprese è possibile ottenere dati su attività principali e secondarie (Tavola 3).

Tavola 3 - Imprese con informazioni in più record per numero di attività diverse svolte - Anno 2008 (valori percentuali)

Numero di attività svolte	Imprese
1	35,45%
2	59,46%
3-20	5,09%

Fonte: IVA rielaborazione per la realizzazione del Registro ASIA

Per le imprese con più codici di attività è stato applicato il metodo dell'ordine decrescente sulla base del volume d'affari. Un primo giudizio di qualità della fonte può essere dato raggruppando le imprese secondo il numero di attività svolte e le caratteristiche del volume d'affari ottenuto. In Tavola 4 è mostrato il peso percentuale delle diverse tipologie ottenute. Oltre il 97% delle imprese dichiara ai fini Iva una sola attività economica, mentre le multiattività con valori di fatturato positivi sono circa 122 mila, pari al 2,5% del totale.

Tavola 4 - Numero di imprese per attività svolte e caratteristiche del fatturato - Anno 2008 (valori percentuali)

Modalità	Imprese
Imprese con un record e una attività	95,68%
Imprese con più record ma una attività	1,53%
Imprese con più record e più attività tutte con fatturato negativo	0,01%
Imprese con più record e più attività e 1 sola attività con fatturato positivo	0,27%
Imprese con più record e più attività, di cui almeno due con fatturato positivo	2,50%

Fonte: IVA rielaborazione per la realizzazione del Registro ASIA

4.3 Camere di Commercio

Il Registro delle Imprese delle Camere di Commercio è un archivio pubblico in cui i soggetti giuridici sono obbligati per legge all'iscrizione (art. 2188 del Codice Civile). Riguardo il campo di osservazione, il settore agricolo è incluso solo parzialmente mentre per lavoratori autonomi e liberi professionisti non sussiste l'obbligo di iscrizione. Inoltre, è da considerare come le iscrizioni siano un atto dovuto a fini amministrativi e non costituiscano un segnale reale di attività, come invece richiede la definizione di impresa prevista nei regolamenti europei¹⁰.

Sono registrate informazioni di diverso tipo e tra queste, a partire dal 2008, i codici di attività principale e secondaria in classificazione Ateco 2007 della sede e delle unità locali.

Dalla fornitura completa delle CCIAA per il 2008 (oltre 8,1 milioni di record) è stato possibile ottenere i codici Ateco principali e secondari (in classificazione 2007), per circa 6,8 milioni di imprese.

Il primo passo nel trattamento dei dati è stato quello di eliminare i record senza alcuna indicazione di codici Ateco in classificazione 2007. Questo ha portato all'esclusione dall'analisi di circa 500 mila record.

Il trattamento è proseguito in maniera distinta per le unità con dati in un solo record (unilocalizzate, oltre 5,5 milioni di imprese) e per quelle presenti in più record (oltre 770 mila plurilocalizzate, con sedi e unità locali per un totale di oltre 2 milioni di record).

Per il primo sottoinsieme di imprese, sono presenti i dati sul codice di attività principale e, eventualmente, su tre attività secondarie. Nelle Tavole 5 e 6 sono indicati rispettivamente il numero massimo di codici Ateco differenti riferiti alle diverse attività economiche svolte da ciascun codice fiscale e il grado di completezza dei codici disponibili riferito alla attività economica principale.

Tavola 5 - Imprese unilocalizzate per numero di attività economiche differenti svolte - Anno 2008 (valori percentuali)

Numero di attività svolte	Imprese
1	69,15%
2	18,35%
3	6,81%
4	5,70%

Fonte: Registro delle Imprese delle CCIAA: rielaborazione per la realizzazione del Registro ASIA

Tavola 6 - Imprese unilocalizzate per grado di completezza del codice di attività economica principale - Anno 2008 (valori percentuali)

Cifre attribuite all'Ateco principale	Imprese
5	86,28%
4	4,67%
3	7,48%
2	1,57%

Fonte: Registro delle Imprese delle CCIAA: rielaborazione per la realizzazione del Registro ASIA

Anche per il secondo sottoinsieme (plurilocalizzate), sono state individuate le attività svolte e valutato il grado di completezza dei codici Ateco per le unità identificate come *sede*.

¹⁰ Regolamento (CEE) N. 696/93, "l'impresa corrisponde alla più piccola combinazione di unità giuridiche che costituisce un'unità organizzativa per la produzione di beni e servizi che fruisce di una certa autonomia decisionale. In particolare per quanto attiene alla destinazione delle sue risorse correnti. Un'impresa esercita una o più attività in uno o più luoghi. Un'impresa può corrispondere a una sola unità giuridica."

Per circa 28.000 imprese (quasi 58.000 record) la fornitura identifica più unità *sede*. Per queste situazioni, la prima scelta delle unità da considerare è stata fatta tenendo conto delle date di inizio e fine attività presenti nella fonte stessa. In assenza di informazioni sulle date, la preferenza è stata accordata ai record con un maggior livello di completezza per il codice Ateco principale.

Infine, per le imprese senza un record identificativo della *sede* l'attività economica è stata assegnata nei soli casi in cui i codici indicati come principali erano gli stessi per tutti i record.

4.4 Anagrafe Tributaria

L'Anagrafe Tributaria è un archivio gestito dal Ministero delle Finanze in cui sono registrate le persone fisiche e giuridiche tenute alla presentazione della dichiarazione per il pagamento delle imposte dirette o indirette. Per tutte le unità giuridiche, identificate dal codice fiscale, sono registrate le informazioni anagrafiche e gli eventi che esse subiscono nel corso della loro vita, ad esempio, modifiche di domicilio fiscale, cambi di natura giuridica, successioni ereditarie, fusioni e scorpori.

Per ciascun anno di riferimento, l'archivio dell'Anagrafe Tributaria costituisce l'universo da cui partire per l'identificazione delle unità statistiche del Registro Asia. Unità per le quali è necessario procedere alla stima dei caratteri, in modo da determinare l'insieme delle sole imprese attive e appartenenti al campo di osservazione del registro. Per l'anno 2008, le unità giuridiche di riferimento erano circa 11,7 milioni, delle quali circa 11,3 milioni di unità presentano le informazioni sulla variabile attività economica in almeno una fonte: circa 400 mila codici fiscali risultano, dunque, non codificabili.

La fornitura dell'Anagrafe Tributaria è da sempre utilizzata dall'Istat nei processi di attribuzione del codice Ateco per le classificazioni precedenti all'Ateco 2007. Riguardo la nuova classificazione, il processo di adeguamento della fonte è ancora in corso e per l'anno di riferimento 2008, sono stati oltre 2,3 milioni i codici forniti in Ateco 2007.

Come già evidenziato, la metodologia adottata privilegia le fonti con risultati numerici (SDS e IVA) e, in seconda battuta, quelle in cui sono presenti almeno le informazioni sulle attività secondarie (CCIAA). I dati dell'Anagrafe Tributaria, dunque, sono stati utilizzati ad integrazione per le unità non coperte da altre fonti.

5. Metodologia Registro Asia 2008

La metodologia utilizzata per l'attribuzione del codice Ateco in classificazione 2007 per il Registro Asia 2008 rappresenta il punto di arrivo della complessa strategia implementata dall'Istituto, avente come obiettivo la doppia codifica della variabile "attività economica", per i quattro anni di riferimento. Salvo adeguamenti ulteriori delle fonti a disposizione o l'ottenimento di nuove banche dati, la metodologia seguente è da considerarsi a regime.

Per le imprese di grandi dimensioni, è stata svolta una continua attività di aggiornamento dei caratteri da parte di revisori esperti mentre, per quelle operanti in particolari settori economici ci si è avvalsi, almeno in parte, di elenchi speciali provenienti da Autorità di Vigilanza, Ministeri e Associazioni di categoria.

Sul resto delle unità, invece, è stato necessario implementare una metodologia di stima del carattere "attività economica principale" secondo la nuova classificazione sulla base delle fonti disponibili, tenendo conto che l'adeguamento alla nuova classificazione Ateco 2007, per tutti gli enti interessati, è stato un processo continuo nel tempo. Questo ha determinato, per il quadriennio considerato, un continuo sviluppo delle procedure per arrivare, con l'edizione del Registro al 2008, ad un processo "chiuso" sia in termini di fonti utilizzate e metodi di trattamento delle stesse, sia in termini di metodologia di assegnazione alle imprese dell'attività economica prevalente e, ove possibile, di quelle secondarie.

Le strategie adottate per il triennio 2005-2007 sono state, dunque, differenti perché basate su un insieme non completo di fonti. In particolare, per gli anni 2005 e 2006, accanto ai risulta-

ti degli Studi di Settore, è stato usato il software ACTR per codificare automaticamente le descrizioni delle attività presenti nell'Archivio delle Camere di Commercio e sono state predisposte alcune indagini speciali per particolari settori di attività economica. Per il 2007, è stato possibile aggiungere alle fonti a disposizione, per la prima volta, una fonte amministrativa con codici Ateco 2007 (Archivio Iva Annuale – anno 2007). Infine, per il 2008, sono state inserite nell'analisi la fornitura completa in Ateco 2007 delle Camere di Commercio e quella parziale dell'Anagrafe Tributaria.

5.1 Regole di scelta

A partire dai dati del 2008, la metodologia a regime per l'attribuzione del codice Ateco 2007 si basa sulle seguenti fonti amministrative: Studi di Settore, Archivio Iva Annuale, Camere di Commercio e Anagrafe Tributaria. Tutte le fonti citate forniscono i propri dati relativi all'attività economica codificati con la nuova versione della classificazione, con la sola eccezione dell'Anagrafe Tributaria, per la quale si è avuto a disposizione, per il 2008, una fornitura parziale.

Come già sottolineato, la metodologia adottata privilegia le fonti che associano ai codici di attività i risultati numerici ordinabili (SDS e IVA), per poi utilizzare quelle di tipo "dichiarativo" come le CCIAA, dove sono presenti i codici di attività secondarie. La fornitura parziale dell'Anagrafe Tributaria è stata usata ad integrazione, per le sole imprese non coperte dalle altre fonti.

Dopo una prima fase di trattamento individuale delle fonti per ricavare i sottoinsiemi con valori validi da analizzare, è stata creata una *base di dati* costituita dall'integrazione di tutte le informazioni a disposizione, per ogni impresa, sull'attività economica. In questo modo è stato possibile associare a ciascuna impresa i codici, principali e secondari, forniti da ciascuna fonte e, ove presenti, i relativi valori percentuali di fatturato e/o di volume d'affari. Come ulteriore variabile per l'assegnazione di un codice univoco ad ogni unità, è stato inserito anche il codice Ateco attribuito per la stessa unità dal processo all'anno t-1. L'universo di riferimento è costituito dalle unità presenti nell'Archivio dell'Anagrafe Tributaria (circa 11,7 milioni di unità nel 2008) ma circa 400 mila unità sono risultate non codificabili.

Una volta predisposta la base di dati, sono stati creati dei sottoinsiemi per fonti disponibili (Tavola 7). In base alla presenza/assenza di informazioni nelle tre fonti principali (SDS, IVA e CCIAA) e nel Registro Asia con riferimento all'anno precedente (ASIA t-1), si sono ottenute 16 diverse combinazioni che hanno dato origine ai relativi sottoinsiemi di riferimento. Quelli in cui esiste almeno una informazione presente in una fonte sono denominati nel seguito con le lettere da A a Z e contengono circa il 67% del totale delle unità. Di queste, per oltre 3,2 milioni di imprese sono presenti dati nelle due fonti principali (SDS e IVA), da poter confrontare anche con il valore già presente nel Registro, mentre circa il 15% del totale è coperto da una sola fonte (quattro gruppi: E, G, I e Z).

L'insieme delle unità giuridiche per cui non esiste alcuna informazione nelle fonti (sottoinsieme "no ateco07"), rappresenta oltre il 33% dell'universo di riferimento. Su queste unità, ad integrazione, è stato possibile attribuire un codice di classificazione 2007 in base alla fornitura dell'Anagrafe Tributaria, nei soli casi in cui questa era l'unica fonte disponibile (5,4% del totale), mentre si è fatto ricorso a delle matrici di conversione a partire dai codici Ateco 2002 già presenti in archivio in tutti gli altri casi: per il 17% del totale delle unità si tratta di codifiche di tipo 1 a 1, mentre circa l'11% di casi si trova in situazioni di codifica 1 a n, con codici spesso relativi al solo livello di divisione Ateco (Tavola 8).

Tavola 7 - Unità giuridiche per fonti disponibili (valori assoluti e percentuali) - Anno 2008

Sottoinsiemi	Fonti				Unità	
	IVA	ASIA (t-1)	SDS	CCIAA	Valori assoluti	Valori percentuali
A	1	1	1	1	2.615.520	23,16%
B	1	1	1	0	627.785	5,56%
C	1	0	1	0	5.198	0,05%
D	1	0	1	1	7.827	0,07%
E	1	0	0	0	116.889	1,04%
F	1	1	0	0	257.684	2,28%
G	0	1	0	0	309.990	2,75%
H	0	1	1	0	276.428	2,45%
I	0	0	1	0	47.143	0,42%
J	0	0	1	1	50.839	0,45%
K	0	1	1	1	421.116	3,73%
W	1	0	0	1	267.427	2,37%
X	1	1	0	1	953.881	8,45%
Y	0	1	0	1	373.360	3,31%
Z	0	0	0	1	1.210.270	10,72%
No ateco07	0	0	0	0	3.750.718	33,22%
Totale					11.292.075	

Fonte: Registro Asia, Anno 2008

Tavola 8 - Unità giuridiche codificate con Anagrafe Tributaria e matrici di conversione (valori assoluti) - Anno 2008

Fonte	Unità
AT	609.767
1 a N	1.923.571
N a M	1.217.380
Totale	3.750.718

Fonte: Registro Asia, Anno 2008

I sottoinsiemi identificati con le lettere E, G, I e Z hanno la caratteristica di avere un'unica fonte informativa e, quindi, un unico codice Ateco 2007 che, di conseguenza, è stato scelto. Aggiungendo a queste circa 1,7 milioni di unità anche le oltre 3,7 milioni codificate con l'Anagrafe Tributaria e le matrici di conversione, risulta come per il 48,1% del totale non sia stato necessario predisporre ulteriori analisi per attribuire il codice Ateco.

Le restanti unità (oltre 5,8 milioni) sono state analizzate prima di tutto in termini di concordanza tra fonti. Successivamente, in presenza di codici diversi sono state predisposte delle regole di scelta tra fonti discordanti.

La nuova metodologia accorda una maggiore preferenza alle fonti che generano codici Ateco ordinabili sulla base di variabili quantitative, quali SDS e IVA. Inoltre, gli SDS sono posti al vertice della gerarchia delle fonti (*rank*) data la disponibilità di informazioni riconducibili alla definizione stessa di attività economica. Le prime regole di scelta si basano su due criteri di carattere generale:

- uso del rank delle fonti stabilito (SDS>IVA>CCIAA), contestualmente al
- confronto con i codici del Registro all'anno t-1, al fine di garantire stabilità nel tempo alla codifica.

Il rispetto del principio di stabilità implica un cambio di codice Ateco presente nel Registro solo nel caso in cui la fonte scelta fornisca, per l'unità economica in esame, un codice diverso per almeno due anni consecutivi.

Le regole adottate con l'indicazione della fonte scelta per i sottoinsiemi con informazioni provenienti da almeno due fonti (il Registro all'anno t-1 è in questo caso visto come fonte) sono evidenziati in Tavola 9. Le unità con il codice di attività principale attribuito sulla base delle prime regole di scelta sono state oltre 4,9 milioni e rappresentano l'84% delle unità analizzate. Quelle per le quali è stato, invece, necessario predisporre delle ulteriori regole di scelta sono indicate, per relativo sottoinsieme di riferimento, nella Tavola 10.

Tavola 9 - Unità risolte con l'applicazione del primo gruppo di regole di scelta (valori assoluti e percentuali), regola applicata e relativa fonte scelta

Sottoinsieme	Unità risolte		Totale unità analizzate	Regole di scelta adottate	Fonte scelta
	Valori assoluti	Valori percentuali			
A	2.117.753	80,97%	2.615.520	ASIA(T-1)=SDS	SDS
B	515.327	82,09%	627.785	ASIA(T-1)=SDS	SDS
C	3.100	59,64%	5.198	SDS=IVA	SDS
D	4.607	58,86%	7.827	SDS=IVA	SDS
F	226.442	87,88%	257.684	ASIA(T-1)=IVA	IVA
H	225.403	81,54%	276.428	ASIA(T-1)=SDS	SDS
J	50.839	100,00%	50.839	SDS	SDS
K	339.667	80,66%	421.116	ASIA(T-1)=SDS	SDS
W	267.427	100,00%	267.427	IVA	IVA
X	807.154	84,62%	953.881	ASIA(T-1)=IVA	IVA
Y	373.360	100,00%	373.360	ASIA(T-1)	ASIA
Totale	4.931.079	84,19%	5.857.065		

Fonte: Registro Asia, Anno 2008

Tavola 10 - Unità residue da trattare (valori assoluti e percentuali)

Sottoinsieme	Unità da trattare		Totale unità analizzate
	Valori assoluti	Valori percentuali	
A	497.767	19,03%	2.615.520
B	112.458	17,91%	627.785
C	2.098	40,36%	5.198
D	3.220	41,14%	7.827
F	31.242	12,12%	257.684
H	51.025	18,46%	276.428
J	-	-	50.839
K	81.449	19,34%	421.116
W	-	-	267.427
X	146.727	15,38%	953.881
Y	-	-	373.360
Totale	925.986	15,81%	5.857.065

Fonte: Registro Asia, Anno 2008

Per i sottoinsiemi A e B, quelli cioè dove sono presenti contemporaneamente i dati di SDS, IVA e ASIA (t-1) ha prevalso prima di tutto una regola in grado di garantire stabilità nel tempo alla codifica: se il codice degli SDS è uguale a quello presente nel Registro t-1 allora è da preferire indipendentemente dal codice IVA. In questo modo, sono stati assegnati i codici di attività prevalente ad oltre l'80% delle unità presenti nei due sottoinsiemi. Per oltre 610 mila unità (Tavola 10) si è, invece, registrata una mancata coerenza tra SDS e ASIA (t-1) che si traduce nella necessità di attribuire in altro modo il codice di attività prevalente.

I sottoinsiemi F e X sono caratterizzati dall'assenza della fonte SDS mentre per quelli H e K non si dispone del codice di provenienza IVA. In questi casi, le regole adottate garantiscono coe-

renza tra la fonte disponibile all'anno t e il valore del Registro dell'anno precedente, attribuendo in questo modo i codici Ateco principali a circa 1,6 milioni di unità. Sempre per questi gruppi restano ancora da trattare, per ottenere l'assegnazione del codice, poco più di 310 mila unità.

Nei casi in cui non è presente alcun dato nel Registro all'anno t-1, si è scelto di attribuire automaticamente il codice Ateco di provenienza IVA o SDS rispetto a quello delle CCIAA (sottoinsiemi J e W). Invece, la contemporanea presenza di IVA e SDS in assenza di ASIA (t-1), ha determinato l'assegnazione del codice Ateco solo nei casi di uguaglianza tra le fonti. Infine, per il gruppo Y (ASIA t-1 e CCIAA), sulla base del principio di stabilità, è stato confermato il codice già presente nel Registro.

Per otto sottoinsiemi è stato necessario determinare ulteriori criteri di scelta. Da sottolineare, comunque, che la metodologia adottata permette di trattare contestualmente i seguenti gruppi: A e B; C e D; F e X; H e K.

In generale, il secondo sottoinsieme di regole di scelta del codice Ateco principale, si basa sulle seguenti decisioni (Tavola 11):

- completamento dei codici SDS a cui le procedure hanno assegnato 2, 3 o 4 digit:
 - viene confermato, se presente, il codice ASIA t-1 concorde rispettivamente in termini di divisioni, gruppi e classi alla fonte SDS;
 - se non è presente il dato t-1, viene scelta tra le fonti IVA e CCIAA quella coerente (per i digit disponibili) con SDS.
- scelta di IVA e CCIAA per le situazioni in cui gli SDS sono considerati di minore qualità (codifiche incomplete o basate su cluster e modalità):
 - se presente il codice ASIA t-1 viene scelto il codice che garantisce stabilità, altrimenti, il giudizio di coerenza è fatto solo sulle fonti disponibili.
- considerare lo stato di attività: per la prima volta nel processo di scelta viene usata un'informazione "terza" rispetto ai codici Ateco; il criterio adottato è quello di preferire il codice legato all'anno in cui l'impresa si attiva (o riattiva) rispetto a quelli eventualmente già registrati:
 - se l'unità si attiva nell'anno in corso, viene scelto automaticamente il codice SDS (indipendentemente dalla presenza di altre fonti o di dati nel Registro) (gruppi A, B, C, D, H e K);
 - se l'unità si attiva nell'anno in corso e non sono presenti gli SDS, si sceglie il codice IVA (anche se non coerente con ASIA t-1) (gruppi F e X).

Tavola 11 - Unità risolte con l'applicazione del secondo gruppo di regole di scelta (valori assoluti e percentuali), regola applicata e relativa fonte scelta

Regola di scelta adottata	Sottoinsiemi			
	A e B	C e D	F e X	H e K
Completamenti del codice SDS	19,10%	19,60%	-	18,60%
Codici altre fonti (SDS minore qualità)	8,80%	4,90%	-	8,90%
Codici SDS	3,80%	75,50%	-	17,00%
Codici IVA	-	-	24,10%	-
Totale risolte	31,70%	100,00%	24,10%	44,50%
Ancora da trattare	68,30%	-	75,90%	55,50%
Totale Unità	610.225	5.318	177.969	132.474

Fonte: Registro Asia, Anno 2008

L'applicazione di questo secondo gruppo di regole è stata sufficiente per risolvere completamente i sottoinsiemi C e D, mentre per il resto è ancora necessario proseguire nella ricerca di modalità di attribuzione del codici Ateco. In pratica, a questo punto, le unità a cui ancora non risulta assegnato un codice Ateco sono quelle per le quali le fonti a disposizione sono discordanti.

Per i sottoinsiemi A e B sono oltre 410 mila le unità ancora da risolvere e si verificano le tre seguenti situazioni:

- codice IVA uguale a codice ASIA t-1 e diverso da quello SDS;
- codice IVA uguale a SDS e diverso da quello presente nel Registro (ASIA t-1);
- tre codici diversi per ASIA t-1, IVA e SDS.

L'ulteriore decisione presa è stata quella di estendere il principio di stabilità riferendola ad ogni singola fonte. I risultati di questa ulteriore verifica sono presentati in Figura 13. In testata è evidenziato il confronto tra le fonti all'anno t, mentre, in fiancata è inserito il giudizio di coerenza per ciascuna fonte rispetto all'anno precedente. In generale, la regola adottata è stata quella di scegliere un codice, anche diverso da quello presente nel Registro, nei casi in cui fosse stato confermato dalla stessa fonte per almeno due anni.

Nel Prospetto 13 sono illustrate le diverse modalità (9) con cui possono essere sintetizzati i giudizi di coerenza "interni" alle fonti. Con "ASIA" e "IVA=ASIA" sono indicate le celle per le quali, alla fine del processo è stato scelto (confermato) il codice dell'anno precedente (circa il 59% dei casi).

Prospetto 13 - Sottoinsiemi A e B - applicazione regole specifiche di scelta

Confronto "interno" per SDS e IVA con l'anno t-1	Confronto tra fonti (anno t)		
	IVA=ASIA e diverso da SDS	IVA=SDS e diverso da ASIA	ASIA diverso da IVA diverso da SDS
1 - no fonti in t-1	ASIA	ASIA	ASIA
2 - solo SDS (t-1) e SDS (t-1) ≠ SDS (t)	ASIA	ASIA	ASIA
3 - solo SDS (t-1) e SDS (t-1) = SDS (t)	SDS	SDS	SDS
4 - solo IVA (t-1) e IVA (t-1) ≠ IVA (t)	IVA=ASIA	ASIA	ASIA
5 - solo IVA (t-1) e IVA (t-1) = IVA (t)	IVA=ASIA	IVA=SDS	IVA
6 - SDS(t)=SDS(t-1) e IVA(t) = IVA(t-1)	SDS	IVA=SDS	SDS
7 - SDS(t)=SDS(t-1) e IVA(t) ≠ IVA(t-1)	SDS	IVA=SDS	SDS
8 - SDS(t) ≠ SDS(t-1) e IVA(t) = IVA(t-1)	IVA=ASIA	IVA=SDS	ASIA
9 - SDS(t) ≠ SDS(t-1) e IVA(t) ≠ IVA(t-1)	IVA=ASIA	IVA=SDS	ASIA

Per le modalità 1, 2, e 4 (*no fonti in t-1; solo SDS (t-1) e SDS (t-1) ≠ SDS (t); solo IVA (t-1) e IVA (t-1) ≠ IVA (t)*) non si hanno a disposizione fonti concordi per due anni e, di conseguenza, è stato confermato il Registro. Uno stesso codice presente in due anni consecutivi per una delle fonti a disposizione è stato scelto indipendentemente dal rank (*riga 5 per IVA e riga 3 per SDS*). Per la modalità 6 - *SDS(t)=SDS(t-1) e IVA(t)=IVA(t-1)*, le regole di base della metodologia hanno determinato il cambio di codice Ateco rispetto all'anno precedente.

Per le ultime tre condizioni sono stati necessari ulteriori approfondimenti:

- nei casi in cui *SDS(t)=SDS(t-1) e IVA(t) ≠ IVA(t-1)* ha prevalso il principio di gerarchia tra fonti: la stabilità degli SDS determina il cambiamento nel Registro;
- la situazione *SDS(t) ≠ SDS(t-1) e IVA(t) = IVA(t-1)* determina il cambiamento solo se gli SDS sono allineati al valore (stabile) di provenienza IVA;
- infine, quando *SDS(t) ≠ SDS(t-1) e IVA(t) ≠ IVA(t-1)* si deroga al principio della stabilità solo quando le due fonti (SDS e IVA) contemporaneamente convergono verso lo stesso valore.

Per i sottoinsiemi F e X (circa 135 mila unità da risolvere) non è presente il dato degli Studi di Settore e, di conseguenza, la prima fonte in ordine gerarchico è costituita dall'IVA. Ai fini della creazione delle regole di scelta, è stata inserita nell'analisi anche la fonte CCIAA. Esaminando la stabilità della fonte, il codice IVA è stato scelto in circa l'80% dei casi.

Infine, per i sottoinsiemi H e K (oltre 73 mila unità residue) si rileva la presenza della fonte SDS ma non di IVA. In questo caso, nel processo di scelta sono intervenute valutazioni che hanno tenuto conto sia del codice CCIAA, sia della stabilità della fonte SDS. Alla fine del processo, a poco più del 38% delle unità, è stato attribuito il codice SDS.

5.2 Analisi per flussi

I metodi di scelta presentati nel paragrafo precedente determinano l'assegnazione del codice Ateco principale di provenienza "amministrativa". I risultati così ottenuti rappresentano una prima base da cui partono delle analisi successive sia sul totale delle unità sia, più nel dettaglio, sul sottoinsieme delle imprese considerate attive e in campo per il Registro Asia nell'anno di riferimento. Inoltre, come già accennato, sulle imprese di grandi dimensioni e su quelle che svolgono particolari attività, l'attribuzione dell'Ateco principale viene comunque validata o da revisori esperti, sulla base anche dei dati statistici a disposizione, e/o da regole di check.

In una situazione in continuo divenire, come quella presentata nel documento, con un'attività di doppia codifica basata su fonti diverse a disposizione per i quattro anni considerati, l'attribuzione del codice Ateco non può prescindere da ulteriori analisi relative, in particolare, ai cambiamenti strutturali rispetto all'anno precedente. Un primo confronto, necessario per completare la fase che può essere definita di attribuzione da fonte amministrativa, è dunque orientato alla valutazione dell'entità delle variazioni di consistenza dei singoli codici Ateco.

Il principio di stabilità presentato nel paragrafo precedente trova la sua base sia nelle *Introductory Guidelines to Nace Rev. 2* ("*Whenever a unit performs two activities both contributing to around 50% of the value added, a stability rule has been established in order to avoid frequent changes no reflecting a substantial change in the economic reality*"), sia nel *Business Register Recommendations Manual* che nell'edizione 2010 par. 17.19 raccomanda l'uso di regole di stabilità "*whereby the change in the principal activity should be made when the current one had been accounting for less than 50 % of the value added for at least two years*".

Si nota, dunque, che il principio di stabilità è da applicare essenzialmente alle situazioni in cui le imprese svolgono più attività economiche e si ha a disposizione il valore aggiunto, o quantomeno un dato *proxy* quale il fatturato, in grado di permettere l'ordinamento dei codici. L'applicazione del principio sulle fonti SDS e IVA presentato nel precedente paragrafo va in parte verso questa direzione.

Accanto a questa situazione "classica", con la presenza di attività principali e secondarie valutabili in termini percentuali, è da considerare, comunque, che la disponibilità di fonti diverse tra un anno e l'altro può generare dei flussi, più o meno consistenti, tra diversi codici Ateco. Ulteriori complicazioni derivano anche dal fatto che ciascuna amministrazione ha svolto le operazioni di passaggio tra Ateco 2002 e Ateco 2007 in maniera autonoma e, soprattutto per le trascodifiche 1 a n, è possibile che i dati di input siano diversi per via del cambio di classificazione. Inoltre, l'evoluzione dei modelli degli Studi di Settore può aver determinato un aggiornamento nel processo di creazione delle regole deterministiche, in grado di generare dei passaggi consistenti tra codici Ateco da tenere sotto controllo. Sulla base di queste considerazioni è stato, dunque, necessario introdurre le ulteriori analisi con l'obiettivo di riuscire a distinguere i passaggi tra codici Ateco da ritenere ammissibili, interpretabili come un cambio sostanziale nella realtà economica, rispetto a quelli derivanti da altri effetti.

Analisi di questo tipo erano state già svolte sulla classificazione Ateco 2002 (Consalvi e Garofalo, 2000) con l'obiettivo non solo di individuare i cambi "reali" ma anche di predisporre regole automatiche con cui correggere eventuali passaggi non ammissibili. Parte di queste regole sono poi confluite nel piano di check per la parte che riguarda Ateco 2002 in relazione ai confronti tra l'anno di riferimento (t) e quello precedente (t-1). Riguardo la nuova versione della classificazione, invece, sono stati usati degli indicatori per evidenziare alcune aree problematiche su cui svolgere ulteriori controlli, mentre restano ancora da elaborare delle nuove regole automatiche di controllo e correzione da aggiungere a quelle già presenti nel piano di check.

Pur con differenze in termini di obiettivi e di fonti disponibili, gli indicatori utilizzati in passato possono costituire un'opportuna base per predisporre ulteriori verifiche sui passaggi di Ateco particolarmente rilevanti. Da sottolineare, però, che l'individuazione dei valori discriminanti per gli indicatori, per selezionare i flussi "anomali" e la predisposizione di regole automatiche di correzione, necessitano chiaramente della disponibilità di dati in serie storica. Ad oggi, dunque, la formalizzazione delle regole di stabilità è in una fase ancora embrionale e gli indicatori sono stati utilizzati al solo fine di evidenziare i flussi principali.

La base per la costruzione degli indici di dissimilarità è costituita dalle tabelle di contingenza a doppia entrata per ciascuna coppia di codici Ateco. Nel Prospetto 14 sono considerati due codici di attività economica, identificati dalle lettere A e B, con l'indicazione delle frequenze assolute (n_{aa} , n_{ab} , n_{ba} , n_{bb}), per gli anni di riferimento t e $t-1$. I valori n_{ab} e n_{ba} identificano i passaggi dal codice A a B e viceversa rilevati all'anno t rispetto all'anno precedente.

L'insieme di queste tabelle di contingenza per tutte le coppie di codici Ateco costituisce la base di partenza per la costruzione di diversi indicatori.

Prospetto 14 - Tabelle di contingenza a doppia entrata

Codice Ateco- anno t-1	Codice Ateco anno t		
	A	B	Totale
A	n_{aa}	n_{ab}	$n_{a.}$
B	n_{ba}	n_{bb}	$n_{.b}$
Totale	$n_{.a}$	$n_{.b}$	$n_{..}$

In passato un primo indicatore utilizzato è stato il seguente:

$$I_{(a,b)} = \frac{\left| \frac{n_{ab}}{n_{ab} + n_{aa}} - \frac{n_{ba}}{n_{ba} + n_{bb}} \right|}{\max \left\{ \frac{n_{ab}}{n_{ab} + n_{aa}}, \frac{n_{ba}}{n_{ba} + n_{bb}} \right\}} \quad (1)$$

dove la differenza al numeratore misura la differenza relativa tra le frequenze dei cambi di Ateco da un anno all'altro, mentre l'espressione al denominatore permette all'indicatore di soddisfare la condizione $0 \leq I_{(a,b)} \leq 1$. L'ipotesi è che i cambi di tipo accidentale siano identificabili con valori prossimi allo 0, perché in qualche modo i flussi tra i due codici coinvolti si compensano. Di contro, valori prossimi all'unità sono in qualche modo indicativi di flussi rilevanti, in particolare, in una direzione. Sono queste ultime le aree che interessa esaminare per capire se effettivamente passaggi così consistenti siano cambi reali nell'economia o derivino da problemi di scelta tra fonti o ancora nascondano situazioni di cambio di prevalenza tra codici diversi.

L'impatto dei passaggi tra i due codici può comunque essere misurato anche nel modo seguente:

$$I'_{(a,b)} = \frac{n_{ab} + n_{ba}}{n_{aa} + n_{bb}} \quad (2)$$

dove a valori elevati dell'indice corrispondono situazioni con alta frequenza di cambiamenti.

Sulla base dei risultati ottenuti per i due indicatori, per diversi livelli di dettaglio della classificazione, sono state effettuate ulteriori analisi e verifiche su gruppi, classi e categorie Ateco 2007 che in alcuni casi hanno determinato delle modifiche nel criterio di scelta e, di conseguenza, nella fonte utilizzata.

A titolo di esempio si illustrano qui di seguito due situazioni che inglobano i principali problemi che determinano i flussi di Ateco, cioè, la scelta tra fonti, le novità della nuova classificazione e la presenza di attività principali e secondarie. La prima riguarda i codici "residuali" delle attività di servizi, mentre, la seconda è riferita alle attività di smaltimento e trattamento rifiuti.

Tra le attività di servizi, nella vecchia classificazione Ateco 2002, in due codici residuali erano classificate tutte quelle unità non altrove classificabili: 74.87.8 – altre attività di servizi alle imprese

n.c.a. (non classificato altrove), e 93.05.0 – altri servizi alle famiglie. Il cambio di classificazione ha determinato per la prima attività una scomposizione che in larga misura riguarda due codici: il primo 74.90.9 – altre attività professionali n.c.a., rappresenta il codice residuale della nuova sezione M (attività professionali, scientifiche e tecniche) e il secondo 82.99.9 – altri servizi di sostegno alle imprese n.c.a., costituisce la categoria generica della nuova sezione N (noleggio, agenzie di viaggio, servizi di supporto alle imprese). Il vecchio codice 93.05.0 corrisponde, invece, ad un'unica categoria (96.09.0 – attività di servizi per la persona n.c.a.) e presenta solo una diversa formulazione della descrizione dell'attività svolta. In definitiva, le attività di tipo residuale dei servizi sono potenzialmente codificate in tre differenti contesti ma in concreto la possibilità di distinguere i differenti ambiti a cui i servizi sono rivolti, attività professionali, non professionali e individui, è certamente difficile da ottenere. Il fatto di provenire poi da un unico codice della versione precedente della classificazione, per quanto riguarda i servizi alle imprese, complica ulteriormente le cose perché sicuramente ogni ente interessato dal cambio di nomenclatura ha applicato regole di trascodifica differenti. Inoltre, lo studio di settore che coinvolge queste attività è unico (SG99U) e pur presentando, come visto nel paragrafo 1.1.2, un elenco di circa 100 attività, in gran parte classificabili con codici Ateco diversi da quelli residuali, determina comunque un'allocazione consistente di imprese nei tre codici in esame (più un'altra categoria residuale la 63.99.0 - Altre attività dei servizi di informazione n.c.a. che ha comunque origine dal vecchio 74.87.8).

Da queste conoscenze teoriche a priori si è passati ad un'analisi dei risultati degli indicatori sulla base delle tabelle di frequenza a doppia entrata che possono essere costruite (Tavola 12¹¹) per i tre codici Ateco.

Tavola 12 - Frequenze a doppia entrata per le generiche attività di servizi

Ateco 2007 – anno t-1	Ateco 2007 – anno t		
	82.99.9	96.09.0	Totale
82.99.9	20.363	7.536	27.899
96.09.0	280	8.043	8.323
Totale	20.643	15.579	36.222

Ateco 2007 – anno t-1	Ateco 2007 – anno t		
	74.90.9	82.99.9	Totale
74.90.9	46.095	215	46.310
82.99.9	13.537	20.363	33.900
Totale	59.632	20.578	80.210

Ateco 2007 – anno t-1	Ateco 2007 – anno t		
	74.90.9	96.09.0	Totale
74.90.9	46.095	275	46.370
96.09.0	508	8.043	8.551
Totale	46.603	8.318	54.921

I risultati degli indicatori sono presentati nella successiva Tavola 13.

Tavola 13 - Indicatori esempio categorie generiche attività di servizi

Codici Ateco		Imprese				Indicatori	
A	B	Nab	Nba	Naa	Nbb	I'(a,b)	I(a,b)
82.99.9	96.09.0	7.536	280	20.363	8.043	0,28	0,8755
74.90.9	82.99.9	215	13.537	46.095	20.363	0,21	0,9884
96.09.0	74.90.9	508	275	8.043	46.095	0,01	0,9002

¹¹ I dati si riferiscono all'anno 2008 e alla versione provvisoria del Registro Asia rilasciato, al solo uso interno, nel mese di gennaio 2010.

Riguardo l'indicatore $I_{(a,b)}$ i valori molto prossimi a 1 denotano una situazione in cui è presente un flusso rilevante in particolare in una direzione, mentre, meno immediato è il giudizio su $I'_{(a,b)}$, non essendo un indice di tipo relativo. I dati empirici ricavati dal passato per Ateco 2002 (necessità per l'indice di superare il valore di 0,0019) non sono utilizzabili come valori soglia ma possono aiutare per avere un metro di giudizio. Valori superiori allo 0,2 sono certamente alti e un semplice rapporto dei dati non presenti sulla diagonale principale rispetto al totale mostra immediatamente come l'entità dei cambi di Ateco incida per circa il 22% per i passaggi dal codice 82.99.9 verso il 96.09.0 e per circa il 17% per quelli tra 74.90.9 e 82.99.9.

La decisione operativa presa, sulla base delle conoscenze a priori e delle successive analisi svolte, è stata quella di annullare i flussi, in ogni direzione per un totale di sei combinazioni, tra i codici Ateco 2007 74.90.9, 82.99.9 e 96.09.0 perché comunque non utili ai fini della rappresentazione complessiva del sistema economico.

Il secondo esempio riguarda una delle principali novità della nuova classificazione Ateco. Nella versione corrente della nomenclatura, infatti, le attività di smaltimento e trattamento rifiuti sono inserite nella nuova sezione E (fornitura di acqua; reti fognarie, attività di gestione dei rifiuti e risanamento) mentre in passato erano codificate nel gruppo 90.0 (smaltimento dei rifiuti solidi, delle acque di scarico e simili) della sezione O (altri servizi pubblici, sociali e personali). Inoltre, nel passaggio nel cambio di classificazione la categoria 90.02.0 – raccolta e smaltimento dei rifiuti solidi è stata scomposta in:

- 38.11.0 – raccolta di rifiuti non pericolosi;
- 38.12.0 – raccolta di rifiuti pericolosi solidi e non solidi;
- 38.21.0 – trattamento e smaltimento di rifiuti non pericolosi;
- 38.22.0 – trattamento e smaltimento di rifiuti pericolosi.

Alla complicazione di aver separato la fase di raccolta da quelle di trattamento e smaltimento, spesso svolte congiuntamente dalle stesse imprese, si è aggiunta anche l'ulteriore distinzione tra rifiuti pericolosi e non pericolosi. Eventuali flussi tra codici relativi al passaggio a livello di tre cifre Ateco (raccolta invece di trattamento e viceversa) potrebbero essere dovuti alla presenza di più attività svolte all'interno delle singole unità. Movimenti interni ai gruppi, invece, in teoria potrebbero essere indicativi un'errata classificazione, precedente o attuale.

In base ai flussi realmente riscontrati, per verificare le ipotesi di partenza, sono state predisposte le tabelle di frequenza e si sono ottenuti i due indicatori scelti (Tavola 14).

Tavola 14 - Indicatori esempio trattamento e smaltimento rifiuti

Codici Ateco		Imprese				Indicatori	
A	B	Nab	Nba	Naa	Nbb	$I'_{(a,b)}$	$I_{(a,b)}$
38.11.00	38.12.00	73	0	1.114	20	0,06	1
38.21.00	38.22.00	8	76	185	117	0,28	0,8947
38.21.00	38.11.00	131	111	1.114	185	0,19	0,7194
38.22.00	38.12.00	1	56	20	117	0,42	0,8529

Anche in questo caso, i risultati degli indicatori risultano alti, confermando l'ipotesi iniziale che le interrelazioni tra i quattro codici fossero piuttosto forti. Nella prima riga si può notare come tra 38.11.0 e 38.12.0 ci siano movimenti in una sola direzione (da raccolta rifiuti non pericolosi a pericolosi) e questo determina il valore massimo nell'indice $I_{(a,b)}$.

Le ulteriori analisi svolte sui quattro flussi analizzati hanno portato all'applicazione delle seguenti regole operative: i flussi tra gruppi Ateco non sono stati accettati, confermando i valori dell'anno precedente, perché il problema è quello della prevalenza tra le due attività; i flussi interni ai gruppi (da non pericolosi a pericolosi e viceversa) sono stati accettati dopo aver verificato che le fonti scelte presentano il nuovo valore da almeno due anni. In base alla metodologia adottata, comunque, nell'anno t+1 i flussi del primo tipo, se confermati, saranno accettati determinando un cambio di Ateco prevalente.

In base alle conoscenze a priori sulle fonti e sulla nuova versione della classificazione e dei ri-

sultati degli indicatori proposti, nel processo di creazione del Registro Asia 2008 sono state analizzate diverse situazioni con flussi tra codici particolarmente rilevanti. Oltre gli esempi proposti, altre specifiche decisioni e regole operative sono state prese per otto divisioni e 22 gruppi Ateco. I controlli effettuati sui codici Ateco sono comunque ulteriori a quelli già previsti eventualmente dalle regole di check, dalle verifiche puntuali sulle grandi imprese e dalle situazioni potenzialmente problematiche su altre variabili chiave del Registro.

In generale, i risultati degli indicatori accanto ad un'altra serie di considerazioni su particolari attività economiche, hanno evidenziato le seguenti problematiche:

- rilevanti flussi interni alla stessa divisione o verso categorie generiche/residuali;
- incremento/decremento notevole, in particolare, da/verso altre attività specifiche;
- nuove entrate: questo significa registrare un notevole incremento di imprese nel settore in esame dovuto principalmente ad unità che si attivano nell'anno in corso;
- attività "particolari" per il Registro: per "particolari" si intendono sia i codici Ateco che, a qualsiasi livello, sono considerati *rari* (poche unità registrate nel settore) o *inesistenti* in riferimento al campo di osservazione considerato, sia le attività soggette a particolari autorizzazioni (elenchi speciali).

Le soluzioni adottate per le ultime due tipologie hanno determinato il controllo puntuale delle unità coinvolte, mentre, per i casi di flussi interni alla divisione, in genere, si è preferito mantenere il codice all'anno t-1. Decisioni *ad hoc* sono state invece adottate per gli incrementi/decrementi valutando principalmente i flussi principali da cui aveva origine l'effetto.

Infine, problemi specifici si sono riscontrati per via di modifiche sostanziali della fonte di riferimento (Studi di Settore) o per l'attribuzione del codice prevalente per le unità della sezione "agricoltura" che svolgono anche l'attività di agriturismo.

6. I Registri Asia 2005-2008

Gli strumenti utilizzati per il quadriennio 2005-2008 per ottenere la codifica in Ateco 2007 sono elencati in Tavola 15. Gli elementi comuni sono rappresentati dagli aggiornamenti da parte dei revisori sulle grandi imprese (ON LINE) e dall'uso della fonte Studi di Settore. Nella categoria residuale "ALTRI METODI" sono inclusi anche le trascodifiche 1 a n risolte con l'ausilio di matrici di conversione. Lo sviluppo della metodologia ha portato ad una drastica riduzione del peso dei metodi macro (coefficienti di conversione) che dall'oltre 10%, in termini di imprese, degli anni 2005 e 2006 è passato a poco più del 4% per le imprese e a circa il 3% riguardo gli addetti. A partire dal 2008, inoltre, non è più significativo il ruolo delle trascodifiche 1 a 1, così come risulta fortemente ridotto l'impatto delle liste speciali, in genere fornite da Enti di previdenza (cfr. par. 2.3.1 tab. 4 e 5), utilizzate per particolari settori.

Gli SDS sono la principale fonte di aggiornamento per oltre il 60% delle imprese, con un peso in termini di addetti superiore al 40%, in tutti gli anni di riferimento. Gli SDS, infatti, sono per loro natura rivolti alle PMI; di conseguenza, il peso per la variabile addetti medi è inferiore rispetto a quello per numero di imprese. Da notare, come a partire dal 2007 i dati ricavati dall'Indagine sulle unità locali delle imprese (IULGI), per la parte validata per le principali variabili del Registro, siano stati utilizzati come fonte di aggiornamento assimilabile ai controlli dei revisori. Questo ha determinato quasi un raddoppio delle imprese con aggiornamento ON LINE (circa 30 mila) con contestuale incremento dell'apporto in termini di addetti, pari a poco meno del 30%.

La fonte IVA, a partire dal 2007, ha fornito il codice Ateco per il 24,5% delle imprese, mentre sul totale degli addetti, nel 2008, ha inciso per oltre il 22%. Infine, da sottolineare l'uso della fonte CCIAA a partire dal 2008, mentre l'impatto della fornitura dell'Anagrafe Tributaria, usata come detto in precedenza solo ad integrazione, è risultato così modesto da essere inglobato nella modalità "ALTRI METODI".

Tavola 15 - Strumenti utilizzati per l'attribuzione della variabile Ateco 2007 nei Registri Asia - anni 2005-2008

Modalità attribuzione	2005		2006		2007		2008	
	Imprese	Addetti	Imprese	Addetti	Imprese	Addetti	Imprese	Addetti
ON LINE	0,3%	15,0%	0,3%	17,2%	0,8%	29,5%	0,6%	27,3%
SDS	66,1%	43,3%	63,1%	41,7%	61,3%	40,6%	67,3%	44,3%
LISTE	2,3%	5,3%	2,6%	5,7%	3,0%	2,4%	0,9%	1,2%
ACTR	1,6%	1,3%	1,1%	1,1%	-	-	-	-
CONVERSIONI 1 A 1	19,3%	26,6%	21,1%	25,0%	2,8%	2,6%	-	-
IVA	-	-	-	-	24,5%	18,8%	24,5%	22,3%
CCIAA	-	-	-	-	-	-	2,4%	2,1%
ALTRI METODI	10,4%	8,5%	11,7%	9,3%	7,7%	6,1%	4,3%	2,8%

Riferimenti bibliografici

- Ambroselli, S. e P. Vicari. 2007. *New economic classification and new instruments for Business Register classification: an opportunity to improve the quality in the Business Register*. Relazione presentata al Convegno: 20th International Roundtable on Business Survey Frames – Wiesbaden Group on Business Registers, Wiesbaden 21-26 ottobre.
- Consalvi, M. e G. Garofalo. 2000. *The real and spurious changes in the Economic Activity Code: methodological treatments and stability criterion*. Relazione presentata al Convegno: 14th International Roundtable on Business Survey Frames, Auckland 29 ottobre-3 novembre.
- Garofalo, G. 2008. *Le regole applicative e l'implementazione nel sistema statistico*. Relazione presentata al Convegno: Ateco 2007. La nuova classificazione delle attività economiche, Roma 23 aprile.
- Eurostat. 2006. *Setting up an implementation plan for NACE Rev. 2 in National Statistical Institutes*. Lussemburgo: Task Force of the implementation of NACE Rev. 2.
- Eurostat. 2006. *Implementation of NACE Rev. 2 in Business Register*. Lussemburgo: Task Force of the implementation of NACE Rev. 2.
- Eurostat. 2006. *Handbook on methodological aspects related to sampling designs and weights estimations*. Lussemburgo: Task Force of the implementation of NACE Rev. 2.
- Eurostat. 2006. *Back-casting handbook*. Lussemburgo: Task Force of the implementation of NACE Rev. 2.
- Eurostat. 2006. *Definitions and rules of classification for units outsourcing their production*. Lussemburgo: Task Force of the implementation of NACE Rev. 2.
- Eurostat. 2010. *Business Register Recommendations Manual*. Lussemburgo: methodologies and Working Papers.
- Kroese, B., H. van Hooff, B. Polak and A. Willeboordse. 1998. *Study on statistical units*. Statistics Netherlands, Voorburg, aprile 1998.
- Istat. 2009. Collana Metodi e norme n. 40 - Classificazione delle attività economiche - Ateco 2007
- Regolamento (CE) n. 696/1993 del Consiglio Europeo. 1993. Regolamento che definisce relativo alle unità statistiche di osservazione e di analisi del sistema produttivo nella Comunità.
- Regolamento (CE) n. 1893/2006 del Parlamento Europeo e del Consiglio. 2006. Regolamento che definisce la classificazione statistica delle attività economiche NACE Revisione 2 e modifica il regolamento (CEE) n. 3037/90 del Consiglio nonché alcuni regolamenti (CE) relativi a settori statistici specifici.
- Regolamento (CE) n. 177/2008 del Parlamento Europeo e del Consiglio. 2008. Regolamento che istituisce un quadro comune per i registri di imprese utilizzati a fini statistici e abroga il regolamento (CEE) n. 2186/93 del Consiglio.
- Vicari, P. 2008. La nuova classificazione delle attività economiche: dall'ISIC all'ATECO 2007. Relazione presentata al Convegno: Ateco 2007. La nuova classificazione delle attività economiche, Roma 23 aprile.

Informazioni per gli autori

La collana è aperta ad autori dell'Istat e del Sistema statistico nazionale, e ad altri studiosi che abbiano partecipato ad attività promosse dal Sistan (convegni, seminari, gruppi di lavoro, ecc.). Da gennaio 2011 essa sostituirà Documenti Istat e Contributi Istat.

Coloro che desiderano pubblicare sulla nuova collana dovranno sottoporre il proprio contributo alla redazione degli Istat Working Papers inviandolo per posta elettronica all'indirizzo iwp@istat.it. Il saggio deve essere redatto seguendo gli standard editoriali previsti, corredato di un sommario in italiano e in inglese; deve, altresì, essere accompagnato da una dichiarazione di paternità dell'opera. Per la stesura del testo occorre seguire le indicazioni presenti nel foglio di stile, con le citazioni e i riferimenti bibliografici redatti secondo il protocollo internazionale 'Autore-Data' del *Chicago Manual of Style*.

Per gli autori Istat, la sottomissione dei lavori deve essere accompagnata da una mail del proprio dirigente di Servizio/Struttura, che ne assicura la presa visione. Per gli autori degli altri enti del Sistan la trasmissione avviene attraverso il responsabile dell'ufficio di statistica, che ne prende visione. Per tutti gli altri autori, esterni all'Istat e al Sistan, non è necessaria alcuna presa visione. Tutti i lavori saranno sottoposti al Comitato di redazione, che valuterà la significatività del lavoro per il progresso dell'attività statistica istituzionale. La pubblicazione sarà disponibile su formato digitale e sarà consultabile on line.

Gli articoli pubblicati impegnano esclusivamente gli autori, le opinioni espresse non implicano alcuna responsabilità da parte dell'Istat. Si autorizza la riproduzione a fini non commerciali e con citazione della fonte.