

logaritmo dei crediti IVA¹¹⁸. Di seguito si riportano le prime stime ottenute col metodo *Difference - in - Differences*, applicato prima senza variabili di controllo e successivamente con un gruppo di covariate (inserite per ragioni di sensitività rispetto a problemi di omissione di variabili rilevanti). La norma determina una riduzione stimata dei crediti IVA pari al 46,1%;

Tabella 4.4.1 Difference-in-Difference: Base Line e Follow up

DID senza covariate							
Outcome Variable(s)	BASE LINE			FOLLOW UP			
	Control	treated	Diff(BL)	Control	treated	Diff(FU)	DIFF-IN-DIFF
ln_iva_credits	7.66	10.615	2.954	7.622	10.116	2.494	-0.461
Std. Error	0.01	0.027	0.029	0.01	0.031	0.032	0.044
t	744.7	116.15	101.49	3.72	-4.37	-11.3	-10.59
P>t	0.000	0.000	0.000***	0.000	0.000	0.000***	0.000***

DID con covariate							
Outcome Variable(s)	BASE LINE			FOLLOW UP			
	Control	treated	Diff(BL)	Control	treated	Diff(FU)	DIFF-IN-DIFF
ln_iva_cre-o	7.749	9.767	2.018	7.721	9.235	1.514	-0.504
Std. Error	0.33	0.33	0.028	0.33	0.331	0.031	0.039
t	23.51	13.86	71.88	7.66	8.21	-14.44	-12.99
P>t	0.000	0.000	0.000***	0.000	0.000	0.000***	0.000***

* Means and Standard Errors are estimated by linear regression; *** p<0.01; ** p<0.05; * p<0.1

Successivamente al metodo DID è stata applicata anche la tecnica del *propensity score matching*¹¹⁹ per meglio separare l'effetto dei provvedimenti normativi da quello del ciclo economico (Cfr.: Tabella 4.3.3). Le stime riportano valori pari a una riduzione del 36% nel caso della procedura di *Kernel Matching*, ad una riduzione del 39,4% nel caso del *Stratification Method*, mentre il metodo del *Nearest Neighbor Matching* è sensibile alla presenza delle variabili di controllo, ma l'impatto risulta superiore al 30% e inferiore al 48%. In generale, la stima del *propensity score matching* evidenzia un impatto della norma inferiore a quanto stimato attraverso la metodologia del DID; ciò nonostante, l'impatto della norma risulta inequivocabilmente significativo.

¹¹⁸ Sono ovviamente esclusi i casi in cui le imprese hanno un saldo debiti-crediti IVA non negativo. Queste imprese non sono interessate dalla norma e non costituiscono oggetto di analisi.

¹¹⁹ Il principale metodo di stima adottato, utilizzando l'informazione dei p-score, è quello del *Kernel Matching*. Tuttavia, per poter effettuare gli opportuni controlli di robustezza e per analisi di sensitività, vengono riportate le stime anche attraverso il *Stratification Method* e il *Nearest Neighbor Matching*.

Tabella 4.4.2 Stime Propensity Score Matching

tipo di stima		treated	untreated	ATT	Std. Err.	t
Kernel Matching	Senza Covariate	3246	3470	-0.361	0.094*	-3.842
Kernel Matching	Con Covariate	3246	3470	-0.361	0.127*	-2.837
Kernel Matching	Con Covariate e Sector Dummies	3246	3470	-0.361	0.117*	-3.1
Stratification Method	Con Covariate e Sector Dummies	3246	3470	-0.394	0.027	-14.381
Nearest Neighbour Matching	Senza Covariate	3246	243	-0.395	0.027	-14.621
Nearest Neighbour Matching	Con Covariate	3246	89	-0.303	0.172	-1.761
Nearest Neighbour Matching	Con Covariate e Sector Dummies	3246	111	-0.474	0.187	-2.539

Nota: * Standard Errors e t-Student calcolati con metodo bootstrap.

A seconda del metodo di stima utilizzato, gli effetti dei provvedimenti legislativi variano dal 30% al 51%, mentre l'impatto del ciclo economico è al massimo dell'11%. Questo risultato è imputabile ad un effetto deterrenza sul comportamento dei contribuenti, come risultato finale dei cambiamenti legislativi avviati dal luglio 2009. Il rafforzamento di controlli preventivi sembra quindi essere efficace come strumento di lotta alle frodi.

4.5 STUDI MICRO SULL'EFFETTO DI DETERRENZA DEGLI ACCERTAMENTI: PRIMI RISULTATI PROVVISORI

L'Agenzia delle Entrate ed il Dipartimento delle Finanze hanno avviato negli ultimi anni un'attività di studio della *tax compliance* incentrata sull'analisi di un panel di imprese individuali e sugli effetti dell'attività di accertamento dell'amministrazione. Si fa presente che tale attività presenta aspetti metodologici di rilevante complessità, considerato anche l'assenza di metodologie standardizzate e consolidate di analisi. Pertanto non può che evidenziarsi l'assoluta valenza provvisoria dei risultati finora raggiunti; nell'ultima Relazione concernente i risultati della lotta all'evasione fiscale al 31/12/2011 presentata al Parlamento il 28 settembre 2012, sono stati già riportati i primi risultati di questa attività di studio (che si riportano nel paragrafo 4.5.1). L'attività descritta nei paragrafi successivi costituisce pertanto un proseguimento di quanto già presentato lo scorso anno.

4.5.1 Studio tax compliance ed effetto dei controlli

Il progetto di studio riguarda l'analisi del comportamento dei contribuenti che hanno subito un controllo fiscale, di cui si evidenziano nel seguito i primi risultati (provvisori).

Si segnala preliminarmente che la metodologia seguita in questo tipo di studi ha carattere sperimentale, di cui si deve tenere conto nell'interpretazione dei risultati ottenuti. Si rammenta, inoltre, che la stima della misura della *compliance* non può prescindere dall'approfondire la metodologia volta ad isolare gli effetti del ciclo economico e dei provvedimenti legislativi mediante il confronto tra soggetti "accertati" e soggetti "non accertati".

Scopo dello studio è di verificare se il comportamento dei contribuenti, dopo aver percepito la presenza attiva del fisco, muta nel senso di una maggiore *compliance* con una significativa variazione (positiva) di alcune variabili fiscali, quali i redditi o gli imponibili dichiarati. Tale tipo di analisi è condotta mediante l'utilizzo di specifici panel di contribuenti, sulla base delle seguenti assunzioni:

- in primo luogo, si ritiene che il controllo possa produrre effetti immediati sul comportamento dei contribuenti; nell'ipotesi, ad esempio, che esso sia effettuato nei primi sei mesi di un anno solare, potrà essere influenzata già la dichiarazione dei redditi presentata nello stesso anno per il periodo d'imposta precedente;
- si prescinde, invece, dal (possibile) effetto deterrenza sui soggetti non controllati, per il quale non si dispone di adeguati strumenti di misurazione. Si assume pertanto che accertamenti e verifiche non producano alcun effetto sui contribuenti "non controllati".

Utilizzando i dati relativi alle dichiarazioni fiscali e ai controlli eseguiti dall'Agenzia delle Entrate, è stata implementata un'apposita banca dati riguardante la popolazione di riferimento, costituita dall'insieme delle ditte individuali con almeno una partita IVA attiva negli anni dal 2004 al 2008 (circa 4 milioni di contribuenti, nella popolazione possono esserci anche soggetti per i quali non risultano dati dichiarativi nel periodo di osservazione).

La costruzione di una banca dati, che avesse al suo interno informazioni di tipo fiscale, anagrafico e dati sui controlli effettuati¹²⁰, ha reso necessaria l'integrazione di dati provenienti da strutture

¹²⁰ Per i controlli e le verifiche sono stati presi in considerazione l'elenco completo (negli anni di riferimento) di tutte le tipologie di accertamenti (Unificato, Studi di Settore, Parametri, 41bis) e di tutte le tipologie di verifiche (Verifiche, Accessi Mirati, Accessi Brevi, Altro).

differenti dell'amministrazione finanziaria mediante la chiave di collegamento rappresentata dal codice fiscale.

Date le informazioni a disposizione, la platea dei soggetti è stata innanzitutto classificata in base ai modelli di dichiarazione presentati nei vari anni (mod. UNICO, mod. 770, mod. IVA, mod. 730) individuando otto possibili "profili di dichiarazione" (incluso anche il caso in cui il contribuente non ha presentato dichiarazione¹²¹).

Definizione del comportamento "virtuoso" dei contribuenti

La fase di avviamento del progetto, di cui si espongono di seguito in forma sintetica i risultati, si è concentrata sulla definizione di alcune metodologie volte a rilevare un cambiamento in senso "virtuoso" del comportamento del contribuente, da interpretare come un miglioramento della *tax compliance*.

Per l'individuazione dei valori da assegnare alla variabile obiettivo sono stati utilizzati diversi criteri in base ai profili di classificazione dei soggetti precedentemente adottati. In primo luogo, per il profilo riguardante i soggetti che hanno presentato sia la dichiarazione Unico che quella Iva, che risulta di maggiore rilevanza sia dal punto di vista numerico¹²², sia con riferimento alle informazioni disponibili dalle dichiarazioni, si è proceduto con una tecnica di *cluster analysis*, per la quale sono state necessarie alcune analisi preliminari al fine standardizzare le variabili considerate (reddituale e fiscali) e per identificare ed eliminare eventuali dati anomali.

Per altri profili dichiarativi, al fine di determinare variazioni virtuose nel comportamento dei soggetti analizzati, è stato costruito un apposito algoritmo deterministico¹²³, che si basa sulle variazioni strutturali di alcuni indici statistici. Per i profili di minore importanza¹²⁴ è stata data rilevanza al confronto tra i modelli di dichiarazione presentati in due anni successivi, con l'individuazione di specifici criteri per identificare il comportamento virtuoso.

Relativamente ai profili di dichiarazione di maggior interesse, per gli anni d'imposta 2005 - 2006 si hanno i seguenti risultati, in termini di cambiamento positivo del comportamento dei soggetti:

Tabella 4.5.1.1
Confronto anni d'imposta 2005 - 2006

Tipo contribuente	Cambiamento		Totale
	Nessuno	Positivo	
UNICO	320.119	1.840	321.959
730	334.138	5	334.143
UNICO+IVA	2.054.879	389.287	2.444.166
730+IVA	736	54	790
Totale contribuenti	2.709.872	391.186	3.101.058

Fonte: elaborazioni su dati anagrafe tributaria

¹²¹ I profili individuati sono: 1. UNICO; 2. 730; 3. 770; 4. IVA; 5. UNICO+IVA; 6. 730+IVA; 7. 770+IVA; 8. Nessuna dichiarazione.

¹²² Si tratta di più di due milioni di partite Iva per ogni anno d'imposta considerato.

¹²³ Utilizzato per i profili 1 (Unico), 2 (730), 4 (solo IVA), 6 (730+IVA), e 7 (770+IVA).

¹²⁴ Profili 3 (solo 770) e 8 (nessuna dichiarazione).

Nelle tabelle seguenti si riporta l'ammontare del reddito derivante dal quadro RN per gli stessi soggetti indicati nella tabella 4.5.1.1:

Tabella 4.5.1.2

Anno d'imposta 2005: reddito RN dei soggetti in tabella 4.5.1.1, milioni di euro

Tipo contribuente	Cambiamento		Totale
	Nessuno	Positivo	
UNICO	8.877	36	8.914
730	5.027	0,06	5.027
UNICO+IVA	59.017	1.900	60.917
730+IVA	10,55	0,71	11,26
Totale RN	72.932	1.937	74.869

Fonte: elaborazioni su dati anagrafe tributaria

Tabella. 4.5.1.3

Anno d'imposta 2006: reddito RN dei soggetti in tabella 4.5.1.1, milioni di euro

Tipo contribuente	Cambiamento		Totale
	Nessuno	Positivo	
UNICO	9.503	105	9.608
730	5.216	0,14	5.217
UNICO+IVA	63.678	4.449	68.126
730+IVA	11,25	0,79	12,04
Totale RN	78.409	4.555	82.963

Fonte: elaborazioni su dati anagrafe tributaria

Concentrando l'analisi sui contribuenti che hanno presentato entrambe le dichiarazioni UNICO+IVA, che rappresentano il sottogruppo di maggiore rilevanza per numerosità e dati disponibili, si può verificare come la variazione del reddito dichiarato dai soggetti considerati "migliorati" dal 2005 al 2006 sia considerevolmente superiore rispetto a quella dei soggetti "non migliorati". Ciò sembrerebbe confermare, sia pure in prima approssimazione, la validità della metodologia usata per l'analisi del comportamento dei contribuenti.

Analizzando l'andamento del reddito medio dichiarato nei due anni d'imposta considerati, si può però verificare che il reddito medio del primo anno in esame (2005) dei contribuenti "migliorati" sia molto più basso rispetto agli altri. In particolare (vedi tabella 4.5.1.5), il reddito medio dichiarato per il 2005 dai contribuenti che hanno fatto registrare un mutamento positivo del proprio comportamento risulta pari al 18,4% del reddito medio dei contribuenti che non migliorano; ovviamente tale percentuale sale al 40,2% nel 2006 dopo aver fatto registrare il miglioramento della *tax compliance*. Da quanto detto emerge quindi che i risultati ottenuti potrebbero essere influenzati da un reddito medio di partenza basso.

Tabella 4.5.1.4
Reddito medio in euro dei contribuenti in tabella 1, anni d'imposta 2005 - 2006

Tipo contribuente	2005			2006		
	Cambiamento		Totale	Cambiamento		Totale
	Nessuno	Positivo		Nessuno	Positivo	
UNICO	27.731	19.782	27.686	29.687	57.052	29.843
730	15.044	12.472	15.044	15.611	29.119	15.612
UNICO+IVA	28.720	4.880	24.923	30.989	11.428	27.873
730+IVA	14.332	13.139	14.251	15.283	14.667	15.241
Totale	26.913	4.952	24.143	28.934	11.643	26.753

Fonte: elaborazioni su dati anagrafe tributaria

Tabella 4.5.1.5
Rapporto reddito medio contribuenti migliorati/non migliorati, anni d'imposta 2005 - 2006

Tipo contribuente	2005	2006
UNICO	71,3%	192,2%
730	82,9%	186,5%
UNICO+IVA	17%	36,9%
730+IVA	91,7%	96%
Totale	18,4%	40,2%

Fonte: elaborazioni su dati anagrafe tributaria

Primi risultati in termini di *tax compliance*

Con l'abbinamento dei contribuenti con i controlli fiscali subiti, si è proceduto ad analizzare gli effetti in termini di *compliance* derivanti dall'attività di accertamento e controllo dell'Amministrazione finanziaria.

Per quanto riguarda l'arco temporale di riferimento, sono stati considerati distinti bienni di osservazione (2004-2005, 2005-2006 e 2006-2007), analizzando per ciascun anno d'imposta i risultati in termini di mutamento del comportamento dei singoli profili di dichiarazione sopra definiti e prendendo in considerazione sia coloro che hanno ricevuto un controllo fiscale, sia coloro che non l'hanno subito.

E' stato definito un indice denominato *effetto del controllo*. Tale indicatore viene calcolato rapportando la percentuale dei contribuenti migliorati tra i controllati (per singolo profilo) rispetto a quella dei migliorati che non hanno subito un controllo nel periodo di tempo precedente la presentazione della dichiarazione¹²⁵.

Un valore dell'indice superiore a 1 indica che l'attività di controllo contribuisce positivamente nello spingere i soggetti verso una maggiore *tax compliance* (rispetto alle unità "non accertate").

Nella tabella seguente si sintetizzano, per ciascun profilo, i risultati riguardanti i tre bienni considerati con una media, sia per ciò che riguarda la percentuale di contribuenti migliorati, sia relativamente all'indice *effetto del controllo*.

¹²⁵ Ad esempio, per il confronto tra gli anni d'imposta 2004-2005 sono stati considerati i controlli effettuati fino a giugno 2006, in tempo quindi per "influenzare" la dichiarazione dell'anno d'imposta 2005.

Tabella 4.5.1.6
Media dei bienni 2004 - 05, 2005 - 06, 2006 - 07 - indice effetto del controllo per profilo di dichiarazione

Tipo contribuente	Non Controllati				Controllati				Effetto del controllo (B)/(A)
	Cambiamento		Tot. Contribuenti	% positivi (A)	Cambiamento		Tot. Contribuenti	% positivi (B)	
	Nessuno	Positivo			Nessuno	Positivo			
UNICO	312.431	1.800	314.231	0,57%	9.361	101	9.462	1,06%	1,86
730	334.138	6	334.144	0,00%	2.045	-	2.045	-	-
770	25.667	-	25.667	-	1.688	-	1.688	-	-
IVA	8.461	1.061	9.522	11,15%	197	34	232	14,82%	1,33
UNICO+IVA	1.943.858	279.936	2.223.794	12,59%	161.157	42.138	203.295	20,73%	1,65
730+IVA	708	63	771	8,13%	8	1	9	11,54%	1,42
770+IVA	3.610	235	3.845	6,12%	101	6	107	5,61%	0,92
No Dichiarazione	720.508	-	720.508	-	15.236	-	15.236	-	-
Cambio profilo	189.315	104.449	293.764	35,56%	6.919	4.902	11.821	41,47%	1,17
Esclusi dall'analisi	86.274	-	86.274	-	5.292	-	5.292	-	-
Totale contribuenti	3.624.971	387.550	4.012.520	9,66%	202.005	47.182	249.187	18,93%	1,96

Fonte: elaborazioni su dati Agenzia delle Entrate ed anagrafe tributaria

I cambiamenti positivi tra i non controllati si presume dipendano da fattori di contesto (ciclo economico, fattori legislativi, ecc). Rappresentano, quindi, una evoluzione che si diffonde in tutto il sistema economico e influenza anche il comportamento dei soggetti controllati.

Se si guarda al profilo UNICO+IVA, il più importante in termini di numerosità dei soggetti, l'indice risulta pari a 1,65 mentre il valore complessivo è quasi pari a 2 (1,96). L'impatto diretto dell'azione di controllo dell'amministrazione sembra quindi produrre effetti positivi in termini di *tax compliance*.

4.5.2 Sviluppi successivi dell'analisi dell'effetto di controllo

Fermo restando l'impianto di base realizzato nel primo anno di lavoro relativamente alla platea di soggetti, variabili fiscali e reddituali di riferimento, l'analisi della *tax compliance* avviata dal Dipartimento e dall'Agenzia è stata orientata verso:

- a) la focalizzazione dello studio sulle persone fisiche con partita IVA e con modelli UNICO ed IVA su cui si è già insistito precedentemente;
- b) un progressivo aggiornamento della base dati;
- c) un continuo miglioramento ed affinamento dei metodi di analisi da utilizzare su questi dati (ad es. tecniche di *data mining* e specifici modelli econometrici), in modo da evidenziare con maggiore accuratezza variazioni della *tax compliance*.

Al riguardo si osserva che nel corso del 2012 la base dati è stata aggiornata all'anno di imposta 2009; la platea dei soggetti è di circa due milioni di persone fisiche con partita IVA (imprese individuali) con dichiarazione UNICO ed IVA per gli anni d'imposta 2006-2009.

Le definizioni adottate nel presente lavoro sono condizionate dall'oggetto dell'analisi che si vuole svolgere, ovvero esaminare il possibile impatto che l'Agenzia delle Entrate può esercitare in termini di deterrenza all'evasione.

Definizione di controllo e anno di percezione

Finalizzare l'analisi statistica alla verifica dell'azione di deterrenza implica l'adozione di specifiche definizioni, riguardanti, in particolare:

- a) l'utilizzo di un generico termine di controllo fiscale, che identifica un'azione tramite la quale il contribuente percepisce di essere controllato dalle autorità competenti;
- b) l'unità di analisi, che è rappresentata dal soggetto controllato e non dall'accertamento svolto;
- c) l'anno di registrazione del fenomeno, che coincide con il momento in cui il soggetto percepisce l'attività di controllo (differente dall'approccio basato sulle annualità accertate).

Le azioni di contrasto all'evasione che concorrono alla definizione di controllo sono le seguenti:

- ✓ Accertamenti Unificati;
- ✓ Studi di Settore Parametri;
- ✓ 41 Bis;
- ✓ Atti di Contestazione;
- ✓ Crediti (Atti di Recupero);
- ✓ Verifiche;
- ✓ Accessi Mirati;
- ✓ Accessi Brevi;
- ✓ Altri Accessi e Verifiche;
- ✓ Crediti (Verifiche ed Accessi).

L'anno di percezione indica l'anno in cui è stato effettuato un controllo indipendentemente dall'annualità d'imposta controllata. L'anno di percezione deve essere messo in relazione al primo anno d'imposta sul quale si può verificare l'effetto deterrente del controllo stesso. Ad esempio si includono nell'anno di percezione 2006 i controlli effettuati tra il 1° luglio 2006 e il 31 giugno 2007. Dopo questa data, infatti, il contribuente presenterà la dichiarazione relativa all'anno d'imposta 2006, sulle risultanze della quale è possibile calcolare l'effetto di deterrenza. Il collegamento tra l'anno di percezione e l'anno d'imposta sul quale verificare l'effetto del controllo è illustrato nella tabella 4.5.2.1

Tabella 4.5.2.1

Anno di percezione del controllo e corrispondenti anni d'imposta sui quali verificare l'effetto deterrente. Anni 2007-2010

Periodo in cui si avverte il controllo	Anno di percezione	Primo anno d'imposta sul quale verificare l'effetto del controllo
Dal 01/07/2007 al 31/06/2008	2007	2007
Dal 01/07/2008 al 31/06/2009	2008	2008
Dal 01/07/2009 al 31/06/2010	2009	2009

La definizione di reddito analizzata per lo studio della deterrenza

La variabile chiave delle analisi è il reddito dichiarato dal contribuente, in quanto è sulla variazione di questa grandezza che si incentra l'inferenza sull'attività di deterrenza esercitata dai controlli.

Nelle dichiarazioni fiscali esistono molteplici definizioni di reddito, che rispondono sia a differenti comportamenti economici (ad esempio lavoro dipendente/indipendente), sia a differenze di trattamento ai fini fiscali (ad esempio il lavoro indipendente in agricoltura e quello in altri settori di attività economica).

L'obiettivo che si intende raggiungere è quello di individuare la nozione di reddito che approssimi maggiormente la base imponibile sulla quale è possibile fare deterrenza tramite gli accertamenti. A questo fine è importante sterilizzare la componente del reddito imponibile che deriva dal lavoro dipendente, la variazione del quale è fortemente influenzata da fattori esogeni all'accertamento quali il cambiamento dello stato del lavoratore (ad es. da occupato e disoccupato) e le dinamiche contrattuali.

Pertanto, al fine di definire il reddito come somma dei redditi da: lavoro autonomo, impresa semplificata e impresa ordinaria (RILA) si è adottata la seguente formula

$$RILA = RE25 + RF46 \text{ col}2 + RG26 \text{ col}2$$

A destra del segno di uguaglianza è indicato il rigo della dichiarazione dei redditi presentata nell'anno 2008 con riferimento all'anno d'imposta 2007; dove¹²⁶:

RE25 = Reddito o perdita da lavoro autonomo, da sommare agli altri redditi del Quadro RNx;

RF46 col 2 = Reddito d'impresa lordo (o perdita);

RG26 col2 = Reddito d'impresa semplificata lordo (o perdita).

Cluster analysis

La cluster analysis precedentemente applicata (cfr. paragrafo 4.5.1) è stata rinnovata con modificazioni nell'approccio e nel metodo.

Il modello statistico di clusterizzazione diventa propedeutico alla fase di comprensione ed individuazione del soggetto fiscale migliorato, ovvero di quel soggetto che presenti specifiche di miglioramento che lo rendano significativamente caratteristico rispetto agli altri soggetti dello stesso gruppo di appartenenza. Sulla base di questo assunto principale, l'intero impianto di analisi è stato revisionato.

Le variabili di analisi prese in considerazione sono le seguenti:

- variabili riguardanti il profilo socio-demografico del contribuente, in particolare l'età, il sesso, la durata della partita IVA;
- variabili geografiche (regione, provincia);
- la popolazione della provincia;
- settore economico (branca 26);
- dimensione dell'impresa (fascia di ricavi);
- variabili legate alla produttività e alla tecnologia dell'impresa¹²⁷.

¹²⁶ La corrispondenza per gli altri anni è la seguente:

A.I. 2008 RE25 + RF44 col2 + RG26 col2;

A.I. 2009 RE25 + RF44 col2 + RG26 col2

¹²⁷ Sono state costruite specifiche variabili indicative di precise caratteristiche delle imprese, quali:

Intensità del lavoro = costo del lavoro / volume d'affari

Costi di funzionamento = totale acquisti / volume d'affari

Sulla base dei primi risultati è stato chiaro sin da subito che il problema principale da affrontare riguarda la polarizzazione indotta dalle variabili di tipo categoriale. La soluzione scelta per risolvere tale problematica consiste nel trasformare le variabili quantitative in qualitative, in modo da ottenere una matrice in forma disgiuntiva completa, sulla quale eseguire una analisi delle corrispondenze multiple¹²⁸, per sviluppare successivamente la *cluster analysis* sulle coordinate fattoriali ottenute.

L'analisi condotta ha dato luogo all'individuazione di una serie di fattori che spiegano la varianza della matrice dei dati iniziale. Ogni fattore estratto è corredato dal relativo autovalore che permette di quantificare la percentuale di variabilità spiegata. La prima componente estratta, per definizione, è quella in grado di spiegare la massima quota di varianza delle variabili originali. Per le altre componenti, gli autovalori decrescono progressivamente.

Tra i metodi generali più utilizzati per determinare il numero di fattori da estrarre ricordiamo:

- *Criterio del tasso d'inerzia*, ovvero della percentuale di variabilità spiegata: bisogna decidere a priori la percentuale soglia che si intende raggiungere;
- *Criterio eigenvalue-one*, in base al quale non avrebbe senso conservare quei fattori con varianza minore dell'unità in quanto ciò significherebbe conservare un fattore che spiega meno di quanto potrebbe spiegare una qualunque variabile originaria;
- *Criterio dello scree-test*: si valuta il grafico degli autovalori e si considerano tutti gli autovalori che precedono la discesa regolare della curva.

Nel nostro caso il criterio dello *scree-test* ci porta a prendere **i primi 5 autovalori**.

Successivamente, si è proceduto a eseguire la *Cluster Analysis*¹²⁹ usando come input i 5 fattori estratti mediante l'Analisi delle Corrispondenze Multiple.

L'applicazione di questa metodologia descritta porta a 6 *cluster*, con una *silhouette* sufficiente (figura 4.4.).

Il *cluster* 3 è il più numeroso (34% della popolazione), con prevalenza del settore commercio, e valori medi di dichiarazione in linea con quelli della popolazione.

Sono presenti due *cluster*, 1 e 5, con un rilevante numero di soggetti appartenenti al Settore Costruzioni, che si differenziano tra loro soprattutto per i valori economici di dichiarazione (il *cluster* 5 presenta valori doppi rispetto al *cluster* 1).

Altri due *cluster*, 4 e 6, caratterizzati soprattutto dal settore commercio, si differenziano per i valori economici di dichiarazione (il 4 è un *cluster* di soggetti con l'età media e durata della partita iva più bassa in assoluto, il 6 è un *cluster* un'età media molto alta e la durata della partita IVA più alta in assoluto).

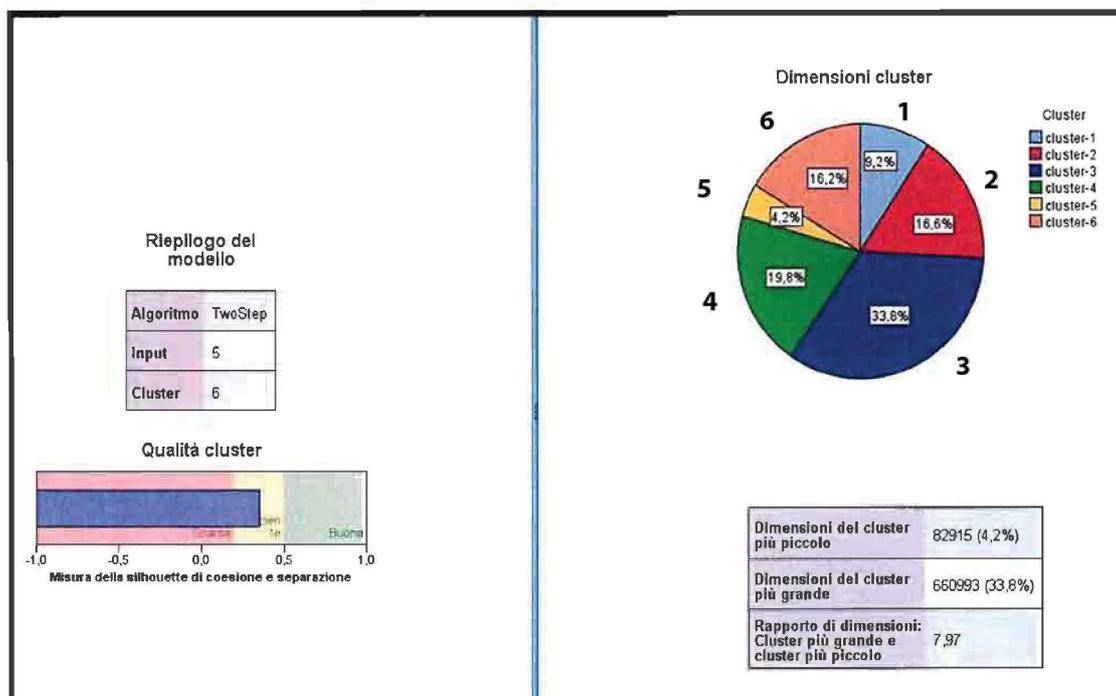
Il *cluster* 2 rappresenta l'agricoltura e mostra valori economici inferiori alla media.

Grado di internazionalizzazione = (volume d'affari – operazioni imponibili) / volume d'affari

¹²⁸ L'analisi delle corrispondenze multiple (ACM) è una tecnica statistica che ha l'obiettivo di esplorare, descrivere e sintetizzare un'intera matrice di dati qualitativi riducendola ad un numero minore di fattori (o assi) che rappresentano combinazioni delle variabili e sintetizzano gli aspetti fondamentali del fenomeno oggetto d'indagine. Per l'applicazione della metodologia, le variabili continue sono state dunque trasformate in variabili discrete, attraverso la costruzione di classi di valori che, di conseguenza, costituiscono le modalità di ogni singola variabile. Il risultato è una variabile categorica ordinale perché tra le modalità create sussiste un ordine di precedenza.

¹²⁹ Tra gli algoritmi di *cluster analysis*, è stato utilizzato il metodo *Twostep* dal momento che, a differenza di altre procedure di raggruppamento, permette la definizione automatica del numero ottimale di *cluster* in base al criterio bayesiano di Schwarz. Il metodo si articola in due fasi principali: la prima fase prevede un pre-raggruppamento sequenziale dei casi in più sotto-*cluster* di piccole dimensioni, nella seconda fase la procedura prevede l'applicazione degli algoritmi gerarchici standard sui sotto-*cluster* risultanti dalla prima fase.

Figura 4.4 - Statistiche descrittive dei cluster ottenuti



Di seguito un dettaglio sui *cluster* ottenuti:

Cluster 1 (180.025 soggetti – 9.2% della popolazione)

È caratterizzato da soggetti del nord (in particolare dalle regioni Lombardia, Veneto e Piemonte), soprattutto del settore costruzioni, con valori alti delle variabili economiche, un grado di internazionalizzazione superiore alla media della popolazione e un'alta percentuale di aziende con dipendenti.

Cluster 2 (325.008 soggetti – 16.6% della popolazione)

Composto soprattutto da soggetti appartenenti al settore Agricoltura, Caccia e Silvicoltura, concentrati nel sud e nel nord est (in particolar modo Puglia, Sicilia, Veneto ed Emilia Romagna), è il *cluster* con la più bassa percentuale di aziende con dipendenti e la più bassa percentuale di aziende a conduzione familiare. I costi di funzionamento e il grado di internazionalizzazione sono più alti rispetto a tutti gli altri *cluster* mentre i valori economici sono i più bassi in assoluto.

Cluster 3 (660.993 soggetti – 33.8% della popolazione)

È il *cluster* più numeroso, caratterizzato da soggetti del Nord Ovest, in particolare dalla regione Lombardia. I settori prevalenti sono Commercio e Altre attività professionali ed imprenditoriali (in particolar modo Servizi di consulenza). I valori economici sono in linea con la media della popolazione, mentre Intensità del lavoro e Costi di funzionamento sono superiori alla media dell'intera popolazione e il Costo del lavoro inferiore.

Cluster 4 (387.435 soggetti – 19.8% della popolazione)

Rappresenta il *cluster* con la più alta percentuale di donne, con l'età media e numero di anni di vita dell'impresa più bassi in assoluto, composto da soggetti del Nord Ovest e Sud (in particolare dalle regioni Lombardia e Campania) appartenenti soprattutto al settore Commercio e altre attività professionali ed imprenditoriali (in particolar modo Servizi di consulenza). Presenta valori economici di dichiarazione inferiori alla media della popolazione.

Cluster 5 (82.915 soggetti – 4.2% della popolazione)

È il *cluster* più piccolo, con età media e numero di anni di vita dell'impresa molto alti, composto da soggetti con la più alta percentuale di aziende a conduzione familiare e con la più alta percentuale di aziende con dipendenti, presenti soprattutto al Nord Ovest e Nord Est (in particolare dalle regioni Lombardia, Veneto ed Emilia Romagna). Il settore prevalente è il quello delle Costruzioni, ma presenta soggetti appartenenti anche alle categorie Altre attività professionali ed imprenditoriali e Trasformazione industriale destinata prevalentemente ad investimenti e beni intermedi. I valori economici di dichiarazione sono i più alti in assoluto.

Cluster 6 (317.141 soggetti – 16.2% della popolazione)

È il *cluster* con età media e numero di anni di vita dell'impresa più alti, caratterizzato da soggetti del Nord Ovest e dal Centro Italia, soprattutto Lombardia e Lazio. Il settore prevalente è il Commercio, ma ci sono diversi soggetti appartenenti anche ad Altre attività professionali ed imprenditoriali. I valori economici sono superiori alla media della popolazione e la percentuale di aziende con dipendenti è piuttosto elevata.

E' interessante osservare, per ciascun *cluster*, il dato medio del reddito da impresa e lavoro autonomo (RILA). Tale grandezza presenta una notevole variabilità tra i gruppi, ad indicare che le caratteristiche strutturali utilizzate per la suddivisione in *cluster* hanno un impatto significativo sui risultati economici che sono dichiarati al fisco. In particolare, il *cluster* 2, caratterizzato prevalentemente da agricoltori, dichiara un reddito medio assolutamente irrisorio, che, sull'intero periodo considerato, non raggiunge mai i 1.000 euro annui. All'estremo opposto troviamo il *cluster* 5 che si connota per i redditi medi più elevati.

Per tutti i *cluster* si osserva una marcata riduzione del reddito nel 2009, sicuramente condizionata dall'avversa fase congiunturale di quell'anno.

Tabella 4.5.2.2

Reddito da impresa e lavoro autonomo per contribuente, suddiviso per *cluster*

	2006	2007	2008	2009
CLUSTER 1	27.634	29.746	28.769	23.471
CLUSTER 2	629	387	506	424
CLUSTER 3	30.905	32.200	31.422	17.983
CLUSTER 4	18.165	19.477	19.536	11.927
CLUSTER 5	46.994	48.663	44.952	30.045
CLUSTER 6	35.983	36.008	33.752	18.223

Valori espressi in euro

Post-stratificazione

Per correggere gli effetti della *selection bias* dovuta alla selezione dei soggetti da controllare (effettuata da Agenzia delle entrate ed altri enti in base al rischio di evasione o pericolosità fiscale) si è adottata una procedura di post – stratificazione del panel che riporta al livello della popolazione i valori di alcune variabili socio – economiche. Questa operazione riduce la distorsione del campione e consente di ottenere una stima più corretta delle variabili reddituali e delle basi imponibili di alcune imposte, in modo che il confronto tra controllati e non controllati risulti più affidabile. Si evidenzia, in ogni caso, che la post - stratificazione consente una correzione parziale. Per affinare tale correzione è necessario adottare l'approccio econometrico illustrato nel successivo paragrafo..

Il metodo di post-stratificazione

Probabilità di inclusione del contribuente nell'insieme degli accertati:

$$\alpha_i = \frac{C_i}{N_i}$$

Dove:

$C_i \rightarrow$ numero di soggetti controllati per la cella i

$N_i \rightarrow$ numero totale di soggetti per la cella i

$i \rightarrow$ cella di elaborazione, identificata dalla tupla Provincia - Fascia Ricavi - Settore Economico (tupla delle variabili di stratificazione)

γ_i fattore di espansione o coefficiente di ponderazione. Tale coefficiente è lo stesso per tutti gli individui appartenenti a quella tupla.

$$\gamma_i = \frac{\begin{matrix} se X \in C \Rightarrow 1 \\ \alpha_i \end{matrix}}{\begin{matrix} se X \notin C \Rightarrow 1 \\ (1 - \alpha_i) \end{matrix}}$$

Dove:

$X \rightarrow$ soggetto di riferimento

$C \rightarrow$ insieme dei soggetti controllati

Calcolo della valore di stima puntuale corretto:

$$\hat{Y}_x = \frac{\begin{matrix} se X \in C \Rightarrow Y_x \\ \alpha_i \end{matrix}}{\begin{matrix} se X \notin C \Rightarrow Y_x \\ (1 - \alpha_i) \end{matrix}}$$

Dove:

$X \rightarrow$ soggetto di riferimento

$C \rightarrow$ insieme dei soggetti controllati

$Y \rightarrow$ valore puntuale per il soggetto della Macrovariabile nella Dichiarazione dei Redditi di riferimento

Calcolo della stima economica stratificata:

$$\bar{Z}_j = \frac{\sum_{x=1}^X \hat{Y}_{x,j}}{\sum_x \gamma_{(i,x),j}}$$

Dove:

$\hat{Y}_{x,j} \rightarrow$ stima puntuale per il soggetto della Macrovariabile nella Dichiarazione dei Redditi di riferimento appartenente allo strato j

$j \rightarrow$ strato di elaborazione, identificato dalla tupla Controllato – Migliorato

$\sum_x \gamma_{(i,x),j} \rightarrow$ sommatoria dei coefficienti di espansione (γ_i) di tutti i soggetti (x) appartenenti a “ j ”.

Nell'applicazione del metodo adottato è cruciale la scelta dei criteri di classificazione che si utilizzano per post-stratificare le osservazioni. Il metodo, infatti, non è invariante a seconda del livello di disaggregazione utilizzato per post-stratificare. L'ipotesi implicita sottostante a questa procedura consiste, infatti, nel presupporre che non esista un errore di selezione sostanziale all'interno di ciascuna cella che rappresenta lo “strato” elementare considerato. E', pertanto, consigliabile adottare un criterio di classificazione il più disaggregato possibile.

Nel nostro caso le variabili di post stratificazione sono:

- 5 settori di attività economica (Agricoltura, Altri Servizi, Commercio, Costruzioni, Industria);
- 109 province;
- 2 fasce dimensionali per ricavi attività (valore di separazione 30.987).

Si sono, pertanto, costruiti 3.488 strati teorici all'interno dei quali è stata suddivisa sia la platea dei controllati sia quella dei non controllati.

L'analisi si concentra sull'andamento del reddito da lavoro autonomo e impresa (RILA), ipotizzando che sia la grandezza più significativa per valutare la dinamica della *compliance*.

Nelle tabelle seguenti si illustrano le variazioni nei tre bienni considerati di RILA, riportato all'universo, distinguendo i non controllati (NC) dai controllati (C) all'interno dei gruppi individuati con la cluster analysis appena illustrata..

Tabella 4.5.2.3

Reddito da impresa e lavoro autonomo in euro per contribuente, suddiviso per *cluster* e distinto tra soggetti controllati e non controllati. Anni 2006-2007

Cluster di appartenenza	Stato rispetto al controllo	2006	2007	Variazione percentuale
Cluster 1	NC	27.701	29.790	7,5%
	C	26.956	29.390	9,0%
Cluster 2	NC	626	384	-38,6%
	C	770	463	-39,8%
Cluster 3	NC	31.506	32.779	4,0%
	C	24.992	26.458	5,9%
Cluster 4	NC	18.486	19.796	7,1%
	C	14.770	15.961	8,1%
Cluster 5	NC	47.101	48.738	3,5%
	C	45.846	47.998	4,7%
Cluster 6	NC	36.741	36.700	-0,1%
	C	29.003	29.673	2,3%
Totale	NC	24.916	25.822	3,6%
	C	21.192	22.310	5,3%

Tabella 4.5.2.4

Reddito da impresa e lavoro autonomo in euro per contribuente, suddiviso per *cluster* e distinto tra soggetti controllati e non controllati. Anni 2007-2008

Cluster di appartenenza	Stato rispetto al controllo	2007	2008	Variazione percentuale
Cluster 1	NC	29.786	28.835	-3,2%
	C	29.217	27.841	-4,7%
Cluster 2	NC	382	502	31,2%
	C	446	810	81,8%
Cluster 3	NC	32.830	32.069	-2,3%
	C	25.633	24.615	-4,0%
Cluster 4	NC	19.884	19.976	0,5%
	C	15.231	14.913	-2,1%
Cluster 5	NC	48.908	45.305	-7,4%
	C	46.229	41.365	-10,5%
Cluster 6	NC	36.803	34.489	-6,3%
	C	28.435	26.979	-5,1%
Totale	NC	25.890	25.059	-3,2%
	C	21.436	20.468	-4,5%

Tabella 4.5.2.5

Reddito da impresa e lavoro autonomo in euro per contribuente, suddiviso per *cluster* e distinto tra soggetti controllati e non controllati. Anni 2008-2009

Cluster di appartenenza	Stato rispetto al controllo	2008	2009	Variazione percentuale
Cluster 1	NC	28.819	23.446	-18,6%
	C	28.232	23.592	-16,4%
Cluster 2	NC	504	418	-16,9%
	C	512	550	7,3%
Cluster 3	NC	31.899	18.137	-43,1%
	C	25.945	16.761	-35,4%
Cluster 4	NC	19.934	12.024	-39,7%
	C	14.727	11.265	-23,5%
Cluster 5	NC	45.192	30.030	-33,6%
	C	42.879	30.638	-28,5%
Cluster 6	NC	34.131	18.334	-46,3%
	C	30.142	17.408	-42,2%
Totale	NC	24.920	14.984	-39,9%
	C	21.672	14.655	-32,4%

L'ipotesi che si intende verificare consiste nel fatto che, a seguito del controllo il contribuente muti il suo comportamento nel senso di una maggiore *tax compliance*, e, quindi, se nell'anno successivo al controllo mostri una variazione nei redditi e negli imponibili dichiarati significativamente "migliore" rispetto ai non controllati. Si è, quindi, calcolata la differenza tra la variazione del RILA per impresa riportato all'universo dei controllati meno l'analoga variazione dei non controllati, assumendo che se il risultato è superiore a zero ci sia stato un effetto deterrenza.

Tale confronto è stato effettuato sia per il totale sia per ciascun *cluster*. La suddivisione in *cluster* è importante poiché raggruppa i contribuenti che hanno caratteristiche strutturali simili e per i quali si può presupporre che abbiano anche dinamiche economiche simili. L'analisi per *cluster*, quindi, dovrebbe sterilizzare, almeno in parte, le condizioni di contesto che possono influenzare la dinamica dei redditi ed isolare il solo effetto del controllo.

La figura 4.5 illustra la differenza tra la variazione dei redditi procapite dei contribuenti controllati e di quelli non controllati, disaggregata per *cluster*, calcolata sui valori riportati nelle tabelle precedenti. Dal grafico sembra emergere un effetto positivo dei controlli, in termini di deterrenza, nei bienni 2007-2006 e 2008-2009; mentre nel biennio 2007-2008 si registra un segnale contrario, poiché la variazione dei redditi dei soggetti controllati è risultata inferiore a quella dei soggetti non controllati.

Il risultato è confermato per tutti i *cluster* ad eccezione del 2 e del 6. Il *cluster 2* raggruppa delle situazioni estremamente peculiari, comprendendo prevalentemente agricoltori e caratterizzandosi con redditi procapite molto bassi. Per queste caratteristiche è plausibile presupporre l'esistenza di fenomeni fortemente anomali e difficilmente interpretabili. Il *cluster 6*, invece è quello che presenta il numero di anni di vita dell'impresa più alto, ovvero la componente di contribuenti più stabilmente presente sul mercato. Per quest'ultimo l'effetto deterrente dei controlli si rileva per tutti e tre i bienni.