



**UNIONE EUROPEA**

Iniziativa a favore dell'Occupazione Giovanile  
Fondo Sociale Europeo  
Investiamo nel tuo futuro



*Ministero del Lavoro  
e delle Politiche Sociali*

DIREZIONE GENERALE PER LE POLITICHE ATTIVE,  
I SERVIZI PER IL LAVORO E LA FORMAZIONE



## **ALLEGATO IV**

### **NOTA METODOLOGICA**

**ISFOL**  
ISTITUTO PER LO SVILUPPO DELLA  
FORMAZIONE PROFESSIONALE  
DEI LAVORATORI

## ALLEGATO IV

### IV.1 RILEVAZIONE SUL GRADO DI SODDISFAZIONE DELLA GARANZIA GIOVANI (Capp. 3.# e 4.3.2)

#### POPOLAZIONE DI RIFERIMENTO DELL'INDAGINE

L'indagine "Rilevazione sul grado di soddisfazione del Programma Garanzia Giovani" è stata realizzata nel corso del mese di luglio 2015 con tecnica CAWI attraverso l'invio di un'apposita e-mail ad un campione di giovani, di età compresa tra i 18 e i 29 anni, che a fine maggio 2015 risultavano iscritti al programma Garanzia Giovani. L'indagine esclude i giovani iscritti e non presi in carico da oltre 6 mesi. Complessivamente si tratta di un universo di 360 mila giovani, suddiviso in due distinte popolazioni di partenza (vedi tabella IV.1):

- A - giovani che alla data di estrazione delle liste avevano stipulato un patto di servizio (Presi in carico);
- B - giovani che alla data di estrazione delle liste non avevano ancora stipulato un patto di servizio (Non presi in carico).

Il secondo gruppo, indicato con la lettera B, ricomprende i giovani che dal momento dell'estrazione delle liste alla loro partecipazione all'indagine hanno sottoscritto il patto di servizio.

**Tabella IV.1 – Popolazione giovani NEET**

A - Presi in carico al 31/05/2015		B - Non presi in carico al 31/05/2015	
Regione Presa in carico	n.	Regione residenza	n.
Piemonte	2.802	Piemonte	4.478
Valle d'Aosta	662	Valle d'Aosta	107
Lombardia	18.724	Lombardia	5.695
P.A. di Trento	1.942	P.A. di Trento	222
Veneto	20.133	Veneto	1.766
Friuli Venezia Giulia	6.377	Venezia Giulia	1.736
Liguria	2.857	Liguria	2.027
Emilia Romagna	18.830	Emilia Romagna	3.556
Toscana	18.421	Toscana	2.544
Umbria	6.948	Umbria	1.647
Marche	12.486	Marche	1.327
Lazio	22.356	Lazio	6.916
Abruzzo	10.508	Abruzzo	3.769
Molise	1.497	Molise	1.261
Campania	24.952	Campania	9.609
Puglia	19.610	Puglia	9.577
Basilicata	8.106	Basilicata	2.020
Calabria	10.911	Calabria	8.516
Sicilia	47.830	Sicilia	15.692
Sardegna	20.027	Sardegna	1.560
	<b>275.979</b>		<b>84.025</b>

Fonte: ISFOL, Rilevazione sul grado di soddisfazione del Programma Garanzia Giovani

## STRATEGIA CAMPIONARIA

Per le due popolazioni di riferimento è stata adottata una diversa strategia campionaria. Per la popolazione dei presi in carico si è proceduto ad un piano di campionamento casuale stratificato di 40.000 unità per regione di presa in carico e condizione nel mercato del lavoro in ingresso, con allocazione di compromesso tra l'allocazione proporzionale, quella ottima secondo i 20 domini di studio "regione di presa in carico" (dominio D1) e quella ottima secondo i 3 domini di studio "condizione in ingresso" (dominio D2) (vedi tabella IV.2). L'allocazione finale di compromesso rappresenta una media semplice delle tre allocazioni che unisce ad un basso effetto del disegno (1,061) un errore relativo massimo nei domini delle regioni di presa in carico inferiore a 0,05.

Ottenuta l'allocazione di compromesso per strato, l'allocazione finale è avvenuta attraverso una successiva post-stratificazione per classe di età. Si è inoltre proceduto ad un sovracampionamento di 4 mila unità per i giovani che, alla data di riferimento, avevano ricevuto un'offerta di una politica attiva.

**Tabella IV.2 – Allocazione del campione dei presi in carico**

Allocazione	Regione presa in carico			Condizione in ingresso			Effetto disegno	Pesi allocazione		
	D1_mean	D1_min	D1_max	D2_mean	D2_min	D2_max		Proporzional	Ottima_d1	Ottima_d2
Proporzionale	0,02972	0,01088	0,09255	0,00826	0,00629	0,00938	1,00000	1,00000	0,00000	0,00000
Ottima D1	0,01868	0,01868	0,01868	0,00823	0,00637	0,00974	1,38546	0,00000	1,00000	0,00000
Ottima D2	0,02965	0,01184	0,08880	0,00777	0,00777	0,00777	1,13164	0,00000	0,00000	1,00000
Compromesso: media	0,02349	0,01286	0,04973	0,00801	0,00673	0,00886	1,06144	0,33333	0,33333	0,33333

Fonte: ISFOL, Rilevazione sul grado di soddisfazione del Programma Garanzia Giovani

Per la popolazione dei giovani non presi in carico (popolazione B) si è proceduto ad estrarre un campione casuale di 10 mila unità stratificato per regione di residenza e allocazione proporzionale negli strati.

## PARTECIPAZIONE ALL'INDAGINE

Il numero effettivo delle unità campionate per la popolazione A – Presi in carico è di 44.137 (vedi tabella IV.3). I rispondenti sono stati 15.386, pari a circa il 35% delle unità campionate, e al 16,4% del numero complessivo di giovani invitati a partecipare durante il mese dell'indagine.

**Tabella IV.3 – Partecipazione indagine Popolazione A-Presi in carico**

Regione presa in carico	A - Presi in carico al 31/05/2015				
	Campione (a)	n. complessivo di invito a partecipare (b)	Rispondenti (c)	% Tasso risposta sulle unità campionate c/a	%Tasso di partecipazione c/b
Piemonte	854	1.881	372	43,6	19,8
Valle d'Aosta	252	385	42	16,7	10,9
Lombardia	3.272	7.471	948	29,0	12,7
P.A. di Trento	579	1.111	248	42,8	22,3
Veneto	3.362	6.383	866	25,8	13,6
Friuli Venezia Giulia	1.496	3.738	609	40,7	16,3
Liguria	782	1.621	433	55,4	26,7
Emilia Romagna	2.778	6.187	1.008	36,3	16,3
Toscana	3.272	7.346	1.006	30,7	13,7
Umbria	1.320	2.893	519	39,3	17,9
Marche	2.590	2.925	332	12,8	11,4
Lazio	3.756	9.074	1.761	46,9	19,4
Abruzzo	1.768	4.164	933	52,8	22,4
Molise	475	935	215	45,3	23,0
Campania	3.311	7.521	1.246	37,6	16,6
Puglia	2.579	7.155	1.225	47,5	17,1
Basilicata	1.772	2.179	273	15,4	12,5
Calabria	1.719	2.923	756	44,0	25,9
Sicilia	5.283	10.882	1.811	34,3	16,6
Sardegna	2.917	6.950	783	26,8	11,3
<b>Totale</b>	<b>44.137</b>	<b>93.724</b>	<b>15.386</b>	<b>34,9</b>	<b>16,4</b>

Fonte: ISFOL, Rilevazione sul grado di soddisfazione del Programma Garanzia Giovani

Il numero effettivo delle unità campionate per la popolazione B – Non presi in carico è di 10.037. Hanno compilato il questionario 2.263 giovani, pari a circa il 22,5% del campione (vedi tabella IV.4). Il tasso di partecipazione complessivo è stato del 13,6%.

**Tabella IV.4 – Partecipazione indagine Popolazione B-Non presi in carico**

Regione residenza	B - Non presi in carico al 31/05/2015				
	Campione (a)	n. complessivo di invito a partecipare (b)	Rispondenti (c)	% Tasso risposta sulle unità campionate c/a	%Tasso di partecipazione c/b
Piemonte	534	756	115	21,5	15,2
Valle d'Aosta	15	22	7	46,7	31,8
Lombardia	680	1.187	165	24,3	13,9
P.A. di Trento	29	52	11	37,9	21,2
Veneto	212	193	22	10,4	11,4
Friuli Venezia Giulia	209	240	32	15,3	13,3
Liguria	243	331	69	28,4	20,8
Emilia Romagna	425	666	69	16,2	10,4
Toscana	302	384	51	16,9	13,3
Umbria	198	210	28	14,1	13,3
Marche	160	214	16	10,0	7,5
Lazio	825	1.494	174	21,1	11,6
Abruzzo	450	719	118	26,2	16,4
Molise	153	279	55	35,9	19,7
Campania	1.145	1.803	231	20,2	12,8
Puglia	1.142	2.261	303	26,5	13,4
Basilicata	242	256	28	11,6	10,9
Calabria	1.016	1.642	276	27,2	16,8
Sicilia	1.870	3.649	455	24,3	12,5
Sardegna	187	252	38	20,3	15,1
	<b>10.037</b>	<b>16.610</b>	<b>2.263</b>	<b>22,5</b>	<b>13,6</b>

Fonte: ISFOL, Rilevazione sul grado di soddisfazione del Programma Garanzia Giovani

### PESO DI CALIBRAZIONE PER MANCATE RISPOSTE TOTALI

Per tenere conto della distorsione per effetto delle mancate risposte totali, si è adottato l'approccio di calibrazione, secondo il quale il peso base campionario di ciascuna unità rispondente, pari all'inverso della corrispondente probabilità di inclusione del primo ordine, è corretto da un fattore scalare determinato applicando un modello di ottimizzazione vincolato. Il sistema dei vincoli rappresenta l'insieme dei totali noti nella popolazione, a cui il campione dei rispondenti è riportato attraverso il processo di calibrazione<sup>1</sup>.

Si definisce:  $\omega_{0,i} = \frac{N_h}{n_h} \forall i \in h$

come il peso base dell'unità i-esima appartenente allo strato h-esimo, che nel piano di campionamento adottato è pari al rapporto tra la popolazione dello strato ( $N_h$ ) e la corrispondente quota del campione allocata ( $n_h$ ). Il peso finale calibrato  $\omega_{c,i}$  può essere così formulato:

$$\omega_{c,i} = \lambda_i \cdot \omega_{0,i}$$

<sup>1</sup> La procedura seguita può essere rintracciata in Särndal, Lundström (2005) a cui si rimanda per gli approfondimenti teorici.

dove il fattore scalare correttivo  $\lambda_i$  è determinato imponendo un sistema di  $m$  vincoli corrispondenti ad altrettanti totali noti nella popolazione e descritti da  $m$  variabili ausiliarie. A questo insieme di vincoli si aggiunge il vincolo di non negatività del peso,  $\omega_{c,i} > 0 \forall i \in R$ , dove l'insieme  $R$  è composto dalle  $r$  unità rispondenti.

Utilizzando grandezze vettoriali, si ha:

$$\boldsymbol{\lambda} = \mathbf{1} + (\mathbf{X}_U - \mathbf{X}_r \boldsymbol{\omega}_0) \cdot (\mathbf{X}'_r \cdot \boldsymbol{\omega}_0 \cdot \mathbf{X}_r)^{-1} \cdot \mathbf{X}'_r$$

I simboli utilizzati hanno il seguente significato:

- $\boldsymbol{\lambda}$ , vettore (colonna)  $n_r$  dimensionale pari alla numerosità delle unità rispondenti all'indagine;
- $\mathbf{1}$  vettore unitario (colonna)  $n_r$  dimensionale;
- $\boldsymbol{\omega}_0$ , come già definito nel testo, rappresenta il vettore (colonna)  $n_r$  dimensionale del peso base di ciascuna unità rispondente;
- $\mathbf{X}_U$ , è il vettore (riga)  $m$ -dimensionale contenente gli  $m$  totali noti nella popolazione  $U$  dati come vincolo;
- $\mathbf{X}_r$ , è la matrice di dimensione  $n_r \times m$  che definisce su ciascuna cella  $i, j$  il valore della variabile ausiliaria  $j$ -esima, appartenente al sistema degli  $m$  vincoli, per l' $i$ -esima unità rispondente;
- $\mathbf{X}_r \boldsymbol{\omega}_0$  è il vettore (riga)  $m$ -dimensionale contenente gli  $m$  totali delle variabili ausiliarie che provengono applicando il peso base alle unità rispondenti;
- $\mathbf{X}_U - \mathbf{X}_r \boldsymbol{\omega}_0$ , rappresenta il vettore (riga)  $m$ -dimensionale dello scostamento tra i totali noti nella popolazione  $U$  e quelli ricavati dalle unità rispondenti all'indagine, ciascuna pesata con l'inverso della rispettiva probabilità di inclusione del primo ordine.

Eseguiti alcuni accorgimenti di natura tecnica legati alla invertibilità della matrice quadrata  $m$ -dimensionale  $\mathbf{X}'_r \cdot \boldsymbol{\omega}_0 \cdot \mathbf{X}_r$ , il vettore  $\boldsymbol{\lambda}$  è determinato di conseguenza.

La difficoltà principale della procedura non è di natura computazionale ma teorica e riguarda la determinazione delle variabili ausiliarie da inserire nella matrice  $\mathbf{X}_r$  e quindi la scelta dei totali noti della popolazione  $U$  da imporre nel sistema dei vincoli. La scelta delle variabili deve essere condotta attraverso un compromesso tra una buona predizione del processo di risposta/partecipazione all'indagine e una buona predizione della variabile outcome (la principale) dell'indagine. Quest'ultima è stata presa considerando una sintesi rispetto ad alcune dimensioni sul grado di soddisfazione. L'insieme delle variabili  $X$  utilizzate sono riportate nella tabella IV.5.

**Tabella IV.5 – Variabili ausiliarie**

Popolazione A - Presi in carico al 31.12.2015	Popolazione B - Non presi in carico al 31.12.2015
Regione presa in carico; quartili distributivi del profiling	Regione residenza; genere
Regione presa in carico; titolo di studio	Età; attesa dalla data di registrazione
Durata di attesa della presa in carico; titolo studio; genere	Genere; mobilità
Età; genere; condizione nel mercato del lavoro	Genere; età
Erogazione misura; genere	

*Nota: il punto e virgola indica l'incrocio delle variabili. Il totale noto imposto è quindi stabilito su ogni cella di intersezione delle variabili che nella tabella IV.5 sono separate dal punto e virgola.*

## ERRORE RELATIVO PER DOMINI DI STUDIO

Per stimare l'errore campionario associato allo stimatore di calibrazione si adotta l'approccio semplificato basato su modelli regressivi<sup>2</sup>. Sia  $\varepsilon^2(\tilde{y})$  il quadrato dell'errore relativo (varianza) della stima  $\tilde{y}$  in un modello di regressione lineare:

$$\tilde{y} = X \cdot \beta + \varepsilon$$

Per diversi valori di  $\tilde{y}$ , distribuiti su tutto il campo di variazione, si costruisce una nuvola di punti  $\tilde{y}$ ,  $\varepsilon^2(\tilde{y})^3$  e si adatta ad essa un modello di regressione log-lineare:

$$\log(\varepsilon^2(\tilde{y})) = a + \beta \cdot \log(\tilde{y}) + \mu$$

Ottenute le stime dei coefficienti  $\hat{\alpha}$ ,  $\hat{\beta}$  è possibile invertire l'equazione e isolare la varianza relativa la cui radice quadrata corrisponde all'errore campionario associato alla stima della frequenza  $p$  (vedi tabella IV.6).

La procedura sopra descritta può essere applicata anche per singoli domini di studio. La tavola IV.7 riporta gli errori relativi associati a diversi valori della frequenza stimata  $\hat{p}$  a livello complessivo e nei domini di studio, rappresentati dalle singole amministrazioni regionali e dalle quattro macro aree: Nord-ovest, Nord-est, Centro e Mezzogiorno.

<sup>2</sup> La metodologia seguita è quella Istat sviluppata per il software Genesees.

<sup>3</sup> Per le stime effettuate in questa rilevazione la nuvola di punti è composta da 42 osservazioni.

**Tabella IV.6 – Stime OLS dei parametri dell'equazione  $\log(\varepsilon^2(\tilde{y})) = \alpha + \beta \cdot \log(\tilde{y}) + \mu$** 

Dominio di studio	$\alpha$	$\beta$	$r^2$
Generale	6,517558	-1,369482	0,967500
Piemonte	3,912891	-1,434116	0,945149
Valled'Aosta*	-93,370979	14,213398	0,216535
Lombardia	6,269664	-1,474040	0,948556
P.A. di Trento	3,873702	-1,470868	0,961625
Veneto	5,914917	-1,395588	0,956630
Friuli Venezia Giulia	4,002604	-1,319968	0,941399
Liguria	3,245779	-1,346889	0,949525
Emilia Romagna	6,056813	-1,438506	0,957447
Toscana	5,294722	-1,355473	0,954417
Umbria	5,185134	-1,448136	0,956272
Marche	5,858586	-1,379071	0,966675
Lazio	4,761415	-1,317518	0,973313
Abruzzo	3,814605	-1,261277	0,966119
Molise	2,515149	-1,334887	0,944400
Campania	4,735570	-1,268637	0,969179
Puglia	4,471255	-1,276348	0,976594
Basilicata	5,004595	-1,340385	0,953554
Calabria	3,468861	-1,196900	0,977798
Sicilia	5,756917	-1,325639	0,974558
Sardegna	5,117439	-1,291628	0,962931
Nord-ovest	6,230954	-1,467575	0,952096
Nord-est	6,398677	-1,426106	0,958864
Centro	5,815382	-1,347710	0,957791
Mezzogiorno	5,818835	-1,328985	0,972173

\*per la Regione Valle d'Aosta le stime non sono significative.

Fonte: ISFOL, Rilevazione sul grado di soddisfazione del Programma Garanzia Giovani

**Tabella IV.7 –  $\varepsilon(\hat{p})$  per alcuni valori della stima di frequenza  $\hat{p}$  per domini di studio**

	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>	<b>0,15</b>	<b>0,2</b>	<b>0,25</b>	<b>0,3</b>	<b>0,35</b>	<b>0,4</b>	<b>0,45</b>	<b>0,5</b>
Generale	10,52%	3,49%	2,17%	1,65%	1,35%	1,16%	1,02%	0,92%	0,84%	0,78%	0,72%
Piemonte	42,82%	13,50%	8,21%	6,14%	5,00%	4,26%	3,74%	3,35%	3,04%	2,79%	2,59%
Lombardia	43,85%	13,39%	8,04%	5,96%	4,82%	4,09%	3,58%	3,19%	2,89%	2,65%	2,45%
P.A. di Trento	76,14%	23,31%	14,00%	10,39%	8,41%	7,14%	6,24%	5,57%	5,05%	4,63%	4,29%
Veneto	45,89%	14,93%	9,20%	6,93%	5,67%	4,86%	4,28%	3,84%	3,50%	3,22%	2,99%
Friuli Venezia Giulia	44,44%	15,36%	9,72%	7,44%	6,15%	5,31%	4,71%	4,25%	3,89%	3,60%	3,36%
Liguria	40,40%	13,67%	8,57%	6,52%	5,37%	4,62%	4,09%	3,69%	3,37%	3,11%	2,90%
Emilia Romagna	44,06%	13,84%	8,41%	6,28%	5,11%	4,35%	3,82%	3,42%	3,10%	2,85%	2,64%
Toscana	38,87%	13,06%	8,16%	6,20%	5,10%	4,39%	3,88%	3,49%	3,19%	2,95%	2,74%
Umbria	57,68%	17,99%	10,89%	8,12%	6,59%	5,61%	4,91%	4,40%	3,99%	3,66%	3,40%
Marche	65,95%	21,74%	13,48%	10,19%	8,36%	7,17%	6,32%	5,68%	5,18%	4,78%	4,44%
Lazio	26,93%	9,33%	5,91%	4,52%	3,74%	3,23%	2,87%	2,59%	2,37%	2,19%	2,05%
Abruzzo	32,24%	11,68%	7,55%	5,84%	4,87%	4,23%	3,77%	3,42%	3,15%	2,92%	2,74%
Molise	48,83%	16,68%	10,50%	8,01%	6,61%	5,70%	5,04%	4,55%	4,16%	3,85%	3,59%
Campania	30,26%	10,90%	7,02%	5,43%	4,52%	3,93%	3,50%	3,17%	2,91%	2,70%	2,53%
Puglia	29,73%	10,65%	6,84%	5,28%	4,40%	3,81%	3,39%	3,08%	2,82%	2,62%	2,45%
Basilicata	61,64%	20,96%	13,17%	10,04%	8,28%	7,13%	6,31%	5,69%	5,20%	4,81%	4,48%
Calabria	30,06%	11,47%	7,58%	5,95%	5,00%	4,38%	3,93%	3,58%	3,31%	3,08%	2,89%
Sicilia	26,99%	9,29%	5,87%	4,48%	3,71%	3,20%	2,83%	2,56%	2,34%	2,16%	2,02%
Sardegna	41,63%	14,72%	9,41%	7,24%	6,01%	5,21%	4,63%	4,19%	3,84%	3,56%	3,33%
Nord-ovest	33,07%	10,15%	6,10%	4,53%	3,67%	3,12%	2,73%	2,43%	2,21%	2,02%	1,87%
Nord-est	28,63%	9,09%	5,54%	4,15%	3,38%	2,88%	2,53%	2,27%	2,06%	1,90%	1,76%
Centro	22,67%	7,66%	4,80%	3,65%	3,01%	2,59%	2,29%	2,06%	1,89%	1,74%	1,62%
Mezzogiorno	13,51%	4,64%	2,93%	2,23%	1,85%	1,59%	1,41%	1,27%	1,16%	1,08%	1,00%

Fonte: ISFOL, Rilevazione sul grado di soddisfazione del Programma Garanzia Giovani

## IV.2 SIMULAZIONE PARTECIPAZIONE ALLA GARANZIA GIOVANI (CAPITOLO 4.3.2)

L'esercizio di simulazione proposto nel paragrafo 4.3.2 del Rapporto ha l'obiettivo di indagare le determinanti alla base del processo di partecipazione alla Garanzia Giovani da parte dei giovani NEET. Il confronto è quindi tra la popolazione in Garanzia Giovani e il totale della popolazione NEET di riferimento. La simulazione consiste nel proiettare la popolazione dei giovani iscritti in Garanzia Giovani all'interno della complessiva popolazione dei giovani NEET.

PopGG e PopNEET identificano le due popolazioni di partenza. La prima è rappresentata dagli oltre 456 mila giovani presi in carico all'interno della Garanzia Giovani di fonte BDPAPL; la seconda è rappresentata dalla popolazione teorica dei giovani NEET 15-29enni costituita dal campione Istat delle forze lavoro rappresentativa di una popolazione complessiva di circa 2,3 milioni di giovani. Si tratta della stessa popolazione utilizzata per la stima del modello econometrico del profiling.

Si considera l'insieme dei giovani presi in carico e non anche l'intera platea dei giovani registrati nella Garanzia Giovani sulla base delle seguenti considerazioni:

- la popolazione dei giovani registrati e non ancora presi in carico rappresenta un insieme che di per sé non dovrebbe essere dissimile rispetto a quanti sono (e saranno) presi in carico, al netto delle cancellazioni per mancanza di requisiti, ripensamento o altro. Rifiutare questa ipotesi varrebbe infatti ad assumere la possibile presenza di selezione (*creaming*) dei Servizi per il lavoro per il primo colloquio;
- è soltanto con la presa in carico che vengono acquisite quelle informazioni minime sulla persona necessarie per analizzare/studiare il fenomeno dell'autoselezione (una su tutte, l'indice di profiling del giovane).

Per le due popolazioni (PopGG e PopNEET) si è costruita una griglia di 88 celle ciascuna delle quali classifica il giovane NEET in base alle informazioni su: genere, età (quattro classi), area geografica di residenza (quattro classi) e titolo di studio (tre classi).

Per ciascuna cella *h-esima*, si sono rilevate le seguenti informazioni:

- la quota di giovani presi in carico ( $gg_h$ )

$$gg_h = \frac{\sum_{i \in h} PopGG_i}{\sum_{i \in h} PopNEET_i}$$

- i valori medi dell'indice di profilazione per le due popolazioni ( $\bar{p}_{gg,h}$  ;  $\bar{p}_{neet,h}$ ).

L'obiettivo è quello di estrarre dal campione delle Forze lavoro dell'Istat un campione "bilanciato" di giovani partecipanti alla garanzia giovani (P) e di giovani non partecipanti (NP). Il primo passo è stato quello di replicare il campione Istat sulla base del peso campionario di ciascun record. In questo modo si è ottenuto un database composto da circa 2,3 milioni di record ciascuno dei quali rappresenta un giovane NEET. Per ciascun record, oltre alle informazioni anagrafiche e a quelle proprie rilevate dalla rilevazione delle Forze lavoro, è riportato il corrispondente indice di profilazione.

Sulla base della quota  $gg_h$  si è dapprima determinato un tasso di campionamento costante (pari a circa il 70%) sia per la popolazione P che per la popolazione NP. Successivamente si sono determinate le

numerosità campionarie, per ciascuno degli 88 strati  $h$  (celle della griglia), dei giovani in Garanzia Giovani ( $n_{p,h}$ ). Per ciascuna cella  $h$ -esima si è proceduto quindi a replicare 200 campioni casuali indipendenti e per ogni campione  $s$ -esimo si è calcolato:

- l'indice di profilazione medio ( $\hat{p}_{s,h}$ );
- l'errore relativo, in valore assoluto, dell'indice di profilazione  $\hat{p}_{s,h}$  rispetto all'indice di profilazione medio della cella  $h$ -esima osservato in PopGG ( $\bar{p}_{gg,h}$ )

$$\epsilon_{s,h} = \frac{|\hat{p}_{s,h} - \bar{p}_{gg,h}|}{\bar{p}_{gg,h}}$$

Per ciascuno strato  $h$ -esimo, tra i 200 campioni si è selezionato il campione  $\bar{s}$  tale che:

$$\bar{s} = \min_{s=1,200} \{\epsilon_{s,h}\} \text{ e } \epsilon_{\bar{s},h} < 0,06$$

cioè, il campione con il più basso errore relativo in valore assoluto nei livelli medi dell'indice di profilazione che soddisfa la condizione che lo stesso errore relativo sia inferiore a 0,06 (6%).

Nei casi in cui l'errore relativo minimo dello strato  $h$  è risultato essere superiore al 6% si è proceduto ad una replica di ulteriori 200 campioni restringendo l'insieme di estrazione ai record che presentano livelli di profilazione più bassi del terzo quartile (nel caso di errore relativo positivo) o più alti del primo quartile (nel caso di errore relativo negativo).

Per quanto riguarda l'estrazione del campione della popolazione NP, questa è avvenuta in maniera casuale sulla popolazione NEET residua, senza alcun vincolo con la griglia precostituita delle celle  $h$ .

Nota. Per costruzione uno stesso record nel database sezionale dell'indagine sulle Forze Lavoro è replicato tante volte quanto è il relativo peso campionario. Questo vuol dire che uno stesso record può comparire  $n$ -volte nel campione P e  $m$ -volte nel campione NP. Il rapporto tra  $n$  e  $m$  è in accordo con  $gg_h$ .

Questa considerazione ha effetti sul calcolo della deviazione standard delle stime.

Il processo appena descritto è stato applicato su tre diversi database:

- database sezionale, utilizzato per le analisi sulle determinanti dell'autoselezione alla partecipazione alla Garanzia Giovani;
- database panel a 3 mesi, utilizzato per le analisi delle differenze negli inserimenti a 3 mesi tra P e NP, sotto l'ipotesi di assenza della Garanzia Giovani;
- database panel a 12 mesi, utilizzato per le analisi delle differenze negli inserimenti a 12 mesi tra P e NP, sotto l'ipotesi di assenza della Garanzia Giovani.

I dati sezionali coprono il periodo che va dal IV trimestre 2011 al III trimestre 2013. I dati panel hanno come base ( $t=0$ ) il database sezionale (vedi tabella IV.8).

**Tabella IV.8 – Periodo di riferimento della simulazione**

	Panel		
	Sezionale (t=0)	3 mesi	12 mesi
	da	a	a
IV_TRIM_2011	I_TRIM_2012	IV_TRIM_2012	
I_TRIM_2012	II_TRIM_2012	I_TRIM_2013	
II_TRIM_2012	III_TRIM_2012	II_TRIM_2013	
III_TRIM_2012	IV_TRIM_2012	III_TRIM_2013	
IV_TRIM_2012	I_TRIM_2013	IV_TRIM_2013	
I_TRIM_2013	II_TRIM_2013	I_TRIM_2014	
II_TRIM_2013	III_TRIM_2013	II_TRIM_2014	
III_TRIM_2013	IV_TRIM_2013	III_TRIM_2014	
<b>Numero record</b>	<b>40.556</b>	<b>17.252</b>	<b>15.034</b>

Fonte: ISFOL, Rilevazione sul grado di soddisfazione del Programma Garanzia Giovani

Per il database sezionale la simulazione ha prodotto un campione di 319 mila giovani partecipanti alla Garanzia Giovani e 1,4 milioni di giovani non partecipanti in Garanzia. Il campione simulato dei partecipanti alla Garanzia presenta un indice di profilazione medio di 0,6746 contro il valore osservato di 0,6739 con un errore relativo complessivo dello 0,08%.

Per i Panel, l'adattamento della distribuzione dell'indice di profilazione del campione simulato rispetto alla popolazione osservata dei giovani presi in carico in Garanzia Giovani appare meno preciso e più soggetto a volatilità soprattutto se considerato per singola area geografica. In effetti a fronte di 40.566 record del dato sezionale, sono 17.252 i record che costituiscono il panel trimestrale e 15.034 quelli del panel annuale.

### IV.3 VALUTAZIONE DI IMPATTO CON APPROCCIO CONTROFATTUALE (CAPITOLO 4.3.3)

#### BREVI CENNI DI TEORIA

Obiettivo della valutazione di impatto (di efficacia) di un programma (o di una policy) è quello di valutare l'effetto netto diretto (causazione) della politica sull'outcome osservato. Definita la variabile risultato "y" (nel rapporto è la variabile dicotomica "Non occupato=0", "Occupato=1") e ottenuta una misura di *outcome* (nel caso in esame il valore medio di "y" che coincide con la probabilità di trovare un lavoro), l'obiettivo è quello di valutare l'effetto netto come differenza tra l'outcome realizzato (situazione fattuale) e quello che si sarebbe realizzato in assenza della politica (situazione controfattuale).

Definita con "D" la variabile trattamento che assume il valore "0" nel caso di assenza della politica e il valore "1" nel caso di presenza della politica, e con  $y_0$  e  $y_1$  il valore dell'outcome rispettivamente in assenza o in presenza di trattamento, si definiscono tre diversi effetti netti della politica:

- effetto medio del trattamento  $ATE = E(y|D = 1) - E(y|D = 0)$
- effetto medio del trattamento sui trattati  $ATET = E(y_1|D = 1) - E(y_0|D = 1)$
- effetto medio del trattamento sui non trattati  $ATENT = E(y_1|D = 0) - E(y_0|D = 0)$

L'ATE misura l'effetto medio del trattamento sulla popolazione complessiva ammissibile al programma; l'ATET misura l'effetto medio della politica sulla popolazione dei trattati; l'ATENT misura l'effetto medio che la politica avrebbe avuto sulla parte della popolazione ammissibile non trattata.

Vale la seguente relazione:

$$ATE = p(D = 1) \cdot ATET + p(D = 0) \cdot ATENT \quad [1]$$

cioè l'effetto medio del trattamento (ATE) è una media ponderata degli effetti ATET e ATENT con pesi pari rispettivamente alla probabilità di essere trattato  $p(D=1)$  e alla probabilità di non essere trattato  $p(D=0)$ . I due termini  $E(y_0|D = 1)$  e  $E(y_1|D = 0)$  nella formulazione rispettivamente dell'ATET e dell'ATENT non sono osservabili: il primo rappresenta il valore dell'outcome in assenza del trattamento ( $y_0$ ) degli individui trattati ( $D=1$ ); il secondo rappresenta il valore dell'outcome in presenza del trattamento ( $y_1$ ) degli individui non trattati ( $D=0$ ).

Si distinguono due categorie di disegni di valutazione:

- disegni sperimentali (a cui possono essere ricondotti anche i cd. disegni quasi-sperimentali);
- disegni non sperimentali.

I primi prevedono una procedura randomizzata che dalla popolazione totale ammissibile definisce l'insieme degli individui sottoposti alla politica ( $D=1$ ) e quelli non sottoposti alla politica ( $D=0$ ) in maniera completamente casuale. Proprio questa procedura di estrazione casuale da un'unica

popolazione assicura l'equivalenza statistica tra il gruppo dei trattati e quello di confronto che a sua volta determina la validità delle seguenti condizioni:

$$\begin{cases} E(y_1|D = 1) = E(y_1|D = 0) \\ E(y_0|D = 1) = E(y_0|D = 0) \end{cases} \quad [2]$$

Ovvero, l'outcome al termine del trattamento dei trattati  $E(y_1|D = 1)$  è pari all'outcome che avrebbero realizzato gli individui non trattati in caso di trattamento  $E(y_1|D = 0)$ ; allo stesso modo per la seconda condizione che esprime l'uguaglianza dell'outcome in assenza del trattamento ( $y_0$ ) sia per il gruppo dei trattati ( $D=1$ ) che per il gruppo dei non trattati ( $D=0$ ). La [2] costituisce la condizione di indipendenza dell'outcome dal trattamento (Independence Assumption, IA).

Questa condizione è chiaramente una conseguenza dell'equivalenza statistica delle due popolazioni dei trattati e dei controlli e consente di sostituire, nelle formulazioni dell'ATET e dell'ATENT i termini non osservabili  $E(y_0|D = 1)$  e  $E(y_1|D = 0)$  con i rispettivi termini osservati  $E(y_0|D = 0)$  e  $E(y_1|D = 1)$ :

$$ATET = E(y_1|D = 1) - E(y_0|D = 1) = E(y_1|D = 1) - E(y_0|D = 0) = E(y_1|D = 0) - E(y_0|D = 0) = ATENT \quad [3]$$

e quindi, in virtù della [1], si ha l'equivalenza tra i tre effetti  $ATE=ATET=ATENT$ .

Nei disegni sperimentali l'effetto netto della politica è allora dato dalla differenza in media (DIM) nei valori degli outcome osservati per i due gruppi:

$$ATE = E(y|D = 1) - E(y|D = 0) = E(y_1) - E(y_0) = DIM \quad [4]$$

Nei disegni non sperimentali l'assegnazione della politica nella popolazione ammissibile non avviene in maniera casuale (randomizzata) ma è il risultato di processi di (auto)selezione da parte del singolo individuo ammissibile.

L'autoselezione è propria di ogni politica (programma) ad adesione volontaria e quindi anche del programma Garanzia Giovani. La selezione è propria del processo di assegnazione o meno alla politica sulla base di pre-requisiti il più delle volte deterministici (punteggio nelle graduatorie, ecc.).

Nei disegni non sperimentali cade l'ipotesi di indipendenza dell'outcome dal trattamento:

$$\begin{aligned} E(y_1|D = 1) &\neq E(y_1|D = 0) \\ E(y_0|D = 1) &\neq E(y_0|D = 0) \end{aligned}$$

La prima conseguenza della violazione della IA è che in generale  $ATE \neq ATET \neq ATENT$ . La semplice differenza delle medie (DIM) non può rappresentare una misura corretta dell'effetto netto del trattamento data la presenza di fattori di distorsione dovuti alla non casualità dei processi di (auto)selezione (selection bias):

$$\begin{aligned} ATE &= E(y_1|D = 1) - E(y_0|D = 1) + Bias_1 = ATET + Bias_1 \\ ATE &= E(y_1|D = 0) - E(y_0|D = 0) + Bias_0 = ATENT + Bias_0 \end{aligned} \quad [5]$$

Il problema della stima degli effetti netti della politica si riduce a quello della stima delle due quantità non osservabili  $E(y_1|D = 0)$  e  $E(y_0|D = 1)$ , che a sua volta può essere ricondotto a un problema di imputazione di dati mancanti (missing value imputation).

Nel caso dell'ATET ad esempio, si ha:

$$ATE_T = \frac{1}{N_1} \cdot \sum_{i=1}^{N_1} (y_{1i} - \hat{y}_{0i}) = E(y_{1i} - \hat{y}_{0i} | D = 1) \quad [6]$$

dove  $N_1$  rappresenta il numero delle unità trattate nella popolazione e  $\hat{y}_{0i}$  rappresenta l'outcome imputato (stimato) che la singola unità  $i$ -esima trattata avrebbe ottenuto in assenza del trattamento. La stima di questo valore non può non essere "ricavata" da quanto è osservabile sul gruppo dei non trattati (gruppo di controllo o di confronto) :

$$\{y_{0,j}; \forall j \in D = 0\} \rightarrow \hat{y}_{0i} \forall i \in D = 1 \quad [7]$$

Lo stesso ragionamento vale per il caso dell'ATENT.

Sono stati sviluppati una serie di modelli diversi per la stima dell'ATE di una politica. Una prima distinzione è quella tra modelli parametrici e non parametrici. I primi si fondano sul presupposto che esiste una determinata forma funzionale che lega l'outcome al trattamento, nonché sono assunte le forme distributive sia dell'outcome che dell'errore. I metodi non parametrici non si fondano sull'assunzione di forme funzionali note o su ipotesi distributive dell'outcome, ma risolvono il problema dell'imputazione attraverso procedure che si basano sul confronto tra unità ritenute "più vicine" appartenenti alle due sottopopolazioni dei trattati ( $D=1$ ) e dei non trattati ( $D=0$ ).

Più importante è invece la distinzione della stima dell'ATE a seconda della natura del *bias* insito nelle [5]. Se la distorsione da (auto)selezione è spiegata completamente da un insieme di variabili ( $X$ ) osservabili (*selection on observables*) oppure se nella selezione agisce anche un insieme ulteriore di variabili non osservabili (*selection on unobservables*) che determinano quindi endogeneità del trattamento.

Se si accetta la prima ipotesi, allora vale la condizione di indipendenza in media dell'outcome condizionata alle variabili  $X$  (Conditional Mean Independence assumption, CMI):

$$\begin{cases} E(y_1|D = 1, X) = E(y_1|D = 0, X) = E(y_1|X) \\ E(y_0|D = 1, X) = E(y_0|D = 0, X) = E(y_0|X) \end{cases} \quad [8]$$

I principali metodi non parametrici (matching, metodi di riponderazione) si basano sulla validità dell'ipotesi di *selection on observables* e dunque sulla validità dell'assunto della CMI. Tra i principali metodi di stima validi in presenza di *selection on unobservables* ci sono gli stimatori IV (*Instrumental Variables*), DID (*Difference-in-differences*), e RDD (*Regression Discontinuity Design*).

## BASE DEI DATI

La banca dati utilizzata per le elaborazioni del capitolo 4 è composta da quattro parti:

1. Anagrafica\_SAP
2. Garanzia\_Giovani
3. Esito\_conclusi a 4 settimane e a 6 mesi dalla conclusione dell'intervento
4. Comunicazioni\_Obbligatorie

### Anagrafica\_SAP

Questo database rappresenta l'insieme "universo" degli individui che sono d'interesse per l'attività di valutazione della Garanzia Giovani, ricomprendendo in questa finalità anche il calcolo degli indicatori di risultato. Esso è costituito da due insiemi di individui:

- gli iscritti a Garanzia Giovani, inclusi i cancellati d'ufficio (U,D,X,C) (insieme 1.A)
- giovani fino a 32 anni non iscritti a Garanzia Giovani che, nel periodo che va dal 01/01/2010 al 30/09/2015 hanno almeno una DID attiva (insieme 1.B)

Questo database raccoglie le informazioni direttamente dalle Schede anagrafiche professionali (SAP) sezioni "0", "1", "2", "5".

Il campo chiave è l'identificativo del soggetto (in relazione 1:1 con il codice fiscale dell'individuo intestatario della SAP)

### Garanzia\_Giovani

Questo database raccoglie le informazioni sul sottoinsieme di individui iscritti a GG (insieme 1.A di cui sopra). Per ciascuna iscrizione dell'individuo (possono esserci più cicli in capo allo stesso identificativo) si rileva lo stato del giovane all'interno del ciclo corrente attraverso le seguenti 4 variabili.

COD_STATO_ADESIONE_PC
IDDATA_STATOADESIONE
FLAG_PRESO_CARICO
CONCLUSO

In particolare la variabile CONCLUSO rappresenta una variabile dummy che identifica (=1) l'individuo che nel periodo 01/05/2014-30/09/2015 ha concluso (data\_di\_fine sez. 6 SAP compresa nel periodo di riferimento) una politica attiva (scheda 2A, 2B, 4A, 4C, 5, 6, 7, 8), ricomprendendo anche la scheda 3 se associata ad una successiva attivazione di un contratto di lavoro (CO), oppure che è uscito dal programma GG successivamente alla presa in carico (stato R, C da P).

Il database oltre ad integrare alcune informazioni di tipo anagrafico non presenti nel database Anagrafica\_SAP, contiene per intero la sezione 6 della SAP legata all'ultima attività, in ordine cronologico, avviata e/o conclusa dal giovane. Ad integrazione della sezione 6 della SAP, il

database raccoglie anche informazioni relative all'avvio di un tirocinio, successivo alla presa in carico, di fonte CO.

#### Esito\_conclusi

Questo database è strettamente legato alla quantificazione degli indicatori di risultato immediato (IRI3\_OCC) e di lungo periodo (ILP2) del PON IOG. L'universo di riferimento del database è costituito dal sottoinsieme di individui che hanno la variabile CONCLUSO=1 nel precedente database (SAP\_GG).

A partire dalla data di conclusione (o di uscita) dalla GG dell'identificativo, si calcolano due date di riferimento (T1) e (T2) pari rispettivamente alla "data di fine/uscita più 4 settimane" e "data di fine/uscita più 6 mesi".

Relativamente a queste due date di riferimento il database registra la presenza o meno di un rapporto di lavoro in essere.

#### Comunicazioni\_Obbligatorie

Questo database è finalizzato alla possibilità di condurre una valutazione di efficacia del PON IOG. Esso mira a ricostruire una storia occupazionale di tutti gli individui ricompresi nel database Anagrafica\_SAP (insieme 1.A e insieme 1.B).

In particolare si richiedono per ogni identificativo:

- informazioni minime sul primo contratto di lavoro attivato a partire dal 01 gennaio 2010
- informazioni minime sul primo contratto di lavoro attivato a partire dal 01 maggio 2015
- informazioni più approfondite sull'ultimo contratto di lavoro attivato (anche se non più in essere).

Il database comprende, inoltre, un insieme variabili calcolate di tipo "contatore" sul numero di attivazioni, cessazioni, trasformazioni e proroghe complessivamente registrate nell'intero arco di riferimento 01/01/2010 – 30/09/2015, fino ad un dettaglio mensile nel periodo di entrata in funzione della GG (maggio 2014-settembre 2015).

**Tabella IV.9 - Tracciato record Anagrafica\_SAP**

<b>cod_SAP</b>	<b>variabile</b>	<b>descrizione</b>
	<b>identificativo</b>	identificativo del soggetto (codice fiscale). Il campo è chiave unica in questa tabella.
0.1.b	data ultimo aggiornamento	
0.1.c	identificativo SAP	
0.1.d	codice ente titolare	
0.1.e	tipo variazione	
0.1.f	data di nascita	
1.1.d	sexso	
1.1.e	data di nascita	
1.1.f	comune (o stato estero) di nascita	
1.1.g	cittadinanza	
1.3.1.a	comune di residenza	
1.3.1.b	CAP residenza	
1.3.2.a	codice del comune di domicilio	
1.3.2.e	CAP domicilio	
2.1.1.a	stato occupazionale	
2.1.1.c	anzianità di disoccupazione (mesi)	
2.1.1.d	data dichiarazione di disponibilità	
2.1.2.a	data ingresso	
2.1.2.b	tipo ingresso	
5.1.a	livello scolarizzazione	
5.1.b	corso di studio	
5.1.c	descrizione	
5.1.d	frequentato in	
5.1.f	se conseguito: anno	
5.1.g	votazione conseguita	
5.1.h	altrimenti: ultimo anno frequentato	
5.1.i	anno di frequenza (in corso)	
5.2.a	titolo corso di formazione	
5.2.d	durata	
5.2.f	certificazioni ed attestazioni	

**Tabella IV.10 - Tracciato record Garanzia\_Giovani**

cod_S AP	variabile	descrizione
	<b>Identificativo</b>	Identificativo del soggetto (codice fiscale). Il campo non è chiave unica del database. Un identificativo compare nel db tante volte quante sono i cicli (n. prese in carico) del giovane all'interno della GG.
	FLAG_PON_IOG	1=PON IOG (tipo progetto=02); 0=NON PON IOG (tipo progetto=01, 03)
	COD_REGIONE_PRIMA_REGISTR	Regione prima registrazione
	COD_STATO_ADESIONE_PC	Valori ammessi D, U, X, A, C, P, R, T e F
	IDDATA_STATOADESIONE	ID Data di aggiornamento dello stato
	FLAG_PRESO_CARICO	Dummy 1 se l'individuo è preso in carico; 0 se non ancora preso in carico
	COD_PROVINCIA_PRESA_CARICO	Provincia di presa in carico
	COND_MERCATO_LAVORO_PROFILING	Condizione nel mercato del lavoro (variabile inserita nel modello di profilazione)
	TITOLO_STUDIO_PROFILING	Titolo di studio (variabile inserita nel modello di profilazione)
	INDICE_PROFILING	Livello di profilazione (4 classi) da dashboard
	PROBAB_PROFILING_ESTR	Valore puntuale dell'indice di profilazione
	TITOLO_STUDIO_DASHBOARD	Titolo di studio dashboard
	STATUS_OCCUP_INGRESSO (DASHBOARD)	Condizione mercato del lavoro alla presa in carico (variabile dashboard)
	FLAG_CANCELLATI	1= se giovane risulta cancellato dal Programma; 0 = se attivo
	CONCLUSO	Dummy = 1 se "data_di_fine" ultima politica attiva <30.09.2015 ; = 1 se stato adesione FLAG_PRESO_CARICO=1 AND COD_STATO ADESIONE_PC=R, C = 0 altrimenti
<b>6.1.a</b>	Attività	Si riferisce alla sezione 6 dell'ultima politica attiva inserita ---> "Data di fine" più recente.
<b>6.1.b</b>	Denominazione	
<b>6.1.l</b>	Data proposta	
<b>6.1.c</b>	Data inizio	
<b>6.1.d</b>	Data di fine	
<b>6.1.e</b>	Durata	
<b>6.1.f</b>	Tipologia Durata	
<b>6.1.g</b>	Descrizione	
<b>6.1.h</b>	Titolo Progetto	
<b>6.1.i</b>	Codice Ente Promotore	
	CO_TIR	Presenza di una CO per un tirocinio avviato successivamente alla presa in carico
	CO_data_avvio_tir	Data avvio del tirocinio
	CO_data_fine_tir	Data fine del tirocinio
	SCHEDA_1C	Flag 1= presenza scheda 1-C; 0=assenza scheda 1-C
	SCHEDA_2A	Flag 1= presenza scheda 2A; 0=assenza scheda 2A
	SCHEDA_2B	Flag 1= presenza scheda 2B; 0=assenza scheda 2B
	SCHEDA_3	Flag 1= presenza scheda 3; 0=assenza scheda 3
	SCHEDA_4A	Flag 1= presenza scheda 4A; 0=assenza scheda 4A
	SCHEDA_4B	Flag 1= presenza scheda 4B; 0=assenza scheda 4B
	SCHEDA_4C	Flag 1= presenza scheda 4C; 0=assenza scheda 4C
	SCHEDA_5	Flag 1= presenza scheda 5; 0=assenza scheda 5
	SCHEDA_6	Flag 1= presenza scheda 6; 0=assenza scheda 6
	SCHEDA_7	Flag 1= presenza scheda 7; 0=assenza scheda 7
	SCHEDA_8	Flag 1= presenza scheda 8; 0=assenza scheda 8
	SCHEDA_9	Flag 1= presenza scheda 9; 0=assenza scheda 9

**Tabella IV.11 - Tracciato record Esito\_conclusi**

variabile	descrizione
identificativo	identificativo del soggetto (codice fiscale). SOLO SE "CONCLUSO==1". Il campo è chiave unica in questa tabella.
data_T1	t1 = ("Data di fine Sap 6.1.d" or ("CO_data_fine_TIR" se "CO_data_fine_TIR" >"Data di fine Sap 6.1.d")) + 4 settimane; 0 altrimenti
data_T2	t2 = ("Data di fine Sap 6.1.d" or ("CO_data_fine_TIR" se "CO_data_fine_TIR" >"Data di fine Sap 6.1.d")) + 6 mesi; 0 altrimenti
CO_RIFERIMENTO_T1	1=se CO attiva (in essere) in data t1
SEDE_LAVORO_T1	sede lavoro (comune) in t1
SETTORE_T1	settore ateco in t1
TIPO_CONTRATTO_T1	tipo contratto in data t1
TIPO_ORARIO_T1	tipologia di lavoro t1
QUALIFICA_L4_T1	qualifica in t1
SEDE_LAVORO_T2	sede lavoro (comune) in t2
SETTORE_T2	settore ateco in t2
CO_RIFERIMENTO_T2	1=se CO attiva (in essere) in data t2
TIPO_CONTRATTO_T2	tipo contratto in data t2
TIPO_ORARIO_T2	tipologia di lavoro t2
QUALIFICA_L4_T2	qualifica in t2

**Tabella IV.12 - tracciato record Comunicazioni\_Obbligatorie**

<b>variabile</b>	<b>descrizione</b>
<b>identificativo</b>	identificativo del soggetto (codice fiscale). Il campo è chiave unica in questa tabella.
IDDATA_ESTRAZIONE	ID Data estrazione
IDDATA_INIZIO_RAPPORTOprima_CO (dal 01.01.2010)	data inizio rapporto prima CO associata all'identificativo a partire dal primo gennaio 2010
cod_prima_CO_attiva	Codice della prima CO associata all'identificativo a partire dal primo gennaio 2010
SETTORE_LAVORO_primaCO	Settore ateco associato alla prima CO dell'identificativo a partire dal primo gennaio 2010
COMUNE_SEDE_LAVORO_primaCO	Sede di lavoro (comune) associata alla prima CO dell'identificativo a partire dal primo gennaio 2010
TIPO_CONTRATTO_primaCO	Tipo di contratto corrispondente alla prima CO dell'identificativo a partire dal primo gennaio 2010
TIPO_ORARIO_LAVORO_primaCO	Tipologia di lavoro associata alla prima CO dell'identificativo a partire dal primo gennaio 2010
QUALIFICA_L4_primaCO	Qualifica associata alla prima CO dell'identificativo a partire dal primo gennaio 2010
n_attivazioni (dal 01.01.2010)	Conteggio del numero delle attivazioni nel periodo 01.01.2010 - 30.09.2015 associate al nominativo
n_cessazioni (dal 01.01.2010)	Conteggio del numero delle cessazioni nel periodo 01.01.2010 - 30.09.2015 associate al nominativo
n_trasformazioni (dal 01.01.2010)	Conteggio del numero delle trasformazioni nel periodo 01.01.2010 - 30.09.2015 associate al nominativo
n_proroghe (dal 01.01.2010)	Conteggio del numero delle proroghe nel periodo 01.01.2010 - 30.09.2015 associate al nominativo
IDDATA_INIZIO_RAPPORTO_primaCO_GG (dal 01.05.2014)	Data inizio rapporto prima CO associata all'identificativo a partire dal primo maggio 2014 (avvio del piano GG)
IDCF_DATORELAVORO_primaCO_GG	Identificativo (codice fiscale) del datore di lavoro associato alla prima CO attiva dell'identificativo nel periodo 01.05.2014-30.09.2015
COMUNE_SEDE_LAVORO_primaCO_GG	Sede di lavoro (comune) associata alla prima CO dell'identificativo dal primo maggio 2014 al 30 settembre 2015
SETTORE_primaCO_GG	Settore ateco associato alla prima CO dell'identificativo dal primo maggio 2014 al 30 settembre 2015
TIPO_CONTRATTO_primaCO_GG	Tipo di contratto corrispondente alla prima CO dell'identificativo dal primo maggio 2014 al 30 settembre 2015
TIPO_ORARIO_primaCO_GG	Tipologia di lavoro associata alla prima CO dell'identificativo dal primo maggio 2014 al 30 settembre 2015
QUALIFICA_L4_primaCO_GG	Qualifica associata alla prima CO dell'identificativo dal primo maggio 2014 al 30 settembre 2015
TOT_nuove_attivazioni_2014-05	Numero complessivo di attivazioni dell'identificativo nel mese di maggio 2014
TIR_nuove_attivazioni_2014-05	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di maggio 2014
INC_nuove_attivazioni_2014-05	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di maggio 2014
TOT_nuove_attivazioni_2014-06	Numero complessivo di attivazioni dell'identificativo nel mese di giugno 2014
TIR_nuove_attivazioni_2014-06	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di giugno 2014
INC_nuove_attivazioni_2014-06	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di giugno 2014
TOT_nuove_attivazioni_2014-07	Numero complessivo di attivazioni dell'identificativo nel mese di luglio 2014
TIR_nuove_attivazioni_2014-07	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di luglio 2014
INC_nuove_attivazioni_2014-07	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di luglio 2014
TOT_nuove_attivazioni_2014-08	Numero complessivo di attivazioni dell'identificativo nel mese di agosto 2014
TIR_nuove_attivazioni_2014-08	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di agosto 2014
INC_nuove_attivazioni_2014-08	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di agosto 2014

<b>variabile</b>	<b>descrizione</b>
TOT_nuove_attivazioni_2014-09	Numero complessivo di attivazioni dell'identificativo nel mese di settembre 2014
TIR_nuove_attivazioni_2014-09	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di settembre 2014
INC_nuove_attivazioni_2014-09	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di settembre 2014
TOT_nuove_attivazioni_2014-10	Numero complessivo di attivazioni dell'identificativo nel mese di ottobre 2014
TIR_nuove_attivazioni_2014-10	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di ottobre 2014
INC_nuove_attivazioni_2014-10	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di ottobre 2014
TOT_nuove_attivazioni_2014-11	Numero complessivo di attivazioni dell'identificativo nel mese di novembre 2014
TIR_nuove_attivazioni_2014-11	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di novembre 2014
INC_nuove_attivazioni_2014-11	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di novembre 2014
TOT_nuove_attivazioni_2014-12	Numero complessivo di attivazioni dell'identificativo nel mese di dicembre 2014
TIR_nuove_attivazioni_2014-12	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di dicembre 2014
INC_nuove_attivazioni_2014-12	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di dicembre 2014
TOT_nuove_attivazioni_2015-01	Numero complessivo di attivazioni dell'identificativo nel mese di gennaio 2015
TIR_nuove_attivazioni_2015-01	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di gennaio 2015
INC_nuove_attivazioni_2015-01	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di gennaio 2015
TOT_nuove_attivazioni_2015-02	Numero complessivo di attivazioni dell'identificativo nel mese di febbraio 2015
TIR_nuove_attivazioni_2015-02	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di febbraio 2015
INC_nuove_attivazioni_2015-02	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di febbraio 2015
TOT_nuove_attivazioni_2015-03	Numero complessivo di attivazioni dell'identificativo nel mese di marzo 2015
TIR_nuove_attivazioni_2015-03	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di marzo 2015
INC_nuove_attivazioni_2015-03	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di marzo 2015
TOT_nuove_attivazioni_2015-04	Numero complessivo di attivazioni dell'identificativo nel mese di aprile 2015
TIR_nuove_attivazioni_2015-04	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di aprile 2015
INC_nuove_attivazioni_2015-04	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di aprile 2015
TOT_nuove_attivazioni_2015-05	Numero complessivo di attivazioni dell'identificativo nel mese di maggio 2015
TIR_nuove_attivazioni_2015-05	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di maggio 2015
INC_nuove_attivazioni_2015-05	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di maggio 2015
TOT_nuove_attivazioni_2015-06	Numero complessivo di attivazioni dell'identificativo nel mese di giugno 2015
TIR_nuove_attivazioni_2015-06	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di giugno 2015
INC_nuove_attivazioni_2015-06	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di giugno 2015
TOT_nuove_attivazioni_2015-07	Numero complessivo di attivazioni dell'identificativo nel mese di luglio 2015
TIR_nuove_attivazioni_2015-07	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di luglio 2015
INC_nuove_attivazioni_2015-07	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di luglio 2015
TOT_nuove_attivazioni_2015-08	Numero complessivo di attivazioni dell'identificativo nel mese di agosto 2015
TIR_nuove_attivazioni_2015-08	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di

<b>variabile</b>	<b>descrizione</b>
	agosto 2015
INC_nuove_attivazioni_2015-08	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di agosto 2015
TOT_nuove_attivazioni_2015-09	Numero complessivo di attivazioni dell'identificativo nel mese di settembre 2015
TIR_nuove_attivazioni_2015-09	Numero di attivazioni con tipo contratto ="TIROCINIO" dell'identificativo nel mese di settembre 2015
INC_nuove_attivazioni_2015-09	Numero di attivazioni "Incentivate", ad esclusione dei tirocini, dell'identificativo nel mese di settembre 2015
n_nuove_cessazioni_2014-5	Numero complessivo di cessazioni dell'identificativo nel mese di maggio 2014
n_nuove_cessazioni_2014-6	Numero complessivo di cessazioni dell'identificativo nel mese di giugno 2014
n_nuove_cessazioni_2014-7	Numero complessivo di cessazioni dell'identificativo nel mese di luglio 2014
n_nuove_cessazioni_2014-8	Numero complessivo di cessazioni dell'identificativo nel mese di agosto 2014
n_nuove_cessazioni_2014-9	Numero complessivo di cessazioni dell'identificativo nel mese di settembre 2014
n_nuove_cessazioni_2014-10	Numero complessivo di cessazioni dell'identificativo nel mese di ottobre 2014
n_nuove_cessazioni_2014-11	Numero complessivo di cessazioni dell'identificativo nel mese di novembre 2014
n_nuove_cessazioni_2014-12	Numero complessivo di cessazioni dell'identificativo nel mese di dicembre 2014
n_nuove_cessazioni_2015-01	Numero complessivo di cessazioni dell'identificativo nel mese di gennaio 2015
n_nuove_cessazioni_2015-02	Numero complessivo di cessazioni dell'identificativo nel mese di febbraio 2015
n_nuove_cessazioni_2015-03	Numero complessivo di cessazioni dell'identificativo nel mese di marzo 2015
n_nuove_cessazioni_2015-04	Numero complessivo di cessazioni dell'identificativo nel mese di aprile 2015
n_nuove_cessazioni_2015-05	Numero complessivo di cessazioni dell'identificativo nel mese di maggio 2015
n_nuove_cessazioni_2015-06	Numero complessivo di cessazioni dell'identificativo nel mese di giugno 2015
n_nuove_cessazioni_2015-07	Numero complessivo di cessazioni dell'identificativo nel mese di luglio 2015
n_nuove_cessazioni_2015-08	Numero complessivo di cessazioni dell'identificativo nel mese di agosto 2015
n_nuove_cessazioni_2015-09	Numero complessivo di cessazioni dell'identificativo nel mese di settembre 2015
n_nuove_trasformazioni_2014-05	Numero complessivo di trasformazioni dell'identificativo nel mese di maggio 2014
n_nuove_trasformazioni_2014-06	Numero complessivo di trasformazioni dell'identificativo nel mese di giugno 2014
n_nuove_trasformazioni_2014-07	Numero complessivo di trasformazioni dell'identificativo nel mese di luglio 2014
n_nuove_trasformazioni_2014-08	Numero complessivo di trasformazioni dell'identificativo nel mese di agosto 2014
n_nuove_trasformazioni_2014-09	Numero complessivo di trasformazioni dell'identificativo nel mese di settembre 2014
n_nuove_trasformazioni_2014-10	Numero complessivo di trasformazioni dell'identificativo nel mese di ottobre 2014
n_nuove_trasformazioni_2014-11	Numero complessivo di trasformazioni dell'identificativo nel mese di novembre 2014
n_nuove_trasformazioni_2014-12	Numero complessivo di trasformazioni dell'identificativo nel mese di dicembre 2014
n_nuove_trasformazioni_2015-01	Numero complessivo di trasformazioni dell'identificativo nel mese di gennaio 2015
n_nuove_trasformazioni_2015-02	Numero complessivo di trasformazioni dell'identificativo nel mese di febbraio 2015
n_nuove_trasformazioni_2015-03	Numero complessivo di trasformazioni dell'identificativo nel mese di marzo 2015
n_nuove_trasformazioni_2015-04	Numero complessivo di trasformazioni dell'identificativo nel mese di aprile 2015
n_nuove_trasformazioni_2015-05	Numero complessivo di trasformazioni dell'identificativo nel mese di maggio 2015
n_nuove_trasformazioni_2015-06	Numero complessivo di trasformazioni dell'identificativo nel mese di giugno 2015
n_nuove_trasformazioni_2015-07	Numero complessivo di trasformazioni dell'identificativo nel mese di luglio 2015
n_nuove_trasformazioni_2015-08	Numero complessivo di trasformazioni dell'identificativo nel mese di agosto 2015
n_nuove_trasformazioni_2015-09	Numero complessivo di trasformazioni dell'identificativo nel mese di settembre 2015
n_nuove_proroghe_2014-05	Numero complessivo di nuove proroghe dell'identificativo nel mese di maggio 2014
n_nuove_proroghe_2014-06	Numero complessivo di nuove proroghe dell'identificativo nel mese di giugno 2014
n_nuove_proroghe_2014-07	Numero complessivo di nuove proroghe dell'identificativo nel mese di luglio 2014
n_nuove_proroghe_2014-08	Numero complessivo di nuove proroghe dell'identificativo nel mese di agosto 2014
n_nuove_proroghe_2014-09	Numero complessivo di nuove proroghe dell'identificativo nel mese di settembre 2014
n_nuove_proroghe_2014-10	Numero complessivo di nuove proroghe dell'identificativo nel mese di ottobre 2014

<b>variabile</b>	<b>descrizione</b>
n_nuove_proroghe_2014-11	Numero complessivo di nuove proroghe dell'identificativo nel mese di novembre 2014
n_nuove_proroghe_2014-12	Numero complessivo di nuove proroghe dell'identificativo nel mese di dicembre 2014
n_nuove_proroghe_2015-01	Numero complessivo di nuove proroghe dell'identificativo nel mese di gennaio 2015
n_nuove_proroghe_2015-02	Numero complessivo di nuove proroghe dell'identificativo nel mese di febbraio 2015
n_nuove_proroghe_2015-03	Numero complessivo di nuove proroghe dell'identificativo nel mese di marzo 2015
n_nuove_proroghe_2015-04	Numero complessivo di nuove proroghe dell'identificativo nel mese di aprile 2015
n_nuove_proroghe_2015-05	Numero complessivo di nuove proroghe dell'identificativo nel mese di maggio 2015
n_nuove_proroghe_2015-06	Numero complessivo di nuove proroghe dell'identificativo nel mese di giugno 2015
n_nuove_proroghe_2015-07	Numero complessivo di nuove proroghe dell'identificativo nel mese di luglio 2015
n_nuove_proroghe_2015-08	Numero complessivo di nuove proroghe dell'identificativo nel mese di agosto 2015
n_nuove_proroghe_2015-09	Numero complessivo di nuove proroghe dell'identificativo nel mese di settembre 2015
IDCF_DATORELAVORO	ID CF Datore lavoro
FLG_DATORE_PA	Flag se Datore è PA (0:NO,1:SI)
SETTORE_LAVORO	Settore lavoro
COMUNE_SEDE_LAVORO	Comune sede lavoro
IDDATA_INIZIO_RAPPORTO	ID Data inizio rapporto
IDDATA_FINE_RAPPORTO	ID Data fine prevista rapporto
DURATA_PREVISTA_INIZIALE	Durata prevista in giorni
TIPO_CONTRATTO	Tipologia Contrattuale
TIPO_ORARIO	Tipologia Orario
QUALIFICA_L4	Qualifica Professionale (livello 4)
CCNL	CCNL applicato
FLG_ASSUNZIONE_LEGGE68	Flag se assunzione ai fini della Legge68 (0:NO,1:SI)
FLG_LAVOROSTAGIONALE	Flag se Lavoro Stagionale (0:NO,1:SI)
FLG_LAVOROAGRICOLTURA	Flag se Datore è PA (0:NO,1:SI)
TIPO_LAVORAZIONE	Tipo lavorazione
IDDATA_CESSAZIONE	ID Data cessazione effettiva (se rapporto cessato all'atto dell'estrazione)
MOTIVO_CESSAZIONE	Motivo cessazione (se rapporto cessato all'atto dell'estrazione)
DURATA_EFFETTIVA	Durata effettiva (se rapporto cessato all'atto dell'estrazione)
FLG_TRASFORMATO_TI	Flag se rapporto trasformato a Tempo Indeterminato (0:NO,1:SI)
IDDATA_TRASFORMAZIONE	ID Data Trasformazione a Tempo Indeterminato
FLG_PROROGATO	Flag se rapporto ha avuto proroghe (0:NO,1:SI)
IDDATA_MAX_PROROGA	ID Data massima fine prorogata
FLG_RAPP_IN_ESSERE	Flag se rapporto è ancora in essere (0:NO,1:SI)

## ANALISI IMPATTO DELLA GARANZIA GIOVANI

La variabile trattamento (D) corrisponde in questa analisi all'aver partecipato e concluso (uscita) uno o più interventi all'interno della Garanzia Giovani. La Garanzia Giovani stessa è considerata come una unica policy.

Per la stima dell'ATET sono stati utilizzati più modelli e metodi.

Metodi non parametrici

- Propensity score matching (procedura Stata **teffects psmatch; psmatch2**)
- Covariate matching (procedura Stata **teffects nmatch**)

Metodi misti parametrici-non parametrici

- Inverse probability weighting with regression adjustment (procedura Stata **teffects ipwra**)

La tabella IV.13 riporta l'insieme delle covariate X utilizzate nei modelli. I valori delle variabili sono prima del trattamento.

**Tabella IV.13 – Covariate utilizzate nei modelli**

N.	Descrizione	tipo	N.	Descrizione	Tipo
1	Dummy 1 = Donne	Dummy	18	Dummy 1 =Lombardia	Dummy
2	Età al quadrato riferimento 30 settembre 2015	Numerica	19	Dummy 1 =Trento	Dummy
3	Dummy 1=Diploma scuola secondaria superiore	Dummy	20	Dummy 1 =Veneto	Dummy
4	Dummy 1=Titolo di istruzione terziaria	Dummy	21	Dummy 1 =Friuli Venezia Giulia	Dummy
5	Distanza anni conseguimento ultimo titolo studio	Numerica	22	Dummy 1 =Liguria	Dummy
6	Inoccupato (nessuna esperienza lavoro pregressa)	Dummy	23	Dummy 1 =Emilia Romagna	Dummy
7	Comune domicilio da 10 a 25 mila abitanti	Dummy	24	Dummy 1 =Toscana	Dummy
8	Comune domicilio da 25 a 100 mila abitanti	Dummy	25	Dummy 1 =Umbria	Dummy
9	Comune domicilio oltre 100mila abitanti	Dummy	26	Dummy 1 =Marche	Dummy
10	Distanza ultimo lavoro superiore a 2 anni	Dummy	27	Dummy 1 =Lazio	Dummy
11	Dummy Uomini in età 15-22 anni	Dummy	28	Dummy 1 =Abruzzo	Dummy
12	Dummy Uomini in età 23-24 anni	Dummy	29	Dummy 1 =Campania	Dummy
13	Dummy Uomini in età 25-27anni	Dummy	30	Dummy 1 =Basilicata	Dummy
14	Dummy Uomini in età 28-32anni	Dummy	31	Dummy 1 =Calabria	Dummy
15	Dummy Donne in età 15-22 anni	Dummy	32	Dummy 1 =Sicilia	Dummy
16	Dummy Donne in età 23-24 anni	Dummy	33	Dummy 1 =Sardegna	Dummy
17	Dummy Donne in età 25-27anni	Dummy			

Le procedure *psmatch2* e *teffects psmatch* realizzano un propensity score matching (PSM). Il Propensity score ( $p(x)$ ) rappresenta la probabilità, dipendente dall'insieme delle variabili X, di essere in trattamento.

$$p(x) = \text{prob}(D = 1|X)$$

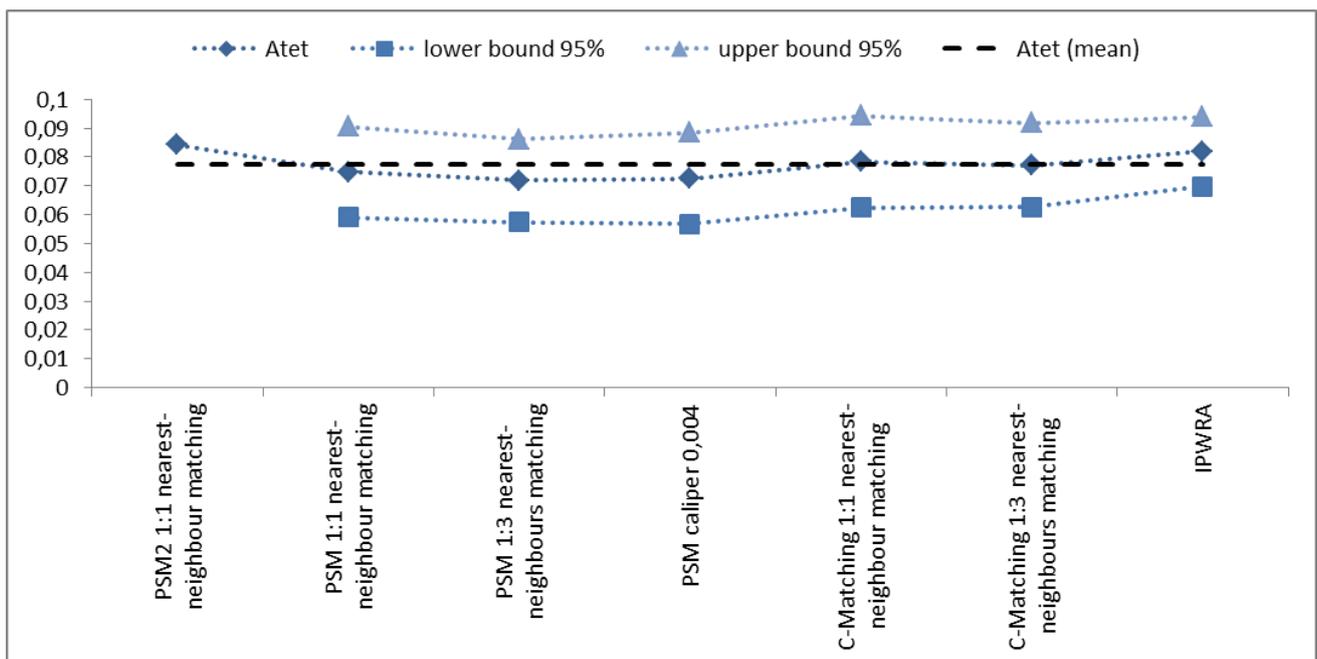
In entrambi i metodi è stato utilizzato un matching 1 a 1 (one-to-one Nearest Neighbour Matching), vale a dire per ogni singola unità trattata è stato abbinato l'individuo con il valore del propensity score più vicino.

La procedura *teffects nmatch* realizza un matching attraverso il calcolo della distanza tra le diverse covariate (distanza di Mahalanobis). Il matching, anche in questo caso 1 a 1, avviene abbinando a ciascuna unità trattata l'individuo appartenente al gruppo di controllo che ha la minima distanza all'interno della stessa regione di domicilio e dello stesso livello di istruzione (stratificazione).

**Tabella IV.14 – Stime ATET**

Comandi Stata	Descrizione	Opzioni	Stima Atet	Std. Error	Z	p>Z
psmatch2	Propensity score matching	1:1 nearest-neighbour matching	0,08424264	0,01163664	7,24	0,000
teffects psmatch	Propensity score matching	1:1 nearest-neighbour matching	0,0749047	0,0080027	9,36	0,000
		1:3 nearest-neighbours matching	0,0718401	0,0073349	9,79	0,000
		caliper 0,004	0,0726131	0,0080595	9,01	0,000
teffects nmatch	Covariate Matching Exact matching (regione e titolo di studio)	1:1 nearest-neighbour matching	0,0784188	0,0081299	9,65	0,000
		1:3 nearest-neighbours matching	0,0772465	0,0073962	10,44	0,000
teffects ipwra	inverse probability weighting regression adjustment (double-robust reweighting)		0,0819393	0,006163	13,3	0,000

**Figura IV.1 – Stime Atet per diverse metodologie**



La procedura *teffects ipwra* stima l'ATET attraverso una procedura mista parametrica e non parametrica. In particolare calcola dapprima un propensity score che utilizzerà per riponderare in modo appropriato le unità trattate e non trattate. Definisce una funzione parametrica per l'outcome e quindi stima tale funzione attraverso una regressione dei minimi quadrati pesati (WLS).

Un matching di qualità deve soddisfare:

- un buon overlap: il propensity score ( $p(x)$ ) deve distribuirsi uniformemente tra trattati e controlli e deve essere strettamente compreso tra 0 e 1 non assumendo mai questi due valori
- un buon bilanciamento tra trattati e controlli in termini sia del propensity score, che dei valori delle singole covariate  $X$
- la validità della CMI
- la validità della SUTVA (stable unit treatment value assumption), ovvero che ciascuna unità trattata non influisce sul valore dell'outcome delle altre unità trattate

Di seguito si riportano alcune analisi e test per verificare la qualità dei metodi di matching nella stima dell'ATET.

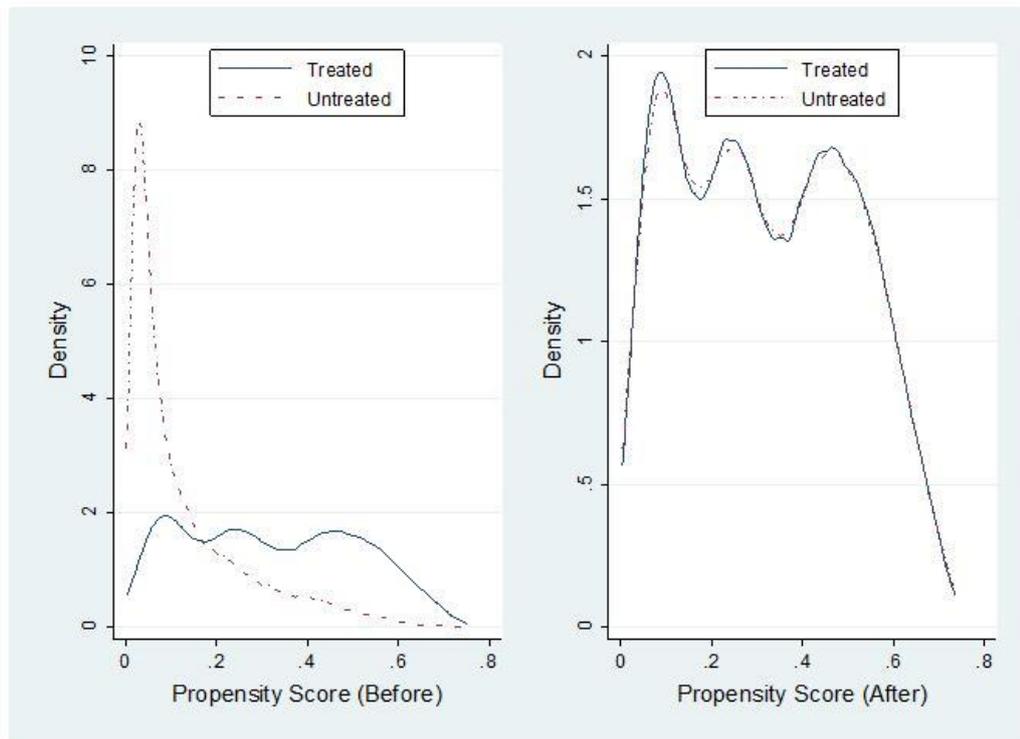
#### **psmatch2 (1-1 Nearest-Neighbour Matching)**

- Overlap

La figura IV.2 mostra la distribuzione del propensity score tra trattati e controlli prima e dopo l'abbinamento.

Le due distribuzioni dopo il matching mostrano un soddisfacente grado sovrapposizione (overlap).

**Figura IV.2 – distribuzione Propensity score trattati/non trattati prima e dopo l’abbinamento**



- Bilanciamento (balancing)

Si riportano qui di seguito i principali risultati relativi al bilanciamento, sia del propensity score che delle singole covariate X. L’ipotesi nulla ( $H_0$ ) alla base del test T-Student è che il *bias*, espresso come differenza tra il gruppo dei trattati e quello dei controlli abbinati (matched) dei valori medi del propensity score (tabella IV.15) e delle covariate X (tabella IV.16), sia pari a zero. Le tavole IV.15 e IV.16 riportano inoltre il valore del rapporto delle varianze del gruppo dei trattati e del gruppo dei controlli: in questo caso un bilanciamento soddisfacente presuppone un valore del rapporto prossimo ad 1.

**Tabella IV.15 – Test bilanciamento del propensity score per trattati e controlli**

Variable	Mean		%bias	t-test		V(T)/V(C)
	Treated	Control		t	p>t	
_pscore	0,32222	0,32222	0,0000	0,0000	1,0000	1,0000

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0,00	0,00	1,00	0,00	0,00	0,00	1,00	0,00

**Tabella IV.16 – Test bilanciamento dei valori medi delle covariate X trattati e controlli**

Variable	Mean			t-test		V(T)/V(C)
	Treated	Control	%bias	t	p>t	
1.sesso2	0,49231	0,49478	-0,5	-0,32	0,748	.
c.eta_300915#c.eta_300915	558,46	563,9	-3,2	-2,15	0,031	1,02
2.titolo3	0,58207	0,58019	0,4	0,25	0,804	.
3.titolo3	0,22152	0,22844	-1,8	-1,08	0,279	.
etaeduc	5,7117	5,7306	-0,5	-0,38	0,705	1.07*
1.esplav1	0,59662	0,5938	0,6	0,37	0,708	.
2.classe_abitanti2	0,23947	0,24663	-1,7	-1,09	0,276	.
3.classe_abitanti2	0,24534	0,24275	0,6	0,39	0,695	.
4.classe_abitanti2	0,16238	0,1483	3,5	2,54	0,011	.
2.genere#2.cleta_2	0,08823	0,08647	0,6	0,41	0,684	.
2.genere#3.cleta_2	0,13094	0,13399	-0,9	-0,59	0,557	.
2.genere#4.cleta_2	0,08894	0,09304	-1,4	-0,93	0,351	.
3.regione_dom	0,19606	0,19887	-0,9	-0,46	0,644	.
4.regione_dom	0,01091	0,0095	1,5	0,91	0,361	.
5.regione_dom	0,17494	0,17775	-1	-0,48	0,63	.
6.regione_dom	0,0149	0,01302	1,4	1,04	0,296	.
7.regione_dom	0,0142	0,0142	0	0	1	.
8.regione_dom	0,03884	0,03755	0,5	0,44	0,66	.
9.regione_dom	0,12284	0,12015	0,9	0,54	0,59	.
10.regione_dom	0,00774	0,00704	0,7	0,54	0,592	.
11.regione_dom	0,06394	0,06594	-0,9	-0,53	0,597	.
12.regione_dom	0,0887	0,08354	2,3	1,2	0,23	.
13.regione_dom	0,00716	0,0061	1,1	0,85	0,396	.
15.regione_dom	0,02112	0,02042	0,3	0,32	0,747	.
17.regione_dom	0,05174	0,05374	-1,1	-0,58	0,56	.
18.regione_dom	0,01361	0,01455	-0,4	-0,52	0,603	.
19.regione_dom	0,04353	0,04388	-0,2	-0,11	0,911	.
20.regione_dom	0,06394	0,06465	-0,3	-0,19	0,851	.
distanza_mag24mesi	0,22398	0,22692	-0,7	-0,46	0,647	.

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0,002	41,92	0,113	1	0,9	9,9	1,02	50

- Validità della CMI – Analisi di sensitività ( )

Questo test sviluppato da Rosenbaum ha come obiettivo quello di verificare la robustezza e l'affidabilità delle stime degli effetti attraverso metodi di propensity score matching, ossia la validità dell'ipotesi CMI. Il test nella sua versione implementata in Stata (comando **rbounds**) è applicabile soltanto a casi di matching "1-1 Nearest-Neighbour" e utilizza una statistica test sviluppata da Wilcoxon.

Il test considera ipotesi di un progressivo allontanamento dal caso dell'esperimento randomizzato in cui la probabilità di trattamento è la stessa per ogni coppia di unità abbinata nel matching.

Sia  $p_{1j}$  la probabilità di essere trattato dell'unità  $j$  appartenente al gruppo dei trattati ( $D=1$ ), e  $p_{0nj}$  la corrispondente probabilità di essere trattato dell'unità  $j$  appartenente al gruppo dei non trattati ( $D=0$ ) abbinata alla precedente.

Il test verifica l'ipotesi, al consueto livello di significatività del 5%, che  $p_{1j} = \text{Gamma} \cdot p_{0nj}$  produce effetti significativi sull'ATET stimato. Se il test risulta significativo per bassi valori di *Gamma* allora

occorre porre attenzione sui risultati raggiunti. Nel caso in esame il test risulta significativo per valori di Gamma compresi tra 1,4 e 1,5, dunque le stime dell'ATET sono robuste fino a considerare situazioni in cui la probabilità di trattamento per i trattati sia 1,5 volte più grande rispetto alla probabilità di trattamento dei non trattati abbinati.

**Tabella IV.17 – Test di sensitività dell'ipotesi CMI**

Gamma	sig+	sig-
1	0	0
1,1	0	0
1,2	8,10E-14	0
1,3	1,50E-07	0
1,4	0,001127	0
1,5	0,129807	0
1,6	0,749924	0
1,7	0,991037	0
1,8	0,999963	0
1,9	1	0
2	1	0

## teffects nmatch (1-1 Nearest-Neighbour Matching)

- Bilanciamento (Balancing)

**Tabella IV.18 – Bilanciamento delle Covariate**

	Standardized differences		Variance Ratio	
	Raw	Matched	Raw	Matched
Donne	0,02427	0,00047	1,00145	1,00002
etasq	-0,13450	-0,00744	0,93091	1,02486
Diploma scuola secondaria superiore	0,09156	0,00000	0,97841	1,00000
Laurea	0,21727	0,00000	1,44490	1,00000
etaeduc	-0,36826	-0,00007	0,74644	1,04038
Inoccupato	0,03036	-0,01436	0,98915	1,00589
Da 10 a 25 mila abitanti	0,11660	0,00936	1,17561	1,01160
Da 25 mila a 100 mila abitanti	-0,01714	0,00245	0,98038	1,00292
Oltre 100 mila abitanti	-0,20259	-0,00095	0,73843	0,99826
Maschi 23-24anni	-0,03044	0,00043	0,91233	1,00136
Maschi 25-27 anni	-0,05518	-0,00198	0,86734	0,99461
Maschi 28-32 anni	-0,09748	0,00000	0,74110	1,00000
Donne 15-22 anni	0,01186	0,00151	1,01985	1,00248
Donne 23-24 anni	0,04154	0,00000	1,13460	1,00000
Donne 25-27anni	0,03613	-0,00104	1,08560	0,99772
Lombardia	0,42685	0,00000	2,92412	1,00000
Trento	0,02895	0,00000	1,34313	1,00000
Veneto	0,49852	0,00000	5,15258	1,00000
Friuli Venezia Giulia	-0,04081	0,00000	0,73937	1,00000
Liguria	-0,29666	0,00000	0,20252	1,00000
Emilia Romagna	-0,20292	0,00000	0,46483	1,00000
Toscana	0,15052	0,00000	1,50307	1,00000
Umbria	-0,07245	0,00000	0,50333	1,00000
Marche	0,09196	0,00000	1,44676	1,00000
Lazio	0,28716	0,00000	3,54981	1,00000
Abruzzo	-0,04447	0,00000	0,62929	1,00000
Campania	-0,43130	0,00000	0,17827	1,00000
Basilicata	0,18710	0,00000	2,83175	1,00000
Calabria	-0,32096	0,00000	0,18064	1,00000
Sicilia	0,02066	0,00000	1,09991	1,00000
Sardegna	-0,06603	0,00000	0,80359	1,00000
Distanza ultima esperienza > 2 anni	-0,09297	0,00535	0,89485	1,00716

## **ANALISI DELL'EFFICACIA DEL TIROCINO**

- Valutazione comparata dell'efficacia dei tirocini IOG e non IOG

Definita la variabile TIOG che assume i valori "0 = Non IOG" e "1 = IOG", sono stati stimati tre equazioni logistiche (procedura Stata **logit**) con variabili dipendenti date dalla condizione di occupato (=1) o non occupato (=0) rispettivamente entro il mese di conclusione del tirocinio, entro il mese successivo a quello di conclusione del tirocinio, ed entro due mesi successivi a quello di conclusione del

tirocinio. I modelli sono stati stimati su di un campione di 5.588 tirocini conclusi tra gennaio 2015 e luglio 2015.

**Tabella IV.19 – Risultati regressione logit per la variabile dipendente Occupato entro il mese di conclusione del tirocinio**

Logistic regression Number of obs = 5.588  
 LR chi2(31) = 451,110  
 Prob > chi2 = 0,000  
 Log likelihood = -2995.4873 Pseudo R2 = 0,070

occ_1mesi	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
TIOG	0,1549317	0,0673002	2,3	0,021	0,0230257	0,2868377
genere						
F	-0,0834072	0,0646489	-1,29	0,197	-0,2101167	0,0433023
cleta_2						
23-24 anni	-0,1911279	0,0981022	-1,95	0,051	-0,3834048	0,001149
25-27 anni	-0,1316903	0,0977116	-1,35	0,178	-0,3232016	0,059821
28-32 anni	-0,1392445	0,1184198	-1,18	0,240	-0,3713431	0,092854
prec_esplav						
Disoccupato con 1 esperienza	0,1477369	0,0876509	1,69	0,092	-0,0240558	0,3195295
Disoccupati con più esperienze lavorative	0,1131933	0,0840699	1,35	0,178	-0,0515805	0,2779672
titolo3						
Diploma e qualifica	0,3857312	0,1029278	3,75	0,000	0,1839964	0,5874661
Laurea	0,6642798	0,1236854	5,37	0,000	0,4218609	0,9066987
regione_dom						
Lombardia	0,3955049	0,1317251	3	0,003	0,1373284	0,6536813
Trentino Alto Adige	-0,3807751	0,5620694	-0,68	0,498	-1,482411	0,7208608
Veneto	0,7182273	0,1267141	5,67	0,000	0,4698723	0,9665823
Friuli Venezia Giulia	0,0060419	0,2362079	0,03	0,980	-0,4569171	0,4690008
Liguria	0,2550757	0,1881249	1,36	0,175	-0,1136423	0,6237938
Emilia Romagna	0,3098268	0,1557638	1,99	0,047	0,0045354	0,6151182
Toscana	0,5252124	0,1220383	4,3	0,000	0,2860217	0,7644031
Umbria	0,7954839	0,2922328	2,72	0,006	0,2227182	1,36825
Marche	0,5288497	0,1263164	4,19	0,000	0,2812742	0,7764253
Lazio	-0,4581859	0,2013035	-2,28	0,023	-0,8527334	-0,0636384
Abruzzo	-0,0627405	0,4275607	-0,15	0,883	-0,9007441	0,7752632
Campania	-0,0173779	0,2584202	-0,07	0,946	-0,5238723	0,4891164
Basilicata	-0,603113	0,2565726	-2,35	0,019	-1,105986	-0,1002399
Calabria	-0,2377347	0,3703149	-0,64	0,521	-0,9635387	0,4880693
Sicilia	-0,4756505	0,4577358	-1,04	0,299	-1,372796	0,4214951
Sardegna	-0,6909171	0,1967356	-3,51	0,000	-1,076512	-0,3053225
t_fine_politica						
Febbraio	-0,1521464	0,139468	-1,09	0,275	-0,4254987	0,1212059
Marzo	0,3373213	0,1282223	2,63	0,009	0,0860101	0,5886324
Aprile	0,305753	0,1288554	2,37	0,018	0,0532011	0,5583049
Maggio	0,4889319	0,1247273	3,92	0,000	0,2444708	0,733393
Giugno	1,505888	0,1241819	12,13	0,000	1,262496	1,74928
Luglio	0,2893352	0,1257802	2,3	0,021	0,0428105	0,5358598
_cons	-2,165763	0,1602594	-13,51	0,000	-2,479865	-1,85166

**Tabella IV.20 – Risultati regressione logit per la variabile dipendente Occupato entro il mese successivo a quello di conclusione del tirocinio**

Logistic regression Number of obs = 5588  
 LR chi2(31) = 480,79  
 Prob > chi2 = 0,000  
 Pseudo R2 = 0,0635

Log likelihood = -3547.5923

occ\_2mesi

	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
TIOG	-0,1730517	0,0608285	-2,84	0,004	-0,2922733	-0,0538301
genere						
F	-0,1562814	0,0578566	-2,7	0,007	-0,2696783	-0,0428846
cleta_2						
23-24	-0,153165	0,0869362	-1,76	0,078	-0,3235568	0,0172267
25-27	-0,0623972	0,0870179	-0,72	0,473	-0,2329492	0,1081548
28-32	-0,1603774	0,1055129	-1,52	0,129	-0,3671789	0,0464241
prec_esplav						
Disoccupato con 1 esperienza	0,1574457	0,0791358	1,99	0,047	0,0023424	0,3125489
Disoccupato con più esperienze lavorative	0,1431342	0,0753144	1,9	0,057	-0,0044792	0,2907476
titolo3						
Diploma e qualifica	0,4301502	0,0884043	4,87	0,000	0,256881	0,6034194
Laurea	0,7686611	0,1079003	7,12	0,000	0,5571804	0,9801419
regione_dom						
Lombardia	0,8244278	0,1157837	7,12	0,000	0,5974959	1,05136
Trentino Alto Adige	0,346486	0,4225617	0,82	0,412	-0,4817198	1,174692
Veneto	0,9136188	0,1134667	8,05	0,000	0,6912282	1,136009
Friuli Venezia Giulia	-0,2064588	0,2137624	-0,97	0,334	-0,6254253	0,2125078
Liguria	0,4352863	0,1632793	2,67	0,008	0,1152647	0,7553079
Emilia Romagna	0,5593752	0,1365392	4,1	0,000	0,2917633	0,8269871
Toscana	0,7062082	0,1078855	6,55	0,000	0,4947565	0,9176599
Umbria	0,7006152	0,2757618	2,54	0,011	0,1601319	1,241098
Marche	0,5428961	0,1122304	4,84	0,000	0,3229285	0,7628637
Lazio	-0,1641659	0,1651052	-0,99	0,320	-0,4877662	0,1594343
Abruzzo	0,8301902	0,3538355	2,35	0,019	0,1366853	1,523695
Campania	0,0871077	0,2222421	0,39	0,695	-0,3484788	0,5226941
Basilicata	-0,4971576	0,2150231	-2,31	0,021	-0,9185951	-0,07572
Calabria	0,0362884	0,2998613	0,12	0,904	-0,551429	0,6240058
Sicilia	-0,148215	0,3545962	-0,42	0,676	-0,8432108	0,5467807
Sardegna	-0,6239093	0,1623259	-3,84	0,000	-0,9420621	-0,3057564
t_fine_politica						
Febbraio	0,0381176	0,1133847	0,34	0,737	-0,1841123	0,2603475
Marzo	0,3157813	0,1096955	2,88	0,004	0,100782	0,5307806
Aprile	0,218026	0,1101855	1,98	0,048	0,0020665	0,4339855
Maggio	0,4409939	0,1077848	4,09	0,000	0,2297396	0,6522483
Giugno	1,055582	0,1128672	9,35	0,000	0,8343661	1,276797
Luglio	-0,0519487	0,1087474	-0,48	0,633	-0,2650897	0,1611922
_cons	-1,37717	0,1358227	-10,14	0,000	-1,643377	-1,110962

**Tabella IV.21 – Risultati regressione logit per la variabile dipendente Occupato entro il secondo mese successivo a quello di conclusione del tirocinio**

Logistic regression Number of obs = 5588  
 LR chi2(31) = 453,67  
 Prob > chi2 = 0,000  
 Pseudo R2 = 0,0587

Log likelihood = -3634.9971

occ\_3mesi

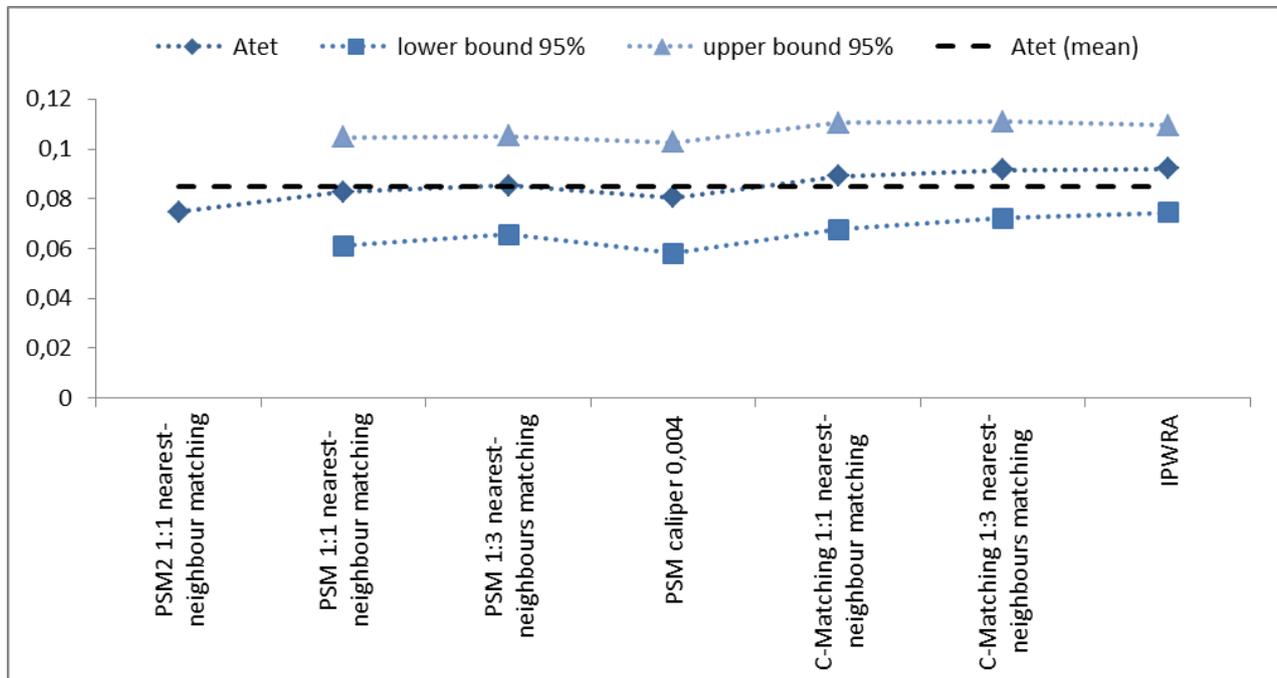
	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]
TIOG	-0,1801557	0,0599357	-3,01	0,003	-0,2976276 -0,0626839
genere					
F	-0,1190068	0,0569421	-2,09	0,037	-0,2306112 -0,0074023
cleta_2					
23-24	-0,1242954	0,0851298	-1,46	0,144	-0,2911467 0,042556
25-27	-0,1023329	0,0856993	-1,19	0,232	-0,2703004 0,0656346
28-32	-0,1869761	0,1036507	-1,8	0,071	-0,3901278 0,0161756
prec_esplav					
Disoccupato con 1 esperienza	0,1027888	0,0781228	1,32	0,188	-0,050329 0,2559066
Disoccupato con più esperienze lavorative	0,1929274	0,074209	2,6	0,009	0,0474804 0,3383743
titolo3					
Diploma e qualifica	0,4851744	0,0857083	5,66	0,000	0,3171892 0,6531596
Laurea	0,8411556	0,1054381	7,98	0,000	0,6345007 1,04781
regione_dom					
Lombardia	0,8410042	0,113854	7,39	0,000	0,6178546 1,064154
Trentino Alto Adige	0,6183906	0,4131467	1,5	0,134	-0,1913621 1,428143
Veneto	1,004062	0,1119524	8,97	0,000	0,7846397 1,223485
Friuli Venezia Giulia	0,1113454	0,1981751	0,56	0,574	-0,2770708 0,4997615
Liguria	0,3892367	0,1597346	2,44	0,015	0,0761625 0,7023108
Emilia Romagna	0,5480327	0,1339025	4,09	0,000	0,2855885 0,8104768
Toscana	0,7186488	0,1054133	6,82	0,000	0,5120425 0,9252551
Umbria	0,9522761	0,278078	3,42	0,001	0,4072533 1,497299
Marche	0,5472421	0,1093833	5	0,000	0,3328548 0,7616295
Lazio	-0,1375879	0,1575226	-0,87	0,382	-0,4463266 0,1711507
Abruzzo	0,6096442	0,3514507	1,73	0,083	-0,0791864 1,298475
Campania	0,0459294	0,2158902	0,21	0,832	-0,3772076 0,4690665
Basilicata	-0,508386	0,204698	-2,48	0,013	-0,9095866 -0,1071853
Calabria	-0,0035709	0,2915988	-0,01	0,990	-0,5750941 0,5679522
Sicilia	-0,1351991	0,338666	-0,4	0,690	-0,7989722 0,5285741
Sardegna	-0,5401019	0,1515009	-3,57	0,000	-0,8370382 -0,2431657
t_fine_politica					
Febbraio	0,1202427	0,1110577	1,08	0,279	-0,0974263 0,3379118
Marzo	0,3716777	0,1082234	3,43	0,001	0,1595637 0,5837917
Aprile	0,3264143	0,1082841	3,01	0,003	0,1141815 0,5386472
Maggio	0,3860027	0,1064914	3,62	0,000	0,1772833 0,5947221
Giugno	0,9376722	0,1121177	8,36	0,000	0,7179256 1,157419
Luglio	0,1642394	0,1057296	1,55	0,120	-0,0429868 0,3714657
_cons	-1,268805	0,1320467	-9,61	0,000	-1,527611 -1,009998

- Analisi di impatto dei tirocini IOG

**Tabella IV.22 – Stime ATET**

Comandi Stata	Descrizione	Opzioni	Stima Atet	Std. Error	Z	p>Z
psmatch2	Propensity score matching	1:1 nearest-neighbour matching	0,07481371	0,0149710	5,00	
teffects psmatch	Propensity score matching	1:1 nearest-neighbour matching	0,0828892	0,0111048	7,46	0,000
		1:3 nearest-neighbours matching	0,0853102	0,0100745	8,47	0,000
		caliper 0,004	0,0804493	0,0113571	7,08	0,000
teffects nmatch	Covariate Matching Exact matching (regione e titolo di studio)	1:1 nearest-neighbour matching	0,0890315	0,0109284	8,15	0,000
		1:3 nearest-neighbours matching	0,0914927	0,0098716	9,27	0,000
teffects ipwra	inverse probability weighting regression adjustment (double-robust reweighting)		0,0919624	0,008875	10,36	0,000

**Figura IV.3 – Stime Atet per diverse metodologie**



**Tabella IV.23 – Bilanciamento metodo Covariate 1-1 Nearest-neighbour Matching**

	Standardized differences		Variance Ratio	
	Raw	Matched	Raw	Matched
Donne	0,09481	-0,00179	0,99890	1,00020
etasq	-0,06401	-0,00028	0,83753	1,01395
Diploma scuola secondaria superiore	0,14633	0,00000	0,94948	1,00000
Laurea	0,27111	0,00000	1,55893	1,00000
etaeduc	-0,38956	0,00894	0,69804	1,04061
Inoccupato	-0,02188	-0,01868	1,00675	1,00547
Da 10 a 25 mila abitanti	0,10139	0,00423	1,15250	1,00532
Da 25 mila a 100 mila abitanti	0,00531	-0,00475	1,00611	0,99486
Oltre 100 mila abitanti	-0,31544	0,00091	0,58541	1,00209
Maschi 23-24anni	-0,00944	0,00108	0,97264	1,00323
Maschi 25-27 anni	-0,07947	-0,00103	0,81420	0,99712
Maschi 28-32 anni	-0,10028	0,00000	0,73131	1,00000
Donne 15-22 anni	0,03009	-0,00153	1,05130	0,99756
Donne 23-24 anni	0,07158	0,00000	1,23020	1,00000
Donne 25-27anni	0,08360	-0,00084	1,19791	0,99834
Lombardia	0,03830	0,00000	1,14735	1,00000
Trento	-0,02517	0,00000	0,66081	1,00000
Veneto	0,47879	0,00000	5,01802	1,00000
Friuli Venezia Giulia	0,00459	0,00000	1,03129	1,00000
Liguria	-0,33623	0,00000	0,12356	1,00000
Emilia Romagna	-0,31059	0,00000	0,23717	1,00000
Toscana	0,25501	0,00000	1,84888	1,00000
Umbria	-0,09686	0,00000	0,30977	1,00000
Marche	0,31390	0,00000	2,74533	1,00000
Lazio	0,44401	0,00000	5,18465	1,00000
Abruzzo	-0,04003	0,00000	0,62557	1,00000
Campania	-0,47127	0,00000	0,14931	1,00000
Basilicata	0,31805	0,00000	4,44449	1,00000
Calabria	-0,30836	0,00000	0,11207	1,00000
Sicilia	-0,16521	0,00000	0,36527	1,00000
Sardegna	-0,03677	0,00000	0,89386	1,00000
Distanza ultima esperienza > 2 anni	-0,06683	0,01193	0,92721	1,01492