

Istituto per la Ricerca Valutativa sulle Politiche Pubbliche



**IRVAPP**  
ISTITUTO PER LA RICERCA VALUTATIVA  
SULLE POLITICHE PUBBLICHE

# TREMODO: un Modello di Microsimulazione per il Trentino

**Davide Azzolini**

**Martina Bazzoli**

**Silvia De Poli**

**Carlo Fiorio**

**Samuele Poy**

Giugno 2014

FBK-IRVAPP Working Paper N. 2014-01

**TREMODO:**  
**un Modello di Microsimulazione per il Trentino**

**Davide Azzolini**  
*FBK-IRVAPP*

**Martina Bazzoli**  
*FBK-IRVAPP*

**Silvia De Poli**  
*FBK-IRVAPP*

**Carlo Fiorio**  
*FBK-IRVAPP e Università di Milano*

**Samuele Poy**  
*FBK-IRVAPP*

FBK-IRVAPP Working Paper N. 2014-01

Giugno 2014



Istituto per la Ricerca Valutativa sulle Politiche Pubbliche  
Fondazione Bruno Kessler  
Via S. Croce 77, 38122 Trento (Italia)

Tel.: (+39) 0461.314209  
Fax: (+39) 0461.314240

Email: [info@irvapp.it](mailto:info@irvapp.it)  
Website: <http://irvapp.fbk.eu>

I *Working Papers* di IRVAPP presentano i risultati preliminari di progetti di ricerca realizzati da IRVAPP o presentati da ricercatori esterni presso l'Istituto medesimo. I *Working Papers* sono diffusi al fine di stimolare la discussione sulle tematiche oggetto di studio. Nel citare tali lavori si prega di tener conto del carattere preliminare degli stessi. Eventuali versioni riviste e aggiornate sono disponibili nella *Reprint Series*, se pubblicate, o direttamente presso IRVAPP.

Le opinioni espresse sono degli autori e non riflettono necessariamente quelle dell'IRVAPP.

# **TREMODO: un Modello di Microsimulazione per il Trentino\***

Davide Azzolini, Martina Bazzoli<sup>†</sup>, Silvia De Poli, Carlo Fiorio, Samuele Poy

## **Abstract**

In questo *Working Paper* viene presentata la prima versione di TREMOD, il modello di microsimulazione sviluppato da FBK-IRVAPP per la provincia di Trento utilizzando la piattaforma europea EUROMOD. TREMOD consente di simulare, con elevata flessibilità, gli effetti di diversi tipi di intervento pubblico su una pluralità di *outcome* di interesse collettivo, quali, ad esempio, la distribuzione dei redditi o le condizioni di benessere delle famiglie e si propone come strumento analitico a supporto di decisioni pubbliche in materia fiscale e di *welfare*. TREMOD si caratterizza, rispetto ad altri modelli di microsimulazione, per l'elevata qualità dei dati utilizzati. La banca dati sottostante il modello è stata ottenuta combinando, mediante un *matching* esatto, dati campionari derivanti dall'indagine sulle Condizioni di Vita delle Famiglie Trentine con dati amministrativi derivanti dalle dichiarazioni dei redditi di fonte Agenzia delle Entrate. TREMOD è un modello di microsimulazione statico. Tuttavia, TREMOD è da considerare come uno strumento in continua evoluzione e pertanto suscettibile di essere aggiornato, esteso, oltre che integrato con nuove basi dati.

**Parole-chiave:** Politiche fiscali, modello di microsimulazione, validazione, dati amministrativi

**Classificazione JEL:** H24, I32, I38.

---

\* Il presente progetto è stato finanziato dal Dipartimento Lavoro e Welfare della Provincia Autonoma di Trento. Il progetto non sarebbe stato realizzabile senza la preziosa collaborazione dell'Ufficio Statistica della Provincia autonoma di Trento e dell'Agenzia del Lavoro di Trento. Si ringraziano per gli utili commenti e per i suggerimenti Erich Battistin, Daniele Checchi, Livia Ferrario, Enrico Rettore, Antonio Schizzerotto, Ugo Trivellato.

<sup>†</sup> Contatto per corrispondenza: FBK-IRVAPP, Via Santa Croce 77. 38122 Trento (Italia). Indirizzo mail: bazzoli@irvapp.it

# 1 Introduzione

In questo studio si presentano e descrivono le finalità e le principali caratteristiche di TREMOD, il modello di microsimulazione per la Provincia autonoma di Trento (PaT). TREMOD è stato sviluppato da FBK-IRVAPP adattando al contesto trentino il modulo italiano di EUROMOD, il modello di microsimulazione fiscale dei paesi dell'Unione Europea (Sutherland e Figari, 2013).

TREMOD è concepito come uno strumento in grado di produrre evidenza empirica a supporto di decisioni pubbliche, per esempio in ambito fiscale e di *welfare*. TREMOD si connota, inoltre, per un'elevata flessibilità e adattabilità di utilizzo, in quanto consente di stimare *ex-ante* gli effetti di diversi tipi di intervento pubblico in ambito fiscale, previdenziale ed occupazionale su una pluralità di *outcome* di interesse, quali, ad esempio, la distribuzione dei redditi o le condizioni di benessere delle famiglie.

Affinchè uno strumento come TREMOD possa adempiere in maniera soddisfacente a tali obiettivi, è indispensabile assicurarsi che questo soddisfi due condizioni. La prima condizione è relativa alla qualità dei dati utilizzati per la costruzione del modello stesso (Sutherland, 1991). Sotto questo punto di vista, TREMOD presenta dei notevoli elementi di miglioramento rispetto alla gran parte dei modelli di microsimulazione esistenti, in quanto esso integra dati campionari con dati di fonte amministrativa. Attualmente i dati dell'Indagine sulle condizioni di Vita delle Famiglie Trentine, di seguito anche "ICFT" vengono combinati con le dichiarazioni dei redditi individuali di fonte Agenzia delle Entrate e la combinazione di queste due fonti garantisce una elevata precisione nella misurazione dei redditi individuali unitamente alla possibilità di analizzare gli effetti redistributivi delle politiche, non solo a livello individuale ma anche a livello familiare. La seconda condizione richiesta per il funzionamento del modello è che questo contenga informazioni dettagliate sulla normativa fiscale e previdenziale in vigore al fine di permettere all'analista di simulare, con il maggior grado di precisione possibile, le diverse grandezze economiche di interesse. Questa condizione è per TREMOD in gran parte garantita dal fatto di utilizzare come punto di partenza per le simulazioni fiscali il modello EUROMOD, che da anni, grazie a cospicui investimenti della DG-EMPL, mantiene aggiornati i modelli in base alla normativa nazionale vigente e dalla possibilità di usufruire di finanziamenti della PaT per sviluppare e mantenere moduli aggiuntivi di EUROMOD per la simulazione di politiche previdenziali ed occupazionali. L'oggetto principale di questo rapporto è dimostrare che queste due condizioni—la qualità dei dati utilizzati e la precisione delle stime di TREMOD—siano verificate.

Il paper è strutturato come segue. Il paragrafo 2 presenta sinteticamente TREMOD, quale declinazione in chiave trentina della piattaforma dei modelli di microsimulazione

EUROMOD. In particolare, sono messe in luce le potenzialità analitiche, le caratteristiche distintive e alcuni potenziali utilizzi pratici del modello concentrandosi sulla descrizione delle potenzialità di TREMOD per l'analisi delle politiche fiscali e lasciando la descrizione dell'uso di TREMOD per l'analisi di politiche occupazionali e previdenziali a futuri rapporti. Il paragrafo 3 presenta la base dati utilizzata per la costruzione di TREMOD e ne testa la qualità, con particolare riguardo alla rappresentatività rispetto alla popolazione trentina. Nel paragrafo 4 viene analizzata la precisione delle stime producibili mediante TREMOD, mettendo a confronto alcune grandezze economiche simulate dal modello con le stesse grandezze osservate nei dati amministrativi. Nel paragrafo 5 si effettua una comparazione tra il modello proposto (che, come detto, integra dati campionari a dati amministrativi) ed un modello alternativo, basato invece solo su dati campionari (SILC, come nel caso di EUROMOD). Infine, il paragrafo 6 conclude riassumendo i principali punti di forza del modello e prospettando le possibili integrazioni ed evoluzioni dello stesso.

## 2 TREMOD

TREMOD, modello di microsimulazione per la Provincia autonoma di Trento (PaT), si basa sulla piattaforma EUROMOD. EUROMOD è un modello di microsimulazione fiscale (*tax-benefit microsimulation model*) sviluppato a partire dal 1998 nell'ambito del programma europeo *Targeted Socio-Economic Research*, e attualmente finanziato dalla DG Emploment, Social Affairs and Inclusion. Il progetto è condotto da un consorzio accademico in rappresentanza di diversi Paesi europei coordinato dall'*Institute for Social and Economic Research* (ISER) ed è regolarmente aggiornato permettendo di rappresentare le discipline fiscali e, in parte, previdenziali vigenti nei 27 paesi membri dell'Unione Europea. Le informazioni contenute nel modello derivano da basi dati individuali, tipicamente indagini di *survey* (ad esempio, EU-SILC), combinate con informazioni simulate dal modello stesso (per esempio, imposte e sussidi). EUROMOD permette di calcolare non solo gli effetti delle misure in discussione (ad esempio, l'introduzione di un dato incentivo, il cambiamento di una determinata aliquota fiscale sul reddito, ecc.) ma anche di stimare la distribuzione degli effetti di cambiamenti simulati nelle politiche di tassazione e di agevolazioni tra diversi soggetti. Il modello consente di fare calcoli di tipo "*before-after*" e, pertanto, permette al ricercatore di simulare i cambiamenti nei redditi indotti da cambiamenti simulati nei livelli di tassazione e nella definizione di agevolazioni e altri *benefit*. Inoltre, con EUROMOD, è possibile rispondere a domande di tipo "*what if*". Ad esempio, è possibile ricostruire il "reddito controfattuale", vale a dire il reddito osservabile qualora venisse introdotta o eliminata una determinata politica, e in questo modo avere una valutazione *ex-ante* degli effetti di

una determinata politica. Come dimostrato dalla crescente letteratura scientifica, EUROMOD si presta a rispondere ad un ampio ventaglio di esigenze conoscitive. Ad oggi è stato utilizzato per costruire modelli di microsimulazione in diversi contesti nazionali, anche *extra*-europei. E' possibile menzionare la versione di EUROMOD per il Sud Africa (Wilkinson, 2009), per la Federazione Russa (Popova, 2012) e la Repubblica di Serbia (Žarković-Rakić, 2010). EUROMOD è inoltre una piattaforma estremamente flessibile che consente, partendo da un'opportuna struttura di dati, la simulazione degli effetti di politiche pubbliche in ambito previdenziale ed occupazionale.

Per definire e costruire TREMOD ci si è basati sul modulo italiano del modello EUROMOD-IT (Ceriani et al., 2013). Questa garantisce la ricostruzione dettagliata della normativa nazionale per quanto riguarda le imposte sui redditi, il patrimonio immobiliare e i contributi sociali, tenendo in considerazione alcune specificità fiscali regionali (es. addizionale Irpef regionale). Tramite la combinazione dei dati amministrativi (fiscali e contributivi) e quelli dell'Indagine sulle condizioni di Vita delle Famiglie (vedi Paragrafo 3), TREMOD garantisce la possibilità di valutare gli effetti di cambiamenti di *policy* riguardanti nello specifico la realtà locale della provincia di Trento.

TREMOD, in questa prima versione, è classificabile come un modello di microsimulazione di tipo "statico". Questa classe di modelli permette di ottenere stime relative all'effetto immediato (c.d. effetto "*morning after*") di cambiamenti di *policy* sugli *outcome* di interesse, ignorando il fatto che le stesse politiche possano alterare i comportamenti individuali e familiari. In prospettiva, l'utilizzo di informazioni longitudinali a livello individuale e familiare potrà fare evolvere TREMOD in un modello di tipo "dinamico" o "comportamentale", che permetterà di tenere in considerazione i cambiamenti indotti dalle politiche pubbliche sui comportamenti degli individui e delle famiglie (Figari et al. 2013).

### **3 La base dati di TREMOD**

Il punto di partenza per la costruzione di un modello di microsimulazione è la disponibilità di dati individuali contenenti informazioni sulle condizioni lavorative e previdenziali. TREMOD, nella versione presentata, utilizza dati relativi all'anno 2010 provenienti da due fonti: l'"Indagine sulle condizioni di Vita delle Famiglie" e le Dichiarazioni dei redditi di fonte Agenzia delle Entrate. Nel paragrafo 3.1 si descrivono brevemente le due fonti e il processo mediante cui sono state collegate. Nel paragrafo 3.2 si presentano i risultati di alcune analisi tese a valutare la rappresentatività statistica della base dati utilizzata. Nel paragrafo 3.3 si descrive l'attività di post-stratificazione svolta per risolvere alcuni problemi di rappresentatività della base dati.

### **3.1 Le fonti dei dati**

La principale fonte dati utilizzata per la costruzione di TREMOD è il Indagine sulle condizioni di Vita delle Famiglie, un'indagine campionaria contenente informazioni di carattere socio-economico a livello individuale e familiare. Si tratta di un *panel* prospettico, basato su interviste a soggetti in età maggiore di 18 anni e residenti in provincia di Trento. Per questa prima versione di TREMOD, si utilizza la sesta ondata ICFT, basata su interviste a 7.200 soggetti realizzate nell'anno 2010.

La seconda fonte utilizzata è invece costituita dai microdati delle Dichiarazioni dei Redditi 2011, contenente quindi i dati individuali relativi all'anno fiscale 2010 (provenienti dai modelli Unico, 730 e 770).

Il collegamento tra le due basi dati è avvenuto attraverso una procedura di *matching* esatto, che ha utilizzato come chiave il codice fiscale individuale. L'intero processo di costruzione della base dati di input per TREMOD è stata supervisionata dal Servizio Statistica della PaT, che ha garantito il rispetto della riservatezza e la successiva anonimizzazione dei dati.

La base dati di TREMOD è composta pertanto di tre sezioni. Nella prima sezione sono comprese le variabili socio-demografiche, che forniscono informazioni a livello individuale su caratteristiche, quali età, genere, titolo di studio conseguito, e a livello familiare, quali la composizione e le relazioni di parentela tra individui della stessa famiglia. Nella seconda sezione sono ricomprese le variabili riguardanti la situazione lavorativa dell'individuo e la sua storia nell'ultimo anno in termini di mesi lavorati, mesi di pensione, mesi di disoccupazione, così come dichiarate al momento dell'intervista. La terza sezione contiene tutte le variabili reddituali che permettono di ricostruire in modo dettagliato il reddito complessivo lordo, quali il reddito da lavoro (dipendente e autonomo), il reddito da pensione, il reddito da immobili e da terreni, i redditi da investimenti, i redditi frutto di altri lavori autonomi, trasferimenti privati ricevuti (come assegni da ex coniugi, per esempio). Tutte le variabili delle prime due sezioni provengono dall'Indagine sulle condizioni di Vita delle Famiglie Trentine, quelle relativa alla terza dagli archivi amministrativi di fonte fiscale.

### **3.2 La validazione della base dati**

L'utilizzo di dati derivanti da fonte campionaria - per definizione rappresentazione di un sotto-insieme di un aggregato più ampio - risulta critico nel momento in cui essi siano utilizzati per trarre inferenza a livello più generale, sulla popolazione di riferimento. Risulta in quel caso opportuno effettuare un'operazione di "validazione" della base dati campionaria al fine di verificare la sua rappresentatività dell'universo di riferimento.

Per compiere tale operazione, mettiamo a confronto le distribuzioni di alcune delle variabili socio-economiche, occupazionali e reddituali rilevate in ICFT con statistiche di fonte esterna (Servizio Statistica della PaT oppure ISTAT) che serviranno come riferimento (in quanto basati su campioni più ampi, e quindi caratterizzati da minore variabilità, o sull'intera popolazione). Se da tale confronto emergessero elevate difformità, vi sarebbero problemi di inferenza statistica, ossia di generalizzazione all'intera popolazione dei risultati ottenuti con l'analisi dei dati campionari. Questo avrebbe chiare implicazioni negative sulle analisi effettuate e, di conseguenza, sulla precisione delle indicazioni di *policy* derivanti. E', inoltre, particolarmente utile verificare la rappresentatività della base dati campionari con riguardo a diversi segmenti socio-demografici della popolazione, in quanto questi possono essere destinatari di politiche pubbliche mirate.

Tra le caratteristiche socio-demografiche individuali rilevate in ICFT e confrontabili con fonti esterne relative vi sono: l'età anagrafica degli individui, il genere, la cittadinanza, la composizione del nucleo familiare, il livello di istruzione e la condizione lavorativa. Le stime delle distribuzioni di queste variabili sono state calcolate utilizzando il peso di riproporzionamento fornito dallo stesso Servizio Statistica della PaT (di seguito "peso originario"). Questo peso è stato costruito utilizzando le seguenti variabili: dimensioni del nucleo familiare, comune di residenza, presenza di cittadini stranieri e persone con più di 65 anni all'interno del nucleo familiare.

La Tabella 1 mette a confronto le stime del numero di individui residenti in Trentino, suddivisi per fasce di età, ottenute da ICFT con le relative statistiche di fonte esterna. Le prime due colonne della tabella indicano le frequenze dei soggetti, per le diverse classi di età, riscontrate rispettivamente nell'indagine campionaria ICFT e da fonte esterna, mentre la terza colonna rappresenta il rapporto tra le numerosità nelle due diverse fonti (moltiplicato per cento). Un valore per quest'ultimo pari a 100 dimostra una perfetta corrispondenza tra il dato ICFT e quello di fonte esterna; valori più bassi del rapporto indicano un problema di sotto-rappresentazione del campione; viceversa, valori superiori a 100, indicano l'esistenza di una sovra-rappresentazione da parte dello stesso. Dalla Tabella 1 emerge come i principali fattori critici riguardino la sotto-rappresentazione dei giovani in età 25-34 anni (che risulterebbero essere pari 51.912 individui anziché 64.328) e da una sovra-rappresentazione dei soggetti in età matura, 55-64enni (75.183 invece di 63.497). I differenziali tra i dati ICFT e le statistiche di fonte esterna, in questo caso, risultano rispettivamente pari a -19,3% e +18,4%. Per quanto riguarda le altre fasce di età, i problemi risultano limitati. La classe di età dei 45-54enni e degli *over-65*enni è molto ben rappresentata, mentre la stima è meno precisa, seppur complessivamente accettabile, per quanto riguarda le rimanenti fasce (con differenziali

di rappresentazione vicini al 10%). Tali problemi di scarsa rappresentatività non sono certamente una novità. Al contrario, questi si riscontrano di sovente in indagini di tipo campionario, dove la partecipazione è volontaria e il tasso di risposta spesso differisce per livelli di età. Tipicamente, per evitare problemi di distorsione delle stime finali, il problema è affrontato alla conclusione della rilevazione mediante l'applicazione di adeguate tecniche di post-stratificazione (questo è quello che verrà fatto anche nel nostro caso, si veda a proposito il paragrafo 3.3).

Tabella 1: Numero di soggetti residenti in provincia per classe d'età. Confronto tra ICFT e fonti ufficiali.

Fasce d'età	Freq. dati Panel (1)	Freq. fonte esterna (2)	Rapporto % (1/2)
Meno di 15	73.300	80.469	91,1
15-24	58.693	52.131	112,6
25-34	51.912	64.328	80,7
35-44	74.440	85.304	87,3
45-54	79.203	77.682	102,0
55-64	75.183	63.497	118,4
Più di 64	106.802	101.415	105,3
Totale	519.533	524.826	99,0

Fonte: elaborazioni FBK-IRVAPP

Note: i dati si riferiscono alla provincia di Trento per l'anno 2010. Le statistiche riferite a ICFT sono ottenute utilizzando i pesi di post-stratificazione rilasciati dal Servizio Statistica (peso originario). I dati da fonte esterna sono forniti dall'ISTAT (demoistat.it)

Altre caratteristiche socio-demografiche delle quali è possibile effettuare la validazione sono: la cittadinanza, il genere, la composizione del nucleo familiare ed il livello di istruzione degli individui. Per la maggior parte di queste variabili, la fonte ICFT si dimostra altamente rappresentativa della popolazione provinciale (Tabella 2). ICFT si dimostra particolarmente in grado di rappresentare la composizione per cittadinanza della popolazione provinciale: la stima della percentuale di individui con cittadinanza non italiana è dell'8,9% rispetto al 9,2% derivante da fonte esterna. ICFT è altamente rappresentativo anche in termini di composizione di genere: la percentuale di donne nel campione è pari al 52,2%, solo di poco superiore al 51,1% derivante dalle statistiche esterne.

Per quanto riguarda le variabili relative alla composizione del nucleo familiare, i soggetti individuati in ICFT risultano rappresentativi, in larga parte, del numero di famiglie con figli (il 38,7% del totale dei nuclei famigliari invece del 36,8%), delle coppie senza figli (il 23,5% invece che il 24,9%) e del numero di persone che vivono sole o che non formano un nucleo familiare (il 30,1% invece che il 32,9%). Una problematicità riguarda, invece, la rappresentatività dei nuclei mono-genitoriali che sono largamente sovra-stimati in ICFT (il 7,7% invece che il 5,4% delle famiglie). E' da ricordare che la misurazione di fenomeni con bassa incidenza nella popolazione porta ad avere stime maggiormente instabili rispetto a fenomeni largamente diffusi.

Risultati molto convincenti riguardano invece la capacità di ICFT di riprodurre la distribuzione dei titoli di studio della popolazione trentina. La rappresentatività di soggetti che hanno conseguito la licenza di scuola media, il diploma maturità e la laurea (e post-laurea) non presenta criticità, con valori molti simili a quelli di fonte Istat, a parte una lieve sovra-stima delle persone con titolo di studio terziario.

Tabella 2: Distribuzioni relative di alcune caratteristiche socio-demografiche. Confronto tra ICFT e fonti ufficiali.

Caratteristiche socio-demografiche	Freq. dati Panel (1)	Freq. fonte esterna (2)	Rapporto % (1/2)
Cittadinanza (non italiana)	8,9	9,2	97,2
Sesso (donne)	52,2	51,1	102,2
Nessun nucleo	30,1	32,9	91,4
Coppia con figli	38,7	36,8	105,2
Coppia senza figli	23,5	24,9	94,6
Nucleo mono-genitore	7,7	5,4	142,2
Elementari o nessun titolo	17,6	18,0	97,9
Licenza media	29,0	28,7	101,1
Maturità o qualifica professionale	41,0	41,5	99,0
Laurea o post laurea	12,3	11,8	104,1

Fonte: elaborazioni FBK-IRVAPP

Note: i dati si riferiscono alla provincia di Trento per l'anno 2010. Le statistiche riferite a ICFT sono ottenute utilizzando i pesi di post-stratificazione rilasciati dal Servizio Statistica (peso originario). I dati di fonte esterna su cittadinanza e genere sono forniti dal Servizio Statistica della PaT, quelli sulle condizioni familiari sono tratti dall'Annuario statistico 2010, quelli sui livelli di educazione dall'Istat (Indagine sulle Forze di Lavoro).

Un ulteriore gruppo di variabili rilevanti, delle quali si intende verificare la rappresentatività in ICFT rispetto ai valori di popolazione, sono quelle riguardanti il mercato del lavoro, in particolare, la situazione lavorativa dei soggetti intervistati. Nella Tabella 3 sono riportati i tassi di occupazione, attività e di disoccupazione (definiti sul totale della popolazione, sui giovani e sulle donne) in considerazione dei soggetti intervistati in ICFT (colonna 1) e derivanti da fonte esterna Istat, tramite l'“Indagine sulle forze di lavoro” (colonna 2). Il confronto tra le due fonti dimostra come in ICFT sia rappresentato un minor numero di persone occupate (il tasso di occupazione riscontrato in ICFT è pari al 58,8% a fronte del 66,0% reale) e di attività (62,6% invece che 69%). Da un'analisi per genere dei tassi appena citati, si nota inoltre una lieve sovra-rappresentazione nel numero di donne intervistate che appartengono alle forze di lavoro.

Il tasso di disoccupazione risulta essere sovrastimato: in ICFT è pari al 5,9% mentre il valore da fonte esterna (Indagine sulle forze di lavoro) è di 4,3%. Le donne risultano disoccupate nel 7,3% dei casi in ICFT contro il 5,2% nei dati di fonte Istat e, analogamente, il tasso di disoccupazione giovanile in ICFT risulta del 21,4% contro il 15,1%. La maggiore frequenza di persone disoccupate intervistate in ICFT potrebbe essere conseguenza della loro maggiore reperibilità e disponibilità a rispondere all'intervista.

Tabella 3: Indicatori di partecipazione al mercato del lavoro. Confronto tra ICFT e fonti ufficiali.

Situazione lavorativa	Freq. dati Panel (1)	Freq. fonte esterna (2)	Rapporto % (1/2)
Tasso di attività	62,6	69,0	90,7
Tasso di occupazione	58,8	66,0	89,1
Tasso di disoccupazione	5,9	4,3	138,0
Tasso di disoccupazione 15-24	21,4	15,1	141,6
Tasso di attività (donne)	64,3	60,5	106,2
Tasso di occupazione (donne)	59,6	57,3	104,0
Tasso di disoccupazione (donne)	7,3	5,2	140,2

Fonte: elaborazioni FBK-IRVAPP

Note: i dati si riferiscono alla provincia di Trento per l'anno 2010. Le statistiche riferite a ICFT sono ottenute utilizzando i pesi di post-stratificazione rilasciati dal Servizio Statistica (peso originario). I dati da fonte esterna concernenti gli indicatori del mercato del lavoro sono tratti da Istat (Indagine sulle Forze di Lavoro).

Infine, un ultimo insieme di variabili di interesse riguarda quelle di natura reddituale. Prendiamo in esame quattro grandezze: reddito da lavoro dipendente, reddito da lavoro autonomo, reddito da pensione e reddito disponibile. Come descritto sopra (Paragrafo 3.1), tali informazioni derivano dall'archivio delle dichiarazioni fiscali dell'Agenzia delle Entrate dell'anno 2011, opportunamente collegati con dati ICFT.

Tabella 4: Numero percettori e ammontare complessivo (in migliaia di Euro) relativo alle principali fonti di reddito dichiarate. Confronto tra ICFT e dati dell'intera popolazione dichiarante, fonte Agenzia delle Entrate (AE).

Fonti di reddito	Freq. Panel (1)	Totale redditi Panel (2)	Freq. AE (3)	Totale redditi AE (4)	Rapporto % (1/3)	Rapporto % (2/4)
Reddito da lavoro dipendente	206.234	4.164.646	239.591	4.426.109	86,1	94,1
Reddito da lavoro autonomo	11.321	201.996	14.022	323.119	80,7	62,5
Pensioni	142.255	2.309.307	135.183	2.036.244	105,2	113,4
Reddito disponibile	373.556	6.668.324	408.046	6.858.059	91,5	97,2

Fonte: elaborazioni FBK-IRVAPP

Note: i dati si riferiscono alla provincia di Trento per l'anno 2010. Le statistiche riferite a ICFT sono ottenute utilizzando i pesi di post-stratificazione rilasciati dal Servizio Statistica (peso originario).

La Tabella 4 mostra le stime ottenute dal database che unisce queste due fonti di dati relativamente al numero di individui (colonna 1) e alla somma dei redditi da essi dichiarati (denominato "complessivo", nella colonna 2). Le colonne 3 e 4 identificano, invece, le stesse informazioni derivanti da fonte esterna, la quale, in questo caso, è rappresentata dall'archivio delle dichiarazioni dei redditi relativo all'intera popolazione trentina e, quindi, non limitato ai soggetti intervistati in ICFT.

Rapportando il numero di soggetti stimati con ICFT al numero di soggetti rinvenuti nell'archivio delle dichiarazioni dei redditi (penultima colonna), si registra una sottosti-

ma da parte di ICFT dei soggetti dichiaranti redditi da lavoro dipendente (circa il 14% in meno rispetto al dato di popolazione) e autonomo (il 20% in meno, circa). Risulta parzialmente sovra-stimata la rappresentatività del numero di pensionati, che sono presenti in misura maggiore (+5,2%) in ICFT.

Considerando invece il rapporto tra la stima ICFT e l'ammontare dei redditi coinvolti (ultima colonna), quello da lavoro dipendente risulta sotto-stimato del 5,9%. Al contrario, quello da pensione è sovra-stimato del 13,4%. La criticità maggiore deriva però dai profili reddituali riguardanti i lavoratori autonomi che, in base a ICFT, risultano essere di molto inferiori rispetto a quelli di popolazione (del 40% circa). In altri termini ICFT, oltre a sotto-rappresentare la numerosità dei lavoratori autonomi, ne sotto-stima altresì largamente i profili reddituali. Un risultato confortante proviene, invece, dal reddito disponibile dei soggetti intervistati: esso risulta essere rappresentativo, con uno scostamento molto limitato (- 2,8%) rispetto al valore riferito all'intera popolazione.

### **3.3 La post-stratificazione**

Come descritto nel paragrafo precedente, ICFT non risulta in grado di rappresentare correttamente la popolazione trentina in alcune delle sue caratteristiche demografiche, sociali e reddituali rilevanti per le microsimulazioni di TREMOD. Il problema risulta particolarmente evidente per quanto riguarda la numerosità e i redditi di alcune fasce di età della popolazione (25-34 e 55-64) e dei lavoratori autonomi. Altre difformità, seppur meno rilevanti, riguardano la condizione familiare e la numerosità dei soggetti attivi occupati e dei disoccupati. Si rende quindi necessaria una correzione delle stime ICFT tramite ribilanciamento da attuarsi per mezzo di un procedimento statistico di post-stratificazione. Quest'ultimo si fonda sulla costruzione di opportuni pesi in grado di riproporzionare le numerosità campionarie al fine di renderle maggiormente rappresentative di quelle relative alla popolazione trentina lungo le dimensioni ritenute rilevanti (Judkins, 1990).

La costruzione dei nuovi pesi è avvenuta utilizzando il programma *survwgt* di Stata. Questo programma permette di calcolare un vettore di pesi che, se applicato ai dati ICFT, minimizza la discrepanza tra questi ultimi e quelli derivanti da fonte esterna. Il punto di partenza del processo di correzione nelle frequenze rilevate in ICFT, è dato dal riproporzionamento peso originario ICFT, rilasciato dal Servizio Statistica della PaT. Sfortunatamente questo tipo di riproporzionamento dei dati campionari genera un problema che possiamo definire di "coperta corta". Infatti, controllare per il bilanciamento lungo una dimensione (variabile) potrebbe dover comportare una correzione di un'altra in una direzione opposta. In altre parole, la correzione lungo una o più distribuzioni marginali non consente la correttezza della distribuzione congiunta. Questo proble-

ma risulta più evidente maggiore è il numero di variabili considerate nell'algoritmo di post-stratificazione.

La Tabella 5 mostra il rapporto percentuale tra le numerosità di individui osservate in ICFT e dalle fonti esterne, secondo le variabili socio-demografiche, di condizione familiare, e di posizione sul mercato del lavoro, quale risulta dall'utilizzazione di diversi pesi di post-stratificazione. In particolare, sono stati calcolati quattro diversi pesi i quali, in via incrementale, coinvolgono un crescente numero di caratteristiche. Il primo peso (Peso 1) tiene conto delle principali variabili socio-demografiche ritenute rilevanti, in particolare l'età, la cittadinanza e il sesso. Il Peso 2 somma al controllo delle appena citate caratteristiche anche quello del livello di istruzione. Il Peso 3 aggiunge al Peso 2 le caratteristiche individuali legate alle posizioni sul mercato del lavoro. Infine, il Peso 4 aggiunge la rappresentatività dei soggetti impiegati in impiego di tipo autonomo.

Tabella 5: Distribuzioni relative di alcune caratteristiche socio-demografiche e indicatori di partecipazione al mercato del lavoro. Rapporto percentuale tra dato ICFT (diversi pesi) e fonti ufficiali.

Variabili	Peso originario (Serv. Stat.)	Peso 1	Peso 2	Peso 3	Peso 4
Meno di 15	91,1	100,0	100,0	100,0	100,0
15-24	112,6	100,0	100,0	100,0	100,0
25-34	80,7	100,0	100,0	100,0	100,0
35-44	87,3	100,0	100,0	100,0	100,0
45-54	102,0	100,0	100,0	100,0	100,0
55-64	118,4	100,0	100,0	100,0	100,0
Più di 64	105,3	100,0	100,0	100,0	100,0
Totale	99,0	100,0	100,0	100,0	100,0
Cittadinanza (non italiana)	97,2	100,0	100,0	100,0	100,0
Sesso (donne)	102,2	100,0	100,0	100,0	100,0
Nessun nucleo	91,4	89,3	89,2	89,8	90,0
Coppia con figli	105,2	110,3	110,1	109,6	109,4
Coppia senza figli	94,6	88,8	89,3	89,2	89,3
Nucleo mono-genitore	142,2	146,0	146,4	146,6	146,2
Elementari o nessun titolo	97,9	91,5	100,0	100,0	100,0
Licenza media	101,1	97,7	100,0	100,0	100,0
Maturità o qualifica professionale	99,0	102,3	100,0	100,0	100,0
Laurea o post laurea	104,1	110,6	100,0	100,0	100,0
Tasso di attività	90,7	94,9	94,4	97,4	97,4
Tasso di occupazione	89,1	93,3	92,7	97,3	97,3
Tasso di disoccupazione	138,0	136,2	138,9	101,5	101,1
Tasso di disoccupazione 15-24	141,6	140,5	141,0	106,2	105,1
Tasso di attività (donne)	106,2	109,7	108,4	111,6	111,5
Tasso di occupazione (donne)	104,0	107,4	106,1	111,5	111,4
Tasso di disoccupazione (donne)	140,2	139,1	141,6	104,2	103,6

Fonte: elaborazioni FBK-IRVAPP

Note: i dati si riferiscono alla provincia di Trento per l'anno 2010. Il Peso 1 riproporziona sulla base della classe d'età, del sesso e della cittadinanza. Il Peso 2 aggiunge la variabile sui livelli di istruzione (licenza media, superiore e università), il Peso 3 quelle relative al mercato del lavoro (tasso di occupazione, disoccupazione e attività) e il Peso 4 controlla, infine, per numero di lavoratori autonomi. Per le fonti si veda Tabelle 1, 2 e 3.

La Tabella 5 mostra quindi il rapporto percentuale tra i valori ottenuti, per le variabili di interesse, applicando il peso originario ICFT e i valori derivanti dall'applicazione dei diversi pesi calcolati con il processo di post-stratificazione.

Se, inizialmente, le diverse classi di età della popolazione trentina non risultavano sempre rappresentate in numerosità corrispondenti a quelle delle fonti esterne (in particolare, nella fascia di età dei 15-24enni e dei 25-34enni) il processo di post-stratificazione consente un netto miglioramento della qualità dei dati a disposizione. L'incremento della corrispondenza tra dati ICFT e dati di fonte esterna risulta essere costante rispetto a tutti i pesi considerati e, nonostante l'inclusione di molteplici variabili, la distribuzione delle diverse classi di età, in ICFT e nella popolazione, sono pienamente coincidenti. I dati dimostrano che, anche per quanto riguarda il sesso e la cittadinanza, le differenze tra le frequenze osservate nei dati ICFT e quelli riscontrati nella popolazione, risultano inoltre essere completamente annullati correggendo i (leggeri) squilibri iniziali.

Per quanto riguarda le caratteristiche familiari si è scelto di non inserirle direttamente nel processo di post-stratificazione del campione, in ragione dell'impossibilità di controllare tutte le variabili pertinenti. A seguito dell'applicazione dei vari pesi considerati, la difformità nella distribuzione per tipologia familiare rispetto ai dati di fonte esterna risulta limitata. Più incisiva è la sovra-rappresentazione dei soggetti che formano famiglia mono-genitoriale, così come già ottenuto con l'applicazione del peso originario. Come già detto in precedenza, tale errore è comunque, in realtà, di limitata entità data la lieve numerosità degli stessi.

Le caratteristiche rispetto al livello di istruzione degli intervistati risultano, così come per quanto riguarda le variabili relative all'età, al sesso e alla cittadinanza, perfettamente rappresentative dei valori reali a partire dall'applicazione del Peso 2 che tiene conto delle variabili di istruzione nel processo di calcolo.

L'operazione di post-stratificazione per le variabili del mercato del lavoro avviene, invece, a partire dall'applicazione del Peso 3 e si dimostra anche in questo caso efficace nel ricondurre gli iniziali squilibri tra le frequenze riscontrate in ICFT e nei di fonte esterna. Il tasso di disoccupazione aggregato (ma anche disaggregato per i gruppi dei giovani, e delle donne) viene ricondotto a valori percentualmente vicini a quelli di fonte esterna. Viene in questo modo ad essere superato un problema che, benché riguardasse valori contenuti, poteva risultare particolarmente rilevante nell'analisi delle politiche trentine dal momento che molte delle politiche allo studio sono specificatamente indirizzate alla categoria dei disoccupati. Per quanto riguarda i valori dei tassi di occupazione e di attività, essi risultano in larga parte corrispondenti ai dati di riferimento, con lievi sovra-stime che permangono solo per quanto riguarda gli indicatori legati al genere.

La Tabella 6 mostra rispettivamente i rapporti percentuali relativi al numero di individui (frequenza) e all'ammontare complessivo dichiarato per tipologia di reddito. Il

Peso 4 corregge in particolare per la distorsione dovuta alla sotto-rappresentazione del numero e del livello medio di reddito dei lavoratori autonomi, inserendo nel processo di ottimizzazione la condizione di bilanciamento per il numero dei lavoratori autonomi. A seguito dell'operazione di post-stratificazione, l'applicazione del Peso 4 consente di aumentare sensibilmente il numero dei lavoratori autonomi rappresentati in ICFT. La numerosità di soggetti dichiaranti altre fattispecie reddituali (lavoratori dipendenti, pensionati e reddito disponibile) rimangono in una discreta approssimazione del dato reale. Nonostante la correzione per le frequenze relative, la stessa tabella in discussione evidenzia un miglioramento incrementale della capacità di rappresentare i redditi dei lavoratori autonomi seppur non in grado di portare la loro rappresentatività a livelli del tutto soddisfacenti. I redditi da lavoro autonomo rimangono quindi sotto-stimati rispetto ai valori di popolazione, mentre quelli dei redditi da lavoro dipendente e pensione, nonché il reddito disponibile, risultano invece in una buona approssimazione dei valori reali.

Tabella 6: Rapporto percentuale tra dato ICFT (diversi pesi) e dati dell'intera popolazione dichiarante, fonte Agenzia delle Entrate (AE). Numero di percettori e ammontare complessivo relativo ai principali redditi dichiarati.

Fonti di reddito	Peso originario (Serv. Stat.)	Peso 1	Peso 2	Peso 3	Peso 4
Rapporti relativi alla frequenza					
Reddito da lav. dip.	86,1	90,0	89,3	91,1	90,8
Reddito da lav. aut.	80,7	82,4	79,2	77,7	99,5
Pensioni	105,2	97,0	97,6	96,6	96,5
Reddito disponibile	91,5	91,7	91,6	92,3	92,3
Rapporti relativi agli ammontari complessivi					
Reddito da lav. dip.	94,1	98,5	96,7	99,9	99,3
Reddito da lav. aut.	62,5	64,0	60,0	60,8	78,0
Pensioni	113,4	104,7	102,7	100,9	100,8
Reddito disponibile	97,2	98,0	96,5	98,1	98,4

Fonte: elaborazioni FBK-IRVAPP

Note: i dati si riferiscono alla provincia di Trento per l'anno 2010. Il peso 1 riproporziona sulla base della classe d'età, del sesso e della cittadinanza. Il peso 2 aggiunge la variabile sui livelli di istruzione (licenza media, superiore e università), il peso 3 quelle relative al mercato del lavoro (tasso di occupazione, disoccupazione e attività) e il peso 4 controlla, infine, per numero di lavoratori autonomi. Per le fonti si veda Tabella 4.

Complessivamente, il processo di post-stratificazione appena descritto, permette, tramite il Peso 4, di ribilanciare i dati di fonte ICFT al fine di renderli particolarmente adatti a rappresentare la popolazione trentina per quasi la totalità delle variabili che sono ritenute fondamentali per le future analisi di TREMOD. In particolare, la rappresentatività di diversi segmenti di popolazione (le fasce di età), la distribuzione per sesso e cittadinanza, il livello di istruzione, la condizione familiare e lo stato occupazionale. Anche le numerosità e la situazione reddituale dei soggetti intervistati risultano in grado di cogliere in maniera efficace le condizioni reali della dimensione trentina, con parti-

colare riferimento ai lavoratori dipendenti e ai pensionati. Rimangono alcune criticità relative all'ammontare del reddito dichiarato da parte dei lavoratori autonomi, i quali rimangono in parte sotto-stimati. D'altronde, le dimensioni coinvolte nel processo di post-stratificazione sono molte e i risultati non possono evidentemente tenere conto di tutto allo stesso momento. Le condizioni di partenza sono state comunque nettamente migliorate e la post-stratificazione ha consentito un elevato cambio di passo verso una reale rappresentazione di ICFT rispetto alle condizioni reali della popolazione trentina.

#### 4 Analisi delle stime prodotte con TREMOD

Il modello TREMOD permette di simulare l'entità delle imposte sul reddito e quindi di calcolare il reddito disponibile a livello individuale e familiare a partire da informazioni relative al reddito complessivo e alla composizione familiare. L'attività di validazione di TREMOD, presentata in questa sezione, è finalizzata a verificare che ogni componente calcolata dal modello (ad esempio le imposte, le detrazioni e le deduzioni) corrisponda ai dati reali. Le componenti di interesse per questa analisi sono diverse e riguardano il reddito complessivo, il reddito imponibile, le imposte lorde e nette, vari tipi di detrazioni e deduzioni.

Tabella 7: Analisi della simulazione delle variabili fiscali utilizzando TREMOD, confronto con i dati amministrativi delle dichiarazioni dei redditi di ciascun contribuente presente in ICFT. Numero percettori (Freq.) e ammontare medio dichiarato (in Euro).

Fonti di reddito	Freq. TRE-MOD (1)	Media TRE-MOD (2)	Freq. AE (3)	Media AE (4)	Rapporto % (1/3)	Rapporto % (2/4)
Reddito complessivo	5.132	21.415	5.132	21.403	100,0	100,1
Altre deduzioni	2.347	1.568	2.343	1.570	100,2	99,9
Reddito imponibile	5.114	21.193	5.072	20.716	100,8	102,3
Imposta lorda	5.079	5.561	4.956	5.705	102,5	97,5
Imposta netta	4.026	4.993	4.105	4.942	98,1	101,0
Detrazioni da lavoro e pensioni	4.768	1.114	4.604	1.127	103,6	98,8
Detrazioni famigliari a carico	1.315	1.085	1.280	1.077	102,7	100,7
detrazioni figli e coniuge a carico	1.308	1.082	1.277	1.072	102,4	100,9
altre detrazioni famigliari	12	948	12	874	100,0	108,5
Altre detrazioni	3.268	647	3.261	645	100,2	100,3
Detrazioni totali	5.029	1.760	4.949	1.701	101,6	103,5
Reddito disponibile	5.114	17.347	5.132	17.451	99,6	99,4

Fonte: elaborazioni FBK-IRVAPP.

Note: i dati si riferiscono alla provincia di Trento per l'anno 2010. Le stime sul campione ICFT utilizzano il Peso 4 (Cfr. Tab. 2).

In primo luogo, mettiamo a confronto le stime ottenute con TREMOD con i valori reali desunti dalle dichiarazioni dei redditi relative ai soggetti facenti parte del campione ICFT. La Tabella 7 mostra che TREMOD riproduce delle stime molto aderenti ai valori reali, per tutte le dimensioni reddituali prese in considerazione. I rapporti tra le stime reddituali calcolate da TREMOD e quelle delle dichiarazioni dei redditi delle persone intervistate in ICFT sono sempre molto vicini a 100, in tutti i casi, dimostrando la bontà di TREMOD nell'applicazione delle normativa fiscale e contributiva di riferimento per gli individui che vivono in Trentino.

In secondo luogo, poichè la finalità di TREMOD è quella di fornire delle stime che siano valide per l'intera popolazione, cerchiamo di verificare in che modo i risultati di TREMOD siano o meno vicini ai dati delle dichiarazioni dei redditi di tutta la popolazione trentina (Tabella 8).

Tabella 8: Analisi della simulazione delle variabili fiscali utilizzando TREMOD, confronto con i dati amministrativi delle dichiarazioni dei redditi dell'intera popolazione dei dichiaranti. Numero percettori (Freq.) e ammontare complessivo dichiarato (in migliaia di Euro).

Fonti di reddito	Freq. TRE-MOD (1)	Totale TRE-MOD (2)	Freq. AE (3)	Totale AE (4)	Rapporto % (1/3)	Rapporto % (2/4)
Reddito complessivo	380.505	8.100.595	412.367	8.218.799	92,3	98,6
reddito imponibile	379.373	7.989.332	407.555	7.873.687	93,1	101,5
Imposta lorda	376.961	2.066.770	399.786	2.110.699	94,3	97,9
Imposta netta	302.892	1.461.191	326.563	1.515.161	92,8	96,4
Detrazioni da lavoro e pensioni	355.729	400.676	370.366	421.017	96,0	95,2
Reddito disponibile	379.373	6.561.434	412.367	6.703.856	92,0	97,9

Fonte: elaborazioni FBK-IRVAPP.

Note: i dati si riferiscono alla provincia di Trento per l'anno 2010. Le stime TREMOD utilizzano il Peso 4 di riporto all'intera popolazione (Cfr. Tab. 2).

Per rapportare i dati campionari a quelli di universo, utilizziamo il peso ritenuto come "ottimale", cioè il Peso 4, in quanto controlla per le diverse dimensioni ritenute rilevanti, a seguito del calcolo di post-stratificazione visto nel sottoparagrafo precedente. Si può osservare come il numero di persone che ha un reddito complessivo (di qualsiasi tipo) sia leggermente sottostimato, il 92,3% rispetto al dato reale di popolazione, mentre l'ammontare del redditi complessivo rappresentato è vicino al 100%.

Questo può significare che sono mediamente sottostimati gli individui con redditi inferiore alla media, errore che va influenzare il rapporto tra le frequenze, ma solo marginalmente quello relativo agli ammontari.

Per quanto riguarda le altre grandezze considerate (reddito imponibile, imposta lorda, netta, detrazioni da lavoro e pensione, reddito disponibile), anche in questo caso

TREMODO si dimostra in grado di rappresentare in maniera buona, seppur con lievi discrepanze, le fattispecie reddituali considerate.

## 5 La comparazione tra TREMOD e EUROMOD-IT

L'aspetto maggiormente innovativo della prima versione di TREMOD rispetto al modulo italiano di EUROMOD (EUROMOD-IT) è costituito dalla banca dati di *input*. Essa, oltre ad avere una dimensione circa dieci volte maggiore del sotto-campione trentino di IT-SILC (il database di input di EUROMOD-IT), è incrociata in modo esatto mediante il codice fiscale con fonti di dati amministrative che forniscono un'informazione significativamente più dettagliata e precisa di quella fornita dall'intervistato nell'indagine campionaria IT-SILC. Questo *matching* esatto ha enormi potenzialità in termini di sviluppi futuri del *database* di partenza e del modello che verranno brevemente discussi nelle conclusioni nel paragrafo 6.

A questo punto risulta comunque importante interrogarci su cosa guadagniamo dall'utilizzare TREMOD invece che EUROMOD-IT per la microsimulazione fiscale relativa ai residenti della PaT, mettendo brevemente a confronto i risultati ottenibili con ciascun modello.

La Tabella 9 mette a confronto le stime del numero di individui che percepiscono redditi e dell'ammontare medio percepito utilizzando TREMOD (e quindi il campione ICFT), EUROMOD limitatamente al sotto campione di SILC riferito al Trentino (SILC-TN), la totalità delle dichiarazioni dei residenti in Trentino presentate all'Agenzia delle Entrate (AE). Le stime mostrano uno scostamento inferiore dai dati dell'AE delle medie del reddito complessivo e disponibile utilizzando TREMOD invece che SILC-TN, mentre l'imposta netta stimata in SILC si avvicina maggiormente ai dati AE. Gli *standard error* nella tabella in discussione evidenziano come le stime di SILC siano più variabili, anche in conseguenza della minore numerosità del campione SILC che conta, per il Trentino, 793 intervistati mentre ICFT ne coinvolge nel 2010 ben 7.200. I risultati ottenuti dai due campioni sono abbastanza soddisfacenti, nonostante le medie relativi alla popolazione talvolta non rientrano negli intervalli di confidenza ottenuti dalle stime ottenute usando i campioni di ICFT e di SILC.

La Tabella 10 riporta i valori del decimo percentile, della mediana e del novantesimo percentile ottenuti dalla distribuzione dei redditi della popolazione, ICFT e SILC. E' da notare che la parte inferiore della distribuzione dei redditi, ed in particolare il decimo percentile, risulta meglio approssimata in SILC-TN rispetto a TREMOD. Questo risultato tuttavia è meno evidente se si suddivide la popolazione per classi d'età e genere. Ciononostante questi risultati suggeriscono cautela nell'analisi di politiche fiscali mirate agli individui con minore reddito.

Tabella 9: Confronto tra i risultati del modello TREMOD, di EUROMOD-IT (sottocampione Trentino - SILC(TN)) e dei dati dell'Agenzia delle Entrate - Numero percettori (Freq.), valori medi dichiarati (in Euro) e s.e.

Fonti di reddito	Freq. TREMOD	Media TREMOD	Freq. SILC (TN)	Media SILC (TN)	Freq. AE	Media AE
Popolazione						
Reddito complessivo	380.505	21.289	400.425	21.845	412.367	19.931
s.e.		290		731		
Reddito disponibile	379.373	17.295	402.686	17.710	412.367	16.257
s.e.		197		539		
Imposta netta	302.892	4.824	331.010	4.519	326.563	4.640
s.e.		120		239		
Classe d'età 25-45						
Reddito complessivo	132.429	21.500	139.731	23.605	143.355	19.569
s.e.		391		967		
Reddito disponibile	132.127	17.744	139.731	19.267	143.355	16.181
s.e.		277		728		
Imposta netta	113.031	4.276	119.976	4.499	117.672	4.129
s.e.		137		306		
Classe d'età 46-65						
Reddito complessivo	125.836	26.379	129.558	27.400	134.800	25.444
s.e.		615		1.615		
Reddito disponibile	125.453	20.943	131.819	21.586	134.800	20.242
s.e.		400		1.168		
Imposta netta	106.833	6.169	113.742	5.806	113.766	6.164
s.e.		257		530		
Uomini						
Reddito complessivo	193.121	26.585	199.697	27.641	213.289	24.196
s.e.		493		1.178		
Reddito disponibile	192.947	21.114	199.697	22.274	213.289	19.302
s.e.		325		849		
Imposta netta	167.834	6.033	174.750	5.707	180.564	5.781
s.e.		198		390		
Donne						
Reddito complessivo	187.384	15.831	200.728	16.079	198.966	15.362
s.e.		257		728		
Reddito disponibile	186.426	13.343	202.989	13.220	198.966	12.994
s.e.		190		560		
Imposta netta	135.058	3.322	156.260	3.191	145.894	3.229
s.e.		93		227		

Fonte: elaborazioni FBK-IRVAPP. Note: i dati si riferiscono alla provincia di Trento per l'anno 2010. Le stime TREMOD utilizzano il Peso 4 di riporto all'intera popolazione (Cfr. Tab. 2). I dati dell'Agenzia delle Entrate sono relativi all'intera popolazione dei dichiaranti nella provincia di Trento. I dati di EUROMOD-IT (sottocampione Trentino) si basano sui redditi dichiarati dagli individui intervistati nell'indagine SILC.

Tabella 10: Confronto tra i risultati del modello TREMOD, di EUROMOD-IT (sottocampione Trentino - SILC(TN)) e dei dati dell'Agenzia delle Entrate - per i percentili 10, 50 e 90 (Valori in Euro).

Fonti di reddito/Percentili	10% TRE- MOD	50% TRE- MOD	90% TRE- MOD	10% SILC (TN)	50% SILC (TN)	90% SILC (TN)	10% AE	50% AE	90% AE
Popolazione									
Reddito complessivo	4.000	18.306	37.729	2.210	20.236	42.161	2.001	16.772	36.027
Reddito disponibile	4.099	15.664	29.240	1.415	16.721	32.930	1.915	14.627	28.445
Imposta netta	600	3.121	9.723	416	3.271	9.863	329	2.839	9.159
Classe d'età 25-45									
Reddito complessivo	5.684	19.997	35.125	6.183	23.014	39.884	1.670	18.346	33.812
Reddito disponibile	5.589	16.908	28.209	5.810	18.961	31.194	1.578	15.863	27.172
Imposta netta	775	3.195	7.646	1.238	3.699	8.292	340	2.966	7.934
Classe d'età 46-65									
Reddito complessivo	5.987	22.131	44.859	5.081	23.185	51.999	3.964	20.818	45.314
Reddito disponibile	5.986	18.654	34.314	2.443	18.753	41.775	3.771	17.812	34.681
Imposta netta	637	3.730	11.650	463	3.664	12.158	547	3.594	12.500
Uomini									
Reddito complessivo	6.754	22.224	45.983	7.475	24.325	50.003	2.093	20.126	43.498
Reddito disponibile	6.583	18.686	34.933	7.239	19.912	38.873	2.000	17.240	33.436
Imposta netta	952	3.672	12.565	746	3.846	11.501	528	3.407	11.656
Donne									
Reddito complessivo	2.545	13.791	29.243	1.186	14.261	31.782	1.953	13.118	28.968
Reddito disponibile	2.639	12.193	23.445	653	12.282	25.403	1.854	11.848	23.532
Imposta netta	399	2.428	6.601	261	2.408	6.627	225	2.124	6.534

Fonte: elaborazioni FBK-IRVAPP. Note: i dati si riferiscono alla provincia di Trento per l'anno 2010. Le stime di TREMOD utilizzano il Peso 4 per riportare le frequenze all'intera popolazione. (Cfr. Tab. 2). I dati dell'Agenzia delle Entrate sono relativi all'intera popolazione dei dichiaranti nella provincia di Trento. I dati di EUROMOD-IT (sottocampione Trentino) si basano sui redditi dichiarati dagli individui intervistati nell'indagine SILC.

Per quanto riguarda la mediana e il novantesimo percentile le stime utilizzando TREMOD si avvicinano maggiormente alla distribuzione della popolazione trentina e il risultato si conferma nei sottocampioni costruiti in relazione all'età e genere. Ciò conferma che ICFT è sufficientemente rappresentativo della distribuzione dei redditi nella popolazione trentina, con qualche limite nella rappresentatività della parte bassa della distribuzione del reddito.

Un vantaggio rilevante di TREMOD rispetto a EUROMOD-IT riguarda il fatto di poter riuscire a catturare con maggiore precisione la distribuzione della parte alta o molto alta della distribuzione. La letteratura economica ha negli ultimi anni sottolineato l'importanza dell'analisi dei *top incomes*, focalizzandosi sulla quota di reddito totale detenuto dal 5%, 1% o addirittura lo 0,01% della popolazione più ricca (si veda, per esempio, Atkinson et al., 2011). Considerato che la rappresentatività per la coda alta della distribuzione del reddito possa essere molto limitata in campioni ristretti, il modello TREMOD, che si basa su un *dataset* che raccoglie oltre 7.000 interviste nella sola PaT, potrebbe avere dei vantaggi strutturali rispetto a EUROMOD-IT.

Tabella 11: Confronto tra i risultati del modello TREMOD, di EUROMOD-IT (sottocampione Trentino - SILC(TN)) e dei dati dell'Agenzia delle Entrate per i percentili più alti della distribuzione (Valori in Euro).

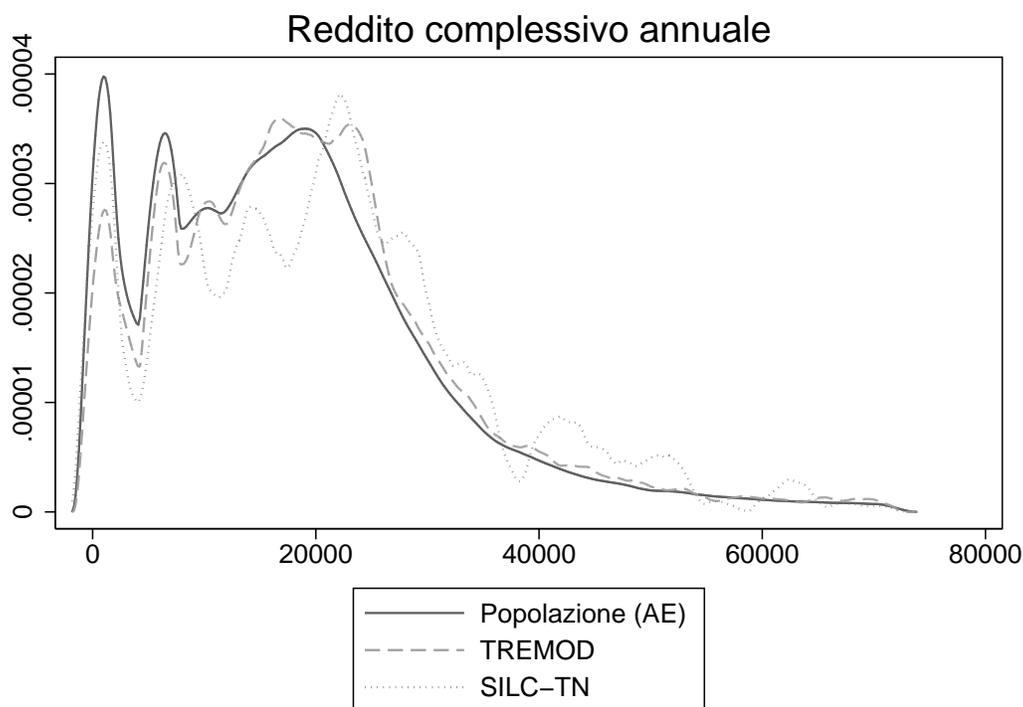
Fonti di reddito/Percentili	95% TRE- MOD	99% TRE- MOD	99.99% TRE- MOD	95% SILC (TN)	99% SILC (TN)	99.99% SILC (TN)	95% AE	99% AE	99.99% AE
<b>Popolazione</b>									
Reddito complessivo	49.813	100.210	472.828	51.047	87.050	217.918	49.202	103.705	534.857
Reddito disponibile	37.331	66.174	293.789	40.217	60.193	160.569	37.095	70.568	324.296
Imposta netta	14.757	35.343	179.039	13.055	27.431	57.349	14.695	37.461	231.011
<b>Classe d'età 25-45</b>									
Reddito complessivo	46.426	87.134	130.396	44.431	62.905	128.550	44.587	82.734	485.502
Reddito disponibile	35.503	59.358	88.369	34.728	42.438	85.008	34.137	58.633	286.031
Imposta netta	12.735	29.102	42.348	11.501	18.532	43.542	11.991	27.488	203.735
<b>Classe d'età 46-65</b>									
Reddito complessivo	61.947	136.149	472.828	76.824	109.211	217.918	66.498	132.706	643.790
Reddito disponibile	45.152	87.719	293.789	56.488	74.716	160.569	48.324	88.402	387.563
Imposta netta	17.822	48.175	179.039	20.379	37.127	57.349	21.130	48.198	276.924
<b>Uomini</b>									
Reddito complessivo	64.741	126.244	472.828	62.872	91.110	217.918	61.859	126.488	680.924
Reddito disponibile	46.354	84.962	293.789	45.889	62.748	160.569	45.318	84.501	404.000
Imposta netta	18.973	42.027	179.039	17.102	34.495	57.349	19.187	46.055	298.537
<b>Donne</b>									
Reddito complessivo	35.760	61.947	162.281	35.904	49.136	97.086	35.540	66.484	302.308
Reddito disponibile	28.116	46.300	150.582	29.366	37.583	64.711	27.909	47.851	182.720
Imposta netta	9.159	21.141	56.730	9.774	14.747	35.920	9.302	22.748	121.001

Fonte: elaborazioni FBK-IRVAPP. Note: i dati si riferiscono alla provincia di Trento per l'anno 2010. Le stime di TREMOD utilizzano il Peso 4 per riportare le frequenze all'intera popolazione. (Cfr. Tab. 2). I dati dell'Agenzia delle Entrate sono relativi all'intera popolazione dei dichiaranti nella provincia di Trento. I dati di EUROMOD-IT (sottocampione Trentino) si basano sui redditi dichiarati dagli individui intervistati nell'indagine SILC.

La Tabella 11 conferma che TREMOD risulta maggiormente preciso nel rappresentare la distribuzione dei redditi più alti e quindi più adeguato per analizzare l'impatto di politiche fiscali che vadano ad incidere sul reddito disponibile dei più ricchi. Fino a questo punto abbiamo osservato alcuni quantili, e poche statistiche descrittive della distribuzione del reddito. Ora mostriamo anche il confronto dell'intera distribuzione del reddito per diverse fonti di dati utilizzate, mediante stime di densità non parametriche (si veda Silverman, 1986).

Le Figura 1 e 2 illustrano la distribuzione del reddito complessivo e disponibile, rispettivamente. La linea continua rappresenta la densità dei redditi per la popolazione (dati AE), la linea tratteggiata per il campione basato su ICFT (TREMODO) e la linea punteggiata per il campione SILC (SILC-TN). Entrambe le figure mostrano come la distribuzione dei redditi della popolazione sia significativamente rappresentata con maggiore precisione utilizzando TREMOD invece che SILC-TN, per altro con esclusione degli individui con redditi bassi.

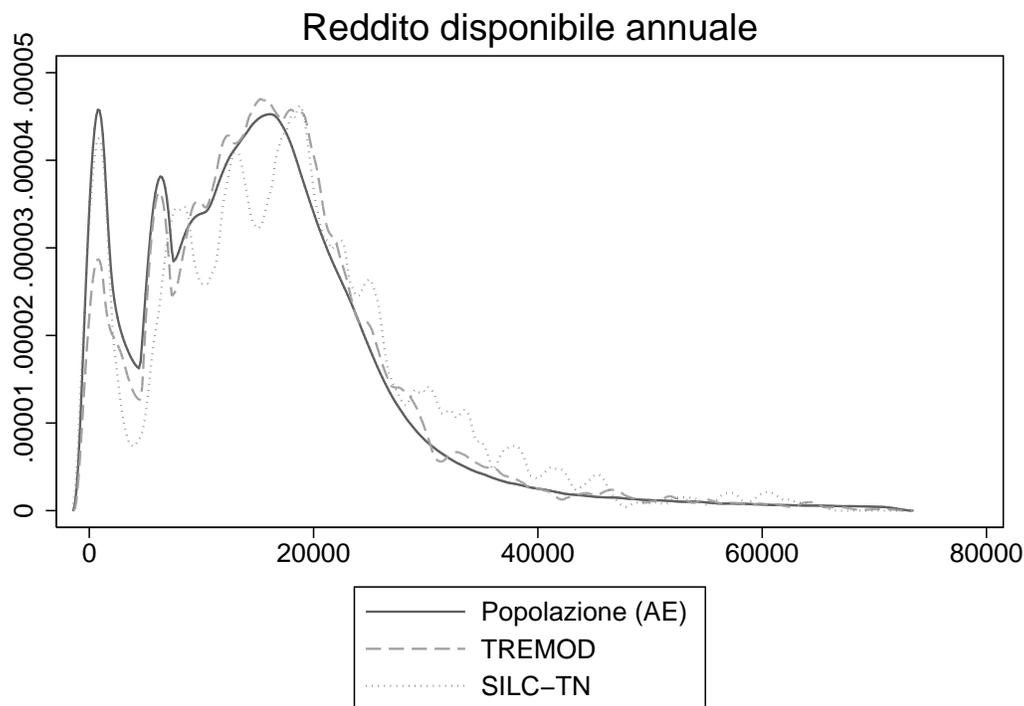
Figura 1: Reddito complessivo: confronto tra i risultati del modello TREMOD, di EUROMOD Trentino e dell'intera popolazione dei dichiaranti. (Trentino, 2010)



Fonte: elaborazioni FBK-IRVAPP.

Note: le stime di TREMOD utilizzano il Peso 4 per riportare le frequenze all'intera popolazione (Cfr. Tab. 2). I dati di IT-EUROMOD (sottocampione Trentino) si basano sui redditi dichiarati dagli individui intervistati nell'indagine SILC.

Figura 2: Reddito disponibile: confronto tra i risultati del modello TREMOD, di EUROMOD Trentino e dell'intera popolazione dei dichiaranti. (Trentino, 2010)



Fonte: elaborazioni FBK-IRVAPP.

Note: le stime di TREMOD utilizzano il Peso 4 per riportare le frequenze all'intera popolazione (Cfr. Tab. 2). I dati di IT-EUROMOD (sottocampione Trentino) si basano sui redditi dichiarati dagli individui intervistati nell'indagine SILC.

Concentriamoci infine su una sintetica analisi distributiva utilizzando il reddito familiare. Si deve innanzitutto notare che non sono disponibili statistiche distributive da fonti esterne e quindi non è possibile procedere ad alcuna validazione dei risultati ottenuti con TREMOD. Ci limitiamo a mostrare alcune popolari statistiche distributive del reddito familiare equivalente mettendo a confronto i risultati ottenuti con ciascun modello.

Il reddito familiare disponibile è ottenuto in modo da tener conto delle diverse dimensioni delle famiglie, sulla base della cosiddetta scala di equivalenza OCSE-modificata che assegna un peso pari ad 1 al capofamiglia, 0,5 agli altri adulti (soggetti con più di 14 anni) e 0,3 ai minori. La Tabella 12 mostra l'indice di Gini e il tasso di povertà individuale e familiare<sup>1</sup>. I due campioni complessivamente stimano livelli di disuguaglianza e di povertà tra loro molto simili.

<sup>1</sup>In armonia con una frequente convenzione, sono stati considerati poveri coloro che possiedono un reddito familiare equivalente inferiore al 60% di quello mediano della popolazione di riferimento.

Tabella 12: Confronto dell'indice di Gini nei risultati di TREMOD e quelli di EUROMOD-IT (sottocampione Trentino - SILC(TN)).

Campione	Indice di Gini	Tasso di povertà individuale (soglia 60%)	Tasso di povertà famigliare (soglia 60%)
TREMOD	0,29	17,88	16,98
SILC-TN	0,27	17,45	17,80

Fonte: elaborazioni FBK-IRVAPP. Note: i dati si riferiscono alla provincia di Trento per l'anno 2010. Le stime di TREMOD utilizzano il Peso 4 per riportare le frequenze all'intera popolazione (Cfr. Tab. 2). I dati di EUROMOD-IT (sottocampione Trentino) si basano sui redditi dichiarati dagli individui intervistati nell'indagine SILC.

## 6 Considerazioni conclusive

In questo rapporto sono state presentate le finalità e le caratteristiche principali di TREMOD, il modello di microsimulazione fiscale sviluppato da FBK-IRVAPP per il Trentino. Tale modello intende essere utilizzato come strumento analitico a supporto di analisti e decisori politici interessati a testare gli effetti, desiderati e indesiderati, di determinati cambiamenti nelle politiche fiscali e/o di *welfare* prima della loro applicazione. Si tratta, pertanto, di un modello altamente flessibile che permette di rispondere, con i dovuti adeguamenti, a diverse domande relative all'efficacia di un ampio spettro di interventi pubblici su diversi tipi di *outcome*.

Come descritto nelle pagine precedenti, TREMOD utilizza la piattaforma europea EUROMOD sfruttando *dataset* amministrativi e campionari relativi alla popolazione residente in provincia di Trento. Al fine di valutare la buona riuscita di questo processo e quindi valutare le capacità di simulazione di TREMOD, sono state effettuate diverse analisi statistiche, i cui risultati sono l'oggetto principale del presente rapporto. Tali approfondimenti metodologici hanno dimostrato che TREMOD è robusto sotto alcuni aspetti ritenuti cruciali per la affidabilità delle stime e delle simulazioni di imposte e benefici producibili con il modello. In primo luogo, la base dati utilizzata è risultata essere altamente rappresentativa dell'intera popolazione trentina con riguardo ad una lunga serie di indicatori demografici e socio-economici. In secondo luogo, TREMOD si è rivelato essere altamente affidabile nella simulazione di svariate grandezze di interesse, che includono varie definizioni di redditi, imposte e detrazioni. In terzo luogo, le analisi presentate in questo report confermano l'importanza di utilizzare misure oggettive dei redditi basate su fonti amministrative (nel nostro caso le dichiarazioni dei redditi) rispetto a fonti di *survey*, poiché le prime garantiscono l'ottenimento di stime meno soggette ad errori di misurazione.

TREMOD, in questa sua prima versione, è già in grado di simulare gli effetti di diverse politiche di rilievo provinciale con elevati livelli di precisione. Va, tuttavia, sottolineato che TREMOD deve essere inteso come uno strumento che deve essere

costantemente mantenuto, aggiornato e possibilmente migliorato. In particolare, è in corso presso FBK-IRVAPP un progetto per includere in TREMOD un modulo “comportamentale”, al fine di produrre stime degli effetti delle politiche che tengano in considerazione anche i cambiamenti nei comportamenti individuali e familiari indotti dalle politiche pubbliche. Inoltre, va considerato che TREMOD è una piattaforma che potrà essere integrata con altre basi dati utili ad arricchirne ulteriormente il potenziale informativo. E’ attualmente in corso presso FBK-IRVAPP un progetto finalizzato ad integrare la base dati utilizzata da TREMOD con dati amministrativi delle comunicazioni obbligatorie e con quelli riguardanti tutte le prestazioni di natura previdenziale degli individui che, una volta concluso, consentirà la simulazione di una pluralità di politiche fiscali, contributive, occupazionali e pensionistiche.

## Riferimenti bibliografici

- Atkinson, A.B., Piketty, T. e Saez, E. (2011) “Top Incomes in the Long Run of History”, *Journal of Economic Literature*, **49**(1): 3-71.
- Ceriani, L., Figari, F. e Fiorio, C. (2013) “EUROMOD country report. Italy”, ISER Technical report.
- Figari, F., Paulus, A. e Sutherland, H. (2013) “Microsimulation and policy analysis”, in Atkinson, A.B. e Bourguignon, F. (a cura di), *Handbook of Income Distribution Volume 2* (in corso di pubblicazione).
- Judkins, D.R. (1990) “Fay's method for variance estimation”, *Journal of Official Statistics*, **6**(3): 223-239.
- Popova, P. (2012) “Constructing the tax-benefit micro simulation model for Russia-RUSMOD”, EUROMOD Working Paper n. 7.
- Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*, Chapman e Hall, London.
- Sutherland, H. (1991) “Constructing a Tax-Benefit Model: What Advice Can One Give?”, *Review of Income and Wealth*, **37**(2): 199-219.
- Sutherland, H. e Figari, F. (2013) “EUROMOD: the European Union tax-benefit microsimulation model”, *International Journal of Microsimulation*, **1**(6): 4-26.
- Wilkinson, K. (2009) “Adapting EUROMOD for use in a developing country: The case of South Africa and SAMOD”, EUROMOD Working Paper n.5.
- Zarković-Rakić, J. (2010) ‘First Serbian tax-benefit microsimulation model-SRMOD’, *The Quarterly Monitor of Economic Trends and Policies in Serbia*, (20).