

**DISTRIBUZIONE DEI REDDITI A LIVELLO LOCALE:
UN ESERCIZIO DI SIMULAZIONE**

PAOLO BRUNORI e ALESSANDRO BONAZZI

JEL Classification: C81, R12

Keywords : distribuzione del reddito, simulazione di Monte Carlo, microsimulazione, redditi locali

Distribuzione dei redditi a livello locale: un esercizio di simulazione

Paolo Brunori^{*} e Alessandro Bonazzi^{**}

Sommario

Negli ultimi anni il processo di riforme fiscali che ha coinvolto il nostro paese ha stimolato l'interesse di accademici e amministratori riguardo alla distribuzione locale dei redditi. La mancanza di dati ha tradizionalmente limitato le possibilità di indagare riguardo alla distribuzione dei redditi in contesti sub nazionali. Le Regioni e gli enti locali solo in alcuni casi hanno risolto il problema finanziando indagini circoscritte al loro territorio, nella maggioranza dei casi il problema è risolto cercando di approssimare la distribuzione delle variabili a livello locale utilizzando sotto campioni di indagini nazionali. Questa soluzione purtroppo raramente porta a stime sufficientemente affidabili. Una parte della letteratura propone di risolvere questo tipo di problemi sfruttando la variabilità di grandezze fortemente correlate al reddito e misurabili con precisione sul territorio. Il metodo che proponiamo segue un approccio leggermente differente: otteniamo una distribuzione locale dei redditi equivalenti attraverso il *matching* di *dataset* provenienti da fonti differenti. Il nostro punto di partenza sono i dati delle dichiarazioni dei redditi pubblicati, disaggregati per comune, dall'Agenzia delle Entrate. Questi redditi sono associati ad un vettore di caratteristiche socioeconomiche tramite l'iterazione di un algoritmo di Monte Carlo. L'algoritmo definisce le probabilità di associazione fra un reddito e un nucleo familiare con determinate caratteristiche, sulla base della distribuzione empirica di redditi lordi e caratteristiche socioeconomiche registrate nei sottocampioni regionali dell'Indagine sui bilanci delle famiglie italiane della Banca d'Italia. La distribuzione dei redditi equivalenti è ottenuta associando a ciascun risultato di iterazione un peso tanto maggiore quanto più la struttura demografica che descrive corrisponde a quella realmente registrata dall'Istat per il territorio considerato. Per mostrare limiti e potenzialità del metodo riportiamo un'applicazione per due province italiane: Bari e Foggia.

JEL: C81, R12

keywords: distribuzione del reddito, simulazione di Monte Carlo, microsimulazione, redditi locali.

^{*}Dipartimento di scienze economiche e metodi matematici, Università di Bari, via C. Rosalba 53, 70124, tel. 080 5049110. paolo.brunori@bayesfor.eu.

^{**}INGV.

Datasets Matching to Simulate Local Income Distributions: an exercise

Paolo Brunori^{*} and Alessandro Bonazzi^{**}

Abstract

Objective: In the last decades in Italy fiscal reforms have fueled policy makers and researchers interests on local income distributions. The lack of subnational data, however, severely challenges the possibility to carry out local income distribution analysis. Local institutions in some cases solve this problem undertaking their own survey, more often local incomes are approximated looking at regional sub-sample of national surveys, which rarely yield statistically reliable estimates. Alternatively methods based on the observation of variables that covariate with income have also been proposed. *Methods and results:* We choose a slightly different approach, obtaining local income information matching dataset. Our starting point are income tax statistics that are matched with households characteristics iterating a Monte Carlo algorithm. The algorithm defines the probability of being matched with a given characteristic using distribution of the same characteristics, conditional on gross taxable income, recorded in Banca d'Italia Survey on Household Income and Wealth sub-samples. Among iterations' output the vector of characteristics that best approximates local population structure is chosen. To show the method potential we applied the algorithm to simulate income distributions in Provincia di Foggia and Provincia di Bari. *Conclusions:* the article presents a simple method to overcome the problem of data availability that arises when attempting to describe the income distribution in small territories in Italy. The method matches data from different sources and different geographical detail. The matching procedure is based on a Monte Carlo algorithm that postulates a fix relationship between income and the demographic structure of the population across the Italian territory. The method has been applied to two Italian Province.

JEL: C81, R12

keywords: income distribution, Monte Carlo simulation, microsimulation, local income distribution.

1 Introduzione

La stagione di riforme fiscali che ha coinvolto il nostro paese a partire dall'inizio degli anni '90 ha influenzato fortemente molti aspetti del sistema fiscale. Uno dei cambiamenti più significativi ha probabilmente riguardato, e più ancora riguarderà stando al dibattito parlamentare attuale, l'allocazione delle risorse fra livelli di governo. L'introduzione della tassa comunale sugli immobili (Ici) e le sue successive modifiche, l'introduzione di meccanismi di compartecipazioni regionali e comunali a imposte tradizionalmente nazionali, assieme al progressivo trasferimento di responsabilità verso il livello sub nazionali di governo nella gestione delle politiche di svariati settori, hanno rafforzato il ruolo dei livelli locali di governo come attori fondamentali del sistema fiscale italiano. I politici e gli amministratori locali oggi hanno continuamente necessità di confrontarsi con le condizioni reddituali dei propri residenti. I bilanci degli enti locali dipendono in misura crescente dal reddito prodotto dalla popolazione residente, e questa tendenza è probabilmente destinata a rafforzarsi visti gli scenari che sembrano emergere dal dibattito politico¹. Una corretta immagine delle condizioni di reddito sul territorio risulta ancor più cruciale dal lato della spesa degli enti. Dal 1997 infatti la maggior parte delle tariffe dei servizi comunali è determinata sulla base dell'Indicatore della Situazione Economica Equivalente (Isee). Un certo numero di autori ha sottolineato come questo processo di riforme renda urgente migliorare la nostra conoscenza riguardo a come il reddito si distribuisca a livello sub nazionale. Questione tradizionalmente trascurata dagli istituti che hanno realizzato indagini riguardo ai redditi² (Baldini et al., 2003; Cannari e D'Alessio, 2003).

¹ Ci riferiamo in particolare al concetto di "capacità contributiva" che a più riprese è stata indicata come una delle componenti dei futuri meccanismi perequativi all'interno del disegno del cd "federalismo fiscale".

² Indagine sui bilanci delle famiglie italiane (Banca d'Italia), Indagine sui consumi delle famiglie (Istat) e European Community Household Panel. Recentemente la

La necessità conoscitiva da parte di amministratori e accademici è stata risolta solo raramente realizzando indagini a livello locale capaci di rappresentare in modo affidabile le condizioni di reddito dei residenti³. Più frequentemente gli enti interessati non hanno possibilità di finanziare un'indagine specifica e devono accontentarsi di metodologie meno raffinate per approssimare le informazioni di interesse. Una strategia possibile consiste nell'isolare sottocampioni regionali o provinciali di indagini nazionali. I campioni ottenuti sono però generalmente piuttosto ristretti e possono essere considerati rappresentativi solamente con un notevole margine di errore. Un secondo approccio consiste nel prendere in considerazione variabili fortemente correlate con il reddito che possono essere reperite in rilevazioni non campionarie, come ad esempio il censimento (Elbers et al. (2003), Demombynes et al. (2007)). In questa prospettiva ad esempio le informazioni riguardo alle abitazioni possono fornire segnali accurati riguardo alle probabili condizioni di vita delle famiglie residenti (un interessante applicazione per la provincia di Modena si trova in Morciano et al. (2006).

L'approccio che presentiamo qui è leggermente differente e si concentra sulle informazioni registrate nelle dichiarazioni dei redditi, anche in questo caso non campionari. Le ragioni di questa scelta sono principalmente due. Per prima cosa le dichiarazioni dei redditi sono generalmente la variabile più rilevante per gli amministratori. In secondo luogo i dati sulle dichiarazioni sono disponibili, se pur con un certo ritardo, e pubblicati in aggregati comunali, che rappresenta il massimo dettaglio territoriale auspicabile.

Il problema principale quando si cerchi di analizzare i dati delle dichiarazioni dei redditi consiste nell'impossibilità di aggregare i differenti redditi in nuclei familiari e di associare a questi una struttura familiare la cui distribuzione risulta essere eterogenea nel paese. Questo limite compromette l'analisi nel caso che la variabile di interesse sia una misura del benessere dei residenti (per il quale il reddito equivalente viene generalmente considerato la componente

diffusione dei dati ITASILC da parte di Eurostat, hanno parzialmente risolto il problema per le analisi a livello regionale.

³ E il caso per esempio della Toscana (Betti et al., 2003).

più essenziale). Una serie di altre questioni deve inoltre essere presa in considerazione fra queste l'esclusione di redditi provenienti da alcune fonti dall'imponibile Irpef.

Proponiamo un semplice metodo per la simulazione della distribuzione dei redditi a livello locale che tenta di affrontare i principali ostacoli derivanti dall'uso di redditi registrati dalla Agenzia delle entrate utilizzando esclusivamente dati liberamente accessibili. L'esercizio è in parte velleitario se si considera che un maggior coordinamento tra Agenzia delle entrate e amministrazioni comunali potrebbe produrre gli stessi risultati eliminando la maggior parte dell'incertezza che noi siamo costretti ad introdurre.

Il punto di partenza dall'esercizio sono i dati registrati e pubblicati dall'Agenzia delle entrate a livello comunale. Il metodo consiste nell'associare i redditi Irpef con caratteristiche demografiche iterando un algoritmo di Monte Carlo nel quale le probabilità di associazione sono condizionate al reddito Irpef lordo dei singoli percettori. Le probabilità condizionate sono ottenute dalle distribuzioni congiunte di reddito lordo e caratteristiche economiche e demografiche registrate nel sottocampione regionale dell'Indagine sui bilanci delle famiglie italiane della Banca d'Italia. I vettori di redditi e caratteristiche demografico-economiche provenienti dalle iterazioni sono poi riassunte in un'approssimazione della distribuzione dei redditi equivalenti che tiene conto della somiglianza della struttura della popolazione ottenuta con le caratteristiche note della popolazione reale. Il resto del testo è organizzato come segue: nel paragrafo 2 si descrive il metodo di simulazione; nel paragrafo 3 si mostra come si passi dai redditi lordi imponibili ai redditi disponibili equivalenti; infine il paragrafo 4 presenta i risultati dell'applicazione del metodo a due province della Regione Puglia: Bari e Foggia.

2 Il metodo di simulazione

Lo scopo dell'analisi è la stima del reddito equivalente I_e per un territorio circoscritto (nel nostro caso l'esercizio sarà applicato a due

province italiane). Il problema bayesiano è così posto:

$$P(I_e | D) \propto P(D | I_e) P(I_e) \quad (1)$$

dove I_e è il reddito equivalente e D rappresenta il data-set disponibile a livello locale (nel nostro caso le dichiarazioni pubblicate dall'Agenzia delle entrate disaggregate per comune).

Il reddito equivalente I_e può essere modellato come $I_e = f(I_f, C)$. Ossia I_e è una funzione del reddito aggregato della famiglia e del numero di componenti C che le compongono⁴.

Inoltre possiamo osservare che I_f è a sua volta una funzione del reddito dei percettori I_p che fanno parte di ogni singolo nucleo familiare. Quindi il problema a priori concerne la definizione della probabilità congiunta $P(I_e, I_f, I_p, C)$.

Questa probabilità può essere così fattorizzata:

$$P(I_e, I_f, I_p, C) = P(I_e | I_f, C) P(C | I_f) P(I_f | I_p) \quad (2)$$

Il primo termine a destra descrive la probabilità del reddito equivalente condizionata a una certa distribuzione di redditi aggregati familiari I_f e del numero di componenti C . Il secondo termine esprime la probabilità del numero di componenti per famiglia dato il reddito aggregato familiare I_f . L'ultimo termine esprime la probabilità del reddito familiare aggregato I_f dato il reddito dei percettori I_p che compongono le singole famiglie. La forma di ciascuna di queste distribuzioni a priori è stimata a partire dai dati Banca di Italia che sono disponibili a livello regionale.

I dati disponibili a livello provinciale sono le dichiarazioni dei redditi Irpef (D_{irpef}) e i dati Istat di censimento (D_{istat}). Il modello dei dati è quindi costituito da due distribuzioni che sono:

- $P(D_{irpef} | I_p)$: dove I_p è il reddito "reale" degli N percettori

⁴ Si noti qui che I_f pubblicati sono redditi imponibili ai fini dell'addizionale comunale Irpef, quindi questa funzione non consiste nella semplice divisione della somma dei redditi per i valori di una scala di equivalenza ma include implicitamente una funzione di simulazione del sistema fiscale.

presenti nel territorio.

• $P(D_{\text{istat}} | C)$: dove C rappresenta il numero di componenti delle M famiglie presenti nel territorio.

Il dato D_{istat} è un'informazione aggregata che è affetto da un errore di misura (per quanto piccolo).

Abbiamo quindi tutti gli elementi per scrivere la distribuzione a posteriori dei redditi:

$$P(I_e, I_f, I_p, C | D_{\text{istat}}, D_{\text{irpef}}) \propto P(D_{\text{irpef}} | I_p) P(D_{\text{istat}} | C) P(I_e | I_f, C) P(C | I_f) P(I_f | I_p) \quad (3)$$

Dove D_{irpef} è la distribuzione dei redditi imponibili ai fini dell'addizionale comunale Irpef e D_{istat} è la composizione demografica registrata dal censimento Istat.

2.1 Le componenti della simulazione

Il punto di partenza della simulazione è la distribuzione del reddito imponibile così come registrato dall'Agenzia delle Entrate. L'Agenzia delle Entrate pubblica, con un ritardo di circa due anni, informazioni relative ai redditi imponibili ai fini dell'addizionale comunale Irpef (I_{irpef}), che comprendono i redditi imponibili al netto di una serie di deduzioni. Per ogni comune i redditi dichiarati sono suddivisi in 20 intervalli di reddito di cui sono noti il numero di percettori e la dichiarazione media⁵. Il primo passo dell'algoritmo consiste nel ricostruire una distribuzione quasi continua a partire da questa distribuzione discreta, risulta infatti irrealistico immaginare che i contribuenti si addensino attorno ai valori medi di scaglioni di reddito arbitrari. Al fine di rendere più credibile la distribuzione delle dichiarazioni si assume una distribuzione quasi uniforme all'interno di ciascun intervallo di redditi. La non uniformità della distribuzione

⁵ Le statistiche soffrono di alcuni omissis nei comuni molto piccoli in quanto l'Agenzia ritiene di tutelare la privacy dei residenti non riportando il numero di percettori negli intervalli di reddito con poche unità di percettori.

è coerente con l'osservazione di una serie di mode locali all'interno delle distribuzioni di reddito, per questo motivo si tiene conto della maggior densità dei redditi attorno ai valori modali imponendo che per ogni intervallo dei redditi la moda e la media siano le stesse⁶.

Otteniamo perciò una distribuzione di redditi in 40 intervalli di reddito che sono funzione della distribuzione discreta proveniente dall'Agenzia delle entrate.

Il secondo passo consiste nel trasformare il reddito I_P in reddito lordo ai fini Irpef. Il fatto che un certo numero di deduzioni venga sottratto dal reddito prima dell'imposizione infatti implica che l'Agenzia non registri una parte del reddito disponibile che deve essere recuperato. Per imputare queste deduzioni si deve ricorrere a dati aggregati a livello nazionale, sempre prodotte dall'Agenzia delle Entrate. In questo caso si attinge ad una distribuzione discreta di deduzioni godute, dato un certo livello di reddito, ancora una volta aggregato in fasce di reddito imponibile. Dai dati aggregati si ottiene la distribuzione di probabilità $p(I_{aD} | I_P)$ di avere un certo reddito al netto delle deduzione I_{aD} dato il reddito campionato I_P .

Per ogni classe di reddito j campioniamo una deduzione utilizzando la distribuzione empirica derivante dai dati aggregati nazionali dove le deduzioni sono registrate condizionatamente all'imponibile, in questo caso assumiamo una distribuzione normale attorno alla media per la frazione dei percettori che gode della deduzione.

$$d_j^i = N(m^j, var^j)$$

$$I_{aD}^i = I_P^i - d^i$$

Dove d_j^i è l'entità della deduzione per l'individuo i , m_j e var_j sono rispettivamente la media e la varianza della distribuzione delle

⁶ Scriviamo la distribuzione come $p(I_P | I_{Irpef})$, dove I_P è un vettore di redditi di dimensione n_p ; dove ogni elemento di I_P rappresenta un reddito percepito da un individuo della popolazione di interesse e n_p è il numero di percettori totale. Per ogni intervallo di reddito campioniamo la distribuzione dei redditi come: $P_j^i = Uniform(x_j^{Irpef}, x_{j+1}^{Irpef})$, per cui il vettore I_P è $I_P = (P_1, \dots, P_{40})^T$.

deduzioni per l'intervallo di redditi j .

Il terzo passo della simulazione consiste nell'associazione dei redditi in nuclei familiari. Il numero totale di nuclei familiari residenti in un certo territorio è noto dalle statistiche demografiche (nel nostro caso ci siamo appoggiati al Censimento dell'Istat). Al fine di poter associare i percettori in nuclei familiari utilizziamo il sottocampione dell'indagine Banca d'Italia sui bilanci familiari. Nel sottocampione regionale troviamo quale sia il numero di percettori per nucleo familiare, queste proporzioni sono utilizzate per generare distribuzioni di probabilità $P(I_F | P_{aD})$ che associano P_{aD} in famiglie I_F con 0, 1, 2 o 3 percettori.

$$p(I_F | P_{afDeduction}) = N(H P, \varepsilon)$$

Dove H è una matrice di n_f righe e n_p colonne. La riga k della matrice H è un vettore di zero e uno di dimensione n_p che definisce la composizione reddituale della famiglia k . La somma degli elementi di H è limitata all'intervallo 0 : 3 visto che in quest'analisi non si prende in considerazione che più di tre percettori possano vivere nello stesso nucleo familiare. Il termine ε è un processo di *white noise* che assumiamo di conoscere.

Il quarto passo della simulazione consiste nell'associare i nuclei familiari, con i loro redditi lordi, a caratteristiche familiari. Questo passaggio si avvale sia della relazione fra redditi e variabili demografiche registrate in Puglia, sia delle informazioni riguardo alla struttura demografica locale che sono conosciute. Basandoci sui dati della Banca d'Italia stimiamo la probabilità che una famiglia abbia una determinata struttura, dato il suo livello di reddito lordo (Irpef complessivo)⁷. In Puglia, ad esempio, una famiglia con un solo

⁷ I redditi lordizzati sono stati ottenuti con una procedura iterativa standard di lordizzazione, a partire dai dati Banca d'Italia (2004). La procedura replica il sistema fiscale vigente nel 2004 a per valori arbitrari di reddito lordo e stabilisce una corrispondenza fra reddito netto e caratteristiche economico-demografiche registrate e reddito lordo.

reddito Irpef di bassa entità tende ad essere associata frequentemente con nuclei unifamiliari (generalmente si tratta di anziani).

Queste informazioni sono utilizzate per costruire la distribuzione di probabilità $p(C | I_F, q)p(q)$ che una famiglia abbia una determinata struttura, dato il suo reddito totale lordo.

Per ciascun intervallo di reddito lordo familiare campioniamo $p(q)$ come:

$$q_c^k = N(p_c^{\text{BancaI talia}}, \Sigma^2) \quad \forall c = 1, \dots, 6$$

I valori normalizzati di q (imponiamo che la somma di q sia pari a 1) sono utilizzati per costruire una distribuzione discreta di probabilità $p(C | I_F, q)$.

Per localizzare l'algoritmo di associazione imponiamo che la struttura demografica C approssimi la struttura demografica registrata dall'Istat per il territorio di interesse.

In altre parole valutiamo la distribuzione di verosimiglianza $p(C_{\text{Istat}} | C)$ come:

$$p(C_{\text{Istat}} | C) = N(\text{hist}(C), \sigma^2)$$

dove σ^2 è una rappresentazione del margine d'errore nel Censimento. Sovrastimiamo σ^2 così da garantire qualche livello di flessibilità all'analisi.

Gli ultimi passi della simulazione consistono: nell'applicare il sistema fiscale, nell'aggiungere redditi non soggetti ad Irpef e utilizzare una scala d'equivalenza per ottenere la misura di benessere desiderata I_D .

Questi passaggi sono leggermente complicati dall'impossibilità di determinare il trattamento fiscale ai quali sono sottoposti i membri di un nucleo familiare sulla sola base della composizione familiare e dei redditi imponibili percepiti. Una serie di altre variabili partecipa a determinare l'imposta dovuta, in questo caso queste variabili sono

indicate con θ e rappresentano un'approssimazione del sistema fiscale vigente. θ in pratica è un modello di microsimulazione nei quali le componenti del trattamento fiscale sono descritte da funzioni di probabilità come brevemente spiegato nel paragrafo 3. La probabilità di un certo trattamento fiscale è condizionata alla classe di reddito e alla struttura demografica, ovvero definiamo una distribuzione a priori $p(I_D|C, I_F, \theta)$. Per fare un esempio: dato il regime fiscale nel nostro paese, una famiglia numerosa, con un solo percettore, ha una probabilità molto elevata di detrarre per carichi familiari, mentre una famiglia formata da un solo individuo ha una probabilità pari a zero di godere di un'analogo detrazione.

3 Dal reddito lordo al reddito netto: l'algoritmo di microsimulazione

Il trattamento fiscale θ (deduzioni escluse) si applica al vettore dei redditi lordi. Questo tipo di procedura di microsimulazione è in linea con altri modelli di microimulazione statici utilizzati in Italia (si vedano a riguardo i paper pubblicati da Pellegrino (2007) o Maitino e Sciclone (2008)). L'algoritmo è in realtà differente in almeno due aspetti rispetto a modelli analoghi: da un lato deve sicuramente essere considerato meno accurato in quanto l'imputazione delle varie componenti della simulazione si basa su probabilità di associazione con determinate caratteristiche e non sulla diretta osservazione di questa. D'altra parte ha il vantaggio di non essere basato su di un campione ma su tutti i redditi dichiarati da residenti nel territorio di interesse.

Al reddito imponibile di ogni membro di ciascun nucleo familiare vengono per prima cosa applicate le aliquote per scaglione, determinando l'ammontare dell'imposta lorda. L'ammontare detraibile viene determinato come la somma di un certo numero di componenti, per tipologia. A seconda delle informazioni disponibili queste componenti sono imputate sulla base di differenti probabilità. Per alcune tipologie di detrazioni le informazioni rilevanti sono registrate nell'indagine della Banca d'Italia, per queste la probabilità di detrarre è ottenuta guardando a come caratteristiche demografiche,

redditi lordi e detrazioni sono associate nel sottocampione regionale. Per tutte quelle detrazioni per le quali non esistono informazioni direttamente osservabili, la probabilità di detrazione è ottenuta guardando alle distribuzioni congiunte registrate dall'Agenzie delle entrate che svelano un legame fra reddito imponibile e detrazioni (la procedura è esattamente quella utilizzata per le deduzioni).

L'assunzione forte di questo metodo di imputazione quindi consiste nel considerare che la distribuzione di alcune variabili di interesse risulti eterogenea ma che vi sia una certa stabilità nella covarianza di queste e del reddito imponibile nelle varie zone del paese e delle regioni.

La Tabella 1 riassume quali variabili sono utilizzate per l'imputazione delle varia componenti delle detrazioni.

Una volta che l'importo Irpef dovuto è determinato vengono aggiunti redditi di altra fonte. Alcune tipologie di reddito in Italia non sono tassate (il caso ad esempio delle borse di studio e dei trasferimenti di assistenza). Redditi di altra tipologia sono tassati alla fonte e non sono inclusi nella base imponibile dell'Irpef (reddito da capitale). L'unica possibilità di tener conto di questi redditi consiste nell'imputarne seguendo, ancora una volta, le relazioni nella distribuzione congiunta delle varie tipologie di redditi, registrate nel sottocampione regionale di Banca d'Italia. Si tratta dell'ultima imputazione, tutti i redditi, le detrazioni e le deduzioni sono applicate a ciascun membro delle famiglie residenti, i redditi ottenuti vengono infine trasformati in redditi equivalenti.

Tabella 1: Componenti delle detrazioni totali

Detrazione	imponibile lordo	componenti	<i>PREGIONALE</i>	<i>PNAZIONALE</i>
tipologia impiego	✓	✓	✓	
familiari a carico	✓	✓	✓	
spese abitazione	✓			✓
spese mutuo	✓			✓
premi assicurativi	✓			✓
affittuari	✓			✓
tutte le altre	✓			✓

4 Un'applicazione: la distribuzione dei redditi equivalenti nelle province di Bari e Foggia.

Il metodo che proponiamo è stato già utilizzato con risultati soddisfacenti in alcuni casi di studio nei quali era essenziale avere una qualche percezione delle condizioni reddituali della popolazione residente (Brunori e Peragine, 2008). Qui proponiamo due applicazioni a due territori della stessa Regione per mostrare le potenzialità e i limiti di questo approccio. La scelta è ricaduta sulle due Province di Bari e Foggia, che pur trovandosi nella stessa Regione, hanno strutture economiche notevolmente differenti⁸.

Simulare la distribuzione dei redditi locali per due territori della stessa Regione, ma con strutture di reddito differenti, può essere considerato un test per la principale debolezza teorica del metodo, cioè il fatto che la maggior parte delle imputazioni si basi su probabilità regionali invece che locali. Se il metodo è in grado di conservare l'eterogeneità delle distribuzioni, allora questa preoccupazione può essere ridimensionata.

L'Agenzia delle Entrate pubblica i redditi Irpef dichiarati disaggregati per ogni Comune in distribuzioni discrete suddivise in 20 intervalli di reddito. La tabella 2 presenta i dati disponibili per il 2005, i redditi provinciali sono stati ottenuti sommando i dati di 64 Comuni della Provincia di Foggia e 48 Comuni della Provincia di Bari⁹.

Il passo successivo consiste nell'associare famiglie e redditi individuali. Le probabilità sono approssimate guardando alle proporzioni riscontrate per il sottocampione regionale dell'indagine

⁸ Nel 2005 la Provincia di Bari risultava la più ricca della Puglia con un valore aggiunto pro capite di 15.383 euro, nello stesso anno la Provincia di Foggia risultava quella più povera con un valore aggiunto pro capite di 12,477 euro.

⁹ I Comuni recentemente confluiti nella Provincia BAT sono considerati come ancora facenti parte delle loro vecchie Province.

della Banca d'Italia riportate nella tabella 3. Il passaggio successivo consiste nell'associare i redditi Irpef alle strutture familiari, anche in questo caso le probabilità di associazione sono approssimate utilizzando le informazioni della Banca d'Italia.

L'algoritmo è iterato (nel caso di questo esercizio si tratta di 1.000 iterazioni), ad ogni iterazione un peso viene associato a ciascun risultato, il peso è tanto più elevato quanto la struttura demografica ottenuta assomigli a quella effettivamente registrata sul territorio dall'Istat (come riportati in tabella 5). A conclusione della procedura otteniamo 1.000 matrici nelle quali il vettore del reddito imponibile è associato ad una struttura familiare ed a un peso di rappresentatività. Per ottenere la variabile di interesse, in questo caso il reddito disponibile equivalente, applichiamo il sistema fiscale vigente (nel caso quello del 2007). L'algoritmo segue il metodo spiegato nel paragrafo precedente. A conclusione di quest'ultimo passaggio otteniamo mille vettori di reddito equivalente, ciascuno associato ad una misura di rappresentatività. I redditi con livello di rappresentatività al di sotto di una certa soglia vengono eliminati e una serie di statistiche di interesse vengono calcolate come media pesata delle distribuzioni più rappresentative, le statistiche sono riportate nella tabella 6. La figura 1 infine rappresenta l'approssimazione della densità delle due distribuzioni considerando i due risultati dell'iterazione con la massima rappresentatività.

Tabella 2: Reddito lordo in Provincia di Bari e Foggia

income	Bari		Foggia	
	frequency	atal	frequency	atal
< 1.000	6.458	31.82.192	2.380	1.198.907
da 1.000 a 2.000	5.665	8.262.066	2.693	3.900.087
da 2.000 a 3.000	5.511	13.434.708	3.933	10.032.373
da 3.000 a 4.000	5.874	20.226.103	2.872	9.867.417
da 4.000 a 5.000	6.915	30.491.872	2.913	14.323.862
da 5.000 a 6.000	7.668	409.568.855	3.795	20.419.756
da 6.000 a 7.500	15.286	100.372.896	7.006	46.330.984
da 7.500 a 10.000	60.5851	517.342.288	28.763	245.092.851
da 10.000 a 15.000	153.622	1.863.010.711	56.292	6.753.185.407
da 15.000 a 20.000	133.770	2.238.978.978	49.239	820.239.575
da 20.000 a 25.000	83.394	1.803.831.376	32.426	696.688.441
da 25.000 a 29.000	43.523	1.134.103.949	17.448	452.103.668
da 29.000 a 32.600	22.393	667.440.930	8.176	242.065.838
da 32.600 a 35.000	9.841	325.404.064	3.329	108.598.165
da 35.000 a 40.000	12.366	447.946.206	3.739	134.102.581
da 40.000 a 50.000	12.161	520.485.001	3.565	151.388.422
da 50.000 a 60.000	6.877	366.484.013	1.940	100.654.056
da 60.000 a 70.000	4.944	319.504.193	1.289	80.591.060
da 70.000 a 100.000	6.257	490.754.163	1.929	155.878.388
100.000 <	3.627	586.209.260	914	134.990.879
toatale	606.737	11.498.421.854	234.969	4.103.785.850

fonte: Agenzia delle Entrate

Tabella 3: Numero di percettori per famiglia in Puglia.

number of records	
0	16%
1	60.22%
2	21.11%
3	2.67%

fonte: Banca d'Italia

Tabella 4: Struttura demografica condizionata al reddito totale lordo

	1 componente	2 componenti	3 componenti	4 componenti	5 componenti	6 o più
intervalli di reddito						
< 10,000	0.4000	0.2889	0.1111	0.1778	0.0000	0.0220
da 10,000 a 14,400	0.4182	0.2909	0.0727	0.2000	0.0180	0.0000
da 14,400 a 18,500	0.2923	0.3077	0.1231	0.2000	0.0615	0.0154
da 18,550 a 21,200	0.1228	0.3860	0.1404	0.2105	0.0877	0.0526
da 21,200 a 25,000	0.1400	0.2800	0.2000	0.2200	0.1400	0.0200
da 25000 a 30,800	0.1042	0.2708	0.2917	0.2292	0.0833	0.0208
da 30,800 a 39,500	0.0870	0.1522	0.2391	0.3261	0.1957	0.0000
da 39,500 a 52,800	0.0000	0.1795	0.2821	0.2821	0.2308	0.0256
above 52800	0.0222	0.2000	0.2000	0.4444	0.1111	0.0222

fonte: Banca d'Italia

Tabella 5: Composizione demografica nella Provincia di Bari e Foggia

Provincia	1 componente	2 componenti	3 componenti	4 componenti	5 componenti	6 o più
Bari	90.300 (17.17%)	124.363 (23.65%)	109.833 (20.89%)	138.565 (26.35%)	50.854 (9.67%)	11.873 (2.23%)
Foggia	47.524 (20.17%)	54.976 (23.33%)	44.690 (18.97%)	55.970 (23.76%)	24.764 (10.51%)	7.656 (3.25%)

fonte: Istat

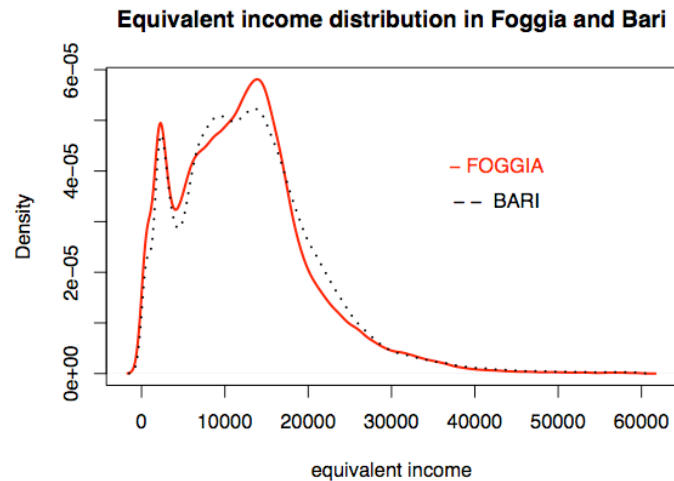


Figura 1: Una rappresentazione grafica dei redditi simulati

5 Conclusioni

Nel nostro paese la comprensione delle condizioni reddituali a livello sub regionale è limitata dalla carenza di dati attendibili. Recentemente questa problematica sembra aver guadagnato di rilevanza soprattutto per via del dibattito che si è sviluppato riguardo all'allocazione delle competenze di spesa fra i livelli di governo. Abbiamo proposto un semplice modello utile ad ottenere una simulazione dei redditi equivalenti nella popolazione residente in territori di piccole dimensioni. Il metodo si basa su dati di fonte amministrativa, gratuitamente disponibili, e quindi rappresenta una valida alternativa per chiunque abbia necessità di ottenere informazioni in tempi rapidi e a costo quasi nullo. Il metodo parte dai dati sulle dichiarazioni dei redditi pubblicati a livello comunale e procede attraverso passaggi nei quali questi redditi sono associati a variabili economiche, demografiche. Tutte le associazioni si basano sull'assunzione che sussista una certa regolarità nella distribuzione

congiunta di redditi Irpef e delle altre variabili di interesse (perlomeno per tutto un territorio regionale). La procedura di simulazione si basa quindi sull'iterazione di un algoritmo di Monte Carlo. Le probabilità di *matching* sono ricavate da altre fonti, nelle quali queste associazioni sono osservate. Abbiamo mostrato come l'algoritmo proposto possa essere formalizzato in termini bayesiani e abbiamo infine presentato un'applicazione ottenendo la distribuzione dei redditi equivalenti in due province del territorio italiano.

Riferimenti bibliografici

- Baldini M., Bosi P., Guerra M.C., Silvestri P., (2003). "Distribuzione del reddito e politiche fiscali in un contesto locale: il caso del Comune di Modena", XV Conferenza SIEP - Università di Pavia, 3 - 4 Ottobre 2003.
- Betti G., Lemmi A. Maltinti G., Sciclone N., (2003). *Indagine sul le condizioni di vita del le famiglie ascane, primi risultati*. Irpet Cridire, Firenze.
- Brunori P., Chuirì M. C., and Peragine V., (2009). "The economic effect of a local minimum income program", Social minima workshop, Università Bocconi, 12-01-2009.
- Cannari L., DAlessio, G., (2003). La distribuzione del reddito e della ricchezza nelle regioni italiane. *Rivista Economica del Mezzogiorno*, 4, 1: 17-26.
- Demombynes G., Elbers C., Lanjouw J. O., and Lanjouw P. ,(2007). "How Good a Map? Putting Small Area Estimation a the Test", World Bank Policy Research Working Paper, 4155.
- Doucet A., De Freitas N. and Gordon (2001). *Sequential Monte Carlo Methods in Practice*. N. Springer.
- Elbers C., Lanjouw J. O., Lanjouw P., (2003). Micro-Level Estimation of Poverty and Inequality. *Econometrica*, 71, 1: 355-364.
- Istat (2008). "Occupazione e valore aggiunto nelle province. Anni 2001 - 2005", *Statistiche in Breve*, 29, Gennaio 2008.
- Istat (2007). "La povertà relativa in Italia nel 2006", *Statistiche in*

Breve, 4 Ottobre 2007.

Maitino M.L. and Sciclone N., (2008). “Il modello di microsimulazione multiregionale dell’Irpet Microreg”, SIEP Working Paper n. 604.

Morciano M., Cantone F., Baruffi A., Bigi G., (2006) “Il matching fra indagini locali e censimento”, Workshop politiche locali e disuguaglianza, Modena 22 Giugno 2006.

Pellegrino S. (2007). “Il modello di microsimulazione Irpef 2004”, SIEP Working Paper N. 583.

Saraceno C. (2005). “I livelli essenziali di assistenza nell’assetto federale italiano”. *Associazione Reforme*, Ottobre 2005.