

Utilizzo di dati amministrativi per la valutazione di policies familiari

Mario Mezzanzanica, Roberto Boselli

CRISP e Dipartimento di Statistica e Metodi Quantitativi
Università degli Studi di Milano Bicocca

Stefano Verzillo

CRISP e Dipartimento di Economia, Management e Metodi Quantitativi
Università degli Studi di Milano

mario.mezzanzanica@unimib.it, roberto.boselli@unimib.it, stefano.verzillo@unimib.it

1. Introduzione

In una situazione di contesto economico difficile come quella attuale il tema della valutazione nella spesa pubblica è sempre più centrale. Nel nostro paese assistiamo alla crescente contrazione dei trasferimenti dello stato alle amministrazioni locali (comuni, province e regioni) ed al conseguente aumento dell'imposizione fiscale locale (si pensi per esempio ai recenti aumenti delle aliquote delle addizionali in molte realtà del nostro paese) al fine di mantenere livelli dignitosi (o almeno essenziali) dei servizi erogati ai cittadini. Una maggior efficienza ed efficacia nell'utilizzo delle risorse pubbliche è certamente in questa ottica un obiettivo primario per qualunque decisore politico.

Poter disporre quindi di strumenti di analisi dei cambiamenti della realtà e di stima dell'impatto dovuto alle variazioni introdotte diventa un *asset* strategico, per non dire essenziale, per una buona gestione e programmazione degli interventi.

È in questo scenario che si introduce l'importanza dell'informazione e quindi dello studio di modelli di analisi di micro-simulazione atti a fornire indicazioni sui cambiamenti che una policy induce nei cittadini ed il conseguente ritorno per il decisore stesso.

In questa sede si vuole contestualizzare e rappresentare alcuni dei primi risultati derivanti dalle attività di studio relative, in particolare, ad un

modello di micro-simulazione di *tax-policies* per le famiglie a livello locale.

2. I modelli di micro-simulazione

Il presente lavoro descrive i risultati di un approccio metodologico (parte di una attività più ampia di ricerca in corso) che prende avvio dalle seguenti tre domande:

- Come varierebbe il prelievo fiscale per una famiglia al variare delle aliquote dell'addizionale individuale Irpef?
- È possibile pensare un sistema che in un contesto come quello attuale – in cui si cercano nuove risorse mediante l'aumento delle imposte- non penalizzi ulteriormente i percettori di reddito con molti familiari a carico?
- È possibile stimare quindi la variazione di gettito per le varie tipologie di famiglie conseguente a variazioni dello schema delle aliquote da parte dell'amministrazione locale?

Per rispondere a queste domande è stato studiato un sistema informativo basato su modelli di micro-simulazione, che permetta ai decisori pubblici di avere uno strumento affidabile di calibrazione delle variazioni dello schema delle aliquote fiscali a livello territoriale locale.

Il lavoro di ricerca è basato sulla ricostruzione dei nuclei familiari dei contribuenti a partire da archivi amministrativi relativi ai redditi individuali con l'obiettivo di verificare sia gli impatti dell'introduzione di politiche che tengano conto dell'ampiezza del nucleo familiare, sia le ricadute sul gettito complessivo derivante per l'amministrazione, sia i cambiamenti che tali novità portano al sistema di welfare nel suo complesso (miglioramento/peggioramento dell'equità fiscale). Questo studio si inserisce nell'ottica, perseguita ormai da alcuni anni dal Crisp, di studiare metodi e modelli per la valorizzazione dell'informazione disponibile negli archivi amministrativi al fine di migliorare il livello di conoscenza dei fenomeni e creare strumenti di analisi e programmazione a supporto del decisore.

L'attività preliminare rispetto alle successive fasi di progettazione ed implementazione di modelli di

simulazione, che necessariamente si deve affrontare nell'utilizzo di dati amministrativi, si è incentrata sulla verifica della qualità del dato disponibile e sulla "trasformazione" dello stesso al fine di permetterne l'utilizzo con finalità statistiche.

Lo schema in figura evidenzia sinteticamente i nodi principali che compongono la struttura dell'intero percorso di lavoro intrapreso per la simulazione di politiche fiscali a favore delle famiglie in contesti locali.

In primis si è affrontato il tema della qualità dei dati su cui fondare qualsiasi successiva analisi. Tale fase del lavoro è stata affrontata al fine di definire le strutture dati secondo criteri di *coerenza delle informazioni* e di *consistenza* tra il dato ivi riportato ed il naturale dominio di appartenenza al fine di raggiungere livelli qualitativi delle informazioni da trattare, funzionali al percorso di analisi ipotizzato.

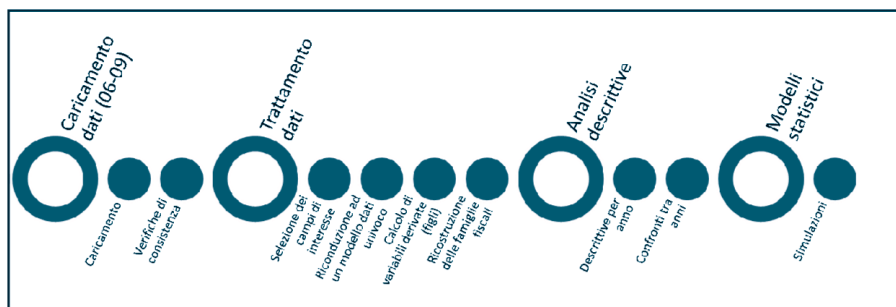


Figura: schema del processo di trattamento, analisi ed implementazione dei modelli di simulazione dei dati.

A tale proposito, nell'implementazione delle opportune verifiche di consistenza, sono state ri-simulate teoricamente alcune delle grandezze chiave tipiche della banca dati in oggetto (calcolo dei valori delle imposte, delle corrispondenti detrazioni per lavoro dipendente e assimilati, delle spettanti deduzioni, degli oneri accessori etc...) così da poter validare il dato a monte nell'archivio stesso. È necessario infatti poter disporre di dati il più possibile privi di "vizi" formali dovuti a possibili errori nel processo di raccolta e caricamento, a tale scopo - accertate alcune delle grandezze - si sono stimate le altre che da essa derivano potendo così sia confrontare il dato stimato con quello reale sia valutare la distanza e le possibili cause.

Inoltre, sempre in questa prima fase, sono state effettuate attività di ricostruzione e stima delle grandezze mancanti (in particolare per certe tipologie di modelli dichiarativi che sono privi di alcune delle informazioni riportate negli altri) tramite algoritmi (modelli di tipo *logit* multinomiale, tecniche di *Propensity Score Matching* e di *Coarsened Exact Matching*) di stima dell'appartenenza individuale dei soggetti alle categorie mancanti, sulla base delle caratteristiche osservabili.

Prevalentemente ciò che è stato al centro della fase di stima dei valori mancanti è stata la dimensione familiare del contribuente e la tipologia di nucleo di appartenenza (appartenenza a famiglie monoreddito o bireddito, con o senza familiari a carico). I modelli dichiarativi infatti riportano solo un numero

prefissato di potenziali componenti a carico del nucleo (è possibile che vi siano contribuenti per cui il modello è incapiente) ed hanno, come da struttura dell'intero sistema fiscale del nostro paese, la caratteristica di essere riferiti all'individuo, e non al nucleo familiare, per cui si rende necessaria una attività di ricostruzione deterministica (per il 75% dei modelli dichiarativi) e di stima (il restante 25% ove il riferimento è mancante) delle tipologie e dei componenti dei nuclei familiari.

In questa fase, in particolare, si è lavorato allo sviluppo, dopo una *review* della letteratura scientifica in merito a tecniche di *matching multidimensionale* per grandi popolazioni (si è trattato infatti di processare e analizzare svariati milioni di record), di algoritmi per l'implementazione di tecniche sia parametriche che non-parametriche di *matching* (si veda il CEM proposto da Iacus et al., 2009). Tale metodologia non parametrica è stata introdotta nel 2009 con lo scopo di poter effettuare operazioni di *matching* tra unità individuali con profili molto simili sulle variabili osservabili (entro un *range* fissato dal ricercatore stesso per ogni covariata) ed ha il vantaggio di mantenere la natura multidimensionale dello spazio dei dati senza ridurla ad uno score unidimensionale (come invece viene implementato da altre tecniche quali il *propensity score matching*).

L'implementazione strutturata di tali tecniche di *matching* per banche dati di considerevoli dimensioni (oltre 5 milioni di record) ha permesso di effettuare un primo studio a livello sperimentale di tale approccio metodologico che permette di superare molti dei limiti finora riscontrati dell'applicabilità di tale tecniche ad elevate quantità di record.

3. Un modello di micro-simulazione dell'addizionale Irpef a livello locale

Per l'analisi d'impatto delle possibili *policies* fiscali a livello familiare è stato elaborato un modello parametrico di simulazione ad un duplice livello. Il primo livello calcola la stima esatta dell'ammontare del tributo locale sulla base delle aliquote fissate

dal ricercatore e dell'ultimo imponibile disponibile in dichiarazione per ciascun individuo presente in archivio; il secondo invece aggrega a livello istituzionale i tributi stimati su base individuale così da ricostruire il gettito totale che ne deriverebbe per l'amministrazione locale.

Tale modello (che rientra nella famiglia dei modelli "statici" poiché basato sulla banca dati all'ultimo anno disponibile in ambito dichiarativo) è stato implementato al fine di permettere al decisore locale di ottenere simulazioni realistiche dei possibili scenari, impostando in ottica parametrica i valori delle variabili di cui può, per autonomia amministrativa, disporre (variazione dei valori dello schema delle aliquote entro i limiti del *range* stabilito dalla normativa, ridefinizione ove consentito dalla normativa vigente degli scaglioni contributivi e delle fasce di reddito escluse dal tributo, esenzione per classi o categorie di contribuenti). Esso in particolare si è focalizzato sull'introduzione di possibili sgravi nel calcolo dell'addizionale locale per quei cittadini che hanno a carico uno o più familiari (siano essi il coniuge o i figli).

La tassazione nel nostro paese è infatti incentrata sull'unità impositiva individuale e qualunque intervento locale si voglia ipotizzare non può snaturarne la struttura (sarebbe altresì necessario un intervento del decisore centrale ben più complesso da realizzare). Incidere invece sulle aliquote individuali di un tributo locale (come l'addizionale Irpef), permetterebbe alle amministrazioni di intervenire in tempi brevi nella riduzione dei carichi fiscali, selezionando puntualmente le categorie di cittadini beneficiari degli stessi.

Alcune delle dimensioni in gioco, quali il gettito totale che l'amministrazione intende ottenere e lo schema degli scaglioni di tassazione Irpef, sono stati fissati a priori (il primo per esigenze di bilancio delle amministrazioni locali, il secondo per via delle precise disposizioni contenute nella legge finanziaria 2011 che uniformano a livello nazionale gli scaglioni dell'addizionale regionale con la struttura degli scaglioni Irpef).

Tale studio ha quindi permesso di effettuare una ipotesi realistica di ri-modellazione di alcuni dei parametri chiave del processo di definizione e

Tabella 1: possibile schema di variazione delle aliquote per classi di reddito, n° figli e coniuge a carico.

Classi Irpef	Coniuge/figli	0	0,5	1	1,5	2	2,5	3	4	5	6
1	0	15,45%	-7,64%	-23,04%	-34,03%	-42,28%	-48,69%	-53,82%	-61,52%	-67,02%	-71,14%
2	0	15,25%	-7,93%	-23,27%	-34,22%	-42,46%	-48,80%	-53,94%	-61,59%	-67,04%	-71,14%
3	0	14,11%	-8,73%	-23,95%	-34,84%	-42,97%	-49,32%	-54,39%	-62,01%	-67,44%	-71,49%
4	0	12,97%	-6,34%	-19,14%	-27,44%	-35,89%	-40,15%	-44,97%	-50,30%	-53,84%	-57,59%
5	0	15,24%	-7,04%	-18,56%	-25,41%	-32,36%	-31,73%	-38,18%	-40,33%	-43,87%	-39,33%
1	1	-42,28%	-48,69%	-53,82%	-58,02%	-61,52%	-64,48%	-67,02%	-71,14%	-74,35%	-76,91%
2	1	-42,41%	-48,79%	-53,91%	-58,06%	-61,60%	-64,50%	-67,06%	-71,16%	-74,35%	-76,93%
3	1	-42,96%	-49,31%	-54,36%	-58,54%	-61,98%	-64,89%	-67,40%	-71,49%	-74,64%	-77,16%
4	1	-35,17%	-39,74%	-44,90%	-45,70%	-52,03%	-53,60%	-56,07%	-58,91%	-61,69%	-63,65%
5	1	-7,65%	-18,77%	-46,16%	-51,02%	-48,31%	-42,78%	-47,58%	-47,72%	-53,06%	-48,05%

riscossione delle addizionali locali. Il modello ipotizza nuovi valori delle aliquote e la progressività dell'imposta individuale al fine di ottenere a parità di gettito per l'amministrazione una maggiore equità nella distribuzione della imposta stessa prevedendo da un lato una progressività che segue gli scaglioni Irpef, dall'altro uno sconto di imposta correlato con il numero di componenti a carico del percettore di reddito (coniuge e/o eventuali figli). Da ultimo esso prevede anche una ripartizione dell'agevolazione fiscale che tenga conto in maniera proporzionale di entrambe le dimensioni di equità e progressività (è stato ipotizzato uno sgravio inferiore per individui con familiari a carico ma con redditi elevati). La formattazione condizionale delle percentuali di sconto in tabella 1 rende, con un effetto semplice ed efficace, l'idea di quali siano le classi di individui con maggiori e minori aggravii di tale modello in termini di incremento/sconto sull'importo medio di addizionale del sistema attualmente in vigore.

Sono altresì implementabili, in questa ottica, modelli più complessi che tengano anche in considerazione le dinamiche evolutive del reddito e la composizione stessa della popolazione in esame. Tali modelli di micro-simulazione dinamica stimano in un primo stadio il cambiamento della popolazione (in termini di caratteristiche dichiarative) che ci aspettiamo al tempo $t+1$ rispetto al tempo t in base ai cambiamenti avvenuti nella popolazione stessa gli anni precedenti (tra t e $t-1$, $t-1$ e $t-2$, etc...) e successivamente stimano sulla nuova popolazione ottenuta i cambiamenti provocati da variazioni nelle *policies*.

4. Conclusioni

Quanto descritto rappresenta un esempio di come sia possibile, con gli strumenti attuali e la disponibilità di banche dati amministrative, ipotizzare ed implementare modelli che permettano al decisore istituzionale di percepire l'impatto degli interventi di policy e conseguentemente di valutare sia il ritorno che ne scaturisce per l'amministrazione sia l'impatto di tali interventi sul sistema del welfare dei cittadini e delle famiglie.

Lo sviluppo di modelli di micro-simulazione basati su dati amministrativi pone le basi, sia a livello metodologico che tecnico, per la costruzione di sistemi informativi statistici a supporto decisionale per la Pubblica Amministrazione in un momento storico in cui certamente ce n'è un grande bisogno. Lo studio di nuove metodologie di supporto decisionale infatti è altresì ampliabile a moltissimi aspetti della relazione tra istituzioni pubbliche e cittadini. Tra questi sono certamente rilevanti ed oggetto di studio presso il Crisp, la valutazione delle *policies* in ambito sanitario, i modelli di ridefinizione delle tariffe di erogazione dei servizi e delle fasce di cittadini esenti, la rimodulazione dei criteri per la definizione degli indicatori equivalenti di accesso ai servizi locali al cittadino.

Riferimenti bibliografici

Iacus, S. M., King, G. and Porro, G., 2009, "CEM: Coarsened Exact Matching Software", Journal of Statistical Software, 30(9), <http://gking.harvard.edu/cem>.