

CAPP

Centro di Analisi
delle Politiche Pubbliche

INTEGRAZIONE TRA DATI CENSUARI E DATI DI INDAGINI
SULLA CONDIZIONE ECONOMICA DELLE FAMIGLIE
PER LA COSTRUZIONE DI MAPPE DELLA POVERTÀ LOCALE
E DI MODELLI DI VALUTAZIONE DELLE POLITICHE SOCIALI LOCALI

Il Rapporto è stato curato dal CAPP, Centro di Analisi delle Politiche Pubbliche, nell'ambito della convenzione con il Comune di Modena relativa alla realizzazione della seconda indagine sulle famiglie finalizzata alla valutazione dell'impatto distributivo sulle famiglie delle politiche fiscali, tariffarie e di offerta di servizi del comune di Modena (ICES $mo2$).

Autori: Marcello Morciano, Paolo Silvestri.

Rapporto intermedio del progetto ***Poverty Mapping* per la provincia di Modena**

Integrazione tra dati censuari e dati di indagini sulla condizione economica delle famiglie per la costruzione di mappe della povertà locale e di modelli di valutazione delle politiche sociali locali

Introduzione

I processi di decentramento amministrativo in atto nel Paese comportano un crescente interesse verso strumenti in grado di fornire dettagliate informazioni sulla distribuzione personale e territoriale del benessere economico nella collettività. L'ampliarsi dell'autonomia tributaria degli enti locali pone poi ulteriori problemi nel monitoraggio degli effetti distributivi delle politiche di tassazione (e di spesa) conseguenti all'imposizione tributaria locale.

Il Sistema Statistico Comunale dispone di un ampio set di informazioni (censuarie ed amministrative) presenti in archivi eterogenei che, anche se ben organizzate, difficilmente possono essere direttamente utilizzate per rispondere ai quesiti sulla distribuzione del benessere e sull'incidenza dei tributi locali tra la collettività. Tali archivi, per contro, hanno il principale vantaggio di avere una copertura totale (o quasi) sul territorio locale.

L'inchiesta diretta e periodica è senza dubbio uno dei migliori metodi per rilevare le informazioni qualitative e quantitative che, se opportunamente elaborate, possono fornire risposte utili agli amministratori per valutare l'efficacia/efficienza e gli esiti distributivi delle politiche locali. Purtroppo, i costi legati alla raccolta delle informazioni "sul campo" costringono a selezionare, campioni di dimensioni limitate. Alle famiglie e individui selezionati si somministra un articolato questionario volto all'acquisizione analitica di informazioni sulle caratteristiche sociali, economiche e sugli stili di vita. Spesso i vincoli di bilancio limitano fortemente la numerosità delle interviste e non consentono, quindi, di disporre di campioni

sufficientemente numerosi per condurre analisi circoscritte ad aree diverse da quelle previste dal piano di campionamento. Inoltre, tutte le indagini campionarie presentano il limite di non riuscire ad intercettare diffusamente alcune categorie sociali (senza fissa dimora, immigrati, persone sole) e non si sottraggono alle possibili reticenze degli intervistati a dichiarare i veri livelli di reddito e di ricchezza.

I pregi e i limiti nell'utilizzo di fonti amministrative ed indagini campionarie sono stati oggetto di numerosi studi. La maggior parte di questi rilevano nell'integrazione delle due tipologie di fonti la strada ottimale da seguire per ottenere un quadro rappresentativo del territorio di analisi.

Da tali considerazioni si sviluppano nuovi metodi econometrici che consentono di integrare in modo corretto ed efficiente le diverse fonti informative. Tra tutti, notevole successo ricevono i metodi che consentono l'integrazione dei dati censuari e dei dati derivanti da indagini campionarie sul reddito, poiché consentono di disporre della "materia prima" necessaria per condurre affidabili stime sulla distribuzione del benessere economico in una collettività, anche ad un dettaglio territoriale molto ampio.

Il presente rapporto, reso possibile grazie alla proficua collaborazione tra alcuni membri del CAPP e del Servizio Statistica del Comune di Modena, espone i risultati intermedi dell'applicazione della procedura di *Poverty Mapping* alla Provincia di Modena. Il lavoro ha permesso di integrare i dati su base individuale del censimento della popolazione 2001 con le informazioni presenti nell'indagine sulla condizione economiche delle famiglie modenesi (ICES_{mo}) realizzata nel 2002 dal CAPP, per realizzare un dataset utile allo svolgimento di analisi sulla distribuzione del benessere economico ad elevati livelli di dettaglio territoriale.

La predisposizione di questo strumento dovrebbe consentire di fare un passo in avanti molto importante nella costruzione di "mappe" della povertà e della disuguaglianza e avere una rilevante ricaduta nella progettazione di politiche sociali a livello locale (costruzione dei piani di zona, ecc.). Su tale banca dati sarà infatti possibile costruire modelli di microsimulazione in grado di affrontare analisi degli effetti distributivi delle politiche sociali a livelli di disaggregazione sino a questo momento non realizzabili.

Di seguito si descrivono i principali risultati intermedi ottenuti nelle diverse fasi del *Poverty Mapping* per la Provincia di Modena con un focus particolare per il comune di Modena. Dopo una sintetica descrizione del metodo e dei dati utilizzati (parr. 1 e 2), si presentano i risultati derivanti dal confronto delle caratteristiche osservate nel campione ICES_{mo} con i dati censuari (par. 3). I paragrafi successivi espongono i risultati econometrici derivanti dalla procedura di stima delle determinanti il livello di benessere economico delle famiglie residenti (par. 4) e alcuni risultati intermedi con un esempio di mappatura possibile per la Città di Modena (par. 5). Nel paragrafo conclusivo si traggono alcune considerazioni sui possibili utilizzi della procedura.

1. Il metodo del Poverty Mapping

A partire dall'inizio del 2000 lo studio territoriale e personale delle condizioni di povertà e della disuguaglianza del reddito ha tratto giovamento dall'applicazione di nuove tecniche di analisi che consentono di realizzare obiettivi più ambiziosi di quelli normalmente raggiungibili mediante l'uso di indagini campionarie. Quasi tutte le indagini di tipo campionario hanno dimensioni troppo ridotte (a causa degli elevati costi di rilevazione) per fornire informazioni a livelli molto dettagliati su sottoinsiemi specifici della popolazione, quali particolari categorie sociali o porzioni territoriali particolarmente circoscritte. Il metodo, noto come *Poverty Mapping*, si basa sull'utilizzo di una fonte censuaria e di un'indagine *ad hoc* sulle condizioni di vita delle famiglie e permette di imputare, mediante l'utilizzo di tecniche statistiche adeguate ed efficienti, le informazioni economiche presenti nell'indagine campionaria su tutte le unità della fonte censuaria, con conseguente stima dei livelli di benessere per tutti gli individui residenti sul territorio modenese.

La ricerca comporta l'esecuzione di diverse fasi. La prima richiede l'individuazione e la standardizzazione delle variabili comuni alle due banche dati. In questa fase vengono selezionate tutte quelle variabili che verosimilmente possono essere buoni predittori del reddito e che risultano comuni alle due fonti di dati. Se le modalità di una variabile rilevata sull'indagine campionaria o censuaria non risultano

confrontabili, si definiscono modalità aggregate corrispondenti. Conclusa la selezione delle variabili si avvia una fase di controllo di conformità tra le distribuzioni campionarie e le distribuzioni ricavate dai dati censuari; anche in questa fase può essere necessario modificare le definizioni delle modalità delle variabili per rendere maggiormente confrontabili le distribuzioni che scaturiscono dalle due fonti (**fase 0**).

La fase successiva riguarda la selezione di un modello econometrico, da stimare sui dati ICES_{mo}, per la previsione della variabile reddito familiare o di una sua trasformata. Il vettore dei parametri stimati e la relativa matrice di varianza-covarianza richiedono il superamento di una serie di test per verificare l'opportunità o meno di utilizzare i pesi campionari, per verificare l'ipotesi di omoschedasticità del termine di errore e la presenza di effetti di gruppo (**fase 1**).

Le stime dei parametri ottenute nella fase precedente sono utilizzate per simulare il reddito di tutte le unità della popolazione di riferimento (unità censuarie) e, mediante procedure di *bootstrap*, si definiscono gli intervalli di confidenza entro cui sono collocate le stime (**fase 2**).

2. I dati utilizzati

Censimento ISTAT

La fonte dei dati censuaria utilizzata per il *Poverty Mapping* della Provincia di Modena è il 14° Censimento della Popolazione e delle Abitazioni (anno 2001). I principali obiettivi di tale rilevazione sono da individuare nel conteggio della popolazione e la rilevazione delle caratteristiche strutturali, l'aggiornamento e la revisione delle anagrafi, la raccolta di informazioni sulla consistenza numerica e sulle caratteristiche strutturali delle abitazioni. Per la prima volta sono stati censiti anche gli edifici: numero e caratteristiche degli edifici ad uso abitativo e, limitatamente ai centri abitati, anche di quelli ad uso non abitativo. Inoltre, la rilevazione ha riguardato anche la struttura demografica e familiare della popolazione straniera residente, le tipologie familiari, il titolo di studio più elevato conseguito in Italia o all'estero, la condizione professionale ed infine informazioni sulle persone non residenti. La rilevazione dei dati è avvenuta nel periodo ottobre–novembre 2001, con riferimento alla data del 21 ottobre 2001.

Per l'utilizzo dei dati censuari sono state rispettate tutte le disposizioni sul segreto statistico del d.lgs. n. 322/89 e le regole a tutela della riservatezza della legge n. 675/96.

L'indagine Campionaria ICESmo

Nel secondo semestre del 2002, il Centro di Analisi delle Politiche Pubbliche ha realizzato la prima Indagine sulle Condizioni Economiche e Sociali delle famiglie della Provincia di Modena (ICESmo). L'indagine ha rilevato, tramite intervista diretta, dettagliate informazioni sulle caratteristiche socio-economiche, sugli stili di vita e sul ricorso ai servizi di cura di un campione rappresentativo della popolazione modenese, selezionato secondo una procedura a due stadi¹. La disponibilità di un variegato *set* di informazioni sulla natura qualitativa e quantitativa delle forme di ricchezza posseduta e dei redditi prodotti nei 12 mesi precedenti l'intervista rendono questa indagine il supporto informativo ideale per lo studio della distribuzione del benessere e per l'analisi della povertà a livello locale. La somiglianza di ICESmo ad altre indagini campionarie nazionali consente di operare dei raffronti tra le caratteristiche socio-economiche locali e le caratteristiche rilevate a livello nazionale o sub-nazionale².

Il campione di ICESmo è composto da 3062 individui appartenenti a 1235 famiglie; per le caratteristiche con cui è stato costruito è rappresentativo sia del Comune di Modena sia del resto della provincia nel complesso. Le famiglie intervistate nel capoluogo di provincia sono state 589. Le restanti interviste si riferiscono a famiglie estratte dalle liste anagrafiche di 16 comuni rappresentativi del rimanente territorio provinciale (Lalla 2004).

¹ Le unità di primo stadio sono i comuni; le unità di secondo stadio sono le famiglie. Una descrizione dettagliata del piano di campionamento adottato è contenuta in Lalla (2004).

² Interessanti analisi sulla distribuzione del reddito e sui livelli di povertà della popolazione modenese e confronti rispetto ai dati rilevati da indagini campionarie nazionali sono contenuti in Baldini e Silvestri (2004a, 2004b).

3. Fase 0: l'analisi dei dati elementari

Le fonti statistiche presentate nel paragrafo precedente si riferiscono allo stesso territorio di analisi e permettono di condurre analisi sia a livello familiare sia a livello individuale. La tabella 1 espone sinteticamente le caratteristiche salienti delle basi di dati. Si ritiene che le distorsioni legate alla differenza nel periodo di rilevazione siano poco rilevanti.

Tabella 1

	<i>ICESmo</i>	<i>Censimento</i>
<i>periodo di rilevazione:</i>	Giugno-Ottobre 2002	ottobre 2001
<i>periodo di riferimento:</i>	12 mesi precedenti	Settimana preced.
<i>territorio di riferimento:</i>	Comune di Modena	Comune di Modena
<i>Unità di riferimento:</i>	Individui/famiglie	Individui/famiglie
<i>Numero famiglie:</i>	589	74.502
<i>Numero individui:</i>	1.388	175.502
<i>% fam ICESmo/Censimento</i>	0.79%	
<i>Numero medio componenti</i>	2.36	2.36

L'analisi iniziale ha visto il controllo della rispondenza delle distribuzioni di frequenza dei dati di *ICESmo* con quelli del censimento. La tabella 2 espone i risultati riferiti al solo comune di Modena, utilizzando come unità di analisi l'individuo. Alcune variabili sono dicotomiche. Queste variabili, definite variabili dummy assumono valore 1 quando l'individuo si trova nella condizione indicata, valore 0 altrimenti.

Tabella 2 Caratteristiche dell'abitazione di residenza

<i>Variabile</i>	<i>Censimento</i>		<i>ICESmo</i>		<i>Kolmogorov-Smirnov Two-Sample Test</i>
proprietà (a1)	<i>Freq.</i>	<i>%</i>	<i>Freq.</i>	<i>%</i>	Pr > KSa
0	45685	26,55	1347	26,74	1
1	126366	73,45	3691	73,26	
n° bagni (a4)	<i>Freq.</i>	<i>%</i>	<i>Freq.</i>	<i>%</i>	Pr > KSa
1			3090	61,33	
2			1541	30,59	
3 e più			407	8,08	
sup. abitaz. (a5)	<i>Freq.</i>	<i>%</i>	<i>Freq.</i>	<i>%</i>	Pr > KSa
(min/59= 1)	14937	8,68	263	5,22	< .0001
(60/74= 2)	22789	13,25	644	12,78	
(75/94= 3)	46120	26,81	1352	26,84	
(95/109= 4)	29167	16,95	873	17,33	
(110/150= 5)	40346	23,45	1253	24,87	
(151/max= 6)	18692	10,86	653	12,96	
posto auto (a6)	<i>Freq.</i>	<i>%</i>	<i>Freq.</i>	<i>%</i>	Pr > KSa
0	23688	13,77	608	12,07	0,1182
1	148363	86,23	4430	87,93	
casa popolare (a7)	<i>Freq.</i>	<i>%</i>	<i>Freq.</i>	<i>%</i>	Pr > KSa
0	166286	96,65	4903	97,32	0,9802
1	5765	3,35	135	2,68	

Tabella 3 Caratteristiche del nucleo familiare

<i>Variabile</i>	<i>Censimento</i>		<i>ICESmo</i>		<i>Kolmogorov-Smirnov Two-Sample Test</i>
n° comp. (f1)	<i>Freq.</i>	<i>%</i>	<i>Freq.</i>	<i>%</i>	Pr > KSa
1			518	10,28	
2			1415	28,09	
3			1552	30,81	
4			1285	25,51	
5			241	4,78	
6 e più			27	0,54	
relazione di parentela (f2)	<i>Freq.</i>	<i>%</i>	<i>Freq.</i>	<i>%</i>	Pr > KSa
intestataro, coniuge/conviv.	113420	65,92	3402	67,53	0,0003
figli	46732	27,16	1429	28,36	
genitori del capof, coniuge	3098	1,8	100	1,98	
altro	8801	5,12	107	2,12	

Tabella 4 Caratteristiche individuali

Variabile	Censimento		ICESm		Kolmogorov-Smirnov Two-Sample Test
	Freq.	%	Freq.	%	
sesto (p1)					Pr > KSa
0	89781	52,18	2605	51,71	0,9999
1	82270	47,82	2433	48,29	
anno nascita (p2)	Freq.	%	Freq.	%	Pr > KSa
(min/1920=1)	8185	4,76	212	4,21	0,0549
(1921/1930=2)	17299	10,05	489	9,71	
(1931/1935=3)	9966	5,79	297	5,9	
(1936/1940=4)	11155	6,48	322	6,39	
(1941/1945=5)	10719	6,23	262	5,2	
(1946/1950=6)	12775	7,43	379	7,52	
(1951/1955=7)	11410	6,63	348	6,91	
(1956/1960=8)	12179	7,08	361	7,17	
(1961/1965=9)	13987	8,13	390	7,74	
(1966/1970=10)	14216	8,26	427	8,48	
(1971/1975=11)	12993	7,55	389	7,72	
(1976/1980=12)	8726	5,07	285	5,66	
(1981/1985=13)	6473	3,76	218	4,33	
(1986/max=14)	21968	12,77	659	13,08	
nascita in prov MO (p3)	Freq.	%	Freq.	%	Pr > KSa
0	56696	32,95	1225	24,32	< .0001
1	115355	67,05	3813	75,68	
nascita in altre prov ER (p4)	Freq.	%	Freq.	%	Pr > KSa
0	45138	26,24	945	18,76	< .0001
1	126913	73,76	4093	81,24	
nascita al nord (p5)	Freq.	%	Freq.	%	Pr > KSa
0	37759	21,95	746	14,81	< .0001
1	134292	78,05	4292	85,19	
nascita al CENTRO-SUD (p6)	Freq.	%	Freq.	%	Pr > KSa
0	144376	83,91	4401	87,36	< .0001
1	27675	16,09	637	12,64	
nascita all'ESTERO (p7)	Freq.	%	Freq.	%	Pr > KSa
0	161967	94,14	4939	98,03	< .0001
1	10084	5,86	99	1,97	
cittadinanza (p8)	Freq.	%	Freq.	%	Pr > KSa
0	7684	4,47	41	0,81	< .0001
1	164367	95,53	4997	99,19	

-segue tab. 4

<i>Variabile</i>	<i>Censimento</i>		<i>ICESm</i>		<i>Kolmogorov-Smirnov Two-Sample Test</i>
	<i>Freq.</i>	<i>%</i>	<i>Freq.</i>	<i>%</i>	
CONUGATO (p9)					Pr > KSa
0	86409	50.22	2481	49.25	0.7383
1	85642	49.78	2557	50.75	
VEDOVO/A (p10)					Pr > KSa
0	156507	90.97	4531	89.94	0.6779
1	15544	9.03	507	10.06	
titolo di studio (p11)					Pr > KSa
1=nessuna	21385	12.43	544	10.8	<0.001
2=lic elementare+avviamento	42736	24.84	1004	19.93	
3=lic media + qualifica (2-3 anni)	40843	23.74	1233	24.47	
4=diplom	48208	28.02	1536	30.49	
6=diploma universitario e laurea, post-	18879	10.97	721	14.31	
OPERAIO (p12)					Pr > KSa
0	152826	88.83	4408	87.5	0.3512
1	19225	11.17	630	12.5	
IMPIEGATO (p13)					Pr > KSa
0	137393	79.86	3965	78.7	0.5321
1	34658	20.14	1073	21.3	
DIRIGENTI (p14)					Pr > KSa
0	167229	97.2	4937	98	0.9143
1	4822	2.8	101	2	
LIBPROF. (p15)					Pr > KSa
0	167279	97.23	4833	95.93	0.384
1	4772	2.77	205	4.07	
IMPRENDITORE (p16)					Pr > KSa
0	169565	98.56	4944	98.13	1
1	2486	1.44	94	1.87	
LAV. AUT. (p17)					Pr > KSa
0	161539	93.89	4773	94.74	0.8714
1	10512	6.11	265	5.26	
DISOCCUPATO (p18)					Pr > KSa
0	167161	97.16	4960	98.45	0.3855
1	4890	2.84	78	1.55	

Variabile	Censimento		ICESmo		Kolmogorov-Smirnov Two-Sample Test
	Freq.	%	Freq.	%	Pr > KSa
STUDENTE (p19)					
0	141623	82,31	4080	80,98	0,352
1	30428	17,69	958	19,02	
CASALINGA (p20)					
0	158083	91,88	4798	95,24	< .0001
1	13968	8,12	240	4,76	
ALTRO+CASALINGA (p20a)					
0	154644	89,88	4618	91,66	0,0897
1	17407	10,12	420	8,34	
PENSIONATO da LAV (p21)					
0	130578	75,89	3824	75,9	1
1	41473	24,11	1214	24,1	
ALTRA CONDIZ (p22)					
0	168612	98	4618	91,66	< .0001
1	3439	2	420	8,34	
TIPO di CONTRATTO (p23)					
0=nessuno	113346	65,88	3208	63,68	0,0173
1=indeterminato	51301	29,82	1638	32,51	
2= determinato+interninale	7404	4,3	192	3,81	
TEMPO PIENO (P24)					
0	103311	60,05	2987	59,29	0,9416
1	68740	39,95	2051	40,71	
ORELAVORO (p25)					
nessuna	96955	56,35	2681	53,22	< .0001
(1/20=1)	6374	3,7	168	3,33	
(21/35=2)	10140	5,89	355	7,05	
(36/50=3)	54761	31,83	1696	33,66	
(51/max=4)	3821	2,22	138	2,74	

-segue tab 4

<i>Variabile</i>	<i>Censimento</i>		<i>ICESm</i>		<i>Kolmogorov-Smirnov Two-Sample Test</i>
	<i>Freq.</i>	<i>%</i>	<i>Freq.</i>	<i>%</i>	
SETTORE AGRICOLTURA (p26)	<i>Freq.</i>	<i>%</i>	<i>Freq.</i>	<i>%</i>	Pr > KSa
0	170880	99,32	5023	99,7	1
1	1171	0,68	15	0,3	
SETTORE EDILIZIA (p27)	<i>Freq.</i>	<i>%</i>	<i>Freq.</i>	<i>%</i>	Pr > KSa
0	167604	97,42	4948	98,21	0,914
1	4447	2,58	90	1,79	
SETTORE INDUSTRIA (p28)	<i>Freq.</i>	<i>%</i>	<i>Freq.</i>	<i>%</i>	Pr > KSa
0	149839	87,09	4235	84,06	0,0003
1	22212	12,91	803	15,94	
SETT. COMMERCIO (p29)	<i>Freq.</i>	<i>%</i>	<i>Freq.</i>	<i>%</i>	Pr > KSa
0	154016	89,52	4594	91,19	0,1307
1	18035	10,48	444	8,81	
SETT. ISTRUZIONE, SANITA', PA (p30)	<i>Freq.</i>	<i>%</i>	<i>Freq.</i>	<i>%</i>	Pr > KSa
0	155429	90,34	4466	88,65	0,121
1	16622	9,66	572	11,35	
SETT ALTRI SERVIZI (p31)	<i>Freq.</i>	<i>%</i>	<i>Freq.</i>	<i>%</i>	Pr > KSa
0	155959	90,65	4599	91,29	0,9883
1	16092	9,35	439	8,71	

I risultati hanno dimostrato una sostanziale capacità dei dati campionari di catturare l'esatta distribuzione per le seguenti variabili individuali:

- genere;
- anno di nascita;
- stato civile;
- condizione professionale;
- tempo di lavoro (PT/FT);
- settore di attività.

Le variabili campionarie *area geografica di nascita, titolo di studio conseguito, ore di lavoro (in classi), tipologia di contratto* presentano invece, distribuzioni di frequenza significativamente diverse da quelle censuarie.

Minori distorsioni si rilevano nelle informazioni relative all'abitazione di residenza. Si accetta la somiglianza delle distribuzioni di frequenza delle variabili *titolo di godimento sull'abpr, dotazione posto auto/garage e dummy su residenza in casa popolare*.

La procedura ha visto sia l'utilizzo dei pesi individuali sia la conduzione di confronti senza l'utilizzo di essi. Si rilevano minori distorsioni dal confronto delle distribuzioni di frequenza pesate con pesi individuali.

La stessa analisi condotta per la Città di Modena può essere condotta anche sull'intera provincia, utilizzando anche l'unità di analisi familiare.

I confronti successivi sono stati condotti sull'intera provincia di Modena utilizzando come unità di analisi la famiglia. Dai risultati emergono distribuzioni di frequenza simili sulle variabili:

- tipologia familiare;
- genere del capofamiglia e dell'eventuale coniuge;
- stato civile del capofamiglia e dell'eventuale coniuge, convivente;
- professione del capofamiglia e dell'eventuale coniuge, convivente;
- tipo contratto di lavoro del capofamiglia;
- settore di attività del coniuge.

Per l'abitazione di residenza si preserva solo la dummy su case popolari, mentre si rilevano significative distorsioni per le altre variabili. Significative differenze si rilevano nelle distribuzioni di frequenza di importanti variabili che la teoria economica identifica come determinanti del livello di reddito. Tra esse:

- anno e luogo di nascita;
- titolo di studio;
- ore di lavoro.

Anche in questo caso, il confronto è stato condotto sia utilizzando i pesi campionari (pesofam e pesonuovo) sia in assenza dei pesi. Informazioni più dettagliate possono essere fornite agli interessati sui risultati ottenuti.

4. Fase 1: un modello di previsione per il reddito

La fase denominata *Fase 1* consiste nella specificazione di un modello di regressione lineare, da stimare sui dati dell'indagine campionaria, per la previsione della variabile reddito familiare o di una trasformata di esso. Le applicazioni svolte fino a questo momento sul tema del *Poverty Mapping* (Elbers *et al.* 2003) specificano un modello di previsione per una trasformata logaritmica della spesa familiare per consumi, quindi – come è già stato anticipato - le applicazioni di questo lavoro costituiscono il primo esercizio di *Poverty Mapping* basato sul reddito. Vari tentativi di specificazione sono stati effettuati, considerando come variabile dipendente del modello, alternativamente, una trasformata logaritmica del reddito familiare equivalente ed una trasformata logaritmica del reddito familiare disponibile.

I risultati ottenuti, soprattutto in termini di capacità esplicativa del modello, hanno portato a scegliere il logaritmo del reddito familiare disponibile come variabile dipendente. Prima di passare alla formalizzazione di questa fase di lavoro, è opportuno precisare che per la scelta delle variabili da utilizzare come esplicative nel modello di regressione, è stata superata la cosiddetta *Fase 0* del *Poverty Mapping*. Nella *Fase 0* vengono selezionate tutte quelle variabili che verosimilmente possono essere buoni predittori del reddito e che

risultano comuni alle due fonti di dati. La formalizzazione del modello di regressione è esposta in appendice.

Per catturare le differenze territoriali, mantenendo al contempo una sufficiente numerosità familiare, abbiamo stimato le determinanti della condizione economica delle famiglie residenti nella Provincia di Modena, dividendo il campione ICES_{mo} in due gruppi: i residenti nella Città di Modena ed i residenti nel resto della provincia.

L'omoschedasticità e l'assenza dell'effetto gruppo sui due sottocampioni ha permesso di stimare i modelli con OLS, dove il logaritmo del livello di benessere economico familiare dipende dalla struttura e dalla composizione del nucleo familiare, dalle caratteristiche dell'abitazione di residenza e dalle condizioni socio-occupazionali del capofamiglia ed infine delle caratteristiche dell'eventuale coniuge/convivente. La variabile utilizzata per misurare il benessere economico familiare è il reddito disponibile (RDF) monetario, dato dalla somma dei redditi, al netto delle imposte, da lavoro (dipendente, indipendente), da pensione, da trasferimenti e da capitale (reale e finanziario) percepiti da tutti i membri del nucleo familiare³. Il modello per la città di Modena e quello per il resto della Provincia presentano una elevata capacità esplicativa, spiegando il 69,6% e il 58,7% della varianza totale. Come si può vedere dalla tab. 5, i segni dei coefficienti sono in linea con quelli attesi. Nel complesso, si rilevano livelli di RDF monetario crescenti al crescere della numerosità familiare. A parità di altre condizioni, la composizione familiare influenza il livello di benessere familiare. Il numero di figli e di anziani presenti in famiglia riduce il livello di RDF monetario. L'effetto è dovuto probabilmente all'attività di assistenza e di cura presente o passata prestata dai componenti familiari autosufficienti, che potrebbe aver precluso o limitato le carriere lavorative di questi ultimi. Il numero di percettori di reddito da lavoro presenti in famiglia influenza positivamente il livello monetario di benessere economico familiare. Il numero di lavoratori registra una relazione non lineare (di secondo grado) con il livello di benessere per le famiglie residenti nel comune capoluogo, mentre si registra una dipendenza lineare per le famiglie residenti nel resto della

³ Il reddito da capitale reale comprende sia gli affitti percepiti sugli immobili dati in affitto sia il reddito figurativo sull'abitazione di proprietà (Baldini e Silvestri 2004a).

provincia. Se escludiamo capofamiglia e coniuge, la presenza di un familiare laureato in una famiglia residente nella Città di Modena, a parità di altre condizioni, aumenta del 25,95% il livello di RDF monetario. Il titolo di godimento sull'abitazione principale è un fattore esplicativo della condizione economica familiare. Le famiglie che vivono in affitto o in una casa popolare registrano livelli di RDF monetario mediamente più basso di quelle proprietarie (o usufruttuarie) dell'abitazione di residenza. Anche la dimensione dell'abitazione di residenza è un fattore che spiega il livello di benessere monetario della famiglia, registrando, a parità di altre condizioni, livelli di reddito crescenti al crescere dello spazio abitativo⁴.

I principali fattori individuali significativi nello spiegare il benessere familiare sono quelli legati alle caratteristiche socio-occupazionali del capofamiglia, che si caratterizza come il principale percettore di reddito nel nucleo familiare (*breadwinner*). Al crescere dell'età⁵ e del livello di scolarizzazione del capofamiglia aumenta il livello di benessere familiare. La laurea del capofamiglia è premiante a Modena (+39,32%) mentre nel resto della provincia assicura un livello medio di RDF monetario più basso di quello registrato per i capofamiglia diplomati⁶. Anche la condizione professionale è un elemento significativo per la quantificazione del benessere economico familiare. A parità di altre condizioni, i livelli più elevati di benessere monetario si rilevano per le famiglie con capofamiglia imprenditore o dirigente.

Infine, la presenza di un coniuge e la sua eventuale attività lavorativa a tempo pieno influenza positivamente il livello di RDF monetario.

⁴ La *dummy* di riferimento per la lettura del coefficiente è la famiglia che dispone di un'abitazione con superficie inferiore a 75 mq. L'effetto marcato della superficie dell'abitazione principale sul RDF monetario è influenzato dall'inclusione, nella definizione del reddito disponibile, del reddito figurativo sull'abitazione di proprietà.

⁵ Il modello per il resto della Provincia non ha rilevato significativa la misura in anni dell'età del capofamiglia. Utilizzando una misura in classi di età del capofamiglia abbiamo rilevato significatività nella *dummy* che caratterizza le famiglie con capofamiglia nato dopo il 1971. A parità di altre condizioni, si rilevano livelli di benessere economico mediamente più bassi del 24,26% per le famiglie residenti nel resto della Provincia con capofamiglia nato dopo il 1971 rispetto ad una stessa famiglia con capofamiglia meno giovane.

⁶ Il risultato è in parte motivato da un minor tasso di rendimento dell'istruzione in Provincia di Modena rispetto al dato nazionale. Questo potrebbe essere l'effetto dell'abbondanza numerica dei laureati, «...che potrebbero svolgere lavori qualitativamente non molto diversi da quelli tipicamente riservati a titoli di studio inferiori (in particolare per i lavoratori dipendenti); inoltre, nel territorio modenese è molto diffusa la categoria del lavoro indipendente, spesso caratterizzata dalla combinazione di titoli di studio medio-bassi e da redditi alti». (Baldini e Silvestri 2004a).

Tabella 5 -Risultati della regressione: coefficienti OLS e livelli di significatività (standard error in parentesi)

	Città di Modena	Resto della Provincia
N_comp	0,1803*** (0,0304)	0,2039*** (0,0386)
N_figli0_14anni	-0,1995*** (0,0358)	-0,2317*** (0,0415)
N_figli15_24anni	-0,1506*** (0,0415)	-0,1477*** (0,0431)
N_over75	-0,0817** (0,0331)	
N_lavoratori	0,3357*** (0,0627)	0,0872** (0,0369)
N_lavoratori ²	-0,0732*** (0,0172)	
LAUREATI_infam ^a	0,2595*** (0,0881)	
Affitto ^a	-0,3554*** (0,0447)	-0,3778*** (0,0642)
SupAB_mq75_109 ^a	0,1221*** (0,0394)	0,1568*** (0,0491)
ABpr_mq110_epiu ^a	0,3268*** (0,0453)	0,2455*** (0,0574)
Vive_casa_pop ^a	-0,2369** (0,0952)	
Cf_età	0,0076*** (0,0019)	
Cf_coorte71_e_più ^a		-0,2426** (0,1231)
Cf_MEDIE ^a	0,1277*** (0,0467)	0,1849*** (0,0561)
Cf_DIPLOMA ^a	0,2519*** (0,0468)	0,2398*** (0,0632)
Cf_LAUREA ^a	0,3932*** (0,0645)	0,1931*** (0,0800)
Cf_OPERAIO ^a	0,1681** (0,0787)	
Cf_IMPIEGATO ^a	0,2099*** (0,0767)	0,2338*** (0,0661)
Cf_DIRIGENTI ^a	0,2900*** (0,1087)	0,6335*** (0,1015)
Cf_LAV_AUT ^a	0,2399*** (0,0921)	0,2401*** (0,0713)
Cf_IMPREND ^a	0,2934** (0,1161)	0,5248*** (0,1086)
Cf_PENS_LAV ^a	0,1786*** (0,0572)	0,1664** (0,0834)
CONIUGE ^a	0,1223*** (0,0389)	0,1309** (0,0602)
Co_TPIENO ^a	0,1102*** (0,0409)	0,1357*** (0,0398)
Costante	80,857*** (0,1462)	9,2560*** (0,0865)
N	589	646
R ² adj	0,6960	0,5867

Livello di significatività: * al 10%, ** al 5%, *** all'1%.

^a Variabile dicotomica che assume valore 1 se l'individuo si trova nella condizione indicata. Le *dummy* di riferimento per la lettura dei coefficienti sono: dimensione abitazione di residenza: meno di 75 mq; livello di scolarizzazione del capofamiglia (CF): nessuno o licenza elementare; coorte di nascita del capofamiglia: nati prima del 1971; condizione professionale del capofamiglia: studente, casalingo, disoccupato o pensionato da non lavoro per il modello «Città di Modena» e operaio, studente, casalingo, disoccupato, pensionato da non lavoro per il modello «resto della Provincia»; lavoro pagato *full-time* del coniuge (Co_TPIENO): rispetto a coniuge che non lavora o lavoratore/ice *part-time*.

5. Fase 2: simulazione delle distribuzioni di reddito

L'ultima fase consiste nella simulazione del livello di benessere economico per ogni individuo residente nella zona di analisi.

Le stime dei parametri ottenute nella fase 1 vengono trattate per simulare il reddito di tutte le unità della popolazione di riferimento (unità censuarie) e, mediante procedure di bootstrap, si definiscono gli intervalli di confidenza entro cui sono collocate le stime.

La costruzione del data base secondo le procedure sopra delineate rappresenta un patrimonio di indubbio valore per le applicazioni di ricerca empirica di studio della distribuzione e anche per la valutazione degli effetti delle politiche pubbliche. Si tratta di tecniche ancora nuove e che non hanno avuto occasioni di essere pienamente sperimentate utilizzando indagini sulla condizione economiche di tipo locale.

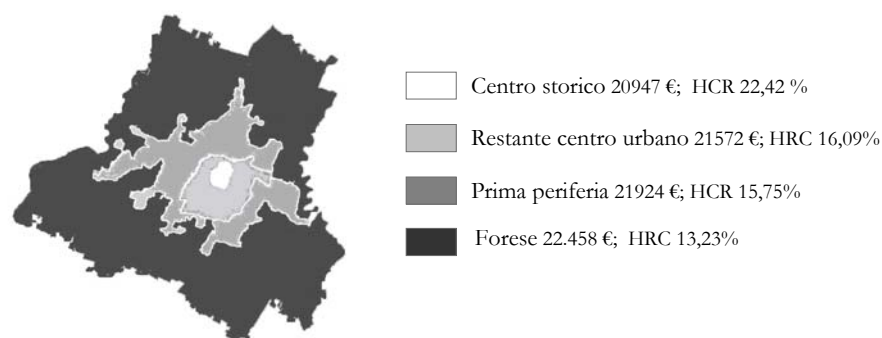
Mappatura

La figura seguente espone un esempio di possibile mappatura della distribuzione territoriale del benessere economico della città di Modena.

Per la generazione di essa è stato ripartito il territorio cittadino in quattro zone concentriche: Centro Storico (9963 individui residenti); Prima Periferia (68.028 individui residenti); Restante Centro Urbano (77.106 individui residenti); Zona Forese (18.321 individui residenti).

Il reddito medio disponibile di un cittadino modenese è stimato pari a 21.768 euro. La fig. 1 mostra come sono suddivise le zone nel territorio e l'incidenza di povertà (HCR, Head Count Ratio) calcolata sulla base del reddito familiare equivalente, l'indicatore del livello di benessere economico.

Figura 1 Distribuzione del livello di reddito equivalente medio e diffusione della povertà nel Comune di Modena



I cittadini residenti nel centro storico registrano un livello di reddito medio equivalente al di sotto della media comunale. Il risultato dipende soprattutto dalla maggiore quota di famiglie in affitto (con un'incidenza del 14,2% della popolazione immigrata sulla popolazione totale) e dalle dimensioni mediamente inferiori delle abitazioni di residenza. La prima periferia rileva l'età media dei capofamiglia attivi più alta di tutta la città. Il 13,2% della popolazione ha un'età superiore ai 75 anni, mentre la quota di immigrati è pari al 5,2% della popolazione totale. Spostandosi verso la zona forese si registra un livello di benessere economico crescente (si veda la fig. 1). Nella zona mediamente più ricca si rileva il più basso tasso di immigrazione (4,6%) e la più ampia presenza di abitazioni in proprietà di dimensioni superiori a 110 mq. Le famiglie residenti nella zona forese sono, in genere, famiglie più numerose, che hanno al loro interno la media più alta di percettori di reddito da lavoro (1,3 per famiglia contro 0,99 registrato nel centro storico)⁷.

⁷ Nella zona forese il 76,4% dei capofamiglia è coniugato e vive in una famiglia di numerosità media pari a 3,3 individui. Il centro storico registra la più bassa numerosità familiare (2,6 individui) e la più bassa quota di capofamiglia coniugati (50,1%).

6. Conclusioni

Sulla base dell'esercizio qui condotto e delle esperienze maturate dal CAPP sia nel campo delle indagini locali sia dei modelli di microsimulazione, è possibile trarre alcune considerazioni utili per procedere nell'analisi delle condizioni di vita delle famiglie modenesi.

1. Ottimizzazione della procedura di costruzione dei pesi dell'indagine ICES_{mo2}

La possibilità di confrontare i risultati di ICES_{mo} con quelli del Censimento ha permesso di quantificare il grado di distorsione dell'indagine rispetto ad una rilevazione che assicura il massimo grado di rappresentatività possibile. E' allo studio una procedura in grado di sfruttare al meglio la ricchezza informativa dei dati censuari, al fine di ottimizzare la procedura di generazione dei pesi campionari di ICES_{mo2}, la nuova indagine campionaria del CAPP.

2. Possibilità di ricostruzione dei dati censuari al 2006

L'occasione di poter sfruttare i dati del Censimento che, come noto, si realizza con cadenza decennale, non va sminuita anche se ormai i dati si riferiscono a 5 anni fa. Tramite opportune procedure di aggiornamento dinamico, sfruttando l'ampia serie di informazioni statistiche a disposizione del servizio statistica del comune, è possibile "ricostruire" le informazioni elementari dei residenti nel comune di Modena al 2006. Dalle fonti anagrafiche è possibile infatti conoscere il numero di iscritti e cancellati e i relativi cambi di residenza. Alcune sofisticate procedure econometriche consentirebbero di aggiornare i microdati disponibili tenendo conto del processo di invecchiamento della popolazione e, tramite l'utilizzo di opportune probabilità di transizione tra i diversi stati è possibile stimare i cambiamenti intervenuti nelle condizioni professionali, ecc.

3. Analisi ad un maggior dettaglio territoriale

L'archivio prodotto con la metodologia del poverty mapping contiene informazioni su tutte le famiglie residenti nella provincia

di Modena. La disponibilità di informazioni con elevato dettaglio territoriale consentirebbe di tracciare un quadro più attendibile delle condizioni economiche e sociali nei territori su cui vengono disegnate le attività di programmazione dei servizi (si pensi ad esempio ai piani sociali di zona).

4. Implementazione del modello di micro simulazione e matching con altre banche dati

Inoltre, trasferendo a questo archivio le tecniche implementate nel modello di micro simulazione costruito sulla base dell'indagine campionaria sarà possibile effettuare analisi di valutazione delle politiche fiscali o tariffarie. L'analisi delle politiche tariffarie potrà risultare più efficiente nel caso in cui si possano agganciare i dati relativi alle public utilities gestite a livello locale.

Bibliografia

- Baldini M., Bosi P., Silvestri P. (2004), *La ricchezza dell'equità. Distribuzione del reddito e condizioni di vita in un'area a elevato benessere*, Il Mulino, Bologna.
- Ballini, F., Betti, G., Neri, L. (2005), "Poverty and Inequality mapping in Transition Countries", *Working Paper n. 52*, Università degli Studi di Siena, Dipartimento di metodi quantitativi, gennaio.
- Betti G., Neri L., Marzadro S., Morciano M. (2006), "Stima della povertà a livello locale. I casi di Toscana, Modena e Trentino" in C. Saraceno (a cura di), *Disuguaglianze economiche e vulnerabilità in Italia*, Il Mulino, Bologna (di prossima pubblicazione);
- Elbers C., Lanjouw J. O., Lanjouw P. (2002a), "Micro-level Estimation of Welfare", *Working Paper n. 2911*, The World Bank, Washington, D.C.
- Elbers C., Lanjouw J. O., Lanjouw P. (2002b), "Micro-level Estimation of Poverty and Inequality", in *Econometrica*.
- Fay, R.E., Harriot, R.A. (1979), "Estimates of income for small places: an application of James-Stein procedures to census data", in *Journal of the American Statistical Association*, vol. 74.
- Hentschel, J., Lanjouw, J.O., Lanjouw, P., and Poggi, J. (2000), "Combining Census and Survey Data to Trace the Spatial Dimensions of Poverty: A Case Study of Ecuador", *World Bank Economic Review* 14(1)147-65.
- Levinson R. (2001), *Sample design for the 2002 Living Standards Measurement Survey (LSMS)*, Final Report to the World Bank, November.
- The World Bank (2002), *Basic Information Document, Living Standard Measurement Study, Albania*, Development Research Group.
- The World Bank (2003), *Construction of the consumption aggregate and estimation of the poverty line, LSMS 2002 – Albania*.

APPENDICE:

Alcuni approfondimenti sul metodo del Poverty Mapping

Se le modalità di una variabile rilevata sull'indagine campionaria o censuaria non sono risultate confrontabili, si è cercato di costruire modalità aggregate corrispondenti, anche a scapito di perdita di informazione. Conclusa la selezione delle variabili si avvia una fase di controllo di conformità tra le distribuzioni campionarie e le distribuzioni ricavate dai dati censuari; anche in questa fase talvolta si modificano le definizioni delle modalità delle variabili per rendere più confrontabili le distribuzioni che scaturiscono dalla fonte campionaria e da quella censuaria.

Passiamo ora alla formalizzazione del modello di regressione.

Sia $\ln y_{ch}$ il logaritmo del reddito disponibile per la famiglia h appartenente al gruppo c , e x_{ch} il corrispondente vettore delle caratteristiche familiari. Il valore atteso del $\ln y_{ch}$ condizionatamente alle caratteristiche familiari può essere espresso dalla seguente approssimazione lineare:

$$\ln y_{ch} = E[\ln y_{ch} | x_{ch}^T] + u_{ch} = x_{ch}^T \beta + u_{ch} \quad [1]$$

Le precedenti esperienze di lavoro nell'ambito del *Poverty Mapping*, hanno mostrato che il modello lineare deve prevedere una struttura del termine di errore complessa, per la possibile presenza di correlazione all'interno dei gruppi e di eteroschedasticità. La specificazione del termine di errore è pertanto $u_{ch} = \eta_c + \varepsilon_{ch}$, dove η è la componente di errore dovuta al gruppo e ε è la componente di errore dovuta alla famiglia; si ipotizza inoltre che le due componenti siano tra loro indipendenti e non correlate con le variabili esplicative del modello. È plausibile che nel modello [1] l'ipotesi di omoschedasticità del termine di errore sia violata; in tal caso la stima dei parametri incogniti del modello dovrebbe avvenire attraverso i minimi quadrati generalizzati (GLS). Per l'applicazione dei GLS, viene stimata, attraverso un'opportuna procedura econometrica, la matrice di varianza-covarianza del termine di errore u_{ch} (Greene 2000).

Il risultato finale della *Fase 1* del *Poverty Mapping*, da utilizzare come input della *Fase 2*, è costituito dal vettore dei parametri stimati e dalla

relativa matrice di varianza-covarianza. Per giungere a tale risultato, è necessario superare diverse fasi intermedie che descriviamo molto sommariamente, mettendo in evidenza che il protocollo seguito è stato lo stesso per i tre contesti territoriali e che i risultati cui si perviene sono analoghi.

In primo luogo è stata messa a punto una procedura per stabilire se è opportuno o meno prevedere l'utilizzo dei pesi nella fase di stima dei modelli. Il *test* applicato (Deaton 1997) prevede di specificare come ipotesi nulla l'ipotesi secondo cui gli stimatori dei parametri ottenuti utilizzando o meno i pesi siano entrambi non distorti. Per tutti i modelli, applicando la procedura di verifica di ipotesi, si giunge a rifiutare l'ipotesi nulla e quindi si procede utilizzando il sistema dei pesi.

Una volta giunti alla specificazione definitiva dei modelli è stata verificata l'ipotesi di omoschedasticità del termine di errore (White 1980); per tutti i modelli si giunge alla conclusione di non rifiutare l'ipotesi nulla di omoschedasticità, per cui non è stato necessario specificare un ulteriore modello per il termine di errore. Inoltre per ognuno dei modelli introdotti nell'analisi è stata verificata la significatività dell'effetto gruppo utilizzando un *test* di tipo moltiplicatore di Lagrange per gli effetti casuali (Breusch e Pagan 1980), in cui l'ipotesi nulla da verificare è $\text{Var}(\eta_c) = 0$. Per i diversi modelli specificati tale ipotesi è stata accettata; in effetti il rapporto tra la variabilità spiegata dall'effetto gruppo e la variabilità totale è risultato inferiore a 0,05.

Le stime dei parametri ottenute nella fase precedente vengono utilizzate in questa fase per simulare il reddito di tutte le unità della popolazione di riferimento (unità censuarie). Il reddito simulato ad ogni replicazione è espresso dalla seguente relazione:

$$\ln \tilde{y}_{ch} = \mathbf{x}_{ch}' \tilde{\beta} + \tilde{\eta}_c + \tilde{\varepsilon}_{ch} \quad [2]$$

Le simulazioni vengono ripetute per un certo numero di replicazioni. Ad ogni replicazione, i parametri vengono estratti casualmente dalla corrispondente distribuzione stimata nella *Fase 1*. Più in dettaglio: i coefficienti $\tilde{\beta}$, sono estratti da una distribuzione Normale multivariata con media $\hat{\beta}$ (il vettore dei coefficienti della stima OLS) e matrice di varianza covarianza uguale a quella associata ai coefficienti $\hat{\beta}$. Per

quanto concerne la simulazione dei residui, una volta verificata la non significatività dell'effetto di gruppo, l'interesse si sposta esclusivamente sulla simulazione del residuo individuale $e_{c,h}$. Per tale simulazione sono state evitate assunzioni parametriche ed i residui $\tilde{\varepsilon}_{c,h}$ si generano direttamente dalla distribuzione dei residui stimati.

L'insieme dei vettori simulati $\tilde{y}_{c,h}$ ⁸ viene utilizzato per il calcolo del valore atteso delle diverse misure di povertà e disuguaglianza che si vogliono calcolare e congiuntamente per la determinazione della deviazione standard per ogni singola misura considerata.

⁸ Per il calcolo delle misure di povertà e disuguaglianza si trasforma il reddito familiare simulato in reddito equivalente. A tal fine si utilizza la scala OCSE modificata che, per il calcolo del numero di adulti equivalenti, utilizza un coefficiente pari a 1 per il capofamiglia, 0,5 per gli altri componenti con 14 anni e più e 0,3 per i soggetti con meno di 14 anni.