

**STIMA DELLA SERIE TEMPORALE 2007-2012  
DEI PARAMETRI DI POVERTÀ E  
DEPRIVAZIONE PER I DISTRETTI SOCIO-  
SANITARI E PER LE PROVINCE DELLA  
REGIONE EMILIA-ROMAGNA**

effettuata sulla base delle informazioni  
raccolte nelle indagini EU-SILC 2008-2013,  
mediante l'impiego di tecniche  
di Stima per Piccole Aree

**Maria Rosaria Ferrante\*, Enrico Fabrizi^, Silvia Pacei\***

\*Dipartimento di Statistica "P. Fortunati", Università di Bologna  
^DISES, Università Cattolica del S. Cuore, Piacenza

Dicembre 2015

# Indice

1. Introduzione	p. 4
2. Le misure di povertà nell'Unione Europea	p. 5
3. Misure di povertà basate sul reddito e misure di povertà basate sui consumi	p. 7
4. La famiglia (household) come unità di indagine	p. 9
5. Gli indicatori di povertà e disuguaglianza stimati, il problema delle soglie di povertà	p. 11
6. I domini di studio	p. 14
7. La base dei dati. L'indagine EU-SILC	p. 17
8. La strategia di stima adottata. Le stime dirette e la valutazione della loro affidabilità	p. 18
9. Le stime basate sui modelli per "piccole aree"	p. 21
10. Analisi dei risultati	p. 23

## **Appendice A - L'indagine EU-SILC, la produzione delle stime dirette e della loro varianza**

p. 49

A1. Indagine EU-SILC. Caratteristiche generali	p. 50
A1.1 Popolazione obiettivo	p. 51
A1.2 Periodo di rilevazione e periodo di riferimento	p. 51
A1.3 Piano di campionamento	p. 51
A1.4 Pesi campionari	p. 54
A1.5 Domini pianificati e domini di interesse	p. 58
A2. Reddito equivalente. Suo trattamento statistico	p. 60
A2.1 Scala di equivalenza e reddito disponibile equivalente	p. 62
A3. Definizione degli indicatori e formule utilizzate nella produzione delle stime dirette	p. 64
A3.1 Misure di povertà basate sul reddito equivalente	p. 65
A3.2 Tasso di deprivazione materiale	p. 66

A3.3	Formule stimatori diretti	p. 67
A4.	Stima della varianza	p. 69
A4.1	Discussione del bootstrap utilizzato e suoi limiti nel contesto attuale (impossibilità di utilizzare il comune come PSU)	p. 69
A4.2	Stima della varianza. Bootstrap approssimato	p. 70
A4.3	Stima della varianza. Effetto del disegno di Kish	p. 73
<b>Appendice B - Produzione delle stime da modello e valutazione della loro affidabilità</b>		p. 81
B1.	Modelli per piccole aree. Introduzione	p. 82
B2.	Inclusione delle provincie delle rimanenti regioni italiane nel processo di stima	p. 83
B3.	Analisi e selezione dell'informazione ausiliaria	p. 84
B.4	I modelli di stima per piccole aree	p. 86
B4.1	Modello per i tassi	p. 86
B4.2	Modello per l'indice di Gini	p. 88
B4.3	Lo "smoothing" della varianza dell'indice di Gini	p. 89
B4.4	Modello per la media del logaritmo del reddito	p. 91
B4.5	Modello per il gap mediano	p. 92
B4.6	Specificazione delle distribuzioni a priori per gli iperparametri	p. 95
B5.	Selezione variabili ausiliarie	p. 96
B6.	Simulazione delle distribuzioni a posteriori	p. 100
B7.	La selezione dei modelli.	p. 101
B8	Le stime da modello ed i relativi CV per i distretti dell'Emilia Romagna	p. 102
<b>Riferimenti bibliografici</b>		p. 112

## 1. Introduzione

L'obiettivo di questa ricerca è la stima di parametri di povertà per i distretti socio-sanitari e per le province dell'Emilia-Romagna, con riferimento agli anni 2007-2012. Al fine di ottenere un quadro completo della condizione delle famiglie in Emilia-Romagna, sono stati considerati parametri in grado di cogliere diversi aspetti del fenomeno, ossia: una misura della diffusione della povertà monetaria, il tasso di povertà; una misura della sua intensità, il "poverty gap" mediano; una misura delle disuguaglianze, l'indice di Gini; e una misura non monetaria della povertà, il tasso di deprivazione materiale. Come dominio oggetto di stima sono stati scelti i distretti, per il ruolo che hanno assunto negli ultimi anni nell'ambito delle politiche regionali volte alla promozione dell'integrazione sociale e sanitaria, e le province. Le stime sono state ottenute utilizzando i dati delle indagini EU-SILC 2008-2013 (reddito 2007-2012) mediante metodi di stima "per piccole aree", poiché il numero delle unità campionate sia nei distretti sia nelle province è troppo piccolo per ottenere stime affidabili con i metodi di stima che l'Istituto Nazionale di Statistica (ISTAT) utilizza correntemente per stimare questi parametri.

Prima di entrare nel merito dei metodi impiegati e dei risultati ottenuti con questo processo di stima, è necessario affrontare alcune questioni preliminari relative alla definizione di povertà ed alla sua misura. Qualsiasi definizione di povertà fa riferimento alla mancanza di qualche cosa, ad una situazione di privazione, ad una limitazione nelle possibilità che l'individuo ha di raggiungere benessere e felicità. Anche restringendo il campo alla povertà economica, cioè ad una situazione di mancanza dei mezzi materiali necessari a raggiungere un livello adeguato di benessere, giungere ad una definizione traducibile in misure empiriche concrete, richiede che vengano effettuate una serie di scelte di fondo.

In primo luogo occorre scegliere uno "spazio valutativo" della povertà, ossia quale variabile o insieme di variabili utilizzare per misurare la deprivazione. Nella letteratura economica, si considerano generalmente due alternative: il consumo o il reddito. In secondo luogo bisogna identificare i poveri. Ovvero, una volta noti i valori della variabile nello spazio valutativo, stabilire chi è povero e chi non lo è, e eventualmente valutare il grado di povertà di coloro che sono stati definiti poveri.

Esistono due tipologie di impostazione per la definizione dello stato di povertà. Si può assumere una "soglia di povertà" che divida in modo netto la popolazione tra poveri e non poveri, oppure rifiutare questa demarcazione e cercare di valutare un grado di povertà che caratterizzi in modo diverso gli individui. Se si adotta un'impostazione basata su una "soglia di povertà", questa va definita e per farlo occorre decidere, prima di tutto, se debba essere assoluta, ossia esprimere un livello minimo di risorse necessarie per soddisfare un insieme minimale di necessità, oppure relativa, strettamente legata al contesto sociale e alle risorse disponibili complessivamente nella società.

Occorre infine stabilire quale debba essere l'unità di analisi, ossia scegliere se riferire le misure delle variabili nello spazio valutativo alle famiglie o direttamente agli individui. Qualora si opti per la famiglia, è evidente che l'insieme delle risorse a disposizione della famiglia si traduce in livelli di benessere diversi per i suoi membri a seconda della dimensione e della composizione per età del nucleo: occorre, in altre parole, far ricorso a scale di equivalenza.

Si tratta di un insieme di questioni di notevole portata, oggetto di dibattito economico, ed in alcuni casi anche filosofico, e per cui non esistono soluzioni universalmente accettate. Una loro trattazione approfondita è ovviamente ben al di là degli obiettivi che ci poniamo in questo lavoro.

Le scelte che noi abbiamo adottato hanno una giustificazione essenzialmente politica ed istituzionale. Si tratta di scelte coerenti con quelle prese in materia di misurazione della povertà dall'Unione Europea. Ci limiteremo quindi a discutere brevemente le motivazioni scientifiche alla base delle scelte adottate a livello comunitario.

## **2. Le misure di povertà nell'Unione Europea**

Una strategia unitaria di misurazione della povertà a livello di Unione Europea si è andata delineando in modo via via più chiaro a partire dall'inizio degli anni '90. Il progetto del panel europeo delle famiglie (European Community Household Panel, ECHP) ha rappresentato un primo tentativo di misurare il reddito disponibile delle famiglie secondo le stesse definizioni e con analoghi metodi di indagine in differenti paesi dell'Unione. Ciò ha consentito di costruire le prime misure di povertà comparabili in senso stretto tra paesi diversi. Misure basate sul reddito come variabile sulla base della quale effettuare la valutazione dello stato di benessere / deprivazione. L'ECHP è stato operativo tra il 1994 e 2001, e nel 2003 è stato sostituito, sulla base del Regolamento (EC) No 1177/2003, dal Progetto EU-SILC (European Union Statistics on Income and Living Conditions) che riprende gli obiettivi della precedente indagine (la misurazione secondo metodi omogenei in tutti i paesi dell'Unione) e li amplia notevolmente arricchendo l'insieme delle variabili misurate. Non solo il reddito, ma anche misure materiali di deprivazione, variabili relative alla condizione sul mercato del lavoro, l'istruzione, l'inclusione sociale. EU-SILC prevede un campione notevolmente più grande rispetto a ECHP e comporta quindi, da parte dei paesi membri, un maggiore sforzo organizzativo.

Il passaggio da un'indagine all'altra riflette precisamente il processo di evoluzione politica e istituzionale avvenuto a partire dal 2000 circa, che possiamo riassumere schematicamente nelle seguenti tappe:

- Marzo 2000, Consiglio di Lisbona: viene posto l'obiettivo di rendere l'economia europea più efficiente, moderna e produttiva facendo leva sulla conoscenza. Gli obiettivi dichiarati di questo tentativo di rilancio competitivo è la crescita sostenibile e una maggiore coesione all'interno dell'Unione, tra i vari stati, le varie regioni, le varie componenti della società;
- Dicembre 2000, Consiglio Europeo di Nizza: individuazione di 4 obiettivi per sostenere l'inclusione sociale: 1) Favorire l'incremento della partecipazione al mercato del lavoro e l'accesso a risorse, diritti, beni e servizi da parte di tutti i cittadini dell'Unione; 2) Prevenire il rischio di esclusione sociale ; 3) Aiutare i più vulnerabili; 4) Mobilitare tutte le più importanti istituzioni nel tentativo di ridurre le disuguaglianze nella distribuzione del benessere all'interno dell'Unione.
- Marzo 2001, Consiglio europeo di Stoccolma: viene dato mandato ad un apposito comitato di individuare un insieme di indicatori sull'inclusione sociale, al fine di migliorare il monitoraggio sulla povertà e sull'esclusione.
- Dicembre 2001, Consiglio Europeo di Laeken: approvazione degli indicatori sviluppati dal Comitato. Si tratta di un insieme molto articolato di indicatori che riguardano quattro grandi aree differenti: a) povertà; b) occupazione e mercato del lavoro; c) salute; d) istruzione.

La misura della povertà è pensata secondo due linee principali. Da una parte una batteria di indicatori di "povertà monetaria" basati sul reddito disponibile familiare come variabile obiettivo per la valutazione delle risorse disponibili. Dall'altra una serie di indicatori che cercano di mettere in luce gli aspetti materiali della deprivazione, tentando di misurare quindi in modo diretto, l'impossibilità degli individui di raggiungere i propri obiettivi ("funzionamenti") a causa dello stato di ristrettezze in cui vivono.

L'indagine EU-SILC è la base di dati, armonizzata a livello di tutti i paesi dell'Unione Europea, sulla base della quale gli indicatori previsti dal Consiglio di Laeken possono essere stimati.

In questa ricerca, ci occupiamo di stimare alcuni indicatori di "povertà monetaria" previsti dal Consiglio di Laeken. Come anticipato, si tratta di tre indicatori che utilizzano il reddito come variabile di valutazione delle risorse disponibili e di un indice che si basa invece sui segnali di deprivazione materiale espressi dalla famiglia (i quattro indicatori sono definiti dettagliatamente nell'Appendice A3). Tali indici adottano la famiglia come unità di indagine. La soglia di povertà è relativa nel caso degli indicatori basati sul reddito, assoluta nel caso di quello di deprivazione materiale. Sono scelte che, sotto un profilo scientifico possono essere discusse e criticate e che, almeno in parte, segnano una frattura con la tradizione italiana di misurazione della povertà incentrata fino a qualche anno fa prevalentemente sui consumi, anche per la carenza di informazioni attendibili sul reddito. Le faremo nostre

nell'ottica di ottenere misure comparabili con quelle del resto dell'Unione, al fine di contestualizzare le nostre regioni nello scenario europeo.

### **3. Misure di povertà basate sul reddito e misure di povertà basate sui consumi**

Nella tradizione degli studi economici le variabili più comunemente utilizzate nello studio della povertà monetaria sono il reddito e i consumi. Le ragioni a favore dell'impiego del reddito possono essere molto schematicamente ricondotte a queste:

- 1) il reddito è un indicatore ex-ante di disponibilità / potere di comando sulle risorse e, in quanto tale, prescinde dalle effettive decisioni di consumo, che risentono degli “stili di vita”;
- 2) le politiche redistributive sono in larga misura basate sul reddito;
- 3) le diseguaglianze tra diversi paesi sono generalmente misurate in termini di reddito, ancorché si tratti di Prodotto Interno Lordo e quindi non dello stesso aggregato a cui fanno riferimento le indagini quali EU-SILC.

Le ragioni a favore dell'impiego dei consumi come spazio valutativo per la misurazione della povertà possono, sempre schematicamente essere ricondotte a queste:

- 1) è il miglior indicatore sintetico dell'utilità;
- 2) è strettamente legata al concetto di ‘reddito permanente’, che rappresenta una “vera” misura del benessere poiché non è influenzata dalle variazioni transitorie.

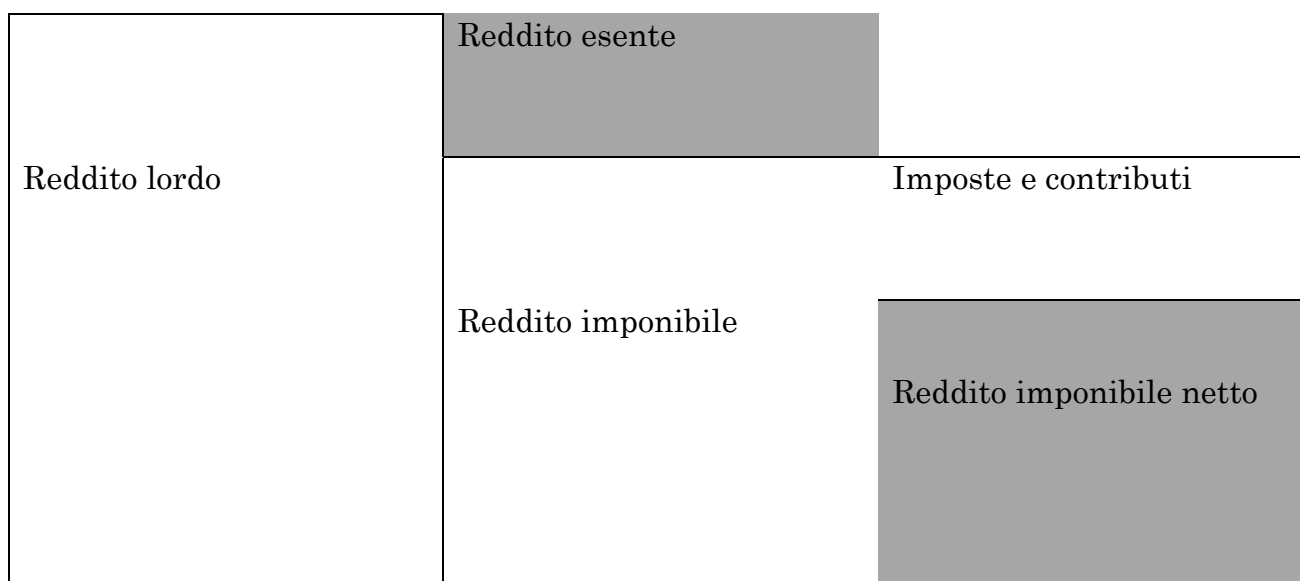
Notiamo che in Italia, le misure di povertà prodotte da ISTAT, fino ad EU-SILC, sono state esclusivamente basate sulla variabile Consumi. Misure di povertà basate su variabili diverse e, lo vedremo, anche su soglie di concezione diversa non sono in alcun modo confrontabili e possono assumere valori anche molto lontani tra loro.

Riguardo alle ragioni riportate a favore dell'impiego del reddito oppure dei consumi, non ci addentriamo nel complesso dibattito economico legato al loro confronto. Notiamo banalmente che in sede UE hanno prevalso quelle a favore del reddito e che i parametri cui facciamo riferimento nel presente lavoro sono quelli stimati da EUROSTAT ad un livello di minor dettaglio territoriale. Aggiungiamo soltanto che a livello teorico generale le misure di povertà basate sul reddito possono essere inquadrare nei tentativi di definizione della povertà nello spazio delle ‘capabilities’ (Sen, 1983), ossia delle possibilità, delle risorse che gli individui hanno effettivamente a disposizione per raggiungere i loro obiettivi in termini di utilità economica. Come vedremo questo assunto teorico è strettamente legato alla scelta di soglie di povertà

relative, in virtù del fatto che una delle ‘capability’ fondamentali è proprio quella di non essere esclusi dalla società in cui si vive.

## Il reddito disponibile

La definizione di reddito a cui fanno riferimento gli indicatori presi in considerazione in questa ricerca è quella di reddito disponibile. Se facciamo riferimento ad aggregati di rilevanza fiscale la definizione di reddito disponibile può essere desunta dallo schema seguente



Il reddito disponibile è dato dalla somma del reddito imponibile netto e del reddito esente. Si tratta di un aggregato che cerca di approssimare al meglio il concetto di risorse disponibili per l'individuo o per la famiglia. Una sua definizione dettagliata di reddito disponibile, quella adottata dall'indagine EU-SILC secondo le linee guida del documento di Camberra è fornita nell'Appendice A2 di questo rapporto.

Vale la pena di ricordare che la misurazione del reddito presso le famiglie, attraverso lo strumento di un questionario, si presenta particolarmente difficile. Vi sono molte fonti di possibili errori per il rispondente, quali dimenticanze, ricordi distorti ed errori di calcolo. Oltre, ovviamente, all'under-reporting deliberato. Per limitare al minimo questi errori l'ISTAT effettua le interviste in ottobre, dopo la consegna delle Dichiarazioni dei Redditi, in modo che i rispondenti possano aiutarsi con queste nel processo di risposta. Inoltre i dati forniti dai rispondenti, sono confrontati con le loro dichiarazioni fiscali rese disponibili dal Ministero delle Finanze.

Alcune componenti del reddito disponibile si presentano tuttavia di difficile valutazione, soprattutto i fitti figurativi che devono essere imputati tra le componenti



positive di reddito per i proprietari di casa e gli affittuari i cui canoni di affitto sono al di sotto dei prezzi di mercato.

Per maggiori dettagli sulla misurazione del reddito e sui meccanismi di correzione delle risposte e di imputazione delle mancate risposte parziali si veda ISTAT (2008).

Notiamo infine che il reddito disponibile totale, stimabile in base all'indagine campionaria EU-SILC non è perfettamente comparabile con il reddito disponibile aggregato del settore Famiglie, che è l'aggregato di Contabilità Nazionale ad esso più prossimo logicamente. L'indagine campionaria EU-SILC riguarda le famiglie residenti registrate nelle Anagrafi comunali, mentre il settore Famiglie di Contabilità nazionale include tutte le famiglie presenti da più di un anno sul territorio nazionale (quindi, per esempio, anche gli immigrati irregolari). In secondo luogo, il reddito disponibile delle famiglie di Contabilità nazionale include una stima dell'economia 'sommersa' che, per ovvie ragioni, non è possibile rilevare attraverso un'indagine campionaria condotta presso le famiglie.

#### **4. La famiglia (household) come unità di indagine**

L'unità di analisi adottata nell'analisi della povertà è la famiglia. Si assume che le risorse disponibili all'interno della famiglia vengano messe in comune e utilizzate per soddisfare i bisogni e incrementare le 'capabilities' di tutti i suoi membri in ugual misura.

La definizione di famiglia adottata è la stessa che ISTAT adotta nelle indagini di argomento sociale: un insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o da vincoli affettivi, coabitanti o aventi dimora abituale nello stesso comune. Si tratta quindi della 'famiglia di fatto' o volendo utilizzare la terminologia inglese della 'household'.

Considerando come unità di analisi la famiglia, il reddito disponibile rilevante per la definizione e lo studio della povertà è quello "complessivo familiare", ossia il reddito disponibile di tutti i membri della famiglia, che si suppone venga messo in comune per il benessere dei membri. L'assunzione implicita è che la ripartizione delle risorse tra i membri della famiglia sia egualitaria. Si tratta di un'ipotesi piuttosto forte, che può avere l'affetto di portare a sottostime anche rilevanti del livello di disuguaglianza nella distribuzione delle risorse all'interno della popolazione, qualora l'egualitarismo all'interno delle famiglie venga meno.

## Scale di equivalenza

Il reddito complessivo familiare non consente di confrontare il livello di risorse individuali effettivamente disponibile per membri di famiglie aventi dimensione e composizione differenti. Una semplice divisione per il numero di componenti (reddito pro-capite) non terrebbe conto delle economie di scala che la messa in comune delle risorse comporta. Poiché queste economie di scala dipendono oltre che dalla dimensione anche dalla composizione del nucleo, una metodologia molto utilizzata per “standardizzare” i nuclei familiari consiste nel calcolarne una “dimensione equivalente” secondo un’opportuna scala. Ciò comporta la divisione del reddito disponibile complessivo familiare per questo coefficiente e l’attribuzione del reddito equivalente risultante a ciascun membro della famiglia in ugual misura.

In questa ricerca è stata adottata la scala di equivalenza OCSE modificata, adottata da Eurostat e consigliata nelle linee guida di analisi di EU-SILC. Questa scala è utilizzata anche dalla Banca d’Italia (Banca d’Italia, 2006). Essa si basa sulla seguente definizione di dimensione equivalente:

$$EQ_{SIZE} = 1 + 0.5(CF_{14+} - 1) + 0.3CF_{13-}$$

Dove  $CF_{14+}$  è il numero di membri della famiglia di età maggiore o uguale a 14 mentre  $CF_{13-}$  è il numero dei rimanenti.

Va sottolineato, come questa scala, rappresenti da un punto di vista scientifico, una scelta tra diverse possibilità e possa essere quindi oggetto di critica. Una scala alternativa, molto popolare in Italia, per la sua applicazione alle misure della povertà basate sull’indagine sui Consumi delle Famiglie condotta da ISTAT è la “scala Carbonaro” (Carbonaro, 1985). Va notato tuttavia, come questa scala sia funzione monotona (crescente a tassi decrescenti) del numero di membri della famiglia ma non consideri la composizione per età. Nello stesso spirito, alcune pubblicazioni OCSE, con riferimento principalmente ai confronti internazionali, fanno riferimento ad una dimensione equivalente definita semplicemente come radice quadrata del numero di membri.

Alcuni riferimenti alla letteratura internazionale sul problema delle scale di equivalenza possono essere Atkinson (1998), Atkinson et al. (1995). Aaberge e Melby (2005) discutono invece più accuratamente il problema della sensitività delle stime dei parametri della distribuzione del reddito alla scelta delle scale di equivalenza.

## **5. Gli indicatori di povertà e disuguaglianza stimati, il problema delle soglie di povertà**

La variabile utilizzata per descrivere le risorse disponibili a livello individuale è data dal “reddito disponibile equivalente” calcolato per la famiglia di cui ciascun individuo è membro. Tutti i membri della stessa famiglia hanno lo stesso valore del reddito equivalente.

In questo paragrafo descriviamo i parametri, scelti tra quelli previsti dal protocollo del Consiglio di Laeken, che formano l’oggetto di questa ricerca. La maggior parte di essi fa riferimento alla definizione di una soglia di povertà che permetta di classificare gli individui in due insiemi disgiunti di “poveri” e “non-poveri”. Prima di illustrare le definizioni degli indicatori affrontiamo brevemente il problema della definizione delle soglie di povertà.

### **La definizione delle soglie di povertà**

La scelta della soglia coinvolge inevitabili margini di arbitrarietà ed ha carattere essenzialmente etico-politico. I principali standard che guidano tale scelta possono essere classificati come segue: standard assoluti; standard relativi; standard soggettivi (Carbonaro, 2002). Nel caso degli standard assoluti le soglie vengono costruite a partire da una stima dei fabbisogni alimentari e abitativi della famiglia. Il livello di risorse necessario per soddisfare questi bisogni viene poi moltiplicato per un coefficiente atto a tenere in conto di tutti gli altri bisogni.

Soglie assolute di povertà vengono correntemente utilizzate negli Stati Uniti, anche in sede di misurazione “ufficiale” della povertà. L’Istituto Nazionale di Statistica Italiano ha avviato un programma per la misurazione della povertà basata su una soglia assoluta all’inizio di questo decennio (ISTAT, 2004), sempre utilizzando i dati provenienti dall’indagine sui consumi. La povertà viene intesa in questo caso come mancato soddisfacimento dei bisogni primari. Gli aspetti più rilevanti delle soglie assolute sono i seguenti: *i*) non sono influenzate dalla distribuzione dei redditi; *ii*) non variano nel tempo. In realtà entrambe queste caratteristiche sono, di fatto, di breve-medio periodo. Nel lungo infatti il livello dei consumi considerato “minimo” tende a seguire la dinamica del reddito e quindi ad essere revisionato nel tempo. Si veda ad esempio Citro e Michaels (1995) per la realtà degli Stati Uniti. Tra revisioni successive, la soglia rimane costante, eccettuata la rivalutazione necessaria a neutralizzare l’effetto dell’inflazione.

Nella tradizione statistica europea le soglie relative sono solitamente preferite a quelle assolute. Queste soglie sono definite come una frazione di un indice di tendenza centrale che sintetizzi la distribuzione dei redditi (o dei consumi). Ad esempio la soglia consigliata da Eurostat per l’analisi dei dati EU-SILC, e che noi adotteremo, è fissata

al 60% del reddito mediano disponibile equivalente, mentre la soglia su cui sono basate le stime ISTAT del tasso di povertà per l'Italia è definita come il 50% del livello medio aritmetico dei consumi. Si tratta di soglie che dipendono direttamente da una misura di tendenza centrale della distribuzione dei redditi (consumi) nella popolazione. Inoltre, sono soggette a variazione di anno in anno.

Alcuni dei vantaggi legati all'adozione di soglie relative possono essere sintetizzati nei punti seguenti.

1) A livello di principio, la definizione di “povero” viene collegata al livello generale di risorse disponibili nella società. Si identifica come povero colui che ha molto poco rispetto agli altri membri della società in cui vive, indipendentemente dal livello assoluto di risorse di cui dispone.

2) Tecnicamente sono più semplici da definire e quindi meno “arbitrarie” delle soglie assolute. Non occorre infatti definire un paniere minimo di bisogni da soddisfare, né quantificare le risorse necessarie per soddisfarlo. Il problema della rivalutazione annuale della soglia non deve essere affrontato. Ciò evita la scelta – o la costruzione – di un opportuno tasso di inflazione (quello calcolato per la popolazione generale potrebbe essere un'approssimazione piuttosto povera del tasso “sperimentato” dalla sottopopolazione dei poveri).

Un possibile svantaggio è legato al suo impiego in politica economica, ad esempio nella valutazione di politiche generali di contrasto della povertà. La soglia varia di anno e questo rende difficile misurare se gli interventi hanno sortito un effetto oppure no. Notiamo inoltre che la soglia di povertà relativa è in qualche misura “ciclica” in periodi di recessione tenderà ad abbassarsi e con essa il numero dei poveri (a parità di tutto il resto); l'opposto accade in periodi di espansione. Variazioni nel tempo della linea riflettono la crescita reale dell'economia e la dinamica dei prezzi (Bellù e Liberati, 2005).

### **Gli indicatori considerati nell'analisi**

Gli indicatori di povertà considerati in questa analisi sono:

- 1) il Tasso di Povertà basato sulla soglia di povertà nazionale pari al 60% del reddito disponibile equivalente mediano (denominato anche Head Count Ratio o At-Risk-of-Poverty Rate);
- 2) il Poverty Gap Mediano basato sulla soglia di povertà nazionale;
- 3) l'Indice di Concentrazione di Gini del reddito disponibile equivalente;
- 4) il Tasso di Deprivazione Materiale, calcolato come la quota di persone che vivono in famiglie che si trovano in uno stato di deprivazione materiale.

Una definizione dettagliata del calcolo di questi parametri è riportata nell'Appendice A3.

I tassi di povertà sono le misure più semplici e diffuse di prevalenza della povertà, la cui interpretazione non richiede particolari considerazioni preventive se non quelle relative alla definizione della soglia su cui sono basati. E' tuttavia opinione unanime in letteratura, che i tassi di povertà non possano essere considerati come unici indicatori della povertà di una popolazione. Essi infatti non godono di alcune proprietà importanti a livello teorico:

- non soddisfano il principio di monotonicità, secondo cui un indice di povertà deve crescere se il reddito anche di un solo individuo povero si riduce;
- non soddisfano il principio dei trasferimenti di Pigou, principio che richiederebbe un incremento dell'indice di povertà qualora si verifici un trasferimento di reddito da un individuo ad un altro con reddito superiore al suo.

In parole povere, sono indici che descrivono la diffusione della povertà senza dire nulla della sua "gravità", ossia non misurano quanto sono poveri gli individui che vivono al di sotto della soglia fissata come discriminante. Per questa ragione ai tassi di povertà si accompagnano comunemente i poverty gap che sintetizzano il divario esistente tra il reddito di ciascun povero e la soglia di povertà. In particolare in questo studio consideriamo il gap mediano, definito come mediana dei gap individuali dati dalla differenza tra la soglia di povertà ed il reddito disponibile equivalente degli individui poveri. Per facilità di interpretazione questo gap mediano viene quindi diviso per la soglia di povertà in modo da ottenere un indicatore variabile tra 0 e 1. Valori elevati indicano una situazione grave in termini di povertà. Idealmente infatti, il valore 1 viene raggiunto quando almeno la metà dei poveri dispongono di un reddito esattamente nullo, mentre un valore 0 significa che almeno la metà dei poveri ha un reddito disponibile equivalente esattamente uguale alla soglia di povertà.

L'indice di concentrazione di Gini è la misura di disuguaglianza più utilizzata nella letteratura sulla distribuzione dei redditi in virtù delle sue buone proprietà formali, infatti:

- i) varia tra 0 (perfetta equidistribuzione del reddito totale) e 1 (massima disuguaglianza);
- ii) è facilmente interpretabile alla luce della rappresentazione grafica della distribuzione dei redditi nota come "curva di Lorenz";
- iii) è sensibile a cambiamenti nella distribuzione del reddito, indipendentemente dal fatto che essi avvengano nella parte superiore, in quella centrale o nella parte inferiore della distribuzione;
- iv) è sensibile a qualsiasi trasferimento concentrativo o perequativo del reddito.

Va tuttavia notato come l'indice di concentrazione calcolato per una certa popolazione (ad esempio l'intero paese) non possa essere espresso banalmente come funzione dello stesso indice calcolato per sottopopolazioni (ad esempio le regioni). L'indice nazionale infatti può essere ottenuto considerando la media degli indici regionali, una componente "tra" le medie delle varie regioni e un terzo addendo che di interazione tra i primi due.

Infine, la deprivazione può essere definita come uno stato di svantaggio dimostrabile e osservabile, relativo alla comunità locale o ad una più ampia parte della società o nazione a cui un individuo, una famiglia o uno gruppo appartengono. In questo senso può essere distinto dall'idea di povertà. La deprivazione materiale poi è definita come una situazione di involontaria incapacità di sostenere spese per determinati beni o servizi. Segnali di uno stato di deprivazione possono considerarsi ad esempio il non potersi permettere: i) di pagare l'affitto o le bollette, ii) di riscaldare adeguatamente la propria casa, iii) di far fronte a spese impreviste, iv) di mangiare ogni due giorni carne, pesce o cibi di tenore proteico equivalente, v) di trascorrere una settimana di vacanza una volta l'anno, vi) un'automobile, vii) una lavatrice, viii) un televisore a colori, o ix) un telefono. Questo indicatore fa quindi riferimento alle condizioni materiali dell'esistenza che influiscono sullo stato di svantaggio degli individui.

## **6. I domini di studio**

### **I distretti**

Come già anticipato nel paragrafo introduttivo, la scelta dei distretti sanitari dell'Emilia-Romagna come ambito territoriale di riferimento per le stime si deve al ruolo che i distretti hanno assunto nella definizione delle politiche regionali per l'integrazione sociale e sanitaria. Il distretto si è venuto a configurare come una macro-struttura aziendale fortemente legata alla comunità locale, che si pone fra l'azienda sanitaria, spesso ampia e lontana dai singoli cittadini, e gli enti locali, per realizzare i programmi per l'integrazione socio-sanitaria (Balduzzi e Moirano, 2011).

Tale ruolo viene stabilito dalla normativa sin dalle riforme degli anni '90 (d.lgs. n. 502 del 1992 come modificato dal d.lgs. n. 517 del 1993) con le quali il distretto diviene centro di governance di una complessa rete di servizi territoriali. Con la riforma sanitaria del 1999 (d. lgs. N. 229/1999) poi, una serie di disposizioni stabiliscono i principi volti a potenziare il distretto, delineandone le funzioni, gli aspetti programmatici ed una responsabilità legata agli obiettivi di salute, nel rispetto delle autonomie regionali che sono lasciate libere di disegnare il modello organizzativo meglio confacente alle caratteristiche dei singoli Servizi sanitari regionali (Balduzzi e Moirano, 2011).

Con la riforma del Titolo V, Parte seconda della Costituzione, ad opera della legge costituzionale n. 3 del 2001, l'iscrizione della tutela della salute nel novero delle materie di legislazione corrente (e cioè regionale) e la valorizzazione della potestà amministrativa hanno rafforzato le responsabilità regionali nella definizione e nelle scelte sul distretto (Balduzzi e Moirano, 2011).

L'Agenas (Agenzia nazionale per i servizi sanitari regionali) ha effettuato negli anni 2005-2006 e nel 2010 un'indagine sulle modalità di funzionamento dei distretti, al fine di comprendere meglio l'evoluzione nel tempo del distretto. Nel 2010 l'indagine si è focalizzata sulla raccolta di informazioni su: la funzione del distretto come punto di riferimento per il cittadino e gli strumenti di comunicazione con gli utenti, il coordinamento della rete di servizi e la presa in carico dei problemi sociosanitari complessi, le modalità di integrazione relazione tra i diversi attori dell'assistenza primaria (Balduzzi e Moirano, 2011). Dei 711 distretti italiani, 681 (96%) hanno partecipato all'indagine. I risultati mostrano un sistema in cui generalmente il distretto programma le sue attività assieme agli enti locali ed in cui alla figura del Direttore vengono affidate complesse responsabilità (Balduzzi e Moirano, 2011). Emergono anche elementi di eterogeneità fra i distretti rispetto agli specifici compiti svolti. Non tutti i distretti, ad esempio, partecipano alla definizione e al monitoraggio del Piano di Zona (il documento di programmazione territoriale con il quale comuni e ULSS mettono a punto le politiche sociosanitarie rivolte alla popolazione). Lo fa circa l'84% dei distretti e con modalità diverse. Il profilo professionale e la specifica funzione del Direttore del distretto poi, non sono omogenee nei distretti, anche perché la normativa stessa non definisce un profilo e una qualificazione accademica ideale per il Direttore.

In Emilia-Romagna i distretti attivi nel 2015 sono 38.

La Regione Emilia-Romagna, per far fronte ai mutamenti in atto nella struttura della popolazione e nella composizione delle famiglie, che vedono un progressivo incremento dell'incidenza delle fasce di popolazione di età avanzata e molto avanzata e l'aumento delle condizioni di solitudine, ha istituito (art. 51, L.R. n. 27/2004) il Fondo regionale per la non autosufficienza (Frna). Il programma per l'impiego dell'Frna è basato su linee strategiche del Piano sociale e sanitario 2008-2010, che riguardano: lo sviluppo dell'assistenza domiciliare integrata (sanitaria e sociale), l'aumento dell'assegno di cura per anziano non autosufficienti (con un'attenzione specifica per coloro che non percepiscono l'indennità di accompagnamento), la qualificazione delle assistenze familiari, i servizi di telesoccorso e teleassistenza, i ricoveri temporanei di sollievo, il sostegno alle reti informali di solidarietà sociale (dal "portierato" al "custode sociale") (Curcetti, 2011). Tale programma viene attuato attraverso il piano annuale di attività elaborato dal Comitato di distretto (Sindaci) e dal Direttore del distretto. Le risorse vengono ripartite fra i distretti dell'Azienda USL, in quanto luoghi deputati alla programmazione e al governo dei servizi necessari alla popolazione di riferimento. Il

50% circa delle risorse assegnate attraverso il Frna e utilizzate per la domiciliarità viene utilizzato per finanziare gli “assegni di cura”.

## **Le province**

Il riferimento alle province (NUTS3 nella *Nomenclatura delle unità territoriali statistiche*) è importante nell’ottica di offrire informazioni statistiche territorialmente disaggregate ai policy makers locali, più aderenti rispetto a quelle regionali o addirittura nazionali alle esigenze informative connesse ai processi di governance delle aree provinciali o metropolitana. Le autorità pubbliche locali hanno la possibilità di delineare azioni politiche in risposta ai bisogni economici, sociali e ambientali della comunità locale e le loro scelte possono avere impatti significativi sulle condizioni di contesto sottostanti al benessere sociale del territorio. Per realizzare tali azioni hanno bisogno di informazioni statistiche riferite a “piccole aree” che, se provenienti da indagini campionarie, la statistica ufficiale non rende usualmente disponibili in quanto le indagini ufficiali non sono costruite per fornire stime affidabili di tale dettaglio. Con riferimento a parametri di povertà e disuguaglianza, ad esempio, se si considera la pubblicazione “Il Benessere Equo e Sostenibile delle Province – 2015”, nella sezione degli indicatori relativa al Benessere Economico mancano informazioni riferite a indicatori di rischio di povertà, povertà assoluta, disuguaglianza e deprivazione (stimati a partire dai dati raccolti mediante indagini campionarie quali EU-SILC e Indagine sulle spese delle famiglie) mentre tali indicatori sono presenti nel set che ha come riferimento il livello nazionale (vedi “Il Benessere Equo e Sostenibile in Italia – 2015”). Nello stesso rapporto sul BES delle Province si afferma che “All’assenza di indicatori appartenenti alla dimensione del benessere soggettivo, si accompagna la mancanza nel sistema informativo del Bes delle province di indicatori soggettivi, (derivati cioè da indagini su individui o famiglie) che informino sulle abitudini e stili di vita, sulla percezione dei fenomeni e che permettano analisi distributive sulla popolazione. Tale situazione appare al momento non superabile, visto che sono pochissime le indagini campionarie presso le famiglie in grado di restituire informazioni statisticamente significative al di sotto del livello regionale”. Infine la scelta del riferimento provinciale vuole seguire le indicazioni che OCSE ha fornito nell’ambito dell’iniziativa “Better life”, per promuovere la disseminazione delle *best practices* di utilizzo degli indicatori di benessere a sostegno delle politiche locali, e le linee guida per l’attuazione a livello locale della strategia “Europa 2020 per le città e i territori”.

Le stime da noi prodotte a livello provinciale potrebbero quindi completare ed arricchire il set di indicatori del benessere già prodotto per le province da ISTAT-SISTAN con un insieme di indicatori riferiti al benessere economico locale, caratterizzati da un adeguato grado di affidabilità, che sappiano cogliere le specificità locali e che siano utili per la programmazione locale.

Nel seguito riportiamo alcune caratteristiche delle province dell’Emilia Romagna:



<b>Provincia</b>	<b>Sigla</b>	<b>Popolazione</b>	<b>Comuni</b>
Bologna	BO	991.924	62
Ferrara	FE	359.994	27
Forlì-Cesena	FC	395.489	30
Modena	MO	700.913	47
Parma	PR	442.120	49
Piacenza	PC	289.875	48
Ravenna	RA	392.458	18
Reggio nell'Emilia	RE	530.343	46
Rimini	RN	329.302	29

## **7. La base dei dati. L'indagine EU-SILC**

Le stime che proponiamo sono basate sui dati dell'indagine Europea sul Reddito e le condizioni di vita (European Survey on Income and Living Conditions; EU-SILC). In particolare abbiamo utilizzato i dati delle ondate 2008-2013 dell'indagine (reddito 2007-2012). Di EU-SILC abbiamo già anticipato che si tratta di un progetto comunitario finalizzato alla produzione, secondo metodologie comuni e condivise, di una serie di indicatori relativi alla distribuzione del reddito, alla povertà, all'istruzione, al mercato del lavoro e alle condizioni di vita delle famiglie con la finalità principale di valutare empiricamente i progressi in tema di integrazione e coesione sociale.

In questo paragrafo ci limitiamo a mettere in luce solo alcune caratteristiche tecniche "essenziali" dell'indagine EU-SILC, rimandando una discussione più dettagliata delle caratteristiche del disegno campionario all'appendice A.

- L'indagine viene ripetuta ogni anno a partire dal 2004. Il periodo di rilevazione, almeno in Italia, è settembre, mentre il periodo di riferimento del reddito (rilevato mediante l'indagine) è l'anno solare precedente. Nel caso delle indagini 2008-2013, che consideriamo in questo rapporto, i redditi rilevati sono quindi quelli degli anni solari precedenti agli anni di indagine.
- Il campione dell'indagine ha una struttura a panel ruotato. Ogni anno solo un quarto del campione viene rinnovato; ciascun individuo che entri a far parte del campione viene quindi intervistato per 4 anni consecutivi, sia che rimanga all'interno della stessa famiglia, sia che ne formi una nuova. Questo permette di studiare, accanto alle dinamiche "macro" anche quelle microeconomiche relative alle tematiche coperte dall'indagine, cercando di limitare i problemi di "attrito" e la necessaria limitatezza nel tempo implicata dai "panel puri".

## 8. La strategia di stima adottata. Le stime dirette e la valutazione della loro affidabilità

Le stime pubblicate da ISTAT per l'indagine EU-SILC, così come per tutte le altre indagini condotte dall'Istituto sono basate su stimatori "da disegno". Si tratta di stimatori ottimizzati rispetto alle proprietà probabilistiche relative alla casualità dovuta alla selezione casuale del campione dalla popolazione. Non viene assunto nessun modello probabilistico che spieghi il modo in cui i dati di popolazione relativi ad una certa variabile siano generati. Questa scelta, adottata anche da tutti gli altri maggiori Istituti Nazionali di Statistica, è motivata dal tentativo di ridurre al minimo le ipotesi necessarie per giungere alle stime, in modo tale che queste possano essere viste, dal pubblico generale a cui sono indirizzate, come il più possibile oggettive, o per lo meno "neutrali".

In particolare l'ISTAT utilizza stimatori per calibrazione (Deville e Särndal, 1992). Questi stimatori permettono di sfruttare informazioni ausiliarie disponibili da fonte amministrativa mantenendo per medie e totali la struttura a somma pesata, che rende particolarmente semplice per gli utenti riprodurre autonomamente le stime prodotte dall'istituto.

Se facciamo riferimento al dominio di studio (sottopopolazione)  $d$  lo stimatore del totale  $Y$  di una data variabile  $y$  potrà essere espresso come:

$$\hat{Y}_d = \sum_{i=1}^{n_d} w_i y_i$$

dove  $w_i$ , con  $i = 1, \dots, n_d$ , è un vettore di pesi campionari o fattori di riporto all'universo.  $n_d$  indica il numero di unità campionarie relative al dominio  $d$ . Questi sono definiti come l'inverso delle probabilità di inclusione moltiplicato per un fattore di correzione (calibrazione) che permette di sfruttare l'informazione ausiliaria e correggere, almeno in parte, gli effetti della non-risposta. In gergo questi stimatori vengono detti "diretti" e "dirette" le stime ottenute sulla loro base.

Premesso questo, la strategia generale di stima che proponiamo per i parametri descritti nei paragrafi precedenti è la seguente:

- produrre le stime dirette dei parametri per tutti i domini di interesse;
- valutare l'affidabilità statistica degli stimatori diretti;
- sostituire le stime dirette con stime ottenute con metodologie più complesse ("da modello") per i parametri le cui stime dirette non risultano sufficientemente affidabili.

Tutte e tre le operazioni appena elencate sono piuttosto complesse.

Le stime dirette ottenute applicando i pesi pubblicati da ISTAT insieme ai dati di EU-SILC per l'Italia presenta aspetti problematici. I domini rappresentati dai distretti non sono infatti “domini pianificati” e sommando i pesi relativi alle unità del campione incluse nel dominio non si ottiene la dimensione esatta della popolazione (noto attraverso le anagrafi); inoltre se si considerano sotto-campioni classificando per sesso, età o altre variabili, il loro peso relativi in termini di somma dei pesi può essere diverso da quello reale nella popolazione del distretto. La calibrazione è infatti effettuata a livello regionale, il più dettagliato per cui ISTAT pubblica ufficialmente le stime .

Al fine di ottenere un sistema di pesi che, all'interno di ciascun dominio riporti alla effettiva dimensione della popolazione, il processo di calibrazione dei pesi può essere modificato introducendo informazioni ausiliarie aggiuntive e facendo in modo che la somma dei pesi all'interno di ciascun dominio sommi alla dimensione della sua popolazione così come risulta dalle anagrafi comunali. Nella stessa direzione si può richiedere che i pesi dei sotto-campioni di distretto ottenuti classificando per sesso, classi d'età e nazionalità dei rispondenti sommino alla dimensione che queste sottopopolazioni hanno a livello di distretto. Questa idea, attraente in linea di principio, richiede in pratica che i campioni per cui si ricalibrano i pesi siano grandi. Una ricalibrazione a livello di piccoli domini può porre i pesi “sotto stress”, aumentandone in modo anche cospicuo la variabilità. L'obiettivo di stimatori con una potenziale minore distorsione, viene così ottenuto al prezzo di un forte incremento della variabilità degli stessi, secondo un “effetto disegno” ben noto in letteratura.

Una soluzione di compromesso, considerata in una versione preliminare di questo rapporto, può essere quella di utilizzare le province come livello di calibrazione, al posto dei distretti. Tecnicamente, ciò potrebbe portare ad un sistema di pesi che riporti ad una popolazione più “vicina” a quella del distretto, limitando tuttavia l'effetto di inflazione della varianza dovuto alla calibrazione di (sotto-)campioni numericamente molto piccoli. Un'analisi accurata ha rilevato tuttavia che questa soluzione non è soddisfacente: le differenze di struttura nelle popolazioni sono in genere più forti tra i distretti della stessa provincia che non tra una provincia e l'altra, per cui si finisce con l'ottenere pesi simili per struttura a quelli ricalibrati a livello regionale ma più variabili e che comunque non riportano ai totali di distretto e non rappresentano accuratamente la struttura della loro popolazione.

La decisione finale è stata quindi quella di calcolare le stime dirette utilizzando i pesi “ufficiali” prodotti da ISTAT, calibrati a livello regionale. Per gli stimatori, come i tassi, che hanno una struttura di media sono stati utilizzati stimatori diretti di tipo Hajék

$$\hat{Y}_d = \frac{\sum_{i=1}^{n_d} w_i y_i}{\sum_{i=1}^{n_d} w_i}$$

la cui dipendenza dalla mancata “calibrazione” dei pesi a livello locale è molto ridotta rispetto a stimatori che assumono nota la dimensione del dominio:  $\hat{Y}_d^* = N_d^{-1} \sum_{i=1}^{n_d} w_i y_i$

Infine, la valutazione dell’affidabilità degli stimatori, in altre parole la stima delle loro varianze, non può essere effettuata con metodi elementari: il disegno campionario è complesso e i parametri oggetto di stima sono funzioni non lineari dei dati. Per questo si è ricorso ad un algoritmo bootstrap di stima della varianza degli stimatori diretti, adattando alle caratteristiche peculiari dell’indagine EU-SILC per l’Italia quanto precedentemente noto in letteratura.

I dettagli relativi a queste due operazioni (calibrazione e stima della varianza) sono descritti nei paragrafi A1.4 e A.4 dell’Appendice A.

### **Valutazione dell’affidabilità delle stime dirette**

Il campione EU-SILC per l’Italia è un campione di grandi dimensioni. Le stime dei parametri descrittivi della povertà e della disuguaglianza distributiva del reddito a livello nazionale, ripartizionale, regionale, sono in generale caratterizzate da un livello di affidabilità sufficiente. L’adeguatezza della precisione di queste stime è coerente con l’obiettivo dichiarato di EU-SILC di produrre stime utili per lo studio e il monitoraggio della coesione sociale a livello regionale.

Come anticipato, nel caso di parametri le cui stime fossero sufficientemente affidabili l’analisi può essere condotta sulle stime “dirette”, mentre occorre introdurre stime “da modello” per gli altri.

Alcune misure dell’affidabilità delle stime dirette a livello di distretto sono riportate, relativamente all’anno di indagine 2013, nella Tabella 1. In particolare vengono riportati, per ciascun parametro oggetto di stima, i valori minimo, medio e massimo del coefficiente di variazione ottenuto per distretto. Non è stato possibile calcolare le stime dirette per i sei distretti in cui non si aveva nessuna famiglia campionata. In linea di principio non è possibile calcolare le misure di povertà (e stimare le loro varianze) anche nel caso in cui vi siano famiglie campionate nel distretto, ma nessuna tra loro risulta povera. Ciò non accade per nessuno dei distretti nel 2013, ma è accaduto in altri anni ed è un caso considerato nel processo di modellazione di cui discuteremo. In particolare, algoritmi di liscio delle varianze permettono di

assegnare un ragionevole livello di incertezza anche alle stime il cui coefficiente di variazione stimato sulla base dell'algoritmo diretto risulta uguale a 0.

Il coefficiente di variazione è stato scelto al posto di altre misure di variabilità (l'errore standard) per rendere più agevoli i confronti fra la variabilità delle stime dei diversi parametri. Nel calcolo dei coefficienti di variazioni medi sono stati esclusi i casi in cui esso è uguale a 0. Dalla tabella 1 è facile notare come le stime del tasso di povertà e del tasso di deprivazione materiale siano caratterizzati da coefficienti di variazione molto elevati e superiori alla soglia del 25% adottata da ISTAT, in linea con altri istituti nazionali di statistica, per discriminare tra stime "pubblicabili" e troppo incerte per essere pubblicate.

Il caso dell'indice di concentrazione di Gini appare nettamente migliore e infatti la media calcolata su tutti i distretti è 21.8%. Tuttavia, la soglia del 25% è superato per un numero elevato di distretti, dieci, ragione per cui il complesso delle stime non è affidabile.

Nella tabella 1 non abbiamo riportato nessuna statistica relativa alla variabilità del poverty gap mediano. Poiché si tratta di una mediana (standardizzata) del reddito dei poveri esso è calcolato solo sui poveri campionati all'interno del distretto e quindi su una frazione minoritaria dei campioni – già piccoli – specifici di ogni distretto. I coefficienti di variazione sono molto elevati. Le stime dirette di questo parametro sono inaffidabili anche per un'altra ragione: esse sono infatti negativamente distorte, in modo anche molto considerevole in piccoli campioni. I coefficienti di variazione non sarebbero quindi molto indicativi della variabilità di questo parametro, le cui stime dirette sono sicuramente non affidabili.

*Tabella 1. Alcune statistiche descrittive relative alla distribuzione dei coefficienti di variazione delle stime dirette per distretto*

Statistiche	Tasso di povertà	Indice di concentrazione di Gini	Tasso di deprivazione materiale
Valore minimo	22.1	9.7	20.8
Valore medio	92.4	21.8	89.6
Valore massimo	276.0	58.1	342.1

## **9. Le stime basate sui modelli per "piccole aree"**

Al fine di risolvere il problema dell'inefficienza delle stime dirette di alcuni parametri, dovuta alla scarsa numerosità campionaria all'interno dei distretti e delle province,

abbiamo proceduto sostituendole con stime ottenute con metodi per piccole aree, seguendo un'impostazione "da modello". Con questa espressione gergale si intende che ci siamo allontanati dal paradigma classico dell'inferenza per quantità descrittive delle popolazioni finite (in cui l'unica sorgente di casualità, in linea di principio, è quella del campionamento probabilistico governato da chi conduce l'indagine) per specificare un modello probabilistico descrittivo della realtà. In altre parole, supponiamo che tra i nostri parametri di interesse esista una "relazione strutturale" descritta da una variabile casuale. Lo sfruttamento di questa ipotizzata relazione strutturale permette di migliorare la precisione della stima in un dominio, utilizzando l'informazione proveniente da tutti gli altri domini.

La relazione strutturale tra i parametri di interesse può consistere semplicemente nell'ipotizzare che i parametri siano la realizzazione di una stessa variabile casuale o, tipicamente coinvolgere informazioni ausiliarie, utili per spiegare la variabilità sottostante ai parametri stessi.

In questa sede non vogliamo addentrarci nei dettagli relativi alla specificazione e alla stima dei modelli e al modo in cui tecnicamente le stime dei parametri descrittivi della povertà sono ottenute sulla base di questi modelli (si rinvia per questo all'Appendice B). Ci limiteremo, schematicamente ad evidenziare alcune scelte che caratterizzano il lavoro che abbiamo svolto.

- Tipo di modelli; i modelli per piccole aree possono essere classificati in due grandi famiglie: "area level" e "unit level" a seconda che il modello probabilistico venga specificato al livello delle stime di dominio oppure di unità statistica (si veda Rao, 2003, cap. 5). In questa ricerca sono stati considerati solo modelli di tipo "area level", poiché permettono una più agevole specificazione di distribuzioni statistiche in linea con la diversa natura dei parametri e di sfruttare informazioni statistiche ausiliarie specifiche per area, in mancanza di informazioni ausiliarie specifiche per unità.

- Informazione ausiliaria; è stata considerata informazione ausiliaria contenuta in varie fonti. La disponibilità di informazioni di carattere fiscale, come il reddito imponibile ed il numero delle dichiarazioni dei redditi a livello comunale, ha consentito di calcolare per distretto tassi importanti come l'imponibile pro-capite, il numero delle dichiarazioni per residente di 15 o più anni ed il reddito imponibile dichiarato in media. Altre informazioni considerate a livello di distretto sono di tipo anagrafico, come la quota di stranieri residenti e la densità abitativa. Sono state inoltre considerate altre informazioni disponibili però a livello di provincia (quindi costanti per i distretti appartenenti ad una stessa provincia) come le stime Istat dei tassi di disoccupazione, ottenute dall'indagine sulle forze di lavoro, e informazioni sul valore aggiunto pro-capite, la cui fonte è la Contabilità Nazionale, il risparmio familiare e la propensione al risparmio (risparmio delle famiglie/reddito delle famiglie) la cui fonte è l'Istituto Tagliacarne.

I criteri utilizzati per selezionare questa informazione sono stati, oltre all'efficacia statistica, la possibilità di ottenere queste informazioni su base regolare anno per anno e l'interpretabilità, anche economica dei modelli specificati. In generale le informazioni di origine fiscale si sono rivelate quelle con maggiore capacità esplicativa nei confronti delle misure di povertà considerate. Per maggiori dettagli sulle informazioni ausiliarie utilizzate per ogni parametro oggetto di stima si veda l'Appendice B3.

- L'impostazione seguita nella stima è quella Bayesiana. Le ragioni dietro questa scelta sono molteplici: i) i recenti progressi nei metodi computazionali, in particolare dei metodi Markov Chain Monte Carlo, permettono la specificazione di modelli complessi senza che ciò comporti un aggravio dei problemi di stima. In particolare è agevole specificare e analizzare modelli che non facciano ricorso all'ipotesi di normalità come in questa ricerca; ii) la specificazione e l'analisi di modelli Bayesiani gerarchici permette di tenere adeguatamente in conto tutte le fonti di incertezza riguardanti i dati e quindi di produrre stime di cui è possibile valutare in modo esaustivo l'affidabilità; iii) al di là delle differenze concettuali di fondo, stime "frequentiste" e Bayesiane, finiscono nei casi più semplici per essere sostanzialmente analoghe. Nel caso dell'analisi di modelli relativamente complessi, come quelli considerati in questa ricerca, i predittori puntuali ottenuti come sintesi delle distribuzioni a posteriori, mostrano in generale buone proprietà frequentiste, così come le varianze a posteriori costituiscono buona approssimazione degli errori quadratici medi frequentisti.

## **10. Analisi dei risultati**

In questo paragrafo ci occuperemo della descrizione dei principali risultati ottenuti con le procedure di stima descritte nei paragrafi precedenti e nelle Appendici A e B. Gli aspetti che metteremo in evidenza non esauriscono tuttavia le informazioni contenute nei dati e altri aspetti potrebbero essere colti da analisti diversi. Si preferisce, inoltre, non avanzare possibili interpretazioni relative alle cause delle diverse ed eterogenee situazioni che si delineano sulla base delle stime o delle relazioni che potrebbero essere ravvisate tra indicatori di povertà e disuguaglianza ed altri fenomeni che caratterizzano il sistema economico. Infine, al fine di non appesantire la trattazione non saranno commentati i risultati relativi agli intervalli di confidenza rappresentati nei grafici seguenti dai segmenti sovrapposti alle barre che riferiscono delle stime puntuali. Il lettore può ovviamente valutare per i casi di interesse l'eventuale grado di sovrapposizione degli intervalli.

Presenteremo le stime per provincia e quelle per i distretti socio-sanitari rispettivamente nelle successive due sottosezioni.

## Le stime per provincia

La Tabella 2 contiene le stime da modello dei quattro parametri di interesse (considerate in percentuale), per le province dell'Emilia Romagna e per gli anni 2007-2012, per la regione Emilia Romagna e per l'Italia. La tabella 3 riporta i rispettivi coefficienti di variazione percentuali. Nei grafici 1-4 riportiamo inoltre le stime dei quattro indicatori per l'ultimo anno considerato (il 2012) ordinati per valore della stima riferita alle province, a cui aggiungiamo le stime riferite all'Emilia Romagna nel suo insieme ed all'Italia. Su ognuna delle barre dei grafici è riportato un segmento che individua l'intervallo di confidenza per la stima.

Riportiamo qui un commento riferito alle stime 2012. Infatti, posto che si osservano differenze tra un anno e l'altro, non pare che la situazione generale (anche considerata in termini relativi) si modifichi in modo sostanziale nel tempo. La prima evidenza (Tab. 2 e Gr. 1-5) mostra che esiste un divario non trascurabile tra le province dell'Emilia Romagna in termini di stima puntuale del tasso (incidenza) della povertà: con riferimento a tutti gli anni considerati, in generale i valori che caratterizzano la provincia di Bologna sono i più bassi mentre quelli riferiti alla provincia di Rimini i più elevati (circa il doppio rispetto ai valori riferiti a Bologna)<sup>1</sup>. In particolare Rimini, Forlì-Cesena e Piacenza presentano valori superiori a quello regionale. In ogni caso per tutte le province il tasso di povertà è decisamente più contenuto rispetto a quello riferito all'intero Paese. Se si guarda al gap relativo emerge che alcune province (Rimini, Forlì-Cesena, Piacenza) oltre ad essere caratterizzate da un elevato tasso di povertà, presentano anche un elevato valore del gap, cioè hanno una quota non trascurabile di popolazione che si colloca nello stato di povertà ed il reddito dei poveri è in mediana anche apprezzabilmente inferiore alla soglia di povertà. Per contro, alcune altre province (Bologna e Parma) presentano un gap non trascurabile, benché caratterizzate da una quota esigua di poveri. E' comunque da notare che la stima del gap riferita a tutte le province dell'Emilia Romagna risulta di gran lunga inferiore a quello del Paese. Bologna, inoltre, assieme a Rimini e Parma vede valori elevati di disuguaglianza dei redditi. Infine, per quel che riguarda il tasso di deprivazione, si osserva che, in generale, risulta superiore al tasso di povertà (a volte decisamente superiore). Fa eccezione la provincia di Rimini che presenta un elevato tasso di povertà ma uno dei più bassi tassi di deprivazione.

Accanto alle stime puntuali, è stato stimato il relativo CV. I CV sono decisamente accettabili per tutte le stime prodotte. La variabilità delle stime del tasso di deprivazione è più elevata rispetto a quella che è stata ottenuta per gli altri parametri a causa del più contenuto potere previsivo delle covariate utilizzate nel modello per il tasso di deprivazione (vedi paragrafo B.5). Nel commentare le stime ottenute per il

---

<sup>1</sup> Ricordiamo che non è corretto effettuare il ranking dei domini di interesse semplicemente ordinando le stime da modello; a questo fine è necessario adottare metodi per la stima del ranking delle aree.



tasso di deprivazione si noti quindi che i relativi intervalli di confidenza sono più ampi rispetto a quelli riferiti ai restanti parametri.

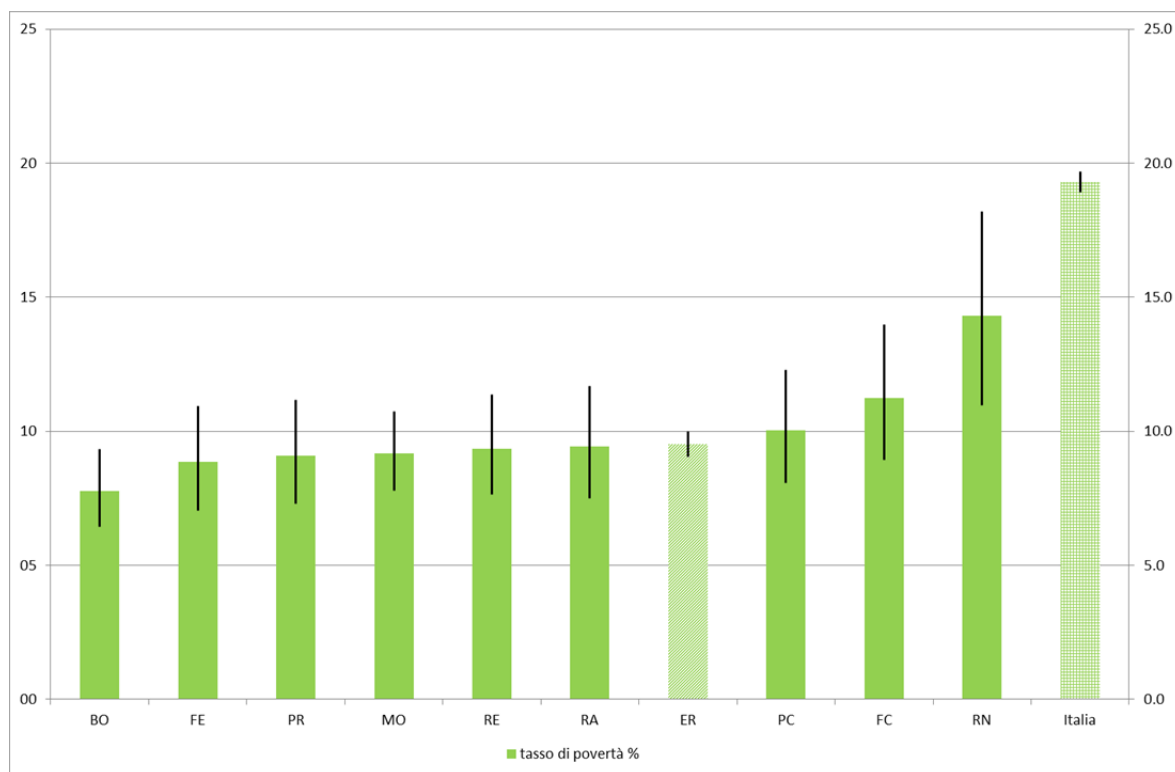
Tab. 2 - Stime da modello dei parametri obiettivo, per le province dell'Emilia Romagna, anni 2007-2012

<b>tasso di povertà %</b>							
codice prov	provincia	2007	2008	2009	2010	2011	2012
33	PC	9.8	10.3	9.2	10.5	10.3	10.0
34	PR	8.9	8.4	8.1	9.1	8.9	9.1
35	RE	8.2	7.8	8.2	9.0	9.2	9.3
36	MO	8.5	7.9	8.1	8.8	9.3	9.2
37	BO	7.3	7.1	6.9	7.7	7.9	7.8
38	FE	8.0	8.6	7.8	8.9	9.0	8.9
39	RA	9.1	9.2	7.7	8.9	8.9	9.4
40	FC	10.4	10.5	9.8	10.7	10.5	11.2
99	RN	13.8	13.5	13.2	13.2	13.1	14.3
	ER	8.9	8.7	8.4	9.2	9.3	9.5
	Italia	18.7	18.4	18.2	19.8	19.5	19.3
<b>gap relativo %</b>							
codice prov	provincia	2007	2008	2009	2010	2011	2012
33	PC	25.7	22.1	21.1	20.2	22.8	21.1
34	PR	25.3	22.0	21.7	20.7	23.4	21.4
35	RE	23.4	20.6	21.3	19.9	22.8	20.6
36	MO	25.6	21.7	20.8	19.5	23.3	20.0
37	BO	25.8	22.1	21.8	19.0	22.2	20.3
38	FE	23.7	20.2	19.2	18.7	21.8	19.8
39	RA	23.5	20.6	19.0	17.7	21.4	19.8
40	FC	24.1	21.5	20.3	20.0	22.7	21.3
99	RN	26.6	23.6	22.0	20.5	23.0	22.0
	ER	24.9	21.5	20.9	19.4	22.5	20.5
	Italia	23.2	23.1	24.8	26.6	26.0	28.2
<b>indice di Gini %</b>							
codice prov	provincia	2007	2008	2009	2010	2011	2012
33	PC	29.0	27.4	27.7	27.0	28.1	28.5
34	PR	28.7	28.2	28.9	28.3	29.6	29.6
35	RE	26.0	26.0	28.1	27.1	27.6	27.8
36	MO	29.2	27.6	27.9	26.8	28.8	27.3
37	BO	30.3	29.1	29.9	27.3	28.4	29.2
38	FE	26.1	25.0	25.7	25.5	26.7	27.2
39	RA	25.8	25.8	25.6	24.5	26.5	27.2
40	FC	26.6	26.5	26.4	26.6	26.9	28.0
99	RN	30.1	28.3	29.8	27.0	28.7	29.4
	ER	30.1	28.3	29.8	27.0	28.7	29.4
	Italia	30.9	31.0	31.0	31.4	31.5	32.0
<b>tasso di deprivazione %</b>							
codice prov	provincia	2007	2008	2009	2010	2011	2012
33	PC	8.0	8.8	10.1	12.4	14.1	16.4
34	PR	8.8	9.1	9.9	11.1	13.0	15.6
35	RE	10.7	12.0	12.9	12.8	14.3	16.5
36	MO	11.0	12.2	12.8	14.2	15.4	17.8
37	BO	9.4	10.4	11.1	12.0	14.5	16.0
38	FE	9.2	9.4	9.9	10.0	11.0	12.3
39	RA	8.5	9.8	10.5	11.7	12.6	13.7
40	FC	10.7	12.0	12.2	14.4	16.0	18.9
99	RN	9.4	10.0	11.1	12.3	13.7	14.5
	ER	9.7	10.6	11.3	12.4	14.1	16.0
	Italia	16.3	16.0	16.8	22.2	25.2	23.8

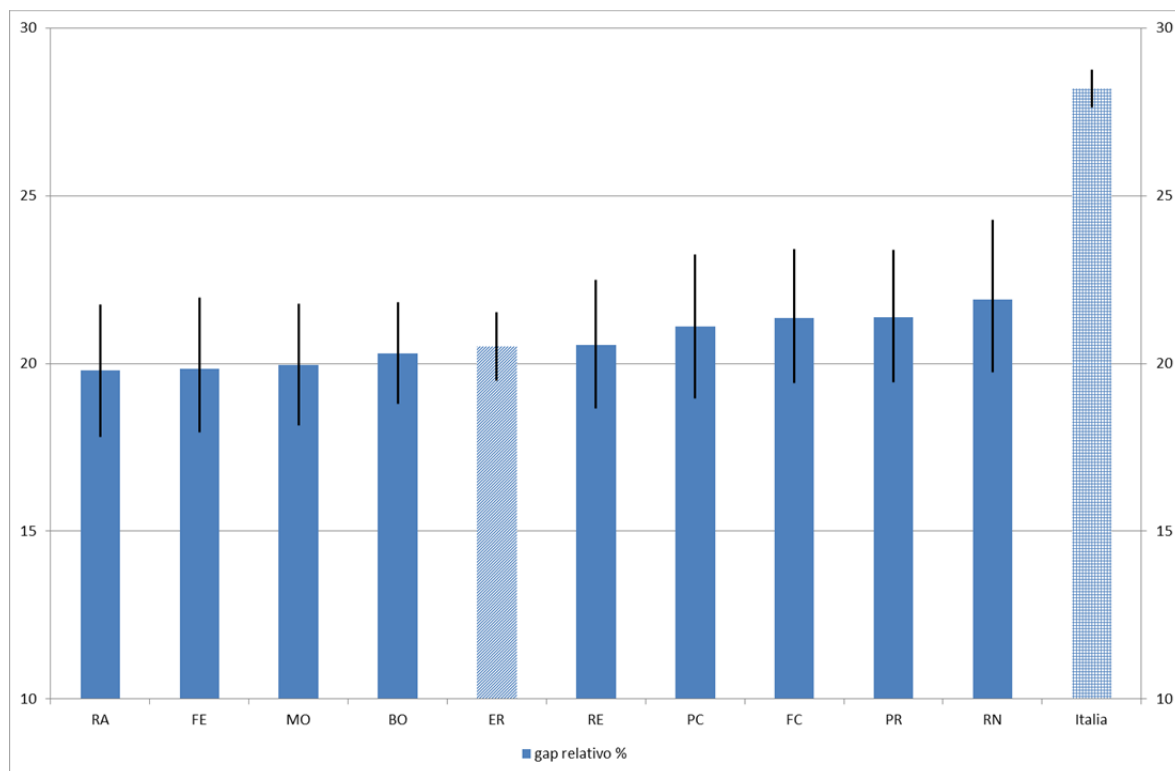
Tab. 3 - Coefficienti di variazione (CV) delle stime da modello, per le province dell'Emilia Romagna, anni 2007-2012

<b>CV % - tasso di povertà</b>							
codice prov	provincia	2007	2008	2009	2010	2011	2012
33	PC	10.8	10.9	11.2	10.9	10.8	10.7
34	PR	10.6	10.6	10.4	11.1	10.6	10.9
35	RE	9.3	8.9	9.3	9.4	9.7	10.1
36	MO	8.2	8.1	8.2	8.0	8.0	8.2
37	BO	9.1	9.2	9.7	9.7	9.6	9.5
38	FE	11.1	11.7	11.7	11.3	11.3	11.3
39	RA	11.4	11.8	11.5	11.4	10.9	11.3
40	FC	10.8	10.9	11.2	10.9	11.1	11.4
99	RN	12.2	12.42	12.8	12.5	13.0	13.0
<b>CV % - gap relativo</b>							
codice prov	provincia	2007	2008	2009	2010	2011	2012
33	PC	3.9	4.4	5.1	5.3	4.4	5.2
34	PR	3.6	4.1	4.1	4.4	4.1	4.8
35	RE	3.7	4.2	4.1	4.5	4.0	4.7
36	MO	3.2	3.6	3.9	4.1	3.6	4.6
37	BO	2.7	2.9	3.3	3.7	3.2	3.9
38	FE	3.6	4.3	4.9	4.8	4.4	5.1
39	RA	3.8	4.1	4.9	5.2	4.5	5.0
40	FC	3.5	4.1	4.7	4.6	4.1	4.6
99	RN	3.7	4.2	4.7	5.2	4.6	5.3
<b>CV % - indice di Gini</b>							
codice prov	provincia	2007	2008	2009	2010	2011	2012
33	PC	4.6	4.6	4.8	4.8	4.6	4.6
34	PR	4.5	4.3	4.0	4.2	4.2	4.4
35	RE	4.5	4.4	4.2	4.2	4.4	4.4
36	MO	4.2	3.9	4.1	4.0	3.9	4.3
37	BO	3.7	3.4	3.8	3.8	3.8	3.8
38	FE	4.5	4.5	4.6	4.7	4.6	4.6
39	RA	4.5	4.3	4.6	4.8	4.6	4.6
40	FC	4.2	4.2	4.6	4.2	4.5	4.4
99	RN	4.4	4.3	4.5	4.7	5.0	4.7
<b>CV % - tasso di deprivazione</b>							
codice prov	provincia	2007	2008	2009	2010	2011	2012
33	PC	18.4	17.6	17.1	17.9	16.0	15.5
34	PR	16.4	16.2	13.9	14.8	14.4	15.1
35	RE	16.2	14.7	14.6	13.6	12.9	13.1
36	MO	13.8	13.2	11.7	11.7	11.2	11.0
37	BO	13.5	12.9	12.8	12.9	12.3	11.6
38	FE	18.5	17.8	18.3	18.1	17.2	16.7
39	RA	17.6	18.3	18.9	19.5	16.8	16.2
40	FC	15.9	15.5	15.3	15.6	15.8	15.6
99	RN	19.3	18.4	19.0	20.6	20.7	19.1

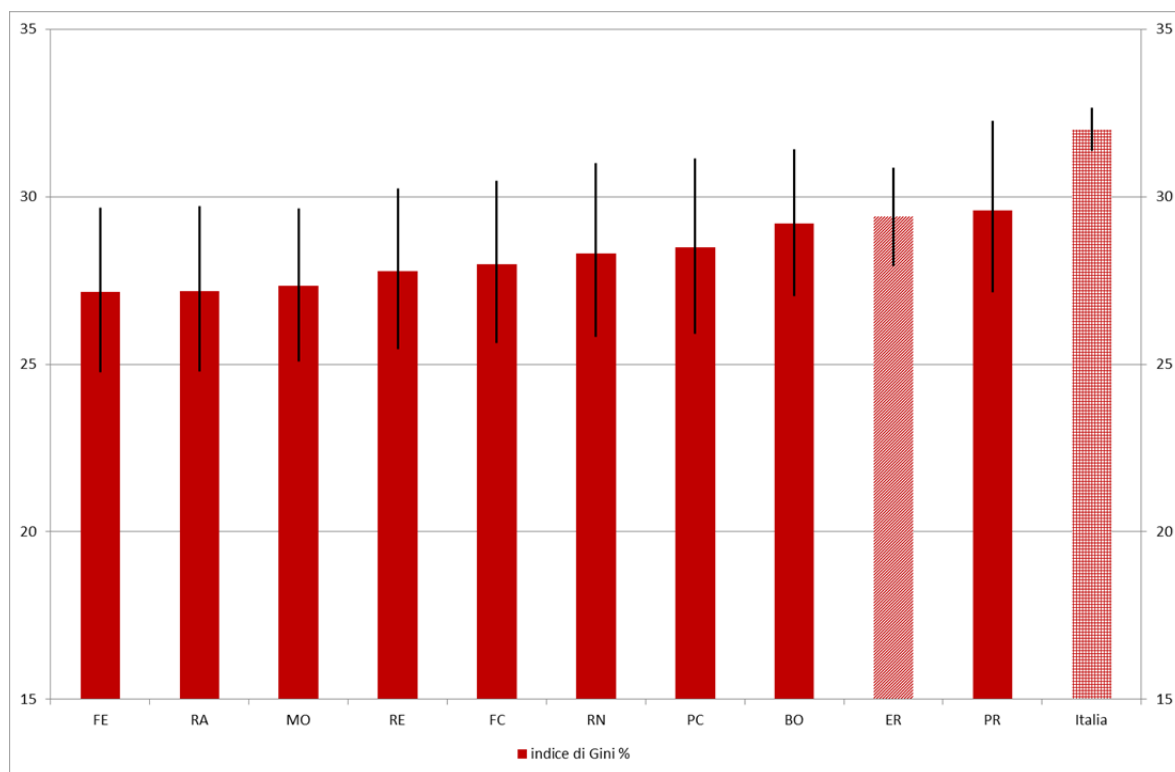
Gr. 1 - Stime da modello del tasso di povertà per le province dell'Emilia Romagna nel 2012 (i segmenti sulle barre rappresentano gli intervalli di confidenza)



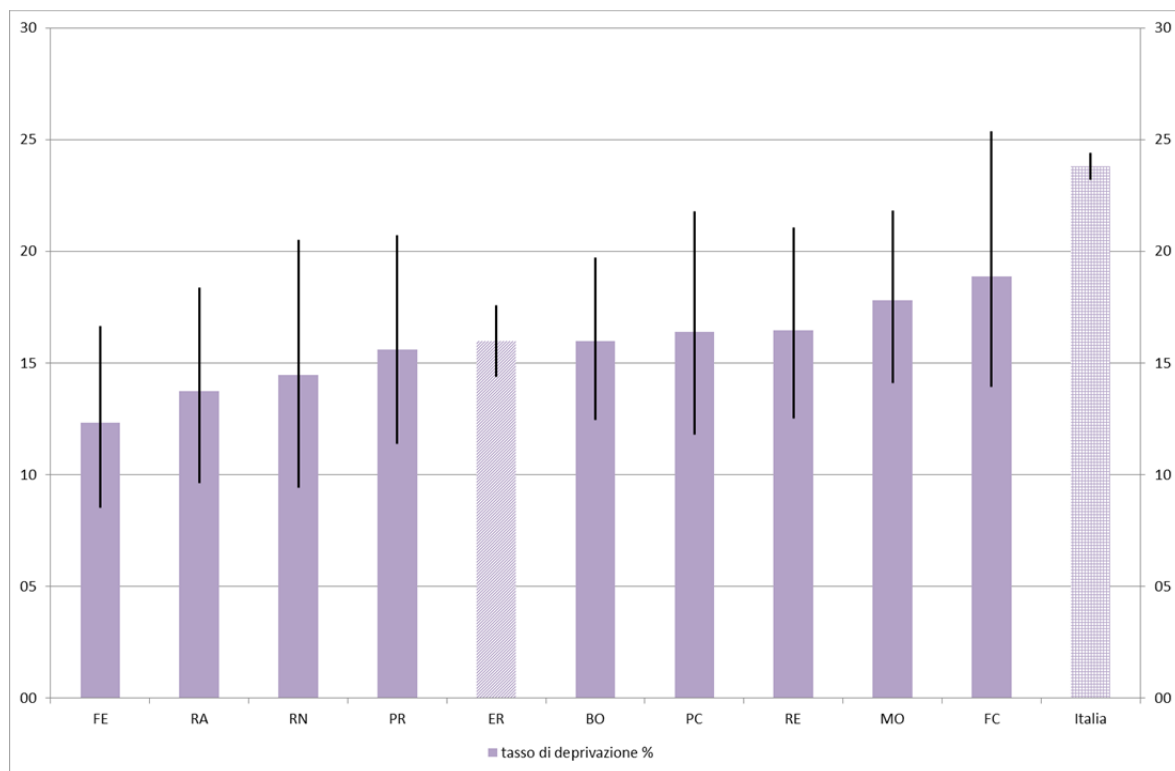
Gr. 2 - Stime da modello del gap relativo per le province dell'Emilia Romagna nel 2012 (i segmenti sulle barre rappresentano gli intervalli di confidenza)



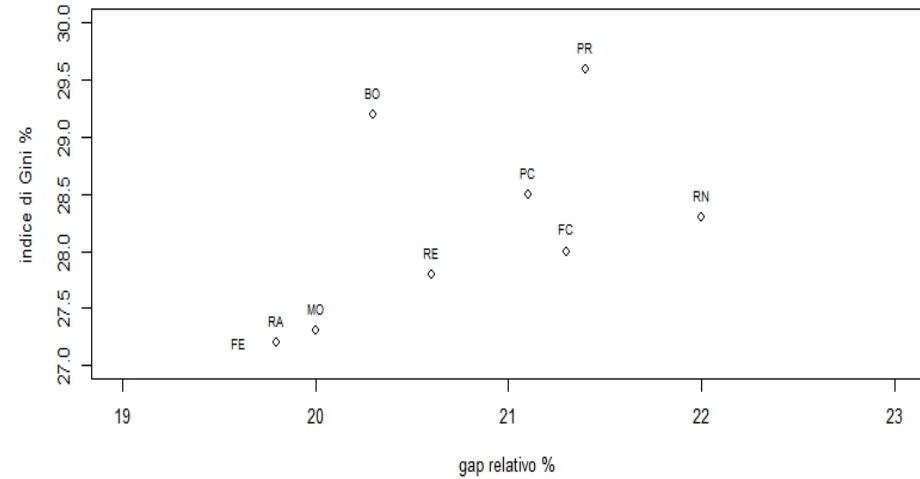
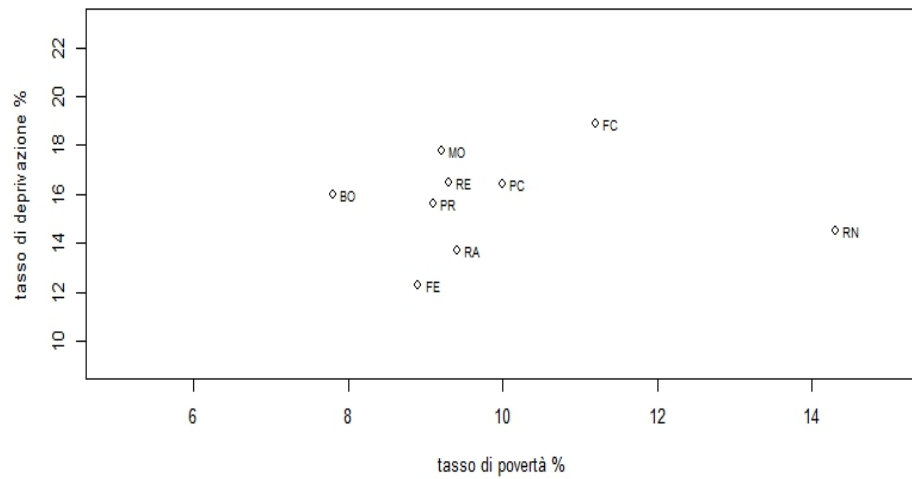
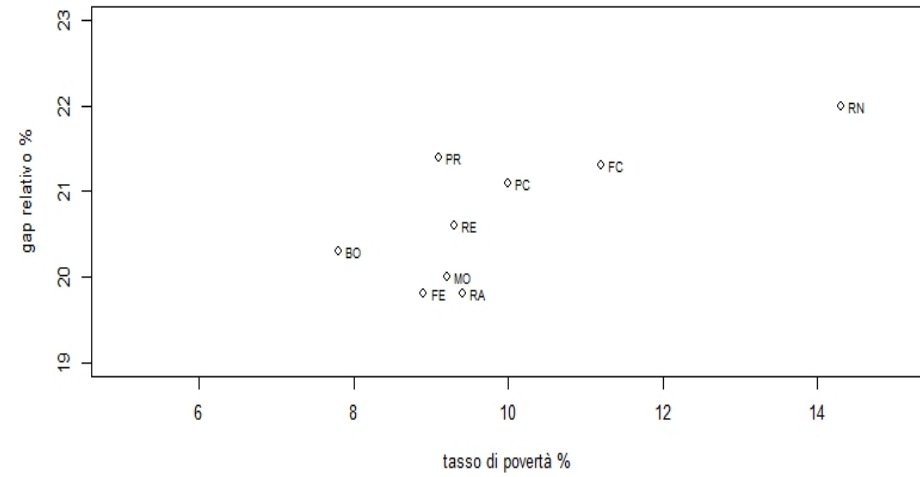
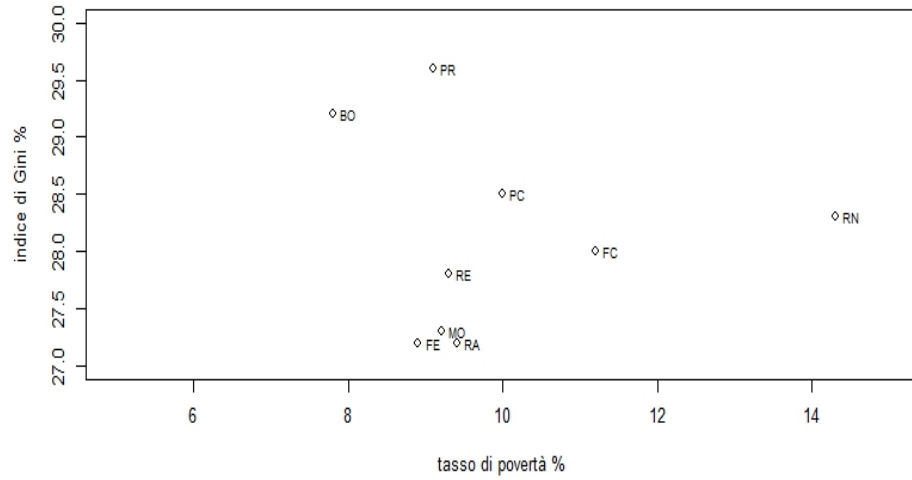
Gr. 3 - Stime da modello dell'indice di Gini per le province dell'Emilia Romagna nel 2012 (i segmenti sulle barre rappresentano gli intervalli di confidenza)



Gr. 4 - Stime da modello del tasso di deprivazione materiale per le province dell'Emilia Romagna nel 2012 (i segmenti sulle barre rappresentano gli intervalli di confidenza)



Gr. 5 - Scatterplot riferiti a combinazioni di tasso di povertà, gap relativo e indice di Gini per le province ER nel 2012



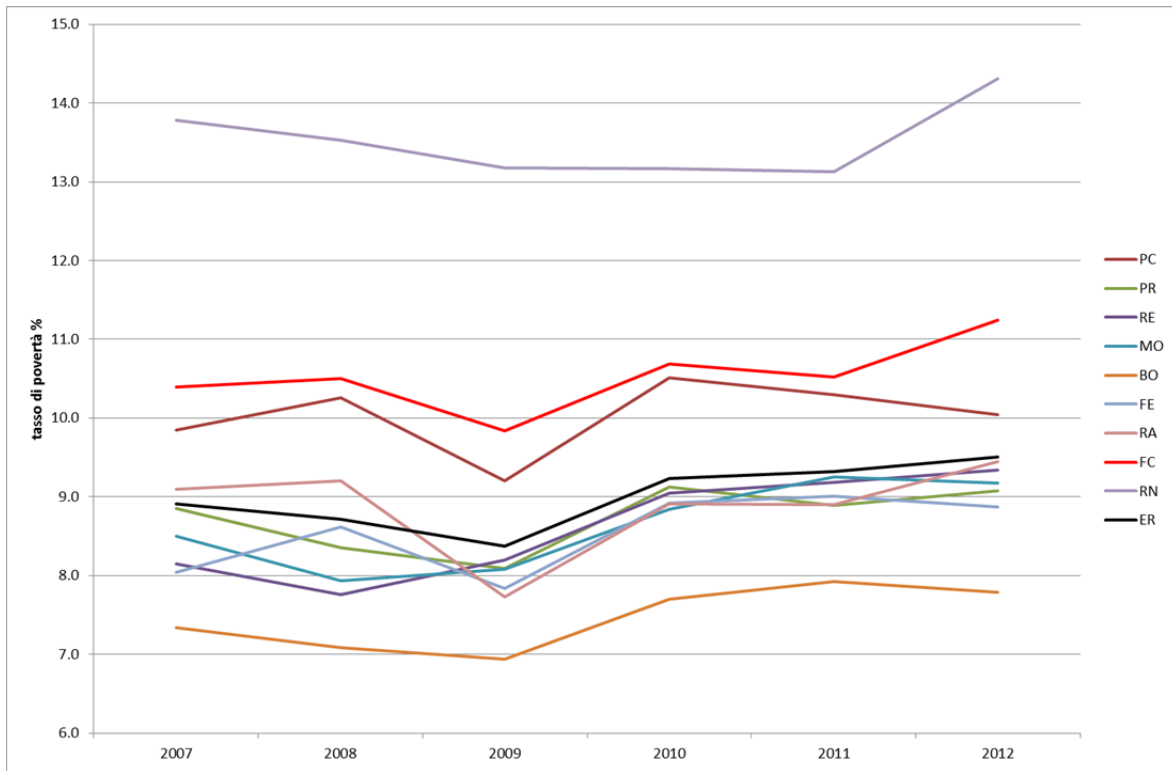
L'andamento temporale delle stime (Gr. 6-9) dei parametri considerati è peculiare di ogni parametro. Con riferimento al tasso di povertà, si registra una diminuzione fino al 2009, un aumento nell'anno successivo ed una sostanziale stabilità negli ultimi anni considerati. Fanno eccezione Rimini e Forlì-Cesena che vedono un aumento rilevante del tasso tra il 2011 e il 2012. Anche per il gap c'è un andamento

Bisogna tuttavia considerare che le stime sopra considerate sono caratterizzate da un certo margine di errore<sup>2</sup> e che, quindi, l'andamento temporale irregolare potrebbe in realtà essere meno marcato laddove considerassimo l'errore di campionamento. Fa eccezione il tasso di deprivazione materiale per cui emerge un chiaro trend in aumento con riferimento a tutto il periodo considerato.

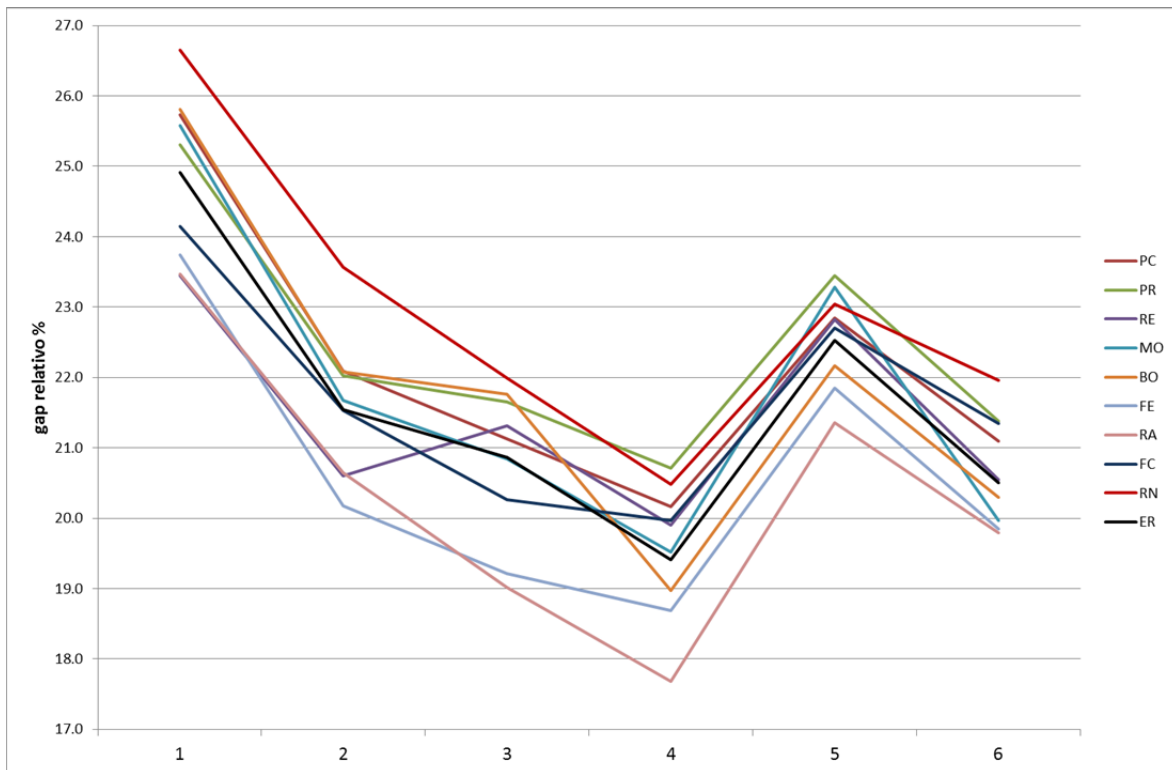
---

<sup>2</sup> Gli intervalli di confidenza non sono stati riportati sui gr. 6-10 che presentano le stime per tutti gli anni al fine di non rendere illegibile il grafico stesso. Essi possono però essere desunti dalla tab. 3 che riporta i CV.

Gr 6 - Andamento temporale delle stime da modello del tasso di povertà per provincia negli anni 2007-2012

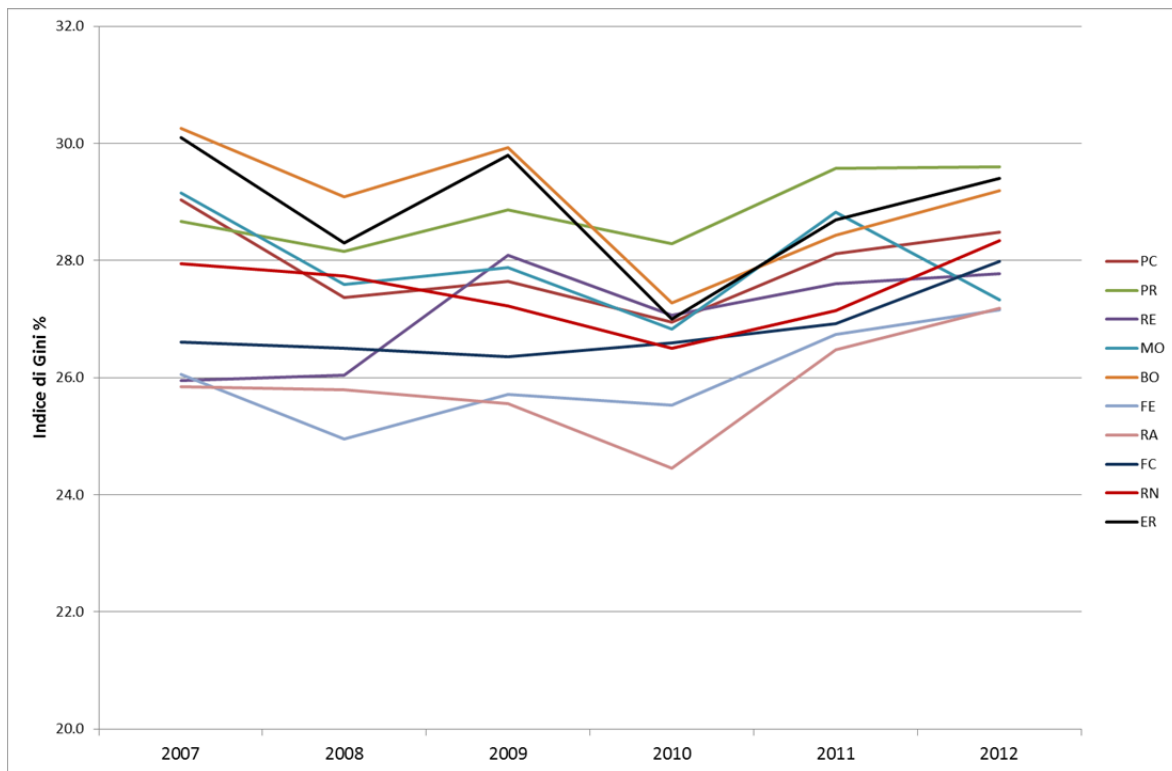


Gr. 7 - Andamento temporale delle stime da modello del gap relativo per provincia negli anni 2007-2012

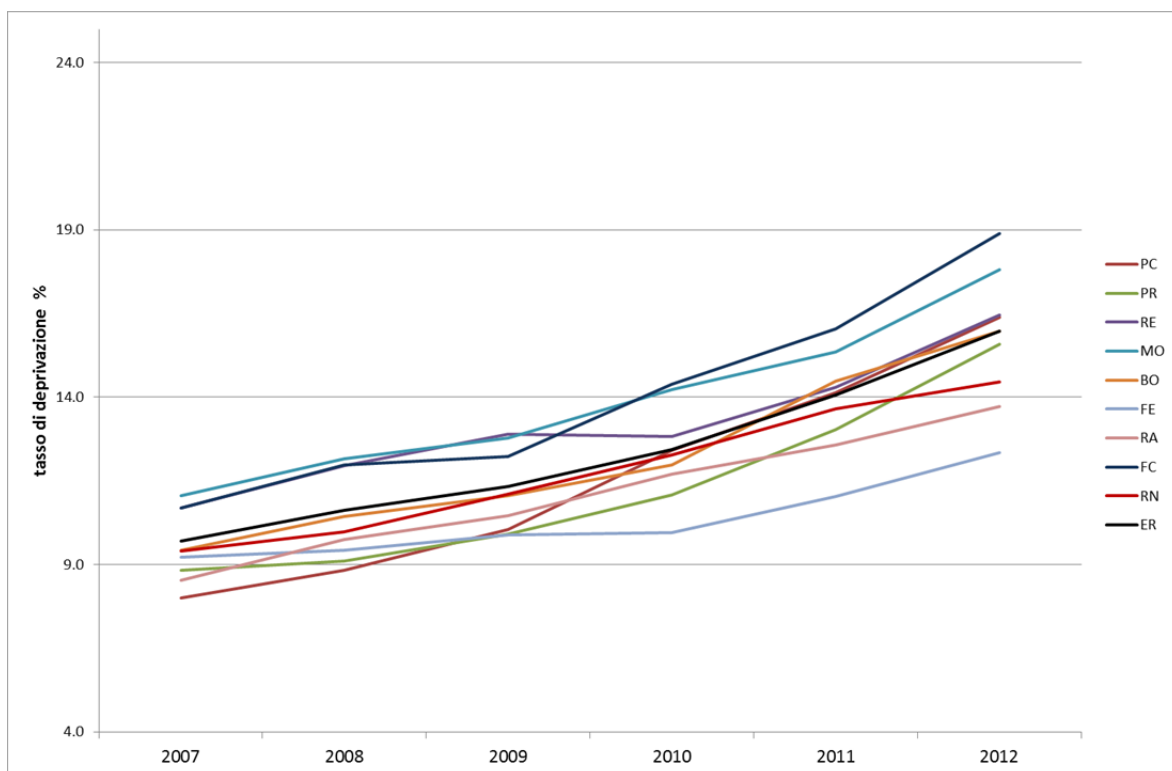




Gr. 8 - Andamento temporale delle stime da modello dell'indice di Gini per provincia negli anni 2007-2012

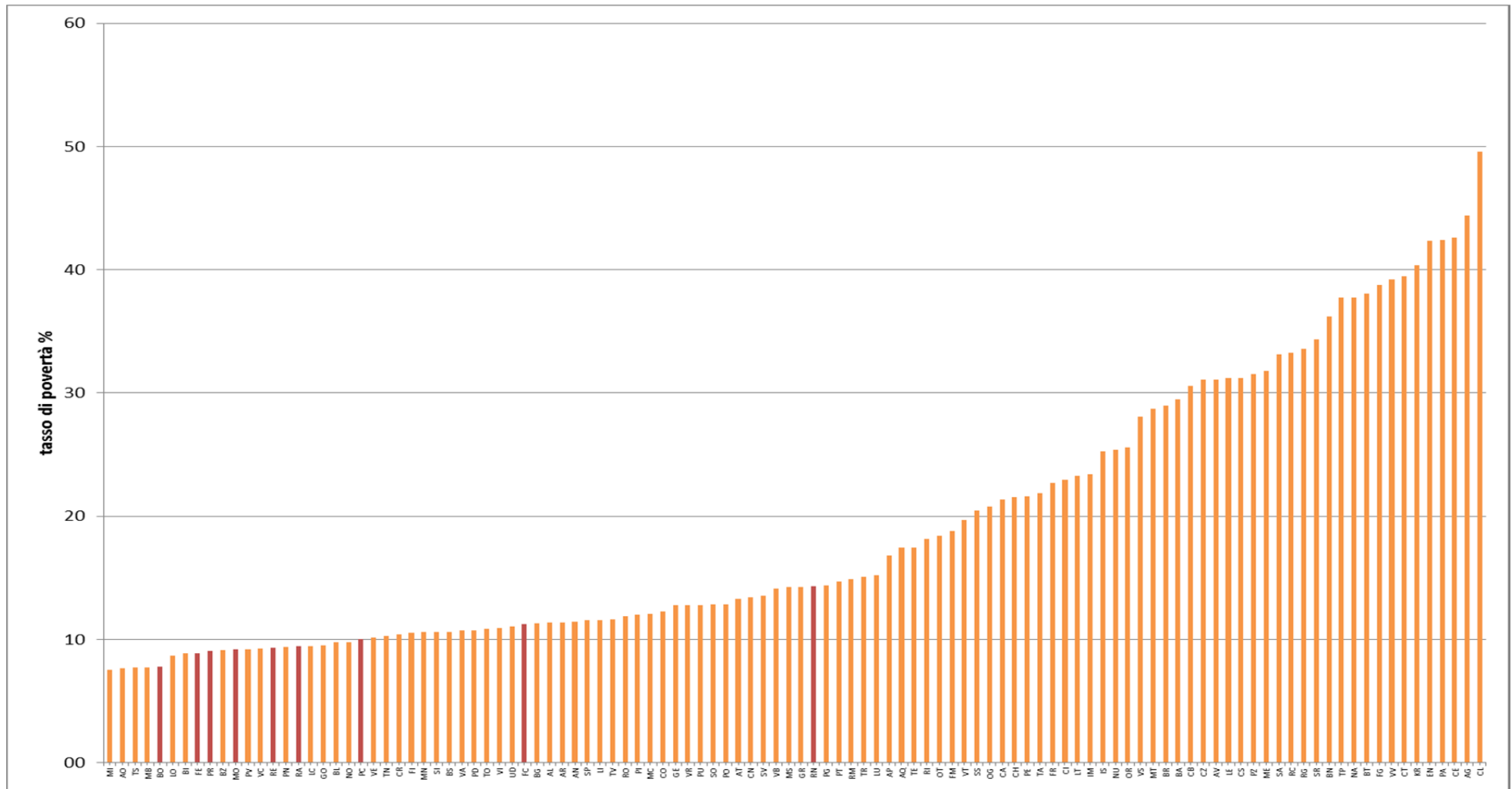


Gr. 9 - Andamento temporale delle stime da modello del tasso di deprivazione mat. per provincia negli anni 2007-2012

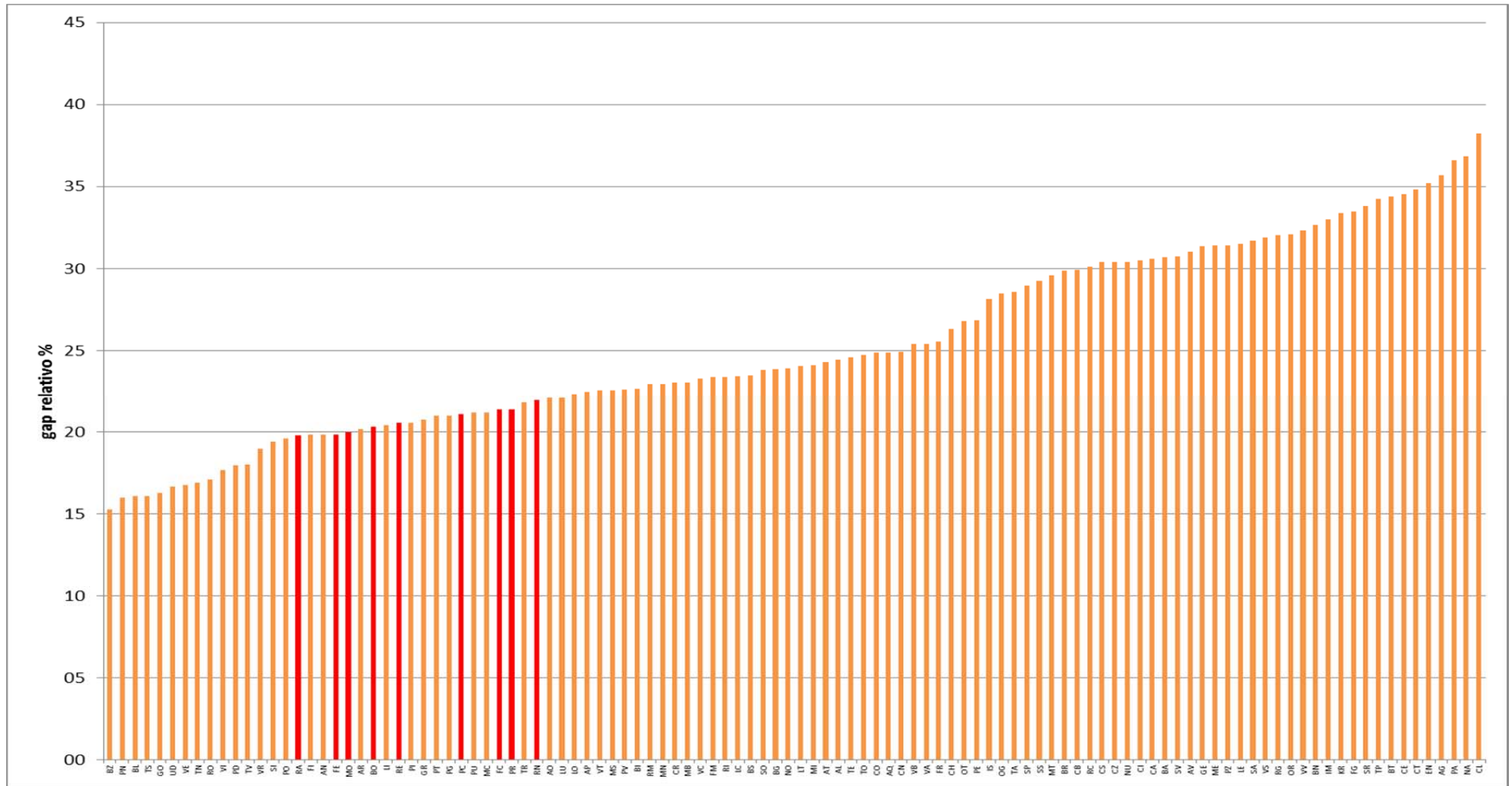


Nei grafici seguenti (Gr. 10-13) si riportano le stime riferite al 2012 di tutte le province italiane e separatamente (Gr. 14-17) anche per le province del Nord, per i quattro parametri considerati, al fine di valutare il “posizionamento” delle province dell’Emilia Romagna. In generale emerge che a) le province dell’Emilia Romagna mostrano parametri di povertà in generale più contenuti rispetto a quelli delle province italiane, infatti le stime di riferimento si collocano sempre nell’area destra del grafico, b) con riferimento al gap la situazione è particolarmente positiva (tutte le province dell’Emilia Romagna sono caratterizzate da valori del gap apprezzabilmente bassi in relazione alle restanti aree), c) se si fa riferimento solo alle regioni del nord il posizionamento positivo che si nota a confronto con tutto il paese si indebolisce, come del resto atteso. In generale, con riferimento a tutti i parametri, le province dell’Emilia Romagna occupano posizioni sia buone sia meno buone in relazione alle restanti province del Nord.

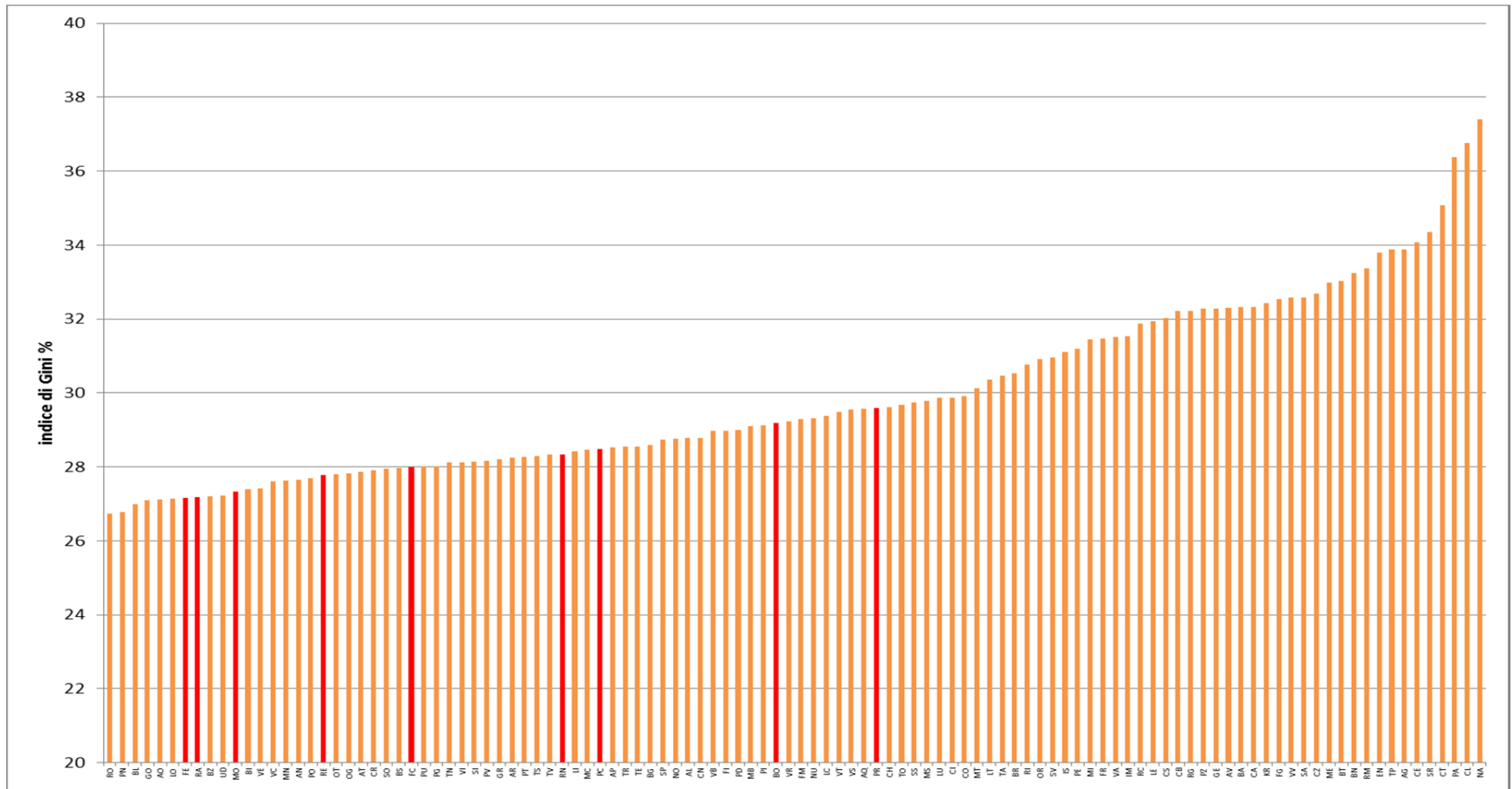
Gr. 9 - Stime da modello del tasso di povertà per le province italiane nel 2012



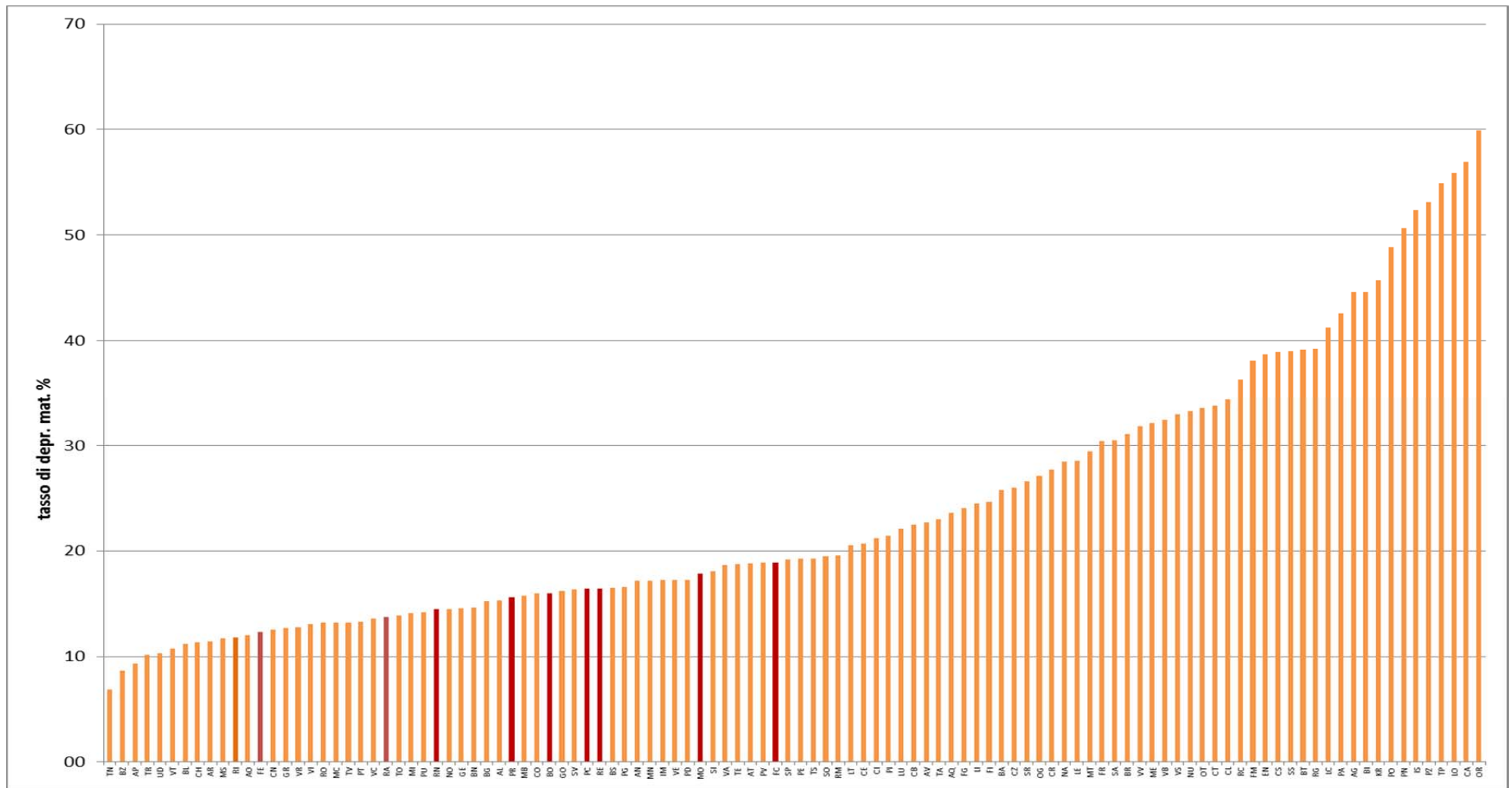
Gr. 10 - Stime da modello del gap relativo per le province italiane nel 2012



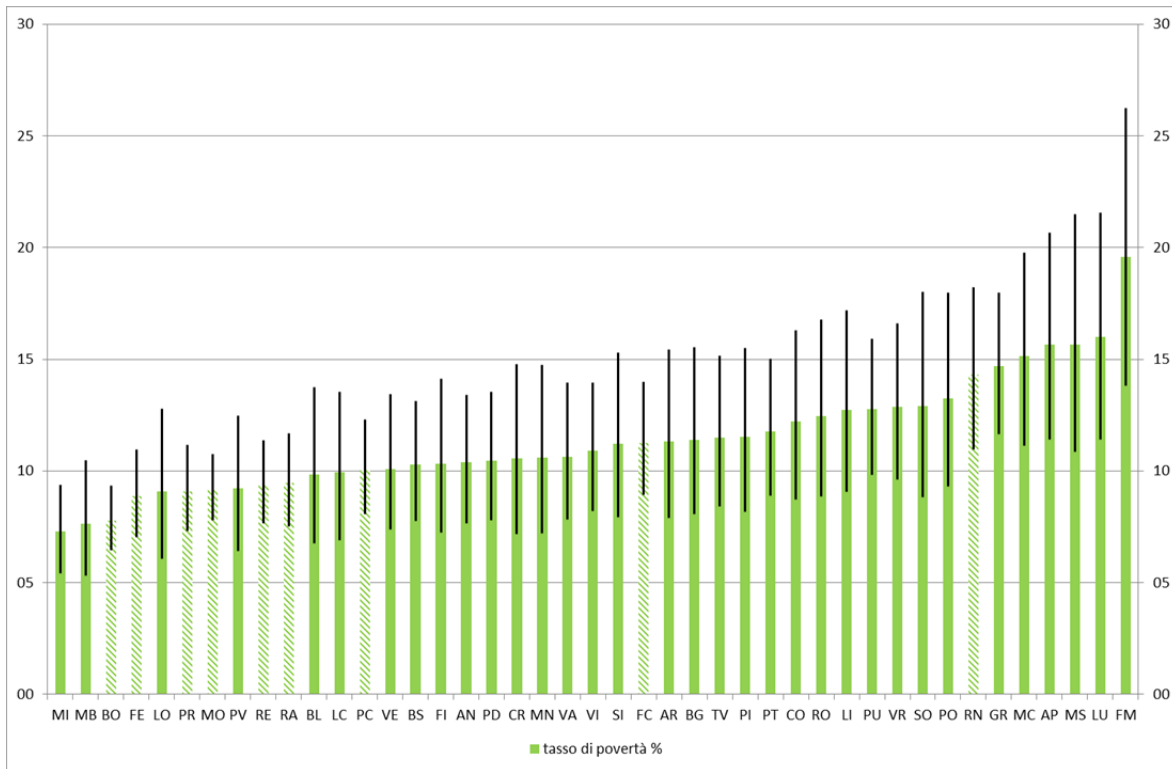
Gr. 11 - Stime da modello dell'indice di Gini per le province italiane nel 2012



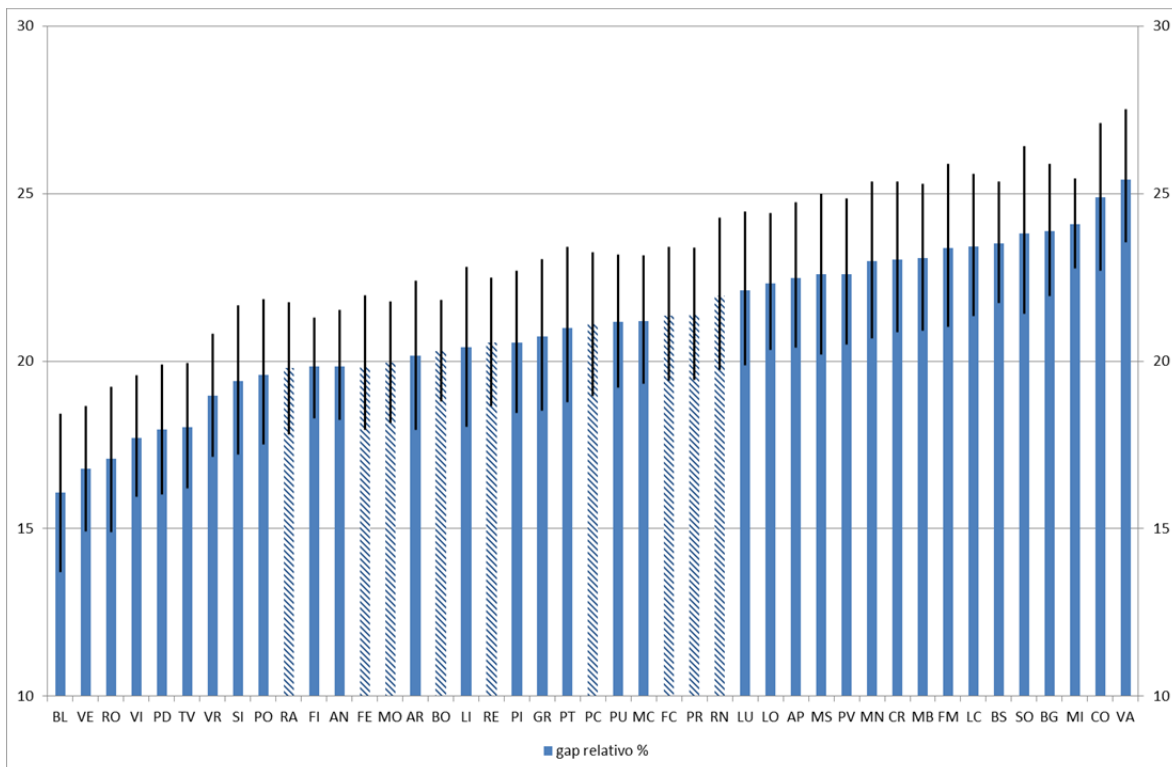
Gr. 12 - Stime da modello del tasso di deprivazione materiale per le province italiane nel 2012



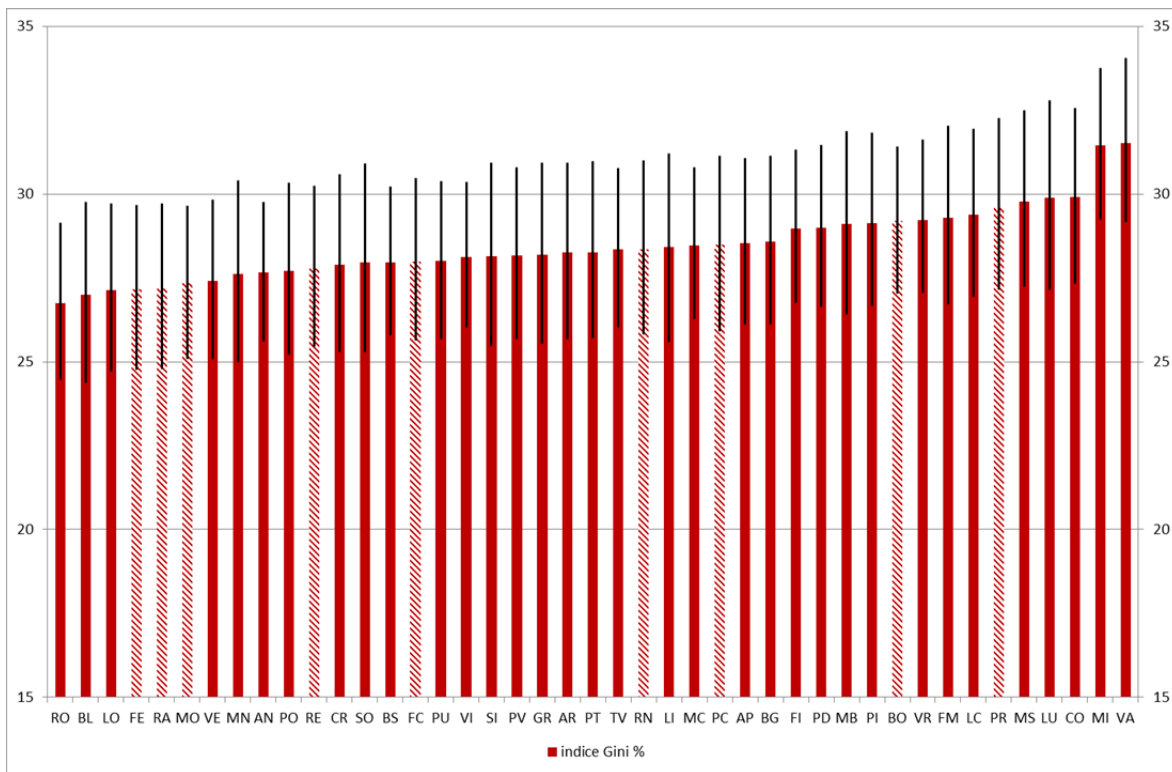
Gr. 13 - Stime da modello del tasso di povertà per le province del Nord Italia nel 2012 (i segmenti sulle barre rappresentano gli intervalli di confidenza)



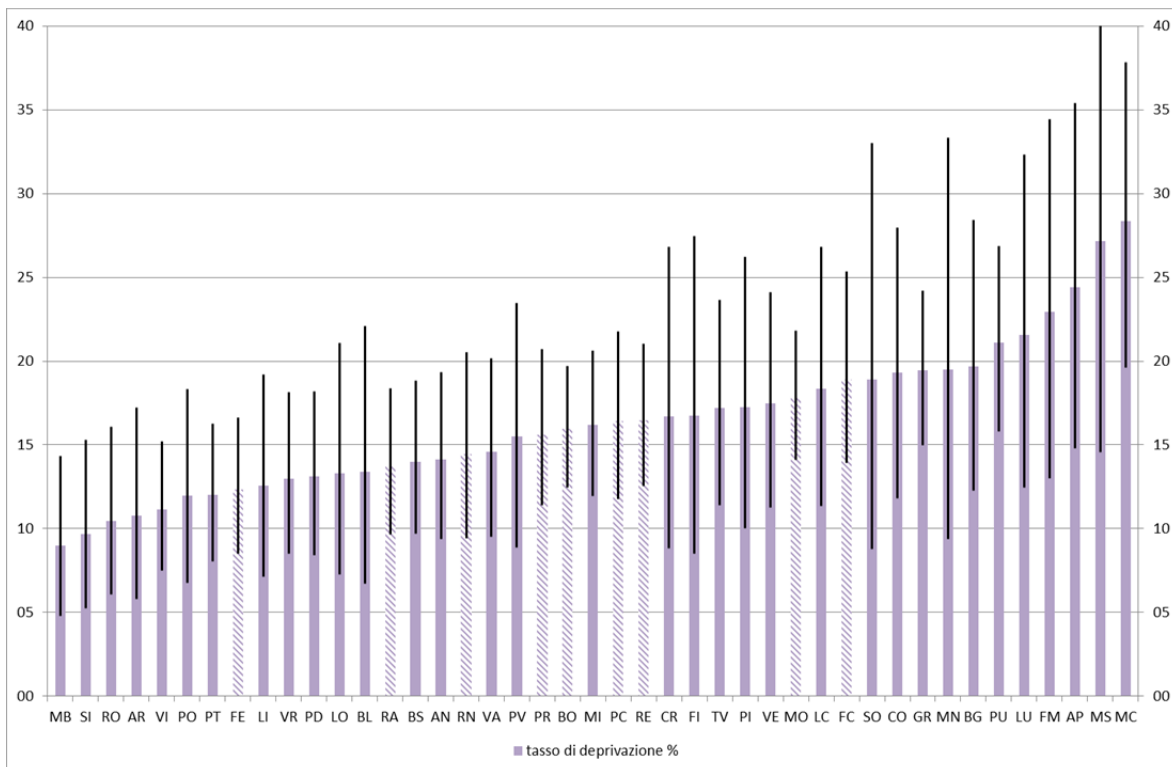
Gr. 14 - Stime da modello del gap relativo per le province del Nord Italia nel 2012 (i segmenti sulle barre rappresentano gli intervalli di confidenza)



Gr. 15 - Stime da modello dell'indice di Gini per le province del Nord Italia nel 2012 (i segmenti sulle barre rappresentano gli intervalli di confidenza)



Gr. 16 - Stime da modello del tasso di deprivazione materiale per le province del Nord Italia nel 2012 (i segmenti sulle barre rappresentano gli intervalli di confidenza)





## Le stime per i distretti socio-sanitari

Le stime dei parametri considerati e dei relativi Coefficienti di Variazione, per gli anni 2007-2012 e per i 38 distretti socio-sanitari dell'Emilia Romagna, sono riportate nell'appendice B, sezione B.8. Qui mostriamo e commentiamo le stime riferite all'anno 2012 a cui aggiungiamo le stime riferite all'Emilia Romagna nel suo insieme ed all'Italia (Gr. 17-20).

La prima evidenza riguarda il divario marcato tra i distretti dell'Emilia Romagna in termini di incidenza della povertà: ci sono distretti che mostrano un tasso di povertà che è circa tre volte quello dei distretti caratterizzati da tassi molto bassi<sup>3</sup>. Le stime vanno da valori intorno al 5% (Pianura Est, Pianura Ovest, San Lazzaro) a valori più elevati e superiori al 14% (Rimini, Riccione). Ben 19 distretti sono caratterizzati da un valore superiore a quello dell'intera regione (per cui il tasso di povertà è pari al 10% circa) benché tutti i distretti mostrino tassi inferiori a quello dell'Italia (19% circa). Tale disparità tra distretti si manifesta anche per i restanti parametri. Con riferimento al Gap, che pure risulta decisamente più basso rispetto a tutta l'Italia, si va da valori inferiori al 18% ( Scandiano, Pianura Est, Castelfranco) a valori superiori al 22% per le Città di Piacenza e di Bologna. Quest'ultimo distretto è poi caratterizzato da un valore dell'indice di Gini che risulta decisamente elevato (intorno al 33%) e più elevato del valore che connota l'Italia. Valori altrettanto elevati caratterizzano Modena e la Città di Piacenza. Infine per la Città di Bologna risulta elevato anche il tasso di deprivazione materiale (intorno al 22%) a fronte dei valori decisamente più contenuti (10% circa) riferiti a Castelnuovo nei Monti, San Lazzaro di Savena, Centro-Nord, Valli Taro e Ceno.

Approfondiamo qui l'analisi delle relazioni tra i quattro indicatori al fine di verificare se ci sono aree in cui più di un indicatore segnala una situazione di disagio. Innanzitutto dal Gr. 21 e dalla matrice che contiene i coefficienti di correlazione lineare tra le stime dei quattro parametri:

$$\begin{array}{c} TP \quad Gap \quad Gini \quad Depr \\ TP \quad \begin{bmatrix} 1 & 0.64 & 0.15 & 0.10 \\ Gap & & 1 & 0.83 & 0.11 \\ Gini & & & 1 & 0.06 \\ Depr & & & & 1 \end{bmatrix} \end{array}$$

emerge, come ci si poteva aspettare, una certa correlazione positiva tra il Tasso di Povertà ed il Gap Mediano (coefficiente di correlazione pari a 0.64) e tra il Gap Mediano e l'indice di Gini (0.83). L'elevata correlazione tra Tasso di Povertà e Gap indica che ci sono vari distretti in cui un tasso elevato (basso) si associa ad un livello

---

<sup>3</sup> Ricordiamo che non è corretto effettuare il ranking dei domini di interesse semplicemente ordinando le stime da modello; a questo fine è necessario adottare metodi per la stima del ranking delle aree.

elevato (basso) di intensità della povertà. Detto in altre parole, vi sono contesti in cui l'incidenza della povertà è elevata che sperimentano anche un livello di reddito distante (e inferiore) dalla soglia di povertà: i poveri sono cioè anche molto poveri. D'altro canto un livello apprezzabile di disuguaglianza associato ad un elevato livello del Gap sono indicazione di aree in cui i poveri sono molto poveri e si manifesta una elevata concentrazione del reddito.

Si noti poi l'assenza di relazione tra il Tasso di Deprivazione Materiale e Tasso di Povertà, una misura di povertà relativa, che cioè tiene conto della distribuzione del reddito, e Tasso di Deprivazione Materiale che, invece, coglie le difficoltà delle famiglie di accedere a determinati beni e servizi. Una debole relazione tra questi due indicatori è stata osservata per molti paesi ed è stata discussa in letteratura (si veda ad esempio Ayala et al., 2011). Va sottolineato come il tasso di povertà misuri in sostanza un basso livello di reddito per l'anno di riferimento, il che non si traduce immediatamente in una condizione di deprivazione materiale per la famiglia. E' possibile infatti che il reddito sia solo occasionalmente inadeguato, ma che la famiglia disponga, o sia in grado di accedere a risorse sufficienti per far fronte ad entrate insufficienti in un singolo anno. Per questa ragione esso è chiamato nella terminologia inglese "at-risk-of-poverty" in quanto identifica una situazione di fragilità e rischio, che non può durare senza far cadere la famiglia in uno stato di deprivazione<sup>4</sup>. In questo senso la misura della deprivazione materiale fa implicitamente riferimento al concetto di reddito permanente. Se infatti una famiglia non dispone di redditi sufficienti per diversi anni, l'incapacità di acquisire beni e servizi porta ad deterioramento cumulativo delle condizioni di vita e pone le basi, soprattutto per i componenti più giovani, per una ridotta capacità reddituale. Inoltre è stato osservato come le famiglie in condizione di deprivazione materiale tendano ad essere caratterizzati da redditi molto variabili di anno in anno (lavori occasionali, occupazioni non stabili, frequenti esperienze di disoccupazione) ed in conseguenza, di sperimentare temporanee uscite dallo stato di povertà, misurato sulla base di una soglia definita in termini di reddito equivalente, senza che questo modifichi il loro stato di deprivazione. Occorre infine tenere in conto l'effetto dei limiti delle misure su cui sono basati i due indicatori: il reddito può essere riportato in modo errato o incompleto; le domande su cui è basata la definizione di deprivazione materiale – espresse volutamente in modo molto semplice nella costruzione del questionario – si prestano ad interpretazioni non univoche da parte dei rispondenti e possono avere valori e rilevanza differenti per segmenti diversi della popolazione. Ricordiamo inoltre che il tasso di deprivazione materiale si basa su una soglia fissa, non variabile di anno in anno; in questo senso la deprivazione potrebbe anche essere interpretata come una misura di povertà assoluta.

I distretti che evidenziano un Tasso di povertà elevato associato ad un Tasso di deprivazione materiale contenuto sono quindi caratterizzati dalla presenza di una

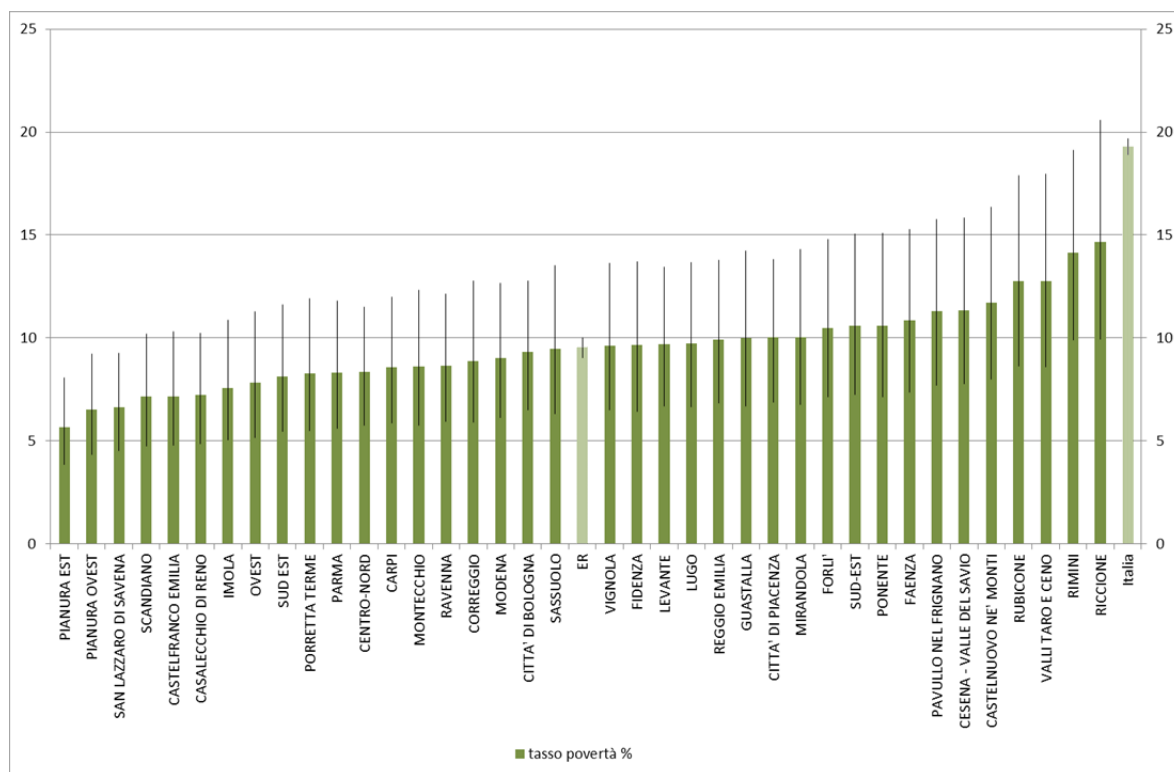
---

<sup>4</sup> In questo documento, quale traduzione di "at-risk-of-poverty-rate", adottiamo l'espressione "tasso di povertà", spesso usata nella letteratura in lingua italiana relativa allo studio della povertà.

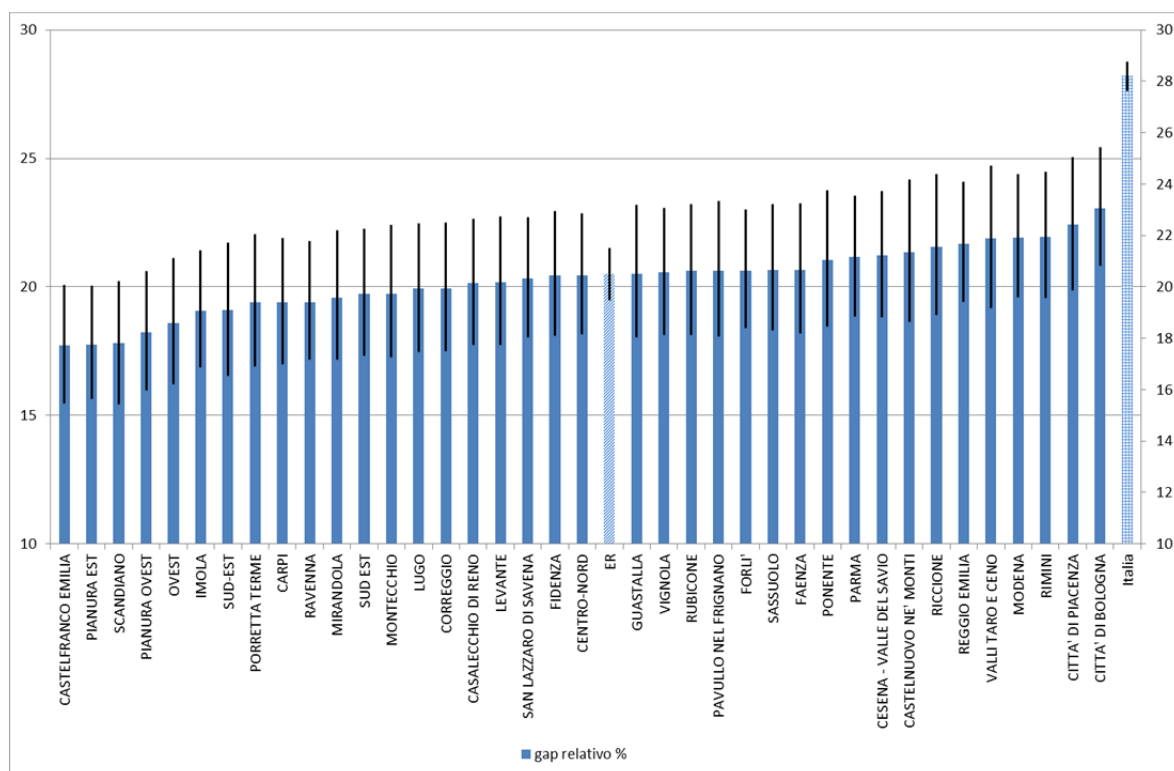
quota elevata di famiglie con reddito inferiore alla soglia di povertà ma, contemporaneamente, da condizioni di vita accettabili per la gran parte della popolazione (ad es. Rimini). I distretti in maggiore difficoltà sono poi quelli caratterizzati da valori elevati di più indicatori. I distretti per cui emergono valori che segnalano situazioni decisamente sfavorevoli su almeno tre dei quattro indicatori considerati sono i distretti: i) Città di Bologna e Sassuolo che, pur essendo caratterizzati da un tasso di povertà basso, manifestano valori elevati sui restanti indicatori (per Città di Bologna sono molto elevati), ii) Rimini e Valli Taro e Ceno per cui ad essere basso è, invece, il tasso di deprivazione.

Qualche ultimo commento sul tasso di variazione % riferita ai quattro indicatori considerati e calcolato con riferimento a due istanti temporali: inizio e fine periodo (quindi 2012 rispetto al 2007). Da tale confronto emerge con chiara evidenza che quasi tutti i distretti sono caratterizzati da un aumento della disuguaglianza, in alcuni casi pari al 6-8%. Una situazione in peggioramento si nota anche per il tasso di povertà: l'aumento interessa un numero cospicuo di distretti e per alcuni di questi è decisamente elevato (tra il 10 ed il 20%). Il tasso di deprivazione aumenta poi per tutti i distretti con un tasso di variazione nella gran parte dei casi uguale o superiore al 50% e per una decina di distretti raddoppia. Infine, in controtendenza rispetto ai restanti parametri, il gap diminuisce per la totalità dei distretti. Se volessimo proporre un commento sintetico che trascura l'eterogeneità tra le aree, potremmo dire che sono aumentate sia la quota di poveri in relazione all'ambiente in cui si vive, sia la quota di coloro che mostrano difficoltà ad accedere a beni e servizi ritenuti "necessari", sia le disparità reddituali. D'altra parte i poveri con riferimento al reddito sono diventati meno poveri (il loro reddito si è avvicinato alla soglia).

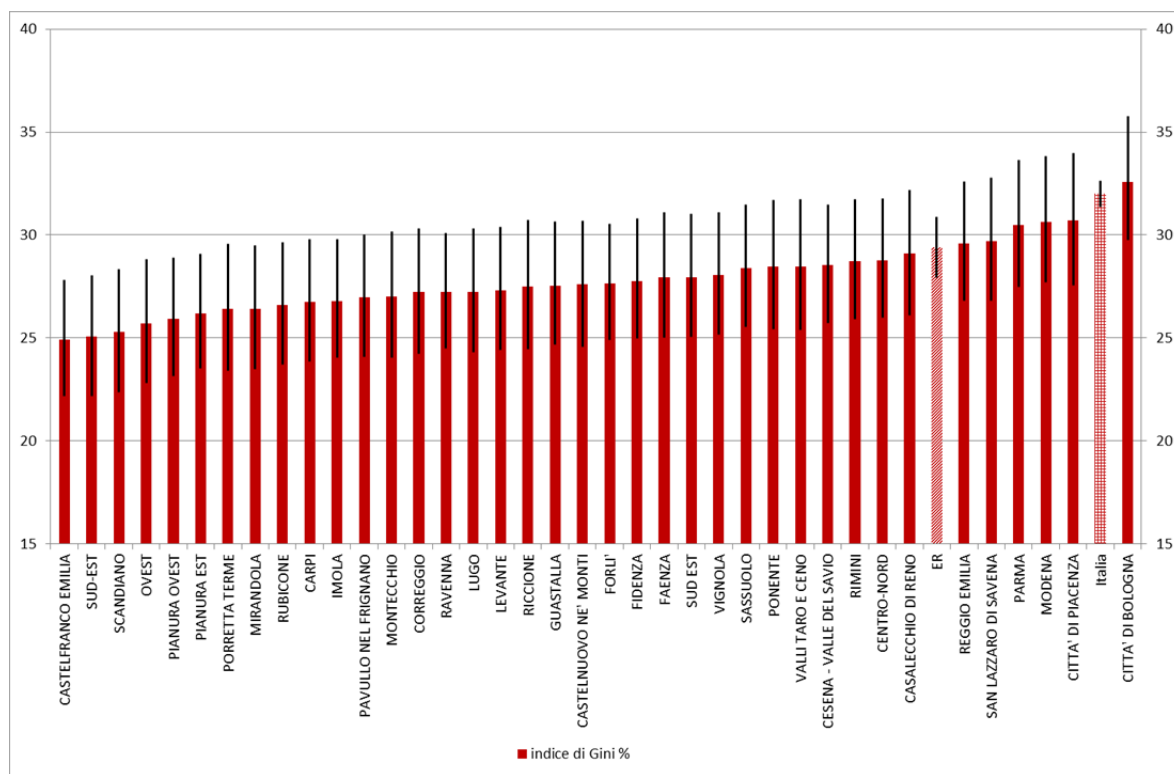
Gr 17 - Stime da modello del tasso di povertà per i distretti socio-sanitari dell'Emilia Romagna nel 2012 (i segmenti sulle barre rappresentano gli intervalli di confidenza)



Gr 18 - Stime da modello del gap relativo per i distretti socio-sanitari dell'Emilia Romagna nel 2012 (i segmenti sulle barre rappresentano gli intervalli di confidenza)



Gr 19 - Stime da modello dell'indice di Gini per i distretti socio-sanitari dell'Emilia Romagna nel 2012 (i segmenti sulle barre rappresentano gli intervalli di confidenza)



Gr 20 - Stime da modello del tasso di deprivazione materiale per i distretti socio-sanitari dell'Emilia Romagna nel 2012 (i segmenti sulle barre rappresentano gli intervalli di confidenza)

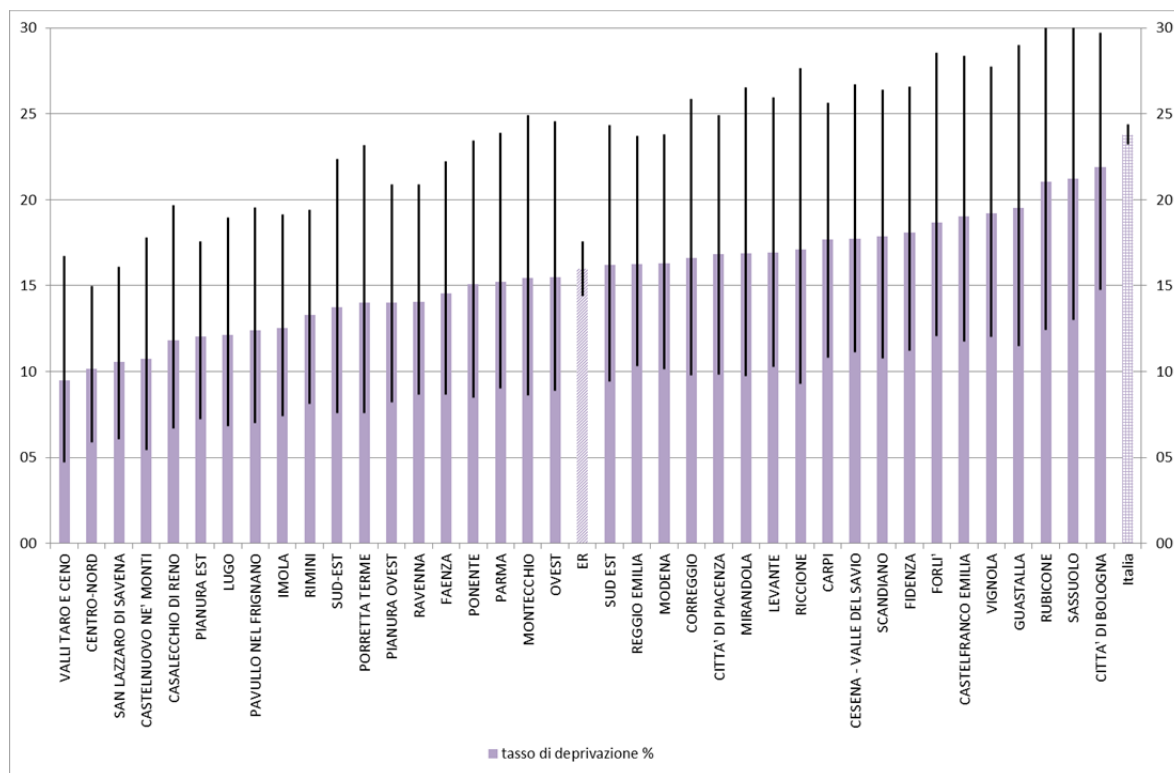
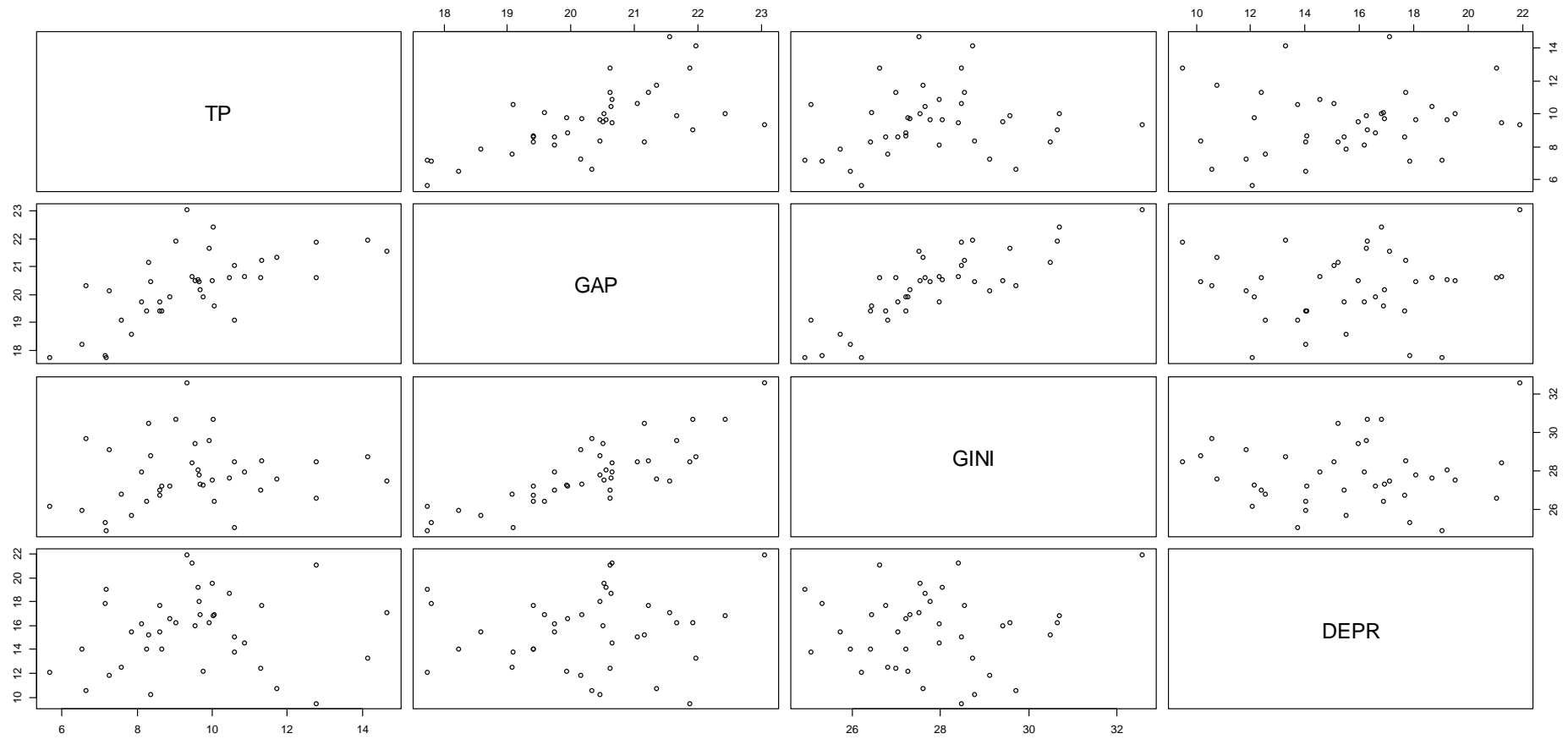
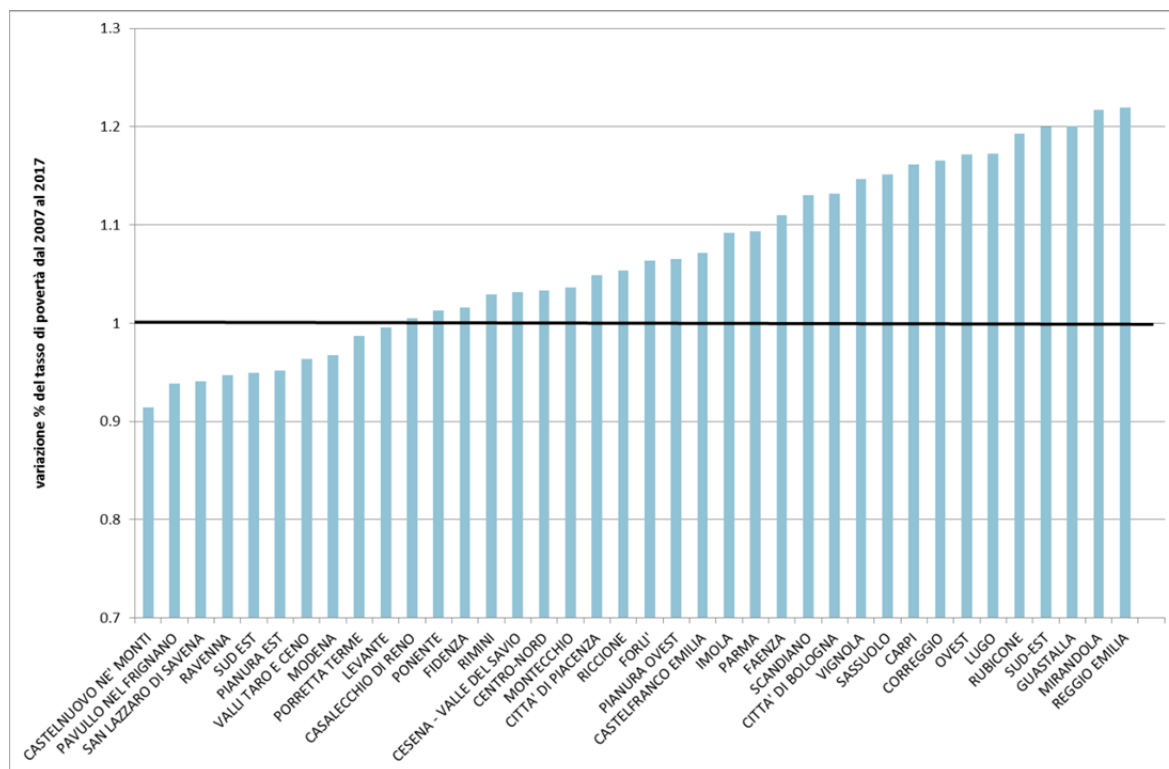


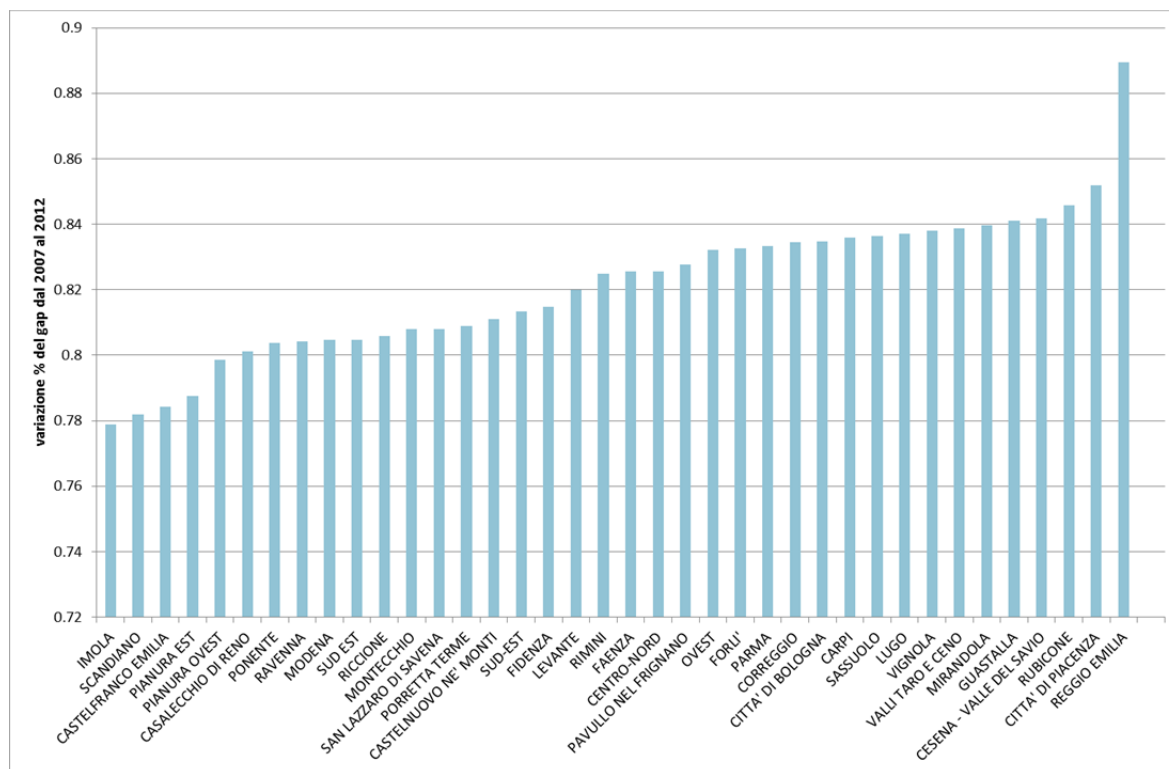
Grafico 21 - *Matrix scatterplot riferito ai quattro indicatori in esame*



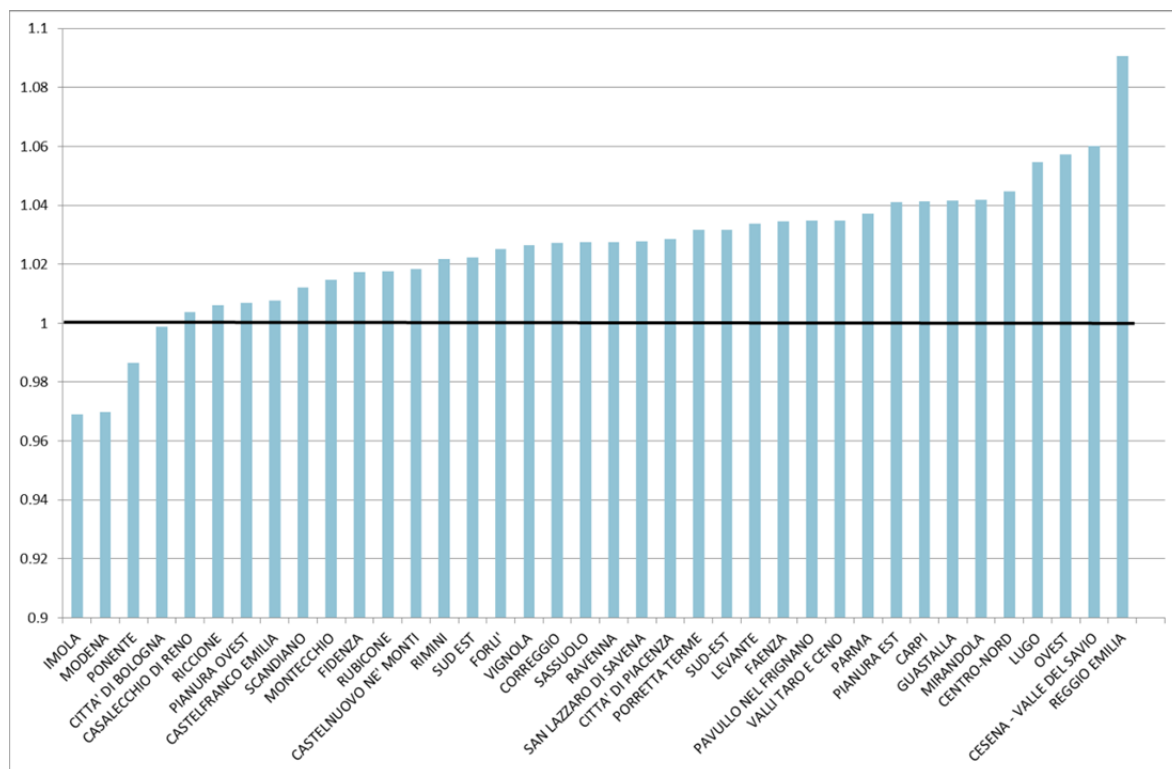
Gr 22 - Variazione % del tasso di povertà nei distretti, dal 2007 al 2012



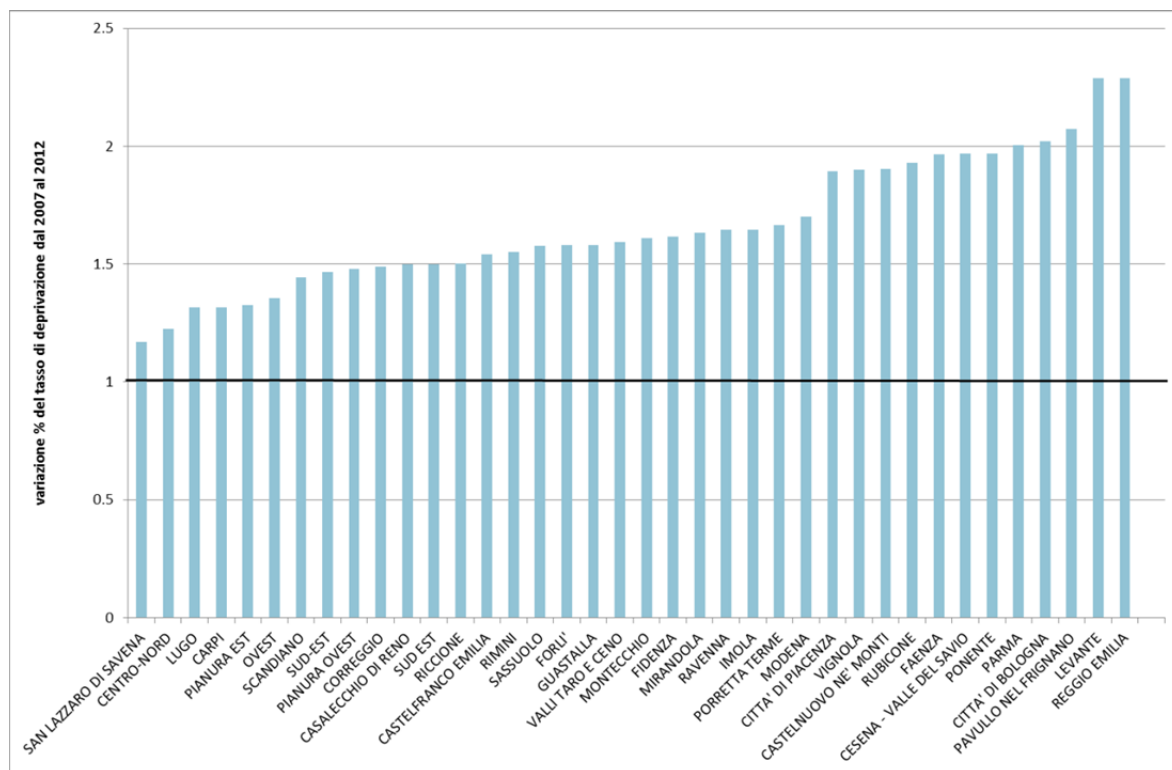
Gr 23 - Variazione % del gap relativo nei distretti, dal 2007 al 2012



Gr 24 - Variazione % dell'indice di Gini nei distretti, dal 2007 al 2012



Gr 25 - Variazione % del tasso di deprivazione nei distretti, dal 2007 al 2012





# **Appendice A**

## **L'indagine EU-SILC, la produzione delle stime dirette e della loro varianza**

## A1. Indagine EU-SILC. Caratteristiche generali

L'indagine EU-SILC (European Statistics on Income and Living Conditions) è stata istituita con il regolamento dell'Unione Europea n. 1177/2003. In Italia viene regolarmente condotta, con cadenza annuale, a partire dal 2004.

L'indagine sui redditi e le condizioni di vita è finalizzata alla produzione di stime annuali su queste tematiche in modo armonizzato per tutti i paesi dell'Unione. Più in dettaglio è pensata per produrre due tipologie di stime annuali: cross-section e longitudinali. Le stime cross-section forniscono la base per confronti coerenti tra le condizioni di vita, l'incidenza e la gravità di povertà ed esclusione sociale tra i paesi membri. Le stime longitudinali, rese possibili dalla struttura del piano di campionamento (gli individui campionati sono "seguiti" per quattro anni) sono finalizzate a rendere possibile l'analisi della dinamica micro-economica di questi indicatori.

La rilevazione del reddito monetario è uno degli obiettivi principali dell'indagine EU-SILC. Il reddito è il mezzo principale attraverso cui gli individui acquistano beni e servizi sul mercato ed è quindi naturalmente impiegabile come *proxy* delle risorse che gli individui e le famiglie hanno a disposizione per soddisfare i propri bisogni. L'indagine EU-SILC costituisce, in questo senso, la base informativa, attraverso cui costruire, in modo omogeneo per tutti i paesi e le regioni UE, indicatori di benessere e povertà basati sul reddito disponibile. Tuttavia l'indagine non è limitata alla sola misurazione del reddito, ma include anche numerose domande sulle condizioni di salute, economiche, nonché di una molteplicità di variabili che misurano le condizioni materiali di vita delle famiglie e che permettono di ricostruire un quadro della diffusione del benessere e dell'incidenza della povertà che prescindono da una misura quantitativa di capacità come è il reddito. Larga parte della comunità scientifica vede infatti come riduttivi gli studi sulla povertà e l'esclusione sociale che si limitino a considerare il solo aspetto monetario di quest'ultima. Un'analisi multidimensionale della povertà richiede che vengano misurate le *capabilities* (possibilità che gli individui hanno di fare, essere o avere ciò che desiderano) e i *funzionamenti* (ossia i risultati effettivamente ottenuti). Il questionario EU-SILC è orientato alla misurazione dei mancati funzionamenti, quali indicatori di un effettivo stato di povertà materiale. La misurazione del reddito è complessa ed errori di calcolo, nonché dichiarazioni per difetto da parte dei rispondenti, sia involontarie, sia deliberate rappresentano una possibile fonte di distorsione difficile da quantificare. Per questa ragione in un'analisi più esaustiva della povertà è bene affiancare stime monetarie e stime non monetarie della povertà.

L'indagine EU-SILC viene condotta in ogni paese dall'Istituto Nazionale di Statistica, sotto il coordinamento di Eurostat. La parte italiana di EU-SILC, a cui faremo riferimento come IT-SILC, è quindi sotto la responsabilità dell'ISTAT. L'indagine viene condotta attraverso interviste personali

## **A1.1 Popolazione obiettivo**

La popolazione obiettivo dell'indagine IT-SILC è costituita da tutti i componenti delle famiglie residenti in Italia, ancorché temporaneamente all'estero. Sono esclusi i membri di convivenze istituzionali e quelli di famiglie residenti all'estero, ancorché temporaneamente in Italia.

Per famiglia si intende la famiglia di fatto, in accordo con la definizione adottata da ISTAT in tutte le altre rilevazioni: un insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o da vincoli affettivi, coabitanti o aventi dimora abituale nello stesso comune.

Tutti i membri delle famiglie campionate di età superiore a 15 anni sono intervistati.

Una volta che un individuo entra a far parte del campione EU-SILC vi rimane per successivi 3 anni (per un totale di 4 rilevazioni), sia che rimanga all'interno della famiglia in cui si trovava al tempo della prima rilevazione, sia che vada a far parte di una nuova. Di conseguenza, il campione annuale IT-SILC è rinnovato ogni anno per un quarto.

## **A1.2 Periodo di rilevazione e periodo di riferimento**

La rilevazione dei dati avviene in settembre/ottobre di ogni anno. La scelta di questo periodo è dettata dalla volontà di permettere alle famiglie di utilizzare nelle risposte le informazioni contenute nella loro dichiarazione annuale dei redditi.

Le domande relative al reddito, fanno riferimento al reddito percepito nell'anno precedente a quello dell'intervista. Possiamo quindi distinguere anno di rilevazione e l'anno di riferimento del reddito. All'anno dell'intervista fanno riferimento le domanda relative alle condizioni di vita della famiglia.

## **A1.3 Piano di campionamento**

Il campione annuale EU-SILC per l'Italia è composto dall'unione di 4 campioni longitudinali. A regime, ogni anno viene rinnovato un quarto del campione; le unità selezionate parteciperanno a quattro successive ondate dell'indagine per poi uscire dal campione. Ovviamente per la prima ondata, nel 2004, l'intero campione è stato selezionato direttamente dalla popolazione. Nel 2005 è iniziato il rinnovamento ad un quarto per volta.

Complessivamente, il piano di campionamento è quindi a panel ruotato, secondo questo schema in cui ciascuna le lettere maiuscole indicano i campioni longitudinali:

	2008	2009	2010	2011	2012	2013	2014
I edizione	ABCD						
II edizione		BCDE					
III edizione			CDEF				
IV edizione				DEFG			
V edizione					EFGH		
VI edizione						FGHI	
VII edizione							GHIL

Per garantire stime sufficientemente precise per tutti i domini pianificati dell'indagine (di cui si dirà) e per ottemperare ai requisiti di numerosità minima stabiliti da Eurostat, ogni campione longitudinale di EU-SILC è composto di 8,000 famiglie.

La numerosità effettiva del campione trasversale per un dato anno  $t$  è quindi costituita:

- dalla frazione delle 8,000 famiglie nuove campionate che hanno aderito all'indagine
- alle famiglie provenienti dai campioni longitudinali entrati negli anni  $t-1$ ,  $t-2$ ,  $t-3$ , al netto ovviamente di quelle cadute per attrito.

Come in tutte le indagini con una componente panel, l'indagine EU-SILC prevede regole di inseguimento delle unità di rilevazione (follow-up rules) che da un'occasione di rilevazione alla successiva mutano di residenza o formano una nuova famiglia. Le regole generali di inseguimento sono state fissate a livello europeo e sono quindi simili per tutti i paesi. In linea di principio, tutte le unità intervistate (membri delle famiglie campionate di età maggiore o uguale a 15) vengono inseguite se formano un nuovo nucleo e/o si spostano all'interno del territorio nazionale nel periodo di vita del panel (le quattro rilevazioni successive). La nuova famiglia entra a far parte del campione e tutti i suoi componenti di età adeguata sono intervistati. Non sono inseguite le unità che si trasferiscono all'estero, se non in casi particolari, ovvero

quando il trasferimento è solamente temporaneo, ossia di meno di sei mesi di durata prevista (Eurostat, 2006).

Il campione longitudinale delle famiglie “nuove” (l’intero campione trasversale per la prima edizione del 2004) è selezionato secondo un piano di campionamento stratificato a due stadi, sostanzialmente analogo a quello adottato da ISTAT per altre grandi indagini di argomento sociale ed economico.

Le unità di primo stadio sono costituite dai comuni, stratificati per regione e dimensione demografica. In particolare all’interno di ciascuna regione i comuni sono divisi in autorappresentativi e non autorappresentativi in ragione della loro dimensione demografica. I comuni autorappresentativi vengono sempre inclusi nel campione mentre quelli non autorappresentativi vengono campionati in modo casuale stratificato con probabilità proporzionale alla dimensione demografica. La soglia che separa i comuni autorappresentativi dagli altri è fissata in modo diverso da regione a regione con l’obiettivo di garantire l’autoponderazione del campione a livello regionale. Complessivamente il numero degli strati è pari a 288 (per maggiori dettagli, si veda ISTAT, 2008)

Le unità di secondo stadio, le famiglie, sono campionate in modo sistematico dalle liste anagrafiche.

Le elaborazioni contenute in questo rapporto sono basate sui dati dell’edizione 2008-2013 dell’indagine. Per quest’ultima la numerosità effettiva del campione trasversale (ossia al netto di non risposta ed attrito) è di 18487 famiglie per un totale di 44622 individui.

La distribuzione di famiglie e individui rispondenti è sintetizzata nella tabella A.1.

Tabella A.1 *Distribuzione di famiglie e individui rispondenti per il campione trasversale dell'edizione IT-SILC, anno di indagine 2013.*

<b>REGIONE</b>	Famiglie	Individui	<b>REGIONE</b>	Famiglie	Individui
Piemonte	1404	3152	Marche	1041	2637
Valle d'Aosta	367	833	Lazio	1406	3346
Lombardia	1846	4437	Abruzzo	424	1090
Trentino A.A.	581	1474	Molise	347	825
Veneto	1395	3501	Campania	991	2660
Friuli V. Giulia	1071	2427	Puglia	797	1992
Liguria	1067	2309	Basilicata	386	945
<b>Emilia-Romagna</b>	<b>1456</b>	<b>3396</b>	Calabria	686	1714
Toscana	1005	2372	Sicilia	957	2455
Umbria	772	1849	Sardegna	488	1208
			<b>Totale</b>	<b>18487</b>	<b>44622</b>

Segnaliamo che la dimensione effettiva del campione EU-SILC per l'Italia si è andata progressivamente riducendo nel corso del tempo, per una molteplicità di cause che includono un attrito (ossia perdita di osservazioni tra occasioni successive nella componente panel) abbastanza severo. Nella wave 2008 – la più remota considerata in questo studio - il numero complessivo di famiglie rispondenti fu di 20928 per un totale di 52433 intervistati. Si è dunque registrata una contrazione del 11.6% in termini di famiglie e del 14.9% in termini di individui nell'arco del periodo considerato. Per la regione Emilia Romagna la contrazione è stata più contenuta rispetto alla media nazionale: -7.9% per le famiglie, -8.8% per gli individui intervistati.

#### **A1. 4 Pesì campionari**

Nel file dei dati relativi ad EU-SILC rilasciato da ISTAT sono pubblicati i coefficienti di riporto all'universo (pesi campionari) sia per le componenti longitudinali che per il campione trasversale nel suo complesso. Poiché in questo lavoro ci interessiamo solamente di stime *cross-section* faremo sempre riferimento a questi ultimi.

Data una variabile obiettivo  $y$  per cui è stato osservato un campione di valori  $(y_1, \dots, y_n)$  i pesi campionari pubblicati in corrispondenza delle unità rispondenti  $(w_1, \dots, w_n)$  permettono di ottenere una stima approssimativamente corretta e consistente del totale di popolazione applicando la seguente formula:

$$\hat{Y} = \sum_{i=1}^n w_i y_i$$

che per la media diventa

$$\hat{Y} = N^{-1} \sum_{i=1}^n w_i y_i$$

dove  $N = \sum_{i=1}^n w_i$ .

Intuitivamente i pesi campionari sono dati dall'inverso delle probabilità di inclusione, in modo tale che se tutte le unità della popolazione avessero la stessa probabilità di essere incluse nel campione, allora i pesi sarebbero costanti e pari all'inverso del tasso di campionamento.

In pratica la loro costruzione è più complessa e il loro contenuto informativo è maggiore. In primo luogo attraverso i pesi campionari si tenta di neutralizzare il più possibile l'effetto distorsivo della *non-risposta*. In secondo luogo, al fine di ottenere stime più efficienti si calibrano i pesi in modo tale che le stime del totale, per variabili il cui totale a livello di popolazione è noto, riproducano esattamente questo valore, secondo il metodo della "calibrazione".

I pesi campionari inclusi nei file IT-SILC rilasciati da ISTAT raggiungono questi obiettivi attraverso una tecnica di riponderazione in due passi. Nel primo passo utilizzando un criterio noto come "criterio delle celle di riponderazione" sono stati corretti, almeno parzialmente, gli effetti della mancata risposta totale. Il secondo passo costituisce un passo di calibrazione vero e proprio.

Nel caso di medie e totali, gli stimatori per calibrazione rappresentano la scelta *standard* di ISTAT che li applica in tutte le grandi indagini di argomento economico e sociale. L'illustrazione dei fondamenti teorici di questi stimatori eccede gli obiettivi di questo rapporto e pertanto si rimanda il lettore alla letteratura sul tema (in primis a Deville e Särndal, 1992 e a Särndal e Lundström, 2005).

Per quanto riguarda la correzione degli effetti della non risposta, alla base del metodo adottato vi sono alcune valutazioni di natura empirica sulle caratteristiche dei non rispondenti. È stato osservato infatti che "la difficoltà ad ottenere l'intervista è crescente al crescere dell'ampiezza demografica del comune di residenza e al diminuire del numero di componenti della famiglia perché spesso non si riesce proprio ad effettuare il contatto; le famiglie con persona di riferimento straniera hanno minori probabilità di partecipare all'indagine, in parte per l'alta mobilità degli stranieri sul territorio, in parte per le difficoltà o la diffidenza che possono avere persone di lingua e cultura differenti rispetto agli intervistatori. Si riscontrano inoltre notevoli differenze nella partecipazione all'indagine a seconda della "regione di residenza" (Ceccarelli e Cuttillo, 2006). Sulla base di queste considerazioni, la popolazione delle famiglie residenti è stata classificata sulla base di questi criteri:

1. dimensione demografica del comune (5 modalità);

2. cittadinanza della persona di riferimento (2 modalità);
3. regione di residenza (21 modalità essendo le province autonome di Trento e Bolzano considerate separatamente);
4. numero di componenti (4 modalità)

Per ciascuna delle celle è stato valutato il numero di rispondenti effettivo  $m_z^r$  e il numero di rispondenti teorico,  $m_z$  ( $z=1,\dots,Z$ ), sulla base del campione estratto. La correzione della non risposta avviene, in linea di principio, moltiplicando il “peso base”  $w_i^{base}$ , definito come l’inverso della probabilità di inclusione ( $\pi_i$ ), per il fattore  $\eta_z = \frac{m_z}{m_z^r}$ .

In pratica quando la numerosità  $m_z^r$  in alcune celle è al di sotto di una certa soglia, si procede ad un collasso mento delle celle stesse per ottenere correzioni numericamente più stabili.

Questa procedura di riponderazione finalizzata alla correzione degli effetti della non-risposta si basa sull’ipotesi che le famiglie all’interno di ciascuna cella abbiano una probabilità di risposta uniforme, ovvero che i non-rispondenti formino, all’interno di ciascuna cella, un sotto-campione casuale. Benché la tecnica delle celle di riponderazione sia stata sviluppata indipendentemente dalla teoria degli stimatori per calibrazioni, può essere ricondotta a quest’ultima, nel caso in cui i totali a cui si calibra siano quelli campionari (si vedano a questo riguardo Särndal e Lundström (2005)).

Riassumendo la procedura di correzione della non risposta porta alla produzione di un

set di “pesi intermedi”:  $w_i^{int} = \frac{1}{\pi_i} \frac{m_z}{m_z^r}$ ,  $i \in s$

dove  $s$  indica l’intero campione trasversale.

Riguardo alla calibrazione a totali noti, ognuno dei quattro campioni longitudinali è vincolato alla:

1. popolazione residente per ripartizione territoriale, sesso e classe d’età al 31/12 dell’anno di riferimento del reddito.
2. il numero di famiglie residenti per regione riferite alla fine dell’anno di rilevazione

Inoltre i pesi relativi all’intero campione trasversale, che sono poi quelli che ci interessano, sono vincolati a rispettare i seguenti totali:



1. popolazione residente per regione, sesso e classi d'età alla fine dell'anno di riferimento del reddito;
2. popolazione residente per regione, sesso e classi d'età alla fine dell'anno di rilevazione;
3. popolazione residente straniera maggiorenne per ripartizione geografica e sesso alla fine dell'anno di riferimento del reddito;
4. popolazione residente straniera maggiorenne per ripartizione geografica e provenienza (UE e non UE) alla fine dell'anno di riferimento del reddito

Le stime dirette ottenute applicando i pesi pubblicati da ISTAT insieme ai dati di EU-SILC per l'Italia presenta aspetti problematici. I domini rappresentati dai distretti non sono infatti "domini pianificati" e sommando i pesi relativi alle unità del campione incluse nel dominio non si ottiene la dimensione esatta della popolazione (noto attraverso le anagrafi); inoltre se si considerano sotto-campioni classificando per sesso, età o altre variabili, il loro peso relativi in termini di somma dei pesi può essere diverso da quello reale nella popolazione del distretto. La calibrazione è infatti effettuata a livello regionale, il più dettagliato per cui ISTAT pubblica ufficialmente le stime .

Al fine di ottenere un sistema di pesi che, all'interno di ciascun dominio riporti alla effettiva dimensione della popolazione, il processo di calibrazione dei pesi può essere modificato introducendo informazioni ausiliarie aggiuntive e facendo in modo che la somma dei pesi all'interno di ciascun dominio sommi alla dimensione della sua popolazione così come risulta dalle anagrafi comunali. Nella stessa direzione si può richiedere che i pesi dei sotto-campioni di distretto ottenuti classificando per sesso, classi d'età e nazionalità dei rispondenti sommino alla dimensione che queste sottopopolazioni hanno a livello di distretto. Questa idea, attraente in linea di principio, richiede in pratica che i campioni per cui si ricalibrano i pesi siano grandi. Una ricalibrazione a livello di piccoli domini può porre i pesi "sotto stress", aumentandone in modo anche cospicuo la variabilità. L'obiettivo di stimatori con una potenziale minore distorsione, viene così ottenuto al prezzo di un forte incremento della variabilità degli stessi, secondo un "effetto disegno" ben noto in letteratura.

Una soluzione di compromesso, considerata in una versione preliminare di questo rapporto, può essere quella di utilizzare le province come livello di calibrazione, al posto dei distretti. Tecnicamente, ciò potrebbe portare ad un sistema di pesi che riporti ad una popolazione più "vicina" a quella del distretto, limitando tuttavia l'effetto di inflazione della varianza dovuto alla calibrazione di sotto-campioni numericamente molto piccoli. Un'analisi accurata ha rilevato tuttavia che questa soluzione non è soddisfacente: le differenze di struttura nelle popolazioni sono in genere più forti tra i

distretti della stessa provincia che non tra una provincia e l'altra, per cui si finisce con l'ottenere pesi simili per struttura a quelli ricalibrati a livello regionale ma più variabili e che comunque non riportano ai totali di distretto e non rappresentano accuratamente la struttura della loro popolazione.

### **A1. 5 Domini pianificati e domini di interesse**

L'obiettivo principale di EU-SILC consiste nello stimare i principali indicatori relativi alla povertà e alla distribuzione del reddito per il paese nel suo complesso e per le sue Regioni (comprese le province autonome di Trento e Bolzano) con un livello di precisione sufficientemente elevato.

Le Regioni e le due province autonome citate rappresentano quindi i domini pianificati dell'indagine EU-SILC.

L'oggetto di studio di questo rapporto sono i distretti socio-sanitari e le provincedella Regione Emilia-Romagna. I distretti sono ripartizioni geografiche di dimensioni ridotte, che non solo non rappresentano domini pianificati per l'indagine ma per cui la dimensione campionaria disponibile per la stima dei parametri è in genere molto ridotta e talvolta del tutto assente.

Nella Tabella A.2 sono riportate le numerosità campionarie (in termini di famiglie rispondenti) per ciascuno dei 38 distretti socio-sanitari della regione Emilia-Romagna per l'anno di indagine 2013. Si va da un minimo di 0 (che è osservato in 6 distretti) ad un massimo di 149. La dimensione campionaria mediana è 35 (primo quartile 14, terzo quartile 60.5) mentre quella media è di 38.3. Si tratta di numeri nella maggior parte inadeguati a garantire stime di efficienza adeguata adottando i dati forniti dall'indagine come sola base informativa. Gli stessi numeri, calcolati per gli anni di indagine precedenti, non si discostano di molto e portano a conclusioni analoghe sulla possibilità di utilizzare EU-SILC come sola base informativa.

Tabella A.2 *Campione IT-SILC Regione Emilia Romagna. Famiglie rispondenti per distretto. anno di indagine 2013.*

Descrizione	n. fam.	Descrizione	n. fam.
Citta' di Bologna	149	Scandiano	35
Ferrara Centro-Nord	94	Bologna Pianura Ovest	33
Bologna Pianura Est	85	Lugo	30
San lazzaro di Savena	84	Faenza	28
Rimini	83	Casalecchio di Reno	25
Reggio Emilia	77	Ferrara Sud-Est	20
Modena	71	Castelnuovo ne' Monti	18
Ravenna	68	Parma Sud Est	16
Forli'	67	Correggio	14
Parma	65	Guastalla	14
Città di Piacenza	47	Pavullo nel Frignano	13
Piacenza Levante	44	Rubicone	12
Cesena - Valle del Savio	40	Mirandola	5
Carpi	39	Ferrara Ovest	0
Castelfranco Emilia	37	Montecchio	0
Vignola	37	Piacenza ponente	0
Fidenza	36	Porretta Terme	0
Imola	35	Riccione	0
Sassuolo	35	Valli Taro e Ceno	0

## **A2. Reddito equivalente. Suo trattamento statistico**

La maggior parte degli indicatori di povertà e disuguaglianza economica che formano l'oggetto di questo studio sono calcolati sulla base del reddito disponibile equivalente. In termini informali, "disponibile" significa effettivamente utilizzabile da parte delle famiglie per l'acquisto di beni e servizi. Calcolare il reddito equivalente significa calcolare un reddito pro-capite per i membri del nucleo familiare in cui però il numero dei componenti non consiste nel semplice conteggio ma viene corretto per le economie di scala che realizzano convivendo e mettendo insieme le risorse. Il reddito equivalente si ottiene dividendo il reddito disponibile totale di una famiglia per un coefficiente (dimensione equivalente) che tiene in conto il numero dei componenti e composizione per età della famiglia. Il reddito disponibile totale della famiglia è la somma dei redditi netti percepiti dai suoi membri ed è la proxy che viene utilizzata per descrivere l'insieme delle risorse che le famiglie hanno a disposizione per raggiungere i propri obiettivi di benessere.

Nella tabella A.3 riportiamo la definizione dettagliata dell'aggregato "reddito disponibile totale familiare", facendo esplicito riferimento a quali variabili nel database di EU-SILC sono state coinvolte nel calcolo. Tale definizione accoglie in modo sostanzialmente fedele le raccomandazioni del manuale di Camberra.

Tabella A.3 *Definizione di reddito netto familiare*

<p>Addendi positivi</p>	<p>reddito lordo da lavoro dipendente di tutti i componenti della famiglia (sia monetario, sia quasi-monetario (PY010G), sia non monetario (PGY020G);                  contributi sociali a carico dei datori di lavoro (PY030G);                  guadagni o perdite lorde da lavoro autonomo compresi diritti e royalties (PG050G);                  valore dei beni prodotti ai fini dell'autoconsumo (PY070G);                  sussidi e assegni di disoccupazione (PY090G);                  pensioni di anzianità (PY100G), pensioni di reversibilità ai superstiti (PY110G), indennità di malattia (PY120G), invalidità (PY130G); borse di studio, premi ed altre sovvenzioni relative all'istruzione (PY140G);                  componenti di reddito lordo attribuibili alla famiglia nel suo complesso: fitto figurativo (HY030G), redditi derivanti dalla locazione di beni o terreni (HY040G), indennità inerenti alla famiglia e ai figli (HY050G), prestazioni sociali erogate a persone a rischio di emarginazione sociale e non classificate altrove (HY060G), indennità di alloggio erogate dagli enti pubblici (HY070G), trasferimenti attivi di denaro tra famiglie quali assegni ricevuti dal coniuge separato (HY080G);                  interessi, dividendi, utili da investimenti di capitale in imprese non costituite in società (HY090G);                  reddito percepito da membri della famiglia di età inferiore ai 16 anni (HY110G)</p>
<p>Addendi negativi</p>	<p>contributi sociali a carico dei datori di lavoro (PY030G);                  interessi pagati sui mutui ipotecari (HY100G);                  imposte correnti sul patrimonio (HY120G);                  trasferimenti passivi di denaro tra famiglie quali assegni ricevuti dal coniuge separato (HY130G);                  imposte sul reddito e contributi sociali a carico del lavoratore al netto di rimborsi/entrate per conguagli fiscali, le imposte sul reddito alla fonte (HY140G)</p>

Occorre notare come ISTAT attui una serie di misure volte a contenere gli effetti della non risposta, degli errori e dell'*under-reporting* sul calcolo del reddito. Nel caso di non-risposta totale da parte di uno dei componenti della famiglia si procede all'imputazione del dato mancante sulla base di modelli statistici (ISTAT, 2008). Per quanto riguarda errori ed *under-reporting*, quest'ultimo per varie ragioni particolarmente rilevante nel caso di percettori di reddito da lavoro autonomo, il reddito risultante dal questionario IT-SILC viene confrontato con quello risultante dalle dichiarazioni fiscali. Nell'ottica della minimizzazione dell'ammontare dell'*under-reporting* viene imputato il valore massimo tra quello che figura nel questionario e quello fiscalmente dichiarato.

Chiaramente questo tipo di controlli è reso possibile dal livello di efficacia raggiunto dalle procedure di *linkage* tra questionari IT-SILC e dati di fonte fiscale.

Notiamo infine che, per effetto della definizione adottata è possibile osservare redditi famigliari netti negativi. Tra gli addendi positivi non sono inclusi poste quali “vincite alla lotteria”, “guadagni in conto capitale”, “aumenti delle disponibilità monetarie in conseguenza di denaro preso in prestito” classificate come variazioni istantanee della ricchezza e non come componenti di reddito. D’altra parte, tra gli addendi negativi sono inclusi trasferimenti ad altre famiglie (e tra essi gli assegni di mantenimento del coniuge separato) nonché le perdite per attività di lavoro autonomo.

## A2.1 Scala di equivalenza e reddito disponibile equivalente

La dimensione equivalente della famiglia è stata calcolata utilizzando la scala di equivalenza OCSE modificata.

$$EQ\_SIZE = 1 + 0.5(CF_{14+} - 1) + 0.3(CF_{13-})$$

dove  $CF_{14+}$  è il numero di componenti della famiglia con età  $\geq 14$  e  $CF_{13-}$  il numero di quelli età inferiore.

Il reddito disponibile equivalente di ciascun individuo incluso nel campione è stato quindi calcolato come rapporto tra il reddito disponibile totale della famiglia di cui è membro diviso per il fattore di equivalenza. Si noti che, come conseguenza di questa definizione tutti i membri della stessa famiglia hanno lo stesso reddito disponibile equivalente. La distribuzione campionaria del reddito disponibile equivalente è caratterizzato da un’asimmetria molto forte. La tabella A.4 riporta alcuni dei percentili di questa distribuzione:

Tabella A.4 *Alcuni percentili della distribuzione campionaria del reddito equivalente (campione nazionale), anno di indagine 2013*

Percentile	Valori reddito.
1	886 415
0.995	85 350
0.99	67 496
0.95	39 883
0.75	23 168
...	...
0.01	542
0.005	0
0	-19 065

E' ragionevole ritenere che i redditi molto alti osservati non siano dovuti ad errori nella misurazione ma rappresentino dati corretti. Tuttavia occorre tenere in conto che la loro influenza sulla stima di parametri quali medie e indici di concentrazione può essere molto grande. In particolare lo sarà quando la stima viene calcolata a livello di domini di piccole dimensioni, in cui la presenza di outliers non è in alcun modo compensata dall'elevato numero di osservazioni. Inoltre proprio per la loro rarità questi valori anomali si distribuiranno in modo casuale in alcune, ma non in tutte le aree, finendo con l'avere un effetto distorsivo su eventuali confronti tra le stime dello stesso parametro in domini diversi.

Riguardo alla produzione delle stime "dirette" si sono poste tre alternative di azione a riguardo di questi *outliers*:

1. non correzione. Si tratta dell'opzione meno "invasiva" e che mantiene inalterato il dataset fornito da ISTAT. Ciò tuttavia comporta il peggioramento delle proprietà statistiche degli stimatori e ne rende più difficoltosa l'interpretazione;
2. fissazione di una soglia arbitraria e: 2a) eliminazione delle osservazioni il cui valore del reddito disponibile equivalente eccede la soglia; 2b) per le osservazioni per cui il valore del reddito disponibile equivalente eccede la soglia, esso viene imputato al valore della soglia (winsorizzazione);
3. fissazione di una soglia arbitraria e modellazione parametrica della coda della distribuzione sia per le aree che presentano outliers, sia per quelle che ne sono prive. Un simile modo di trattare il problema, con riferimento all'indagine EU-SILC è descritto ad esempio in Balog e Thornburn (2007).

Nel presente lavoro abbiamo scelto la via del trattamento degli outliers mediante una winsorizzazione della coda destra della distribuzione e una modellazione paretiana della coda della distribuzione. Sono stati imputati a zero 140 redditi equivalenti (individuali) pari a 0 e 143 redditi equivalenti che risultavano negativi.

La modellazione della coda destra della distribuzione consiste nell'individuazione di una soglia oltre la quale si assume che la distribuzione di Pareto fornisca un modello adatto per i dati e successivamente nella stima del parametro di forma che caratterizza questa distribuzione. Per una motivazione dettagliata di questa scelta si rimanda alla letteratura e in particolare ad Van Kerm (2007) e Alfons e Templ (2013) che l'hanno proposta proprio nell'analisi dei dati EU-SILC. La soglia è stata individuata – a livello nazionale, separatamente per ciascun anno di indagine – utilizzando il metodo di Van Kerm (Van Kerm, 2007), mentre il parametro di forma che caratterizza la distribuzione è stato stimato con il metodo del "partial density component" (Vandewall et al., 2007). Per i calcoli è stato utilizzato il pacchetto Laeken di R (Alfons e Templ, 2013). Alcuni aspetti tecnici relativi alla modellazione della coda destra sono riassunti nella tabella A.5

Tabella A.5 Dimensione del campione (individui), soglia oltre cui si è modellata la distribuzione utilizzando la distribuzione di Pareto, n. di osservazioni oltre la soglia, per ciascun anno di indagine.

Anno	Dim. campione (nazionale)	Soglia imputazione Pareto	Parametro shape	N. di osservazioni oltre la soglia
2008	52 433	44 218	3.10	611
2009	51 196	45 046	4.25	665
2010	47 551	46 990	2.95	536
2011	47 841	45 663	4.63	713
2012	47 365	46 186	3.24	753
2013	44 622	48 286	2.79	554

### A3. Definizione degli indicatori e formule utilizzate nella produzione delle stime dirette.

L'obiettivo di questo rapporto è stimare, per ciascuno dei distretti e per ogni provincia della regione Emilia-Romagna misure della prevalenza della povertà monetaria e della sua intensità, una misura non monetaria della povertà, nonché una misura di concentrazione della distribuzione del reddito. Più nel dettaglio gli indicatori di povertà e disuguaglianza considerati in questo rapporto sono:

- Incidenza del rischio di povertà (at risk of poverty rate o “tasso di povertà”);
- Il poverty gap mediano;
- Il tasso di deprivazione materiale;
- L'indice di concentrazione di Gini.

I primi tre rappresentano misure di povertà, incluse nell'insieme che Eurostat chiede ai paesi membri di monitorare proprio sulla base dell'indagine EU-SILC. Il quarto, in quanto misura di concentrazione e quindi di disuguaglianza, fornisce una misura indiretta di coesione sociale. I primi due e il quarto sono basati sul reddito equivalente, mentre il terzo è un indicatore non monetario della povertà in quanto prescinde dal reddito. Per praticità illustreremo prima le misure basate sul reddito equivalente per tornare successivamente sul tasso di deprivazione materiale.



### A3.1 Misure di povertà basate sul reddito equivalente

Per semplicità indichiamo la variabile reddito equivalente con  $Y$  e con  $y_{dj}$  il valore che essa assume in corrispondenza dell'individuo  $j$ -esimo incluso nel dominio  $d$ -esimo. Ricordiamo che  $d = 1, \dots, 38$ .

I parametri considerati sono definiti nel modo seguente:

- Tasso di povertà (at risk of poverty rate):

$$TP_d = \frac{\sum_{j=1}^{N_d} \mathbf{1}(y_{jd} \leq PTN)}{N_d}$$

La soglia di povertà relativa  $PTN$  coincide con il 60% del reddito disponibile equivalente mediano nazionale. Notiamo che  $\mathbf{1}(y_{jd} \leq PTN)$  è una funzione indicatrice che assume valore 1 se  $y_{jd} \leq PTN$ , 0 altrimenti.

- Poverty gap mediano:

$$GM_d = \frac{PTN - RPCMe_d(y_{POOR,N})}{PTN} \cdot 100$$

$y_{POOR,N} = \{y_{di} \mid y_{di} \leq PTN\}$  è il vettore dei redditi degli individui con reddito disponibile equivalente inferiore alla soglia di povertà e  $RPCMe_d(y_{POOR,N})$  è la mediana dei loro redditi. In generale il reddito pro-capite mediano è definito come  $RPCMe_d = y_{\left(\frac{N_d+1}{2}\right)}$ . Dove  $y_{(1)}, y_{(2)}, \dots, y_{(N_d)}$  rappresenta la successione dei redditi disponibili equivalenti del dominio di studio ordinati in modo non decrescente, ovvero:  $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(N_d)}$ .

- Indice di concentrazione di Gini

$$G_d = \frac{1}{N_d^2 RPC_d} \sum_{i=1}^N (2i-1-n) y_{(i)}$$

dove  $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(N_d)}$  è la successione in modo crescente dei valori di  $y$  all'interno del dominio.

### A3.2 Tasso di deprivazione materiale

La deprivazione materiale è definita come una situazione di involontaria incapacità di sostenere spese per determinati beni o servizi. Si considerano i seguenti segnali di deprivazione, rilevati tramite l'indagine EU-SILC:

1. essere in arretrato nel pagamento di almeno una tra le bollette di gas, luce, acqua, oppure essere in arretrato nel pagamento dell'affitto, mutuo o altro prestito;
2. non potersi permettere di riscaldare adeguatamente l'abitazione di residenza;
3. incapacità di affrontare spese impreviste ( fissate nella misura di 800 euro) con risorse proprie;
4. incapacità di fare un pasto adeguato almeno una volta ogni due giorni;
5. impossibilità di permettersi una settimana di ferie lontano da casa una volta l'anno;
6. non potersi permettere un televisore a colori;
7. non potersi permettere il frigorifero;
8. non potersi permettere l'automobile;
9. non potersi permettere il telefono;

Il concetto di "pasto adeguato ogni due giorni" si traduce nella domanda: *“La sua famiglia, se volesse, potrebbe permettersi di mangiare carne, pollo o pesce (o equivalente vegetariano) almeno una volta ogni due giorni?”*

Il tasso di deprivazione materiale è definito come la frazione di persone che vivono in famiglie con almeno 3 segnali di deprivazione materiale. I principi che ispirano la definizione di questo indicatore sono: *i)* costruire un indicatore che prescinda dai problemi di misurazione implicati dal riferimento ad una misura quantitativa della capacità come il reddito; *ii)* introdurre una forte componente soggettiva nella misurazione della povertà. E' infatti evidente come la risposta si/no alle domande a cui si riferiscono i vari segnali di deprivazione facciano riferimento a come il soggetto rispondente percepisca la propria situazione, più che ad un tentativo di approfondire se questi segnali di deprivazione corrispondano effettivamente alla realtà. Notiamo infine che il tasso di deprivazione materiale mantiene la famiglia come orizzonte fondamentale per la definizione dello stato di povertà: le risposte vengono fornite da un solo rispondente tra i componenti del nucleo intervistato.

Formalmente:

$$\text{TDM}_d = \frac{1}{N_d} \sum_{j=1}^{N_d} \mathbf{1}(r_{dj} \geq 3)$$

dove  $r_{dj}$  è il numero di segnali di deprivazione pari a 1 per la famiglia di cui è componente l'individuo  $j$ -esimo nel dominio  $d$ .

### A3.3 Formule stimatori diretti

Passiamo ora ad illustrare gli stimatori “diretti” dei parametri ora introdotti. Per stimatore “diretto” si intende, nel gergo delle piccole aree uno stimatore di un parametro descrittivo di un dominio di interesse che utilizza l'informazione sulla variabile obiettivo  $y$  relativa al solo dominio in questione. In pratica con stimatori diretti si indicano stimatori con proprietà ottimali rispetto al disegno campionario, ossia rispetto alla distribuzione di probabilità indotta dalla selezione casuale del campione. Gli stimatori che ora introdurremo sono tutti consistenti rispetto al disegno, nel senso che a questo termine viene attribuito nella letteratura sulla teoria dei campioni (si veda ad esempio Isaki e Fuller, 1982).

Assumiamo di denotare con  $\tilde{w}_{dj}$  il peso “finale” ricalibrato secondo la procedura illustrata nel paragrafo A1.4 per l'individuo  $j$ -esimo nel dominio di interesse  $d$ -esimo. Riportiamo di seguito le formule degli stimatori “diretti”; con un abuso di notazione utilizziamo gli stessi simboli per denotare il parametro da stimare e lo stimatore diretto che utilizziamo per stimarlo.

- Tassi di povertà:  $\hat{\text{TP}}_d = \frac{\sum_{j \in s_d} \tilde{w}_{dj} \mathbf{1}(y_{jd} \leq PTN)}{\sum_{j \in s_d} \tilde{w}_{dj}}$

(indichiamo con  $s_d$  l'insieme delle unità campionate nel dominio  $d$ )

- Poverty gap mediano:

$$\hat{GM}_d = \frac{PTN - \text{med}(y_{POOR,N})}{PTN} \cdot 100$$

dove con  $RPCMe_d(y_{POOR,N})$  si indica lo stimatore consistente rispetto al disegno della mediana nazionale del reddito equivalente disponibile. Più nel dettaglio Reddito pro-capite mediano:  $RPCMe_d = y_{(j)}$  tale che  $j$  è l'indice minimo per cui

$$\frac{\sum_{i=1}^j \tilde{w}_{di}}{\sum_{i \in s_d} \tilde{w}_{di} y_{di}} \geq 0.5$$

- Indice di concentrazione di Gini:

$$\hat{G}_d = \left( \frac{2 \sum_{i=\text{poorest person}}^{\text{richest person}} \left( \tilde{w}_{dj} y_{dj} \cdot \sum_{i=\text{poorest person}}^{\text{person } i} \tilde{w}_{dj} \right) - \sum_{i=\text{poorest person}}^{\text{richest person}} \tilde{w}_{dj}^2 y_{dj}}{\left( \sum_{j=\text{poorest person}}^{\text{person } i} w_{dj} \right) \cdot \sum_{j=\text{poorest person}}^{\text{richest person}} \tilde{w}_{dj} y_{dj}} - 1 \right) \cdot 100$$

- Tasso di deprivazione materiale:

$$\hat{\text{TDM}}_d = \frac{\sum_{j \in s_d} \tilde{w}_{dj} \mathbf{1}(r_{dj} \geq 3)}{\sum_{j \in s_d} \tilde{w}_{dj}}$$

dove  $r_{dj}$  è il numero di segnali di deprivazione pari a 1 per la famiglia di cui è componente l'individuo  $j$ -esimo nel dominio  $d$ .

## **A4. Stima della varianza.**

### **A4.1 Discussione del bootstrap utilizzato e suoi limiti nel contesto attuale (impossibilità di utilizzare il comune come PSU).**

Il piano di campionamento di EU-SILC è complesso. Si tratta di un panel ruotato in cui il 75% delle famiglie campionate è composto da famiglie selezionate per la prima volta 3, 2 e l'anno precedente, mentre il 25% proviene da una nuova selezione. Questa struttura è stata pensata per poter permettere studi di tipo longitudinale, di cui però questo studio non si occupa. La parte nuova del campione ("fresh" nel gergo tecnico) è selezionata con un campionamento stratificato a due stadi, le unità di primo stadio sono i comuni, quelle di secondo stadio le famiglie. I comuni sono stratificati per regione amministrativa e dimensione demografica, come in altre indagini ISTAT. I comuni oltre una certa soglia dimensionale sono considerati auto-rappresentativi (campionati con probabilità 1). Il quadro del campionamento è complicato ulteriormente dalle regole di inseguimento delle famiglie che rimangono nel campione anche se cambiano comune di residenza e delle famiglie gemmate da quelle che fanno parte della parte che rimane tra due waves successive. Le probabilità di inclusione sono poi modificate per calibrazione per correggere la potenziale distorsione indotta da non-risposta e attrito. Per queste ragioni nessuna tecnica, né di stima diretta, né di ricampionamento standard può essere utilizzata per ottenere stime affidabili delle varianze. In linea con il precedente rapporto l'idea è quella di utilizzare un bootstrap semplificato, validarlo e utilizzare le stime ottenute come input per algoritmi di lisciamento della varianza. La principale novità rispetto al rapporto precedente consiste nel considerare la provincia e non il distretto come riferimento nella costruzione del bootstrap semplificato.

## A4.2 Stima della varianza. Bootstrap approssimato

Abbiamo utilizzato un bootstrap in cui i campioni replicati sono stati ottenuti semplicemente campionando con ripetizione all'interno delle province. Questa approssimazione trascura alcuni aspetti del disegno e principalmente la sua struttura a due stadi: si lavora infatti come se le province coincidessero con gli strati dell'indagine EU-SILC e si campionano direttamente famiglie al loro interno, senza tener conto che il disegno prevede il campionamento di comuni e delle famiglie all'interno dei comuni campionati. Un potenziale sorgente di distorsione nell'utilizzo di stimatori ottenuti sulla base di questa impostazione consiste nel trascurare l'effetto di correlazione intra-cluster. In altre parole, non si tiene in conto che le famiglie campionate all'interno dello stesso comune possono essere tra loro più simili tra loro rispetto all'insieme delle famiglie dell'intera provincia. Ci sono fondate ragioni tuttavia, per ritenere che questo effetto sia molto moderato.

La stima *bootstrap* della varianza di un generico parametro  $\theta$  stimato da  $\hat{\theta} = \hat{\theta}(\text{data})$  si effettua estraendo  $B$  campioni *bootstrap*, calcolando lo stesso stimatore per ciascuno dei campioni replicati:  $\hat{\theta}^*(b)$ ,  $b = 1, \dots, B$  e calcolando quindi la media delle stime così ottenute:  $\hat{\theta}^* = B^{-1} \sum_{b=1}^B \hat{\theta}^*(b)$ . La varianza bootstrap dello stimatore è quindi ottenuta applicando questa formula:

$$\hat{V}_{boot}(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^B \left( \hat{\theta}^*(b) - \hat{\theta}^* \right)^2$$

Sotto le ipotesi riportate in precedenza, il campione *bootstrap* è ottenuto estraendo con ripetizione un campione casuale di  $n_d - 1$  famiglie dalle  $n_d$  presenti nel campione estratto per il  $d$ -esimo dominio.

Un applicazione coerente della formula dello stimatore  $\hat{\theta}$  richiede che, se questo coinvolge pesi campionari calibrati a totali noti, questa operazione di ricalibrazione venga effettuata nuovamente per i pesi *bootstrap* base. Questi ultimi sono definiti in questo modo:

$$w_{dj}(b) = w_{hzi}^{\text{int}} \frac{n_d}{u_d - 1} n_j(b)$$

$w_{hzi}^{\text{int}}$  è il peso campionario intermedio (ossia quello base aggiustato per la sola non-risposta) descritto nel paragrafo A.2, relativo alla famiglia  $j$  all'interno del dominio  $d$  e  $n_d(b)$  è il numero di volte in cui essa figura nel campione *bootstrap*  $b$ .

Lo stimatore utilizzato non è direttamente riconducibile a risultati formali noti in letteratura, che ci permettano di descriverne le proprietà. Per validare le stime delle varianza ottenute con il metodo *bootstrap* descritto presentiamo una semplice analisi

relativa al parametro  $TP_d$  (at-risk-of-poverty rate). Sappiamo che se la soglia di povertà fosse nota in modo esogeno e non stimata sui dati e se il piano di campionamento fosse casuale semplice con reintroduzione allora

$$V\left(\hat{TP}_d\right) = V_d = \frac{TP_d(1-TP_d)}{n_d^*}.$$

dove  $n_d^*$  è il numero di individui campionate nel dominio. Ricordiamo infatti che il tasso di rischio di povertà è definito in termini di individui e non già di famiglie (che sono invece le unità di riferimento per il campionamento).

Se consideriamo le varianze degli stimatori che calcoliamo sulla base delle formule descritte nel paragrafo A3.3, possiamo ancora aspettarci che approssimativamente

$$\frac{\hat{TP}_d\left(1-\hat{TP}_d\right)}{\hat{V}_{d,boot}} \propto n_d^*$$

Tuttavia varie ragioni renderanno questa relazione soltanto approssimativa: *i)* il campionamento è per cluster (le famiglie) e non per individui. Tutti i membri di una famiglia sono poveri o non poveri allo stesso modo, ma non tutte le famiglie hanno la stessa dimensione e la distribuzione delle famiglie per dimensione può variare da un dominio all'altro; *ii)* nel calcolo degli stimatori usiamo i pesi; *iii)* le soglie sono calcolate sulla base del campione stesso, ancorché a livello nazionale; *iv)* sia le stime puntuali  $\hat{TP}_d$ , sia  $\hat{V}_{d,boot}$  sono caratterizzate da un errore campionario non trascurabile.

Nella figura A.1  $n_d^*$  e  $\frac{1}{\hat{V}_{d,boot}}\left[\hat{TP}_d\left(1-\hat{TP}_d\right)\right]$  sono rappresentati in un diagramma a dispersione in cui ciascun punto corrisponde ad una provincia. L'anno di indagine a cui si riferiscono questi dati è il 2013. Alla nuvola dei punti è sovrapposta una retta di equazione  $y = \alpha x$  che descrive l'adattamento di una funzione lineare ai dati. Le stime del parametro  $\alpha$  e una misura di adattamento del modello lineare ai dati sono riportati nella tabella A.6

Dalla tabella A.6 possiamo notare in primo luogo come l'adattamento della retta ai dati sia in tutti gli anni di indagine ampiamente soddisfacente. Il coefficiente angolare, ampiamente inferiore a 1 (approssimativamente oscilla tra 0.24 e 0.28)

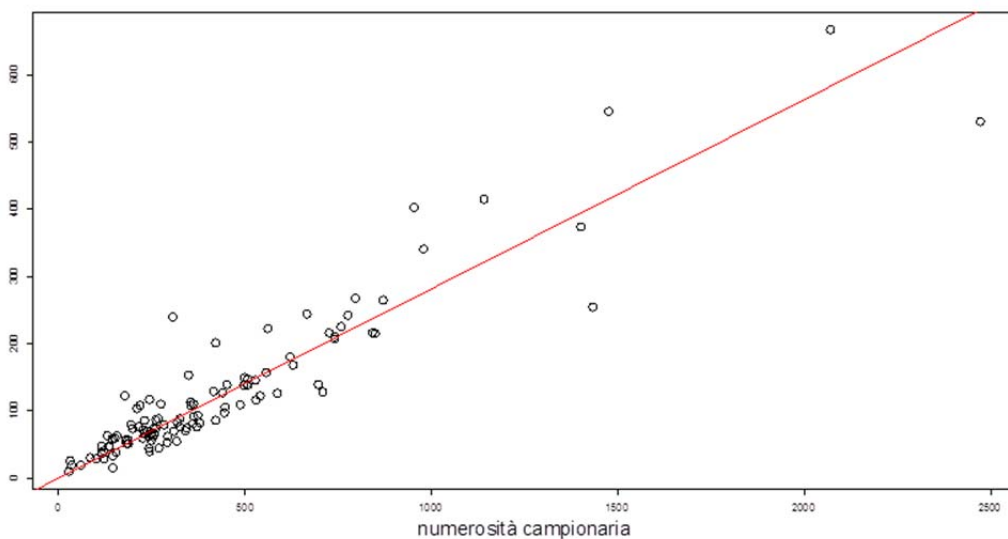
significa che  $\hat{V}_{d,boot} > \frac{\hat{TP}_d\left(1-\hat{TP}_d\right)}{n_d}$ ; ciò è coerente in primo luogo con la struttura a cluster del campionamento.

Tabella A.6 *Stima del parametro della retta passante per l'origine che esprime  $\frac{1}{\hat{V}_{d,boot}} \left[ TP_d \left( 1 - TP_d \right) \right]$  in funzione di  $n_d^*$  separatamente per ciascun anno di indagine*

Anno	Stima parametro	R <sup>2</sup>
2008	0.283	0.77
2009	0.269	0.79
2010	0.282	0.86
2011	0.251	0.87
2012	0.240	0.87
2013	0.236	0.80

In ogni famiglia ci sono in media 2.4 individui, ma per definizione l'informazione relativa alla povertà è la stessa. Se moltiplichiamo il coefficiente per la dimensione media della famiglia otteniamo un numero di poco inferiore a 0.7. La differenza tra questo e 1 (che avremmo se il campionamento fosse casuale semplice) è attribuibile al fatto che il disegno è comunque meno efficiente di quello casuale semplice per via del campionamento a due stadi e che soprattutto, l'introduzione dei pesi nelle formule degli stimatore se da un lato li rende potenzialmente meno distorti, dall'altro ne inflaziona varianza. Infine, una piccola componente aggiuntiva di incertezza è attribuibile alla soglia stimata sul campione (nazionale) e non completamente esogena.

Figura A.1 *Stima del tasso di povertà, numerosità campionaria di dominio  $n_d$  e rapporto  $\frac{1}{\hat{V}_{d,boot}} \left[ TP_d \left( 1 - TP_d \right) \right]$ . Anno di indagine 2013.*





Notiamo come il modello di variance smoothing adottato in questo studio sia, seppur simile, un po' diverso da quello implementato nei rapporti precedenti. Questo riflette le caratteristiche parzialmente diverse dello studio attuale e l'estensione della popolazione oggetto di stima a tutte le province italiane.

### A4.3 Stima della varianza. Effetto del disegno di Kish

Kish (1965) ha introdotto l'effetto disegno per studiare l'efficienza relativa di piani di campionamento complessi rispetto al campionamento casuale semplice. L'effetto disegno è dato dal rapporto tra la varianza di uno stimatore associato ad un piano di campionamento complesso e quella associata allo stimatore della stessa quantità ma basato sul campionamento casuale semplice. Alcune caratteristiche dei disegni campionari utilizzati nella pratica delle indagini, *in primis* la stratificazione, tendono a ridurre l'effetto disegno, mentre altre, quali il campionamento per clusters, hanno un effetto tipicamente, ancorché non necessariamente, inflattivo. Il ruolo dei pesi campionari è ambiguo: se essi sono correlati positivamente alla variabile oggetto di studio, riducono l'effetto disegno, mentre se sono incorrelati o correlati negativamente con essa, ingrandiscono la varianza. Correlazioni positive molto deboli o negative tra pesi e variabili d'interesse sono la regola nelle grandi indagini sociali: il piano di campionamento è basato su una stratificazione dettata da motivi pratici e dalla necessità di disporre di numerosità campionare adeguate in tutti i domini pianificati e solo raramente da criteri di efficienza degli stimatori. I pesi incorporano anche l'effetto della correzione per la non risposta, che è necessario per ridurre la distorsione degli stimatori, ma che introduce nei pesi una variabilità aggiuntiva e solitamente riduce la correlazione tra pesi e variabile oggetto di interesse.

Se consideriamo il reddito equivalente come variabile di riferimento, questa situazione caratterizza anche il campione EU-SILC 2013 per l'Italia.

Kish (1987) propone la seguente formula per determinare l'effetto disegno di uno stimatore lineare (di tipo Horwitz-Thompson) in presenza di un disegno campionario con stratificazione, campionamento a due stadi e probabilità di inclusione non costanti:

$$deff = [1 + CV^2(\mathbf{w})][1 + (\bar{b}_1 - 1)\rho_1][1 + (\bar{b}_2 - 1)\rho_2].$$

$CV^2(\mathbf{w})$  è il coefficiente di variazione al quadrato dei pesi campionari,  $\bar{b}_1$  la dimensione media dei clusters di primo livello (i comuni nel caso dell'indagine EU-SILC) e  $\rho_1$  il coefficiente di correlazione intracluster, una misura di similarità tra le unità di secondo stadio (nel nostro caso famiglie) campionate all'interno della stessa unità di primo stadio. Analogamente  $\bar{b}_2$  è la dimensione media dei cluster di secondo livello (le famiglie nel caso di EU-SILC) e  $\rho_2$  il relativo coefficiente di correlazione

intra-cluster. Per costruzione (tutte le famiglie hanno lo stesso reddito equivalente)  $\rho_2 = 1$ .

Sui nostri dati abbiamo proceduto a stimare il coefficiente di correlazione intra-cluster  $\rho_1$  sulla base di un semplice modello per il logaritmo del reddito equivalente familiare:

$$\log y_{dcj} = \mu_d + v_c + e_{cj}$$

dove  $y_{dcj}$  indica il reddito equivalente della  $j$ -esima unità del campione (famiglia) nel comune  $c$  all'interno del distretto socio-sanitario/della provincia  $d$ .  $\mu_d$  è un effetto fisso, mentre  $v_c$  e  $e_{cj}$  sono quantità casuali per cui si assume  $v_c \stackrel{ind}{\sim} N(0, \sigma_v^2)$ ,  $e_{cj} \stackrel{ind}{\sim} N(0, \sigma_e^2)$ ,  $E(v_c e_{cj}) = 0$ . Sulla base di questo semplice modello abbiamo stimato  $\hat{\rho}_1 = \hat{\sigma}_v^2 (\hat{\sigma}_v^2 + \hat{\sigma}_e^2)^{-1}$  dove le varianze  $\sigma_v^2$  e  $\sigma_e^2$  sono stimate con il metodo REML. La forma dello stimatore del parametro  $\rho_1$  si giustifica notando che: *i*) la covarianza tra due unità all'interno dello stesso comune è data da

$$\begin{aligned} Cov(\log y_{dcj}, \log y_{dcj'}) &= E(\log y_{dcj} \times \log y_{dcj'}) - E(\log y_{dcj}) \times E(\log y_{dcj'}) \\ &= E(\mu_d^2 + v_c^2 + e_{cj} e_{cj'} + \mu_d v_c + \mu_d e_{cj} + \mu_d e_{cj'} + v_c e_{cj} + v_c e_{cj'}) - \mu_d^2 \\ &= \mu_d^2 + \sigma_v^2 - \mu_d^2 = \sigma_v^2; \end{aligned}$$

*ii*) per la covarianza tra due unità in comuni diversi abbiamo che  $Cov(\log y_{dcj}, \log y_{dc'j'}) = 0$ ; *iii*)  $V(\log y_{dcj}) = \sigma_v^2 + \sigma_e^2$ . Possiamo quindi notare che

$$\rho_1 = Corr(\log y_{dcj}, \log y_{dcj'}) = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2}.$$

Abbiamo lavorato sulla trasformata logaritmica dei redditi per avvicinarci all'ipotesi di normalità distributiva richiesta dagli stimatori REML utilizzati nella stima delle varianze. La stima di questo parametro sui dati del campione EU-SILC per l'Emilia-Romagna è risultata estremamente modesta (inferiore a 0.01). Un valore così piccolo della correlazione tra osservazioni sulla scala logaritmica ci permette di affermare che questa sarà estremamente modesta anche sulla scala originale dei dati. Per questa ragione possiamo assumere che approssimativamente  $deff \cong [1 + CV^2(\mathbf{w})] \bar{b}_2$ . Il coefficiente di variazione dei pesi è stimato pari a  $est.deff = 4.05$  per l'anno di indagine 2013.

Notiamo che nel paragrafo precedente abbiamo dimostrato come fosse accettabile assumere  $\frac{1}{\hat{V}_{d,boot}} \left[ \hat{TP}_d (1 - \hat{TP}_d) \right] = \hat{\alpha} n_d^* + e_d$ .

Poiché  $\hat{\alpha} = \frac{1}{\hat{V}_{d,boot}} \frac{\left[ \hat{TP}_d (1 - \hat{TP}_d) \right]}{n}$ , il coefficiente  $\hat{\alpha}$  è interpretabile come una stima del reciproco dell'effetto disegno. Ora, la nostra stima di  $\alpha$  per il 2013:  $\hat{\alpha} = 0.236$  è in sostanziale accordo con  $\frac{1}{est.deff} \cong 0.25$ . Il fatto che il reciproco dell'effetto disegno stimato sia leggermente più grande è coerente con il fatto che la stima di *est.deff* sia per difetto; seppur modesto, abbiamo infatti trascurato l'effetto del clustering delle unità di primo livello.

Quest'analisi sull'effetto disegno ci fornisce una sostanziale validazione dell'algoritmo bootstrap descritto nel paragrafo A4.2 e un metodo di stima alternativo per la varianza degli stimatori che non richiede il ricorso a metodi di ri-campionamento. Tuttavia poiché l'utilizzo del bootstrap permette di ottenere agevolmente stime degli intervalli di confidenza senza dover assumere la normalità distributiva degli stimatori nello spazio dei campioni, le varianze contenute in questo rapporto sono basate sulla metodologia illustrata nel paragrafo A4.2.

Tavola A1. *Stime dirette: tasso di povertà* , anno di indagine 2013

Codice	Descrizione distretto	Stima puntuale	CV
4	Parma	12.3	38.5
5	Fidenza	32.6	22.1
6	Valli Taro e Ceno	--	--
7	Parma Sud Est	5.2	178.7
8	Montecchio	--	--
9	Reggio Emilia	13.7	35.5
10	Guastalla	8.9	109.5
11	Correggio	21.8	42.6
12	Scandiano	6	89.1
13	Castelnuovo ne' Monti	1	--
14	Carpi	2.5	216.9
15	Mirandola	6.3	276.0
16	Modena	7.9	61.3
17	Sassuolo	25	26.7
18	Pavullo nel Frignano	1	--
19	Vignola	11.4	55.8
20	Casalecchio di Reno	4	186.2
21	Porretta Terme	--	--
22	San lazzaro di Savena	3.4	109.7
23	Imola	15.2	41.5
25	Bologna Pianura Est	1.9	176.8
26	Bologna Pianura Ovest	2.2	257.0
27	Citta' di Bologna	9.4	35.5
30	Ferrara Ovest	--	--
31	Ferrara Centro-Nord	4.3	90.3
32	Ferrara Sud-Est	9.4	101.4
35	Ravenna	7.5	60.8
36	Lugo	12	65.1
37	Faenza	18.2	46.8
38	Forli'	19.9	26.8
39	Cesena - Valle del Savio	20.6	32.4
40	Rimini	14.9	34.6
41	Riccione	--	--
44	Castelfranco Emilia	5.7	94.6

Tavola A2. *Stime dirette: Gap mediano, anno di indagine 2013*

Codice	Descrizione	Stima puntuale	CV
4	Parma		
5	Fidenza		
6	Valli Taro e Ceno		
7	Parma Sud Est		
9	Reggio Emilia		
11	Correggio		
12	Scandiano		
13	Castelnuovo ne' Monti		
14	Carpi		
15	Mirandola		
16	Modena		
17	Sassuolo		
18	Pavullo nel Frignano		
19	Vignola		
20	Casalecchio di Reno		
21	Porretta Terme		
22	San lazzaro di Savena		
23	Imola		
25	Bologna Pianura Est		
26	Bologna Pianura Ovest		
27	Citta' di Bologna		
31	Ferrara Centro-Nord		
32	Ferrara Sud-Est		
35	Ravenna		
36	Lugo		
37	Faenza		
38	Forli'		
39	Cesena - Valle del Savio		
40	Rimini		
41	Riccione		
44	Castelfranco Emilia		
49	Rubicone		
50	Citta' di Piacenza		
51	Piacenza Levante		

Tavola A3. *Stime dirette: Indice di concentrazione di Gini, anno di indagine 2013*

Codice	Descrizione	Stima puntuale	CV
4	Parma	32	13.4
5	Fidenza	32.7	16.5
6	Valli Taro e Ceno	--	--
7	Parma Sud Est	27.4	29.8
8	Montecchio	--	--
9	Reggio Emilia	34.7	11.6
10	Guastalla	26.8	28.4
11	Correggio	25.4	29.0
12	Scandiano	17.2	25.8
13	Castelnuovo ne' Monti	17	46.3
14	Carpi	22.1	20.8
15	Mirandola	22.3	58.1
16	Modena	30.5	14.4
17	Sassuolo	30.5	17.0
18	Pavullo nel Frignano	26	34.8
19	Vignola	29.6	17.4
20	Casalecchio di Reno	28.7	24.8
21	Porretta Terme	--	--
22	San lazzaro di Savena	32.3	12.2
23	Imola	26.8	19.7
25	Bologna Pianura Est	24.8	13.7
26	Bologna Pianura Ovest	22.1	23.3
27	Citta' di Bologna	32.2	9.7
30	Ferrara Ovest	--	--
31	Ferrara Centro-Nord	30	12.1
32	Ferrara Sud-Est	18.1	36.8
35	Ravenna	27.2	14.1
36	Lugo	26.1	23.2
37	Faenza	31.8	20.2
38	Forli'	29.5	13.3
39	Cesena - Valle del Savio	34.1	14.6
40	Rimini	30.8	11.9
41	Riccione	--	--
44	Castelfranco Emilia	21	21.2

Tavola A4. *Stime dirette: Tasso di deprivazione materiale, anno di indagine 2013*

Codice	Descrizione	Stima puntuale	CV
10	Guastalla	7.2	175.2
11	Correggio	14.4	83.7
12	Scandiano	15.2	49.7
13	Castelnuovo ne' Monti	4.6	279.6
14	Carpi	14	55.3
15	Mirandola	13.6	164.4
16	Modena	17.9	37.2
17	Sassuolo	27.1	32.6
18	Pavullo nel Frignano	4.3	342.1
19	Vignola	17.3	48.9
20	Casalecchio di Reno	23.1	45.3
21	Porretta Terme	--	--
22	San lazzaro di Savena	8.5	58.7
23	Imola	7	106.2
25	Bologna Pianura Est	6.7	69.9
26	Bologna Pianura Ovest	6.2	121.5
27	Citta' di Bologna	22.9	20.8
30	Ferrara Ovest	--	--
31	Ferrara Centro-Nord	4.8	95.2
32	Ferrara Sud-Est	9.4	123.0
35	Ravenna	13.2	44.1
36	Lugo	5	185.8
37	Faenza	10.2	98.7
38	Forli'	31.4	23.3
39	Cesena - Valle del Savio	22.3	37.0
4	Parma	23.5	27.1
40	Rimini	10.6	50.7
41	Riccione	--	--
44	Castelfranco Emilia	20.5	39.1
49	Rubicone	27.4	52.0
5	Fidenza	15.9	52.5
50	Città di Piacenza	13.3	56.0
51	Piacenza Levante	23.8	34.3
52	Piacenza ponente	--	--

Tavola A5. *Stime dirette: Reddito equivalente medio.*

Codice	Descrizione	Stima puntuale	CV
4	Parma	20867	8.3
5	Fidenza	20540	7.7
6	Valli Taro e Ceno	19384	11.0
7	Parma Sud Est	20010	12.3
9	Reggio Emilia	22868	7.4
11	Correggio	16787	13.4
12	Scandiano	21890	11.8
13	Castelnuovo ne' Monti	16286	14.5
14	Carpi	25273	15.3
15	Mirandola	23648	20.0
16	Modena	21937	9.6
17	Sassuolo	22199	12.5
18	Pavullo nel Frignano	21202	8.6
19	Vignola	22143	8.9
20	Casalecchio di Reno	24808	21.0
21	Porretta Terme	19715	14.2
22	San lazzaro di Savena	19555	12.3
23	Imola	19892	9.7
25	Bologna Pianura Est	23852	5.1
26	Bologna Pianura Ovest	24754	7.7
27	Citta' di Bologna	23178	6.2
31	Ferrara Centro-Nord	23906	6.4
32	Ferrara Sud-Est	18998	7.1
35	Ravenna	21783	5.9
36	Lugo	20835	7.2
37	Faenza	19122	5.5
38	Forli'	20776	6.5
39	Cesena - Valle del Savio	19417	9.4
40	Rimini	17766	6.8
41	Riccione	17073	7.8
44	Castelfranco Emilia	20030	7.1
49	Rubicone	17893	13.0
50	Citta' di Piacenza	21444	7.5
51	Piacenza Levante	19057	10.6

Nota: mancano le stime relative ai distretti socio-sanitari Montecchio, Guastalla, Piacenza ponente, Ferrara ovest in quanto non abbiamo per questi distretti nessun famiglia campionata.



## **Appendice B**

### **Produzione delle stime da modello e valutazione della loro affidabilità**

## **B1. Modelli per piccole aree. Introduzione**

L'inadeguata rappresentatività dei campioni EU-SILC disponibili per la maggior parte dei distretti socio-sanitari e delle province della regione Emilia-Romagna rende necessario l'utilizzo di modelli di stima per piccole aree in cui l'informazione fornita dal campione venga integrata con informazioni di natura amministrativa.

In questa breve introduzione, discutiamo alcune scelte di fondo da cui dipendono quelle più specifiche che saranno illustrate nei paragrafi successivi.

I modelli per piccole aree possono essere classificati in "area level" e "unit level". Nei modelli "area level" le stime dirette ottenute per ciascuno dei domini in questione (vedi appendice A) vengono messe in relazione con informazione ausiliaria riferita allo stesso livello di disaggregazione geografica o ad un livello superiore. Nei modelli "unit level" il valore delle variabili oggetto di studio rilevati per ciascuna unità statistica nel campione viene messa in relazione con valori di variabili anch'esse note a livello di ciascuna unità del campione.

A seconda dei problemi e delle informazioni disponibili possono essere preferibili gli uni o gli altri. Noi lavoreremo con modelli "area level" per le seguenti ragioni: i) l'insieme di informazioni disponibili a livello di individuo o famiglia e per cui i totali siano noti a livello di distretto o di provincia sono relativamente poche e non particolarmente esplicative; ii) i modelli "area level" permettono di ottenere stime consistenti rispetto al disegno. In altre parole, nei casi dei pochi domini per cui la dimensione del campione è adeguata, lo stimatore ottenuto sulla base del modello converge allo stimatore diretto; iii) i modelli "area level" sono, in termini di informazioni richieste, più parsimoniosi di quelli "unit level" quando le relazioni tra variabili obiettivo e variabili ausiliarie sono di tipo non lineare. Ai primi basta infatti che sia nota la media delle variabili ausiliarie a livello di dominio; ai secondi serve invece che il valore delle variabili ausiliarie sia noto per ciascun individuo / famiglia della popolazione.

Per quanto riguarda la produzione delle stime per i parametri di interesse per l'Emilia Romagna sia a livello sia di provincia che a livello di distretto, è necessario fare una premessa. In generale, come è motivato nel paragrafo B2, si è scelto di includere nella specificazione e stima dei modelli stime dirette e variabili ausiliarie relative alle province delle rimanenti regioni italiane.

Questa seconda appendice è organizzata nel modo seguente. Nel paragrafo B2 viene discussa la scelta di includere nella specificazione e stima dei modelli stime dirette e variabili ausiliarie relative alle province delle rimanenti regioni italiane. Nel paragrafo B3 presenteremo l'informazione ausiliaria presa in considerazione nel processo di specificazione dei modelli e il processo di selezione delle variabili che ha portato alla scelta delle variabili effettivamente incluse. I paragrafi da B4 a B6 contengono una descrizione dei modelli che sono stati utilizzati per produrre le stime

per piccole aree. Nel paragrafo B7 vengono presentate una serie di misure diagnostiche tese a validare i modelli scelti.

## **B2. Inclusione delle provincie delle rimanenti regioni italiane nel processo di stima.**

Sul piano logico, possiamo pensare ai modelli di tipo “area level” come a modelli in cui le osservazioni sono costituite dalle quantità descrittive dei domini oggetto di studio. Il numero dei domini definisce un sostanziale analogo della dimensione del campione nei problemi classici di statistica. Inoltre le quantità descrittive dei domini non sono note, ma soltanto stimate; in un certo senso possiamo pensare all’errore campionario associato alle stime dirette come ad un errore di misurazione che influenzi la nostra osservazione delle quantità descritte di popolazione.

Per queste ragioni, è semplice comprendere che nei modelli di tipo “area level” maggiore è il numero di aree su cui vengono stimati i parametri di interesse, maggiore sarà la precisione con cui questi ultimi sono stimati. In Emilia-Romagna vi sono 38 distretti socio-sanitari; ma le stime dirette sono disponibili solamente per 34, i campioni relativi ai rimanenti quattro distretti essendo vuoti. Ciò implica che per stimare un modello “area level” sono disponibili solamente 34 “osservazioni” caratterizzate da un elevato errore di campionamento. Un discorso analogo vale per le provincie dell’Emilia Romagna.

Abbiamo pertanto deciso di coinvolgere nel processo di specificazione e stima dei modelli anche le 101 provincie delle rimanenti regioni italiane.

Vi è più di una ragione che motivano tale decisione:

- quando l’obiettivo è la stima a livello di distretto, va osservato che in molti casi i distretti-socio sanitari istituiti in altre Regioni non sono paragonabili a quelli della regione Emilia-Romagna per dimensioni e ruolo
- le provincie rappresentano aree per le quali è in generale disponibile un’informazione campionaria maggiore di quella che è disponibile per i distretti; le stime dirette a livello provinciale risultano quindi più affidabili di quelle a livello di distretto socio-sanitario;
- la relazione tra parametri di interesse e informazione ausiliaria utilizzata nei modelli è più evidente se si utilizzano aree su tutto il territorio nazionale di quanto lo sia concentrando l’attenzione solo su sotto-aree del territorio (ad esempio i distretti dell’Emilia-Romagna); più è forte tale relazione tanto più le stime ottenute a livello di singole aree saranno precise.

Gli stimatori, sia diretti che da modello, del tasso di povertà e del tasso di deprivazione materiale sono additivi nel senso che la stima a livello provinciale si può ottenere come opportuna media aritmetica ponderata delle stime dei distretti che appartengono alla

stessa provincia; la proprietà di additività non vale nel caso dell'indice di Gini e del GAP mediano. Se vale la proprietà di additività si possono seguire due strade alternative per la produzione di stime da modello a livello provinciale: 1) utilizzare un modello per le sole province italiane; 2) utilizzare un modello stimato sui dati dei distretti della regione Emilia Romagna e delle province delle rimanenti regioni e poi ottenere le stime per provincia dell'Emilia Romagna sfruttando la proprietà di additività. Si è preferito seguire questa seconda strada sia perché garantisce un livello di coerenza maggiore tra stime dei distretti e quelle delle province, sia perché gli stimatori da modello ottenuti sono risultati più precisi di quelli ottenuti utilizzando modelli specificati per le sole province. Per i parametri che non godono della proprietà di additività sono stati utilizzati sia modelli stimati sui soli dati provinciali (al fine di ottenere stime da modello per le province della regione Emilia Romagna), sia modelli stimati sui dati dei distretti della regione Emilia Romagna e delle province delle rimanenti regioni provinciali (al fine di ottenere stime da modello per i distretti della regione Emilia Romagna).

### **B3. Analisi e selezione dell'informazione ausiliaria**

In un modello di stima per piccole aree del tipo che consideriamo qui ("area level") l'informazione ausiliaria deve essere definita allo stesso livello di dettaglio geografico dei domini considerati oppure ad un livello di aggregazione maggiore. Le fonti che abbiamo preso in considerazione in questo studio sono:

1. Ministero dell'Economia e delle Finanze. Sono stati utilizzati i file annuali contenenti i principali dati fiscali di tutti i comuni italiani, resi disponibili dal Dipartimento delle Finanze. ([http://www1.finanze.gov.it/finanze2/pagina\\_dichiarazioni/dichiarazioni.php](http://www1.finanze.gov.it/finanze2/pagina_dichiarazioni/dichiarazioni.php))
2. I registri anagrafici dei comuni, accessibili attraverso i database ISTAT, accessibili attraverso il sito web: <http://demo.istat.it/>. Per gli anni precedenti al Censimento (2011) si è fatto riferimento alla ricostruzione intercensuaria della popolazione.
3. Database ISTAT sulla superficie dei comuni e delle province italiane (<http://www.istat.it/it/archivio/82599>).

Le variabili costruite e la loro descrizione è contenuta nella tabella B1. Le informazioni disponibili a livello comunale sono state aggregate per ottenere le covariate da inserire nei modelli a livello di Distretto, per l'Emilia-Romagna, e a livello di provincia per le rimanenti regioni.

Tabella B.1 *Descrizione delle informazioni ausiliarie considerate nella costruzione de modelli.*

Etichetta	Descrizione	Fonte
f_dich	Numero dichiaranti/popolazione totale	Fiscale/anagrafe
rmedio_pop	Reddito totale/popolazione totale	fiscale/anagrafe
rmedio_dich	Reddito totale/Numero dichiaranti	fiscale
dens	Densità di popolazione	anagrafe
freq_giov	Popolazione 0_14/popolazione totale	anagrafe
dip_giov	Popolazione 0_14/popolazione 15_64	anagrafe
dip_sen	Popolazione 65 e oltre /popolazione 15_64	anagrafe
freq_stra	Popolazione straniera/popolazione totale	anagrafe
masc_stra	Numero stranieri maschi/numero stranieri femmine	anagrafe
medialog_f*	Media del log-reddito	fiscale
hr_f*	Frequenza di individui con reddito<10000	fiscale
logit_hr*	Logit della variabile precedente	fiscale
gini_f*	Indice di Gini calcolato sulla base delle covariate fiscali	fiscale
logit_gini_f*	Logit della variabile precedente	fiscale

\*si tratta di un'approssimazione perché i dati fiscali sono disponibili in forma raggruppata con le seguenti classi:

da 0 a 10000

da 10000 a 15000

da 15000 a 26000

da 26000 a 55000

da 55000 a 75000

da 75000 a 120000

oltre

120000

## B.4 I modelli di stima per piccole aree

### B4.1 Modello per i tassi

Indichiamo per semplicità con  $p_{it}$ ,  $i=1,\dots,m$ ,  $t=1,\dots,T$ , la stima diretta del tasso di povertà nell'area  $i$  nell'anno  $t$  (o, in modo del tutto equivalente, la stima diretta del tasso di deprivazione materiale nell'area  $i$  nell'anno  $t$ );  $m$  è il numero delle aree considerate,  $T=6$  il numero degli anni. L'indice  $i$  identifica aree che ricoprono l'intero territorio nazionale. Poiché  $p_{it} \in (0,1)$  possiamo specificare un modello di tipo Beta per questo parametro:

$$\begin{aligned} p_{it} &\sim \text{Beta}(\alpha_{it,p}, \beta_{it,p}) \\ \alpha_{it,p} &= \pi_{it}(f_{it} - 1) \\ \beta_{it,p} &= (1 - \pi_{it})(f_{it} - 1) \end{aligned} \tag{B.1}$$

che implica  $E(p_{it}) = \pi_{it}$ ,  $V(p_{it}) = \frac{\pi_{it}(1 - \pi_{it})}{f_{it}}$ .  $\pi_{it}$  è il tasso (di povertà o di deprivazione materiale) per la popolazione del dominio  $i$  e l'anno  $t$ , il parametro che stiamo modellando.

Poiché  $\frac{V_{ccs}(p_{it})}{V(p_{it})} = \frac{f_{it}}{n_{it}^*}$ ,  $f_{it}$  definisce una sorta di numerosità campionaria effettiva che varrà  $n_{it}^*$  (numero di individui campionati nell'area  $i$  e nell'anno  $t$ ) se il campionamento ha la stessa efficienza di quello casuale semplice, sarà più bassa se, come è il caso il piano di campionamento è meno efficiente. Sulla base delle varianze da disegno possiamo stimare il parametro  $\kappa_{t,des}$  nell'equazione

$$\frac{p_{it}(1 - p_{it})}{\hat{V}_{boot}(p_{it})} = \kappa_{t,des} n_{it}^* + e_{it} \tag{B.2}$$

utilizzando il metodo dei minimi quadrati. Potremo allora porre  $f_{it} = \hat{\kappa}_{t,des} n_{it}^*$  con  $\hat{\kappa}_{t,des}$  interpretabile direttamente come misura di efficienza del piano di campionamento nell'anno  $t$  (se  $\hat{\kappa}_{t,des} < 1$  il disegno sarà meno efficiente di quello casuale semplice). Le stime  $\hat{\kappa}_{t,des}$  sono riportate nell'appendice A4.2 (Tabella A.6) per quanto riguarda il tasso di povertà.

Nella tabella seguente sono riportate quelle relative al tasso di deprivazione materiale:

Tabella B.1 valori stimati del parametro  $\kappa_{t,des}$  relativo al tasso di deprivazione materiale

Anno	Stima parametro	R <sup>2</sup>
2008	0.243	0.74
2009	0.237	0.84
2010	0.244	0.84
2011	0.233	0.84
2012	0.214	0.92
2013	0.201	0.90

Per modellare  $\pi_{it}$  specifichiamo il seguente modello:

$$\text{Probit}(\pi_{it}) = \Phi^{-1}(\pi_{it}) = \alpha_{t,\pi} + \mathbf{x}_{it,\pi}^t \beta_\pi + v_{it,\pi} \quad (\text{B.3})$$

dove  $\mathbf{x}_{it,\pi}$  è un vettore di variabili esplicative dell'area  $i$  al tempo  $t$ ,  $v_{it,\pi}$  è un effetto casuale,  $\alpha_{t,\pi}$  è un'intercetta specifica di anno. Il modello assume inoltre che il parametro di regressione  $\beta_\pi$  sia lo stesso per ogni anno; questa scelta è stata fatta allo scopo di favorire il meccanismo di "borrowing strength" tra anni diversi ed è supportata da un'accurata analisi di scelta tra modelli alternativi (si veda il paragrafo B.7).

Riguardo al vettore  $\mathbf{a}_\pi$  delle intercette specifiche di anno si assume che

$$\begin{aligned} \alpha_{t,\pi} | \mathbf{a}_{\setminus t,\pi} &\sim N(\alpha_{t+1,\pi}, \sigma_{\alpha\pi}^2) & t = 1 \\ \alpha_{t,\pi} | \mathbf{a}_{\setminus t,\pi} &\sim N\left(\frac{\alpha_{t-1,\pi} + \alpha_{t+1,\pi}}{2}, \frac{\sigma_{\alpha\pi}^2}{2}\right) & t = 2, \dots, T-1 \\ \alpha_{t,\pi} | \mathbf{a}_{\setminus t,\pi} &\sim N(\alpha_{t-1,\pi}, \sigma_{\alpha\pi}^2) & t = T \end{aligned} \quad (\text{B.4})$$

dove  $\mathbf{a}_{\setminus t,\pi}$  rappresenta l'insieme delle intercette escluso quella del tempo  $t$ . In altre parole le  $T$  intercette non sono a priori indipendenti, ma seguono un processo *random walk* di ordine 1.

Per gli effetti casuali  $v_{it,\pi}$  si assume a priori una distribuzione Variance Gamma (VG), ovvero:

$$v_{it,\pi} | \alpha_{t,\pi}, \lambda_{t,\pi} \sim VG\left(\alpha_{t,\pi}, \frac{1}{\sqrt{2\lambda_{t,\pi}}}, 0, 0\right) \quad (\text{B.5})$$

(si veda Bibby and Sørensen, 2003 per maggiori dettagli su questa distribuzione). Tale specificazione implica che la distribuzione a priori degli effetti casuali sia simmetrica, con  $E(v_{it,\pi} | \alpha_{t,\pi}, \lambda_{t,\pi}) = 0$  e  $V(v_{it,\pi} | \alpha_{t,\pi}, \lambda_{t,\pi}) = \alpha_{t,\pi} / \lambda_{t,\pi}$ . L'uso di tale distribuzione presenta notevoli vantaggi rispetto alla tradizionale ipotesi di normalità degli effetti casuali (si veda Frühwirth-Schnatter and Wagner 2010 e Fabrizi and Trivisano, 2016): essa rappresenta un modello opportuno quando è di interesse tenere conto contemporaneamente della presenza di outliers e di effetti casuali piccoli in valore assoluto a causa, per esempio, di variabili ausiliarie con forte potere esplicativo.

## B4.2 Modello per l'indice di Gini

Indichiamo per semplicità con  $g_{it}$  la stima diretta dell'indice di concentrazione di Gini nell'area  $i$  e l'anno  $t$ . Poiché  $g_{it} \in (0,1)$  possiamo specificare un modello di tipo Beta anche per questo parametro:

$$g_{it} \sim \text{Beta}(\alpha_{it,g}, \beta_{it,g})$$

$$\alpha_{it,g} = \frac{2\phi_{it}}{1 - \gamma_{it}} - \gamma_{it} \tag{B.6}$$

$$\beta_{it,g} = \frac{2\phi_{it} - \gamma_{it}(1 + \gamma_{it})}{1 + \gamma_{it}} \frac{(1 - \gamma_{it})}{\gamma_{it}}$$

dove  $\gamma_{it}$  indica il parametro oggetto principale di stima, ovvero l'indice di concentrazione in popolazione. Notiamo che, pur essendo basato sulla distribuzione Beta, il modello (B.6) è diverso da (B.1) in quanto, nonostante il valore atteso coincida con il parametro oggetto di stima, la varianza ha una struttura diversa (Fabrizi and Trivisano, 2016). Infatti il modello (B.6) implica  $E(g_{it}) = \gamma_{it}$ ,

$$V(g_{it}) = \gamma_{it}^2 (1 - \gamma_{it}^2) (2\phi_{it})^{-1}. \tag{B.7}$$

I parametri  $\phi_{it}$  sono assunti noto nel modello; essi sono stimati esogenamente con un modello di “variance smoothing” illustrato nella sezione B4.3.

Assumiamo inoltre analogamente alla (B.3):

$$\text{Probit}(\gamma_{it}) = \Phi^{-1}(\gamma_{it}) = \alpha_{t,\gamma} + \mathbf{x}_{it,\gamma}^t \beta_\gamma + v_{it,\gamma} \tag{B.8}$$

dove  $\mathbf{x}_{it,\gamma}$  è un vettore di variabili esplicative dell'area  $i$  al tempo  $t$ ,  $v_{it,\gamma}$  è un effetto casuale,  $\alpha_{t,\gamma}$  è un'intercetta specifica di anno. Il modello assume inoltre che il parametro di regressione  $\beta_\gamma$  sia lo stesso per ogni anno. Per i parametri  $\alpha_{t,\gamma}$  e  $v_{it,\gamma}$  si è



scelta una specificazione uguale a quella utilizza per gli analoghi parametri nel modello per i tassi, quindi rispettivamente le specificazioni (B.4) e (B.5).

### B4.3 Lo “smoothing” della varianza dell’indice di concentrazione di Gini

Per semplicità espositiva verrà illustrato il caso della stima dei parametri  $\phi_{it}$  relativamente ad un solo anno; quindi in questa sezione il pedice  $t$  è soppresso. Assumiamo come punto di partenza che per il logaritmo del reddito di un individuo incluso nel campione relativo all’area  $i$  valga l’ipotesi di normalità distributiva:

$$y_{ij} \sim N(\theta_i, \sigma_i^2) \quad (\text{B.9})$$

Sotto l’ipotesi (B.9) possiamo sfruttare le proprietà della distribuzione log-normale (Aitchinson e Brown, 1963) e scrivere

$${}_{LN}\gamma_i = 2\Phi\left(\frac{\sigma_i}{\sqrt{2}}\right) - 1 = \text{erf}\left(\frac{\sigma_i}{\sqrt{2}}\right) \quad (\text{B.10})$$

Se assumiamo il campionamento casuale semplice una stima diretta (consistente) di

$${}_{LN}\gamma_i \text{ è data da } g_{2i} = \text{erf}\left(\frac{s_i}{\sqrt{2}}\right) \quad (\text{B.11})$$

$$\text{in cui } s_i = \sqrt{(n_i^* - 1) \sum_{j=1}^{n_i^*} (y_{ij} - \bar{y}_i)^2}.$$

Per calcolare la varianza di  $g_{2i}$  sfruttiamo il fatto che  $\text{erf}(\ )$  è una funzione differenziabile e per ogni trasformazione  $f$  (differenziabile) di una variabile casuale  $Z$  possiamo scrivere  $V[f(Z)] \cong V(Z)\{f'[E(Z)]\}$ .

E’ quindi possibile dimostrare che

$$V(g_{2i}) \cong \frac{\sigma_i^2(1 - c_{4i}^2)}{\pi} \exp\left\{-\frac{c_{4i}^2 \sigma_i^2}{2}\right\} \quad (\text{B.12})$$

$$\text{dove } c_{4i} = \sqrt{\frac{2}{n_i^* - 1}} \frac{\Gamma\left(\frac{n_i^*}{2}\right)}{\Gamma\left(\frac{n_i^* - 1}{2}\right)}$$

e  $\pi$  indica qui semplicemente la costante matematica 3.141592...

Utilizzando un paio di approssimazioni giustificate dal fatto che  $\sigma_i^2$ ,  $c_{4i}^2 \sigma_i^2$  sono numeri piccoli che legittimano sviluppi in serie di McLaurin delle funzioni di cui sono argomento, e utilizzando inoltre il fatto che  $(1 - c_{4i}^2) \simeq (2n_i^*)^{-1}$ , abbiamo

$$V(g_{2i}) \simeq \frac{\sigma_i^2 (1 - c_{4i}^2)}{\pi} \exp\left\{-\frac{c_{4i}^2 \sigma_i^2}{2}\right\} \simeq \frac{\sigma_i^2}{2n_i^* \pi} \left(1 - \frac{\sigma_i^2}{2}\right) \quad (\text{B.13})$$

Dalla (B.10) è facile ottenere che

$$\sigma_i = \sqrt{2} \Phi^{-1}\left(\frac{LN \gamma_i + 1}{2}\right) \quad (\text{B.14})$$

Per praticità possiamo riscrivere (B.14) come  $\sigma_i \simeq \frac{\sqrt{\pi}}{2} \text{logit}\left(\frac{LN \gamma_i + 1}{2}\right)$  sfruttando la relazione  $\text{logit}(x) \simeq \sqrt{\frac{8}{\pi}} \Phi^{-1}(x)$ . La relazione può essere ulteriormente semplificata notando che, poiché in un intorno di 1  $\log(1+x) \simeq 1+x$ ,  $\text{logit}\left(\frac{LN \gamma_i + 1}{2}\right) \simeq 2 LN \gamma_i$  e quindi

$$\sigma_i \simeq LN \gamma_i \sqrt{\pi}. \quad (\text{B.15})$$

Dalle formule (B.15) e (B.13) si ottiene

$$V(g_{2i}) \simeq (2n_i^*)^{-1} LN \gamma_i^2 (1 - 0.5 \pi LN \gamma_i^2) \quad (\text{B.16})$$

Poiché la relazione (B.16) può portare, per valori elevati di  $LN \gamma_i$ , a valori negativi di  $V(g_{2i})$ , essa è sostituita da

$$V(g_{2i}) \simeq (2n_i^*)^{-1} LN \gamma_i^2 (1 - LN \gamma_i^2) \quad (\text{B.18})$$

Quest'ultima relazione motiva la funzione di varianza (B.7) e giustifica euristicamente il seguente modello di smoothing per le stime della varianza ottenute con il metodo bootstrap:

$$\hat{V}_{boot}(g_i) = \rho_{des} (2n_i^*)^{-1} g_i^2 (1 - g_i^2) + e_i \quad (\text{B.19})$$

Abbiamo stimato la relazione (B.19) utilizzando il metodo dei minimi quadrati ordinari ottenendo per ciascun anno le seguenti stime del parametro  $\rho_{des}$ :

**Tabella B.2** *valori stimati del parametro  $\rho_{des}$  per ogni anno*

Anno	Stima parametro	R <sup>2</sup>
2008	0.160	0.68
2009	0.226	0.91
2010	0.189	0.79
2011	0.207	0.90
2012	0.173	0.80
2013	0.146	0.66

Infine, per i parametri  $\phi_{it}$  nella (B7), alla luce della (B19) si assume  $\phi_{it} = \frac{\hat{\rho}_{t,des}}{n_{it}^*}$ .

#### **B4.4 Modello per la media del logaritmo del reddito**

In questa sezione presentiamo un modello per la media del logaritmo del reddito equivalente. Questo parametro, pur non essendo esplicitamente di interesse, sarà utilizzato nella prossima sezione per la stima del GAP mediano.

Sia  $\hat{y}_{it}$  la stima diretta la media del logaritmo del reddito equivalente.

Assumiamo che

$$\hat{y}_{it} \sim N(\theta_{it}, \varphi_{it} \sigma_{it}^2) \quad (\text{B.20})$$

Se vale la (B.9) allora l'assunzione (B.20) è esatta, altrimenti essa è supportata dal teorema del limite centrale.

Si assume inoltre che

$$\theta_{it} = \alpha_{t,\theta} + \mathbf{x}_{it,\theta}^t \beta_{\theta} + v_{it,\theta} \quad (\text{B.21})$$

dove  $\mathbf{x}_{it,\theta}$  è un vettore di variabili esplicative dell'area  $i$  al tempo  $t$ ,  $v_{it,\theta}$  è un effetto casuale,  $\alpha_{t,\theta}$  è un'intercetta specifica di anno. Il modello assume inoltre che il parametro di regressione  $\beta_{\theta}$  sia lo stesso per ogni anno. Per i parametri  $\alpha_{t,\theta}$  e  $v_{it,\theta}$  si è scelta una specificazione uguale a quella utilizza per gli analoghi parametri nel modello per i tassi e per l'indice di Gini, quindi rispettivamente le specificazioni (B.4) e (B.5).

I parametri  $\varphi_{it}$  rappresentano il reciproco della numerosità campionaria effettiva; se il campionamento avesse la stessa efficienza di un campionamento casuale semplice,

allora si avrebbe  $\varphi_{it} = 1/n_{it}^*$ . Riguardo alle varianze  $\sigma_{it}^2$ , in base alla l'ipotesi (B.9) si può assumere

$$\sigma_{it}^2 = \pi \gamma_{it}^2 \quad (\text{B.22})$$

dove  $\gamma_{it}^2 = E(g_{it})$  sotto il modello per l'indice di Gini della sezione B4.2. Il modello per l'indice di Gini e quello per la media del logaritmo del reddito equivalente sono dunque esplicitamente legati dalla relazione (B.22) e vanno stimati congiuntamente.

I parametri  $\varphi_{it}$  sono assunti noti e posti uguali alla stima ottenuta mediante il seguente modello di "variance smoothing" giustificato ancora dalla (B.22):

$$\hat{V}_{boot}(\hat{y}_{it}) = \mathcal{G}_{t,des}(n_i^*)^{-1} \pi g_{it}^2 + e_{it} \quad (\text{B.23})$$

dove  $\hat{V}_{boot}(\hat{y}_{it})$  è una stima bootstrap della varianza dello stimatore media aritmetica del logaritmo del reddito equivalente. Abbiamo stimato la relazione (B.23) utilizzando il metodo dei minimi quadrati ordinari ottenendo le seguenti stime del parametro  $\varphi_{t,des}$ :

**Tabella B.3** *valori stimati del parametro  $\mathcal{G}_{t,des}$  per ogni anno*

Anno	Stima parametro	R <sup>2</sup>
2008	4.506	0.64
2009	4.544	0.86
2010	4.368	0.68
2011	5.767	0.68
2012	4.769	0.66
2013	7.204	0.68

Infine, per i parametri  $\varphi_{it}$  si assume  $\varphi_{it} = \frac{\hat{\mathcal{G}}_{t,des}}{n_{it}^*}$ .

### B4.5 Modello per il gap mediano

Le stime dirette per il gap mediano sono molto imprecise. La ragione è che si tratta di un parametro definito sulla popolazione dei poveri (è la differenza tra la soglia di povertà e il loro reddito mediano). I poveri sono una minoranza della popolazione e dei campioni disponibili per ciascun distretto o ciascuna provincia. Ciò rende le informazioni disponibili per le stime dirette del gap mediano inadeguate. Inoltre la correlazione tra stime dirette del gap mediano e covariate disponibili è molto bassa (la

correlazione massima tra stime dirette e covariate è pari a -0.21). Per queste ragioni il modello che specifichiamo per il gap mediano si basa essenzialmente sull'ipotesi euristica (B.9). E' inoltre possibile dimostrare che le stime dirette del gap mediano sono positivamente distorte, a causa della distorsione negativa con cui viene stimata la mediana dei poveri. Questa distorsione decresce al crescere della dimensione campionaria; tuttavia proprio perché i sotto-campioni su cui questa stima viene calcolata sono di dimensioni estremamente ridotta essa può non essere trascurabile.

Per il gap mediano (che indicheremo con il parametro  $\zeta_i$ ; per semplicità di esposizione tratteremo il problema con riferimento a un solo anno di calendario) abbiamo che, sotto l'ipotesi di log-normalità (B.9)

$$\log \zeta_i = Me(y_{ij} | y_{ij} \leq \log(PTN)) \quad (B.24)$$

e la distribuzione di  $y_{ij} | y_{ij} \leq \log(PTN)$  è in forma nota ( $PTN$  indica la soglia nazionale di povertà che è assunta nota). Si tratta di una Normale troncata

$$y_{ij} | y_{ij} \leq \log(PTN) \sim TN(\theta_i, \text{var}_i, l, u) \quad (B.25)$$

dove  $l = -\infty$  e  $u = \log(PTN)$  definiscono l'estremo inferiore e superiore di troncamento.

Usando le formule standard per questa distribuzione otteniamo che il gap mediano può essere ri-definito dall'equazione:

$$\frac{\Phi\left(\frac{\zeta_i - \theta_i}{\sigma_i}\right)}{\Phi\left(\frac{u - \theta_i}{\sigma_i}\right)} = \frac{1}{2} \quad (B.26)$$

Risolvendo e tornando sulla scala naturale dei redditi otteniamo

$$\zeta_i = \exp\left[\theta_i + \sigma_i \Phi^{-1}\left\{\frac{1}{2} \Phi\left(\frac{u - \theta_i}{\sigma_i}\right)\right\}\right] \quad (B.27)$$

Come già notato in precedenza, il parametro incognito  $\sigma_i$  può essere assunto come stocastico e la distribuzione a posteriori di  $\zeta_i$  simulata congiuntamente a quella di  $\sigma_i$ . Un modello per  $\sigma_i$  è proposto in (B.15). Tuttavia, possiamo notare, sulla base di (B.9), che (B.27) può essere riscritta come

$$\zeta_i = \exp\left[\theta_i + \sigma_i \Phi^{-1}\left\{\frac{1}{2} \pi_i^{LN}\right\}\right]$$

(infatti sotto l' ipotesi (B.9)  $\pi_i^{LN} = \Phi\left(\frac{u - \theta_i}{\sigma_i}\right)$ , quindi  $\pi_i^{LN}$  denota il tasso di povertà sotto ipotesi di log-normalità), ossia (B.27) esprime il poverty gap mediano in funzione della stima del tasso di povertà basato sull'ipotesi euristica di log-normalità dei redditi. Uno stimatore dipendente in modo meno restrittivo dall'ipotesi di log-normalità di questo parametro può essere ottenuto sostituendo  $\pi_i^{LN}$  con il  $\pi_i$  ottenuto sulla base della relazione (B.3). L'inferenza sul poverty gap mediano sarà quindi basata sulla relazione

$$\zeta_i = \exp\left[\theta_i + \sigma_i \Phi^{-1}\left\{\frac{1}{2}\pi_i\right\}\right] \quad (\text{B.28})$$

Il parametro  $\zeta_i$  è quindi espresso in funzione di tre parametri stocastici:  $\theta_i, \pi_i, \sigma_i$ . La distribuzione a posteriori di  $\zeta_i$  può essere esplorata utilizzando i metodi MCMC. Una volta generati valori campione dalle distribuzioni a posteriori di  $\theta_i, \pi_i, \sigma_i$ , valori campione della distribuzione di  $\zeta_i$  possono essere generati sulla base di (B.28).

Purtroppo le medie a posteriori  $\hat{\zeta}_i^B = E(\zeta_i | \text{dati})$  ottenute mediante la (B.28) sono stimatori negativamente distorti. La ragione, seppur non molto evidente dal punto di vista intuitivo, è legata al considerare una trasformazione esponenziale dei parametri  $\theta_i, \sigma_i, \pi_i$  che sono definiti sulla scala dei logaritmi dei redditi. La riduzione della distorsione di stimatori ottenuti attraverso una trasformazione esponenziale di stimatori approssimativamente corretti è un problema ampiamente dibattuto nella letteratura statistica. In linea con Fay e Herriot (1979), un lavoro di estrema importanza nella letteratura sulla stima per piccole aree, abbiamo scelto di affrontarlo attraverso una tecnica di *benchmarking*. In sostanza, si tratta di vincolare un certo funzionale di un insieme di stime ad essere uguale ad una costante data. Poiché

$$\xi_i = \log(\zeta_i) = \theta_i + \sigma_i \Phi^{-1}\left\{\frac{1}{2}\pi_i\right\}$$

sono valori attesi di una distribuzione normale troncata, e quindi additivi, è ragionevole vincolare la somma pesata di queste quantità a riprodurre la stima diretta ottenuta per l'intera regione alla quale appartiene il dominio (provincia o distretto che sia). A questo riguardo, sappiamo che benché le stime dirette del gap mediano siano positivamente distorte, il campione regionale è sufficientemente grande da rendere la distorsione trascurabile e la precisione dello stimatore soddisfacente. Imponiamo quindi il vincolo

$$\sum_{i \in R} w_i \hat{\xi}_i^B = \widehat{LMP}_R \quad (\text{B.29})$$

dove  $\widehat{LMP}_R$  è la stima diretta della mediana del logaritmo del reddito dei poveri della regione  $R$  cui appartiene il dominio  $i$ . Il sistema dei pesi nella (B.29) dev'essere tale che  $w_i > 0$ ,  $\sum_{i \in R} w_i = 1$ . In questa applicazione li abbiamo fissati pari alla popolazione delle province/distretti che risulta dalle anagrafi. Esistono impostazioni alternative per ottenere stimatori Bayesiani (medie a posteriori) che soddisfino un vincolo del tipo (B.29). Tra queste, abbiamo scelto quella di Datta et al. (2010) che è ben nota in letteratura e, per la sua flessibilità ci ha dato la possibilità di valutare differenti opzioni e scegliere quella più appropriata. Descriviamo brevemente la soluzione adottata, tralasciamo le possibili alternative. Le stime di benchmarking sono quei valori  $e_i$  che minimizzano il valore atteso a posteriori della funzione di perdita

$$\sum_{i \in R} w_i E[(\xi_i - e_i) | \hat{\xi}] \quad (\text{B.30})$$

rispettando il vincolo (B.29). Un problema non direttamente trattato da Datta et al. (2010) è quello del calcolo della varianza a posteriori degli stimatori “benchmarkati”, ottenuti calcolando la trasformazione esponenziale degli  $e_i$  nella (B.30). Poiché associare una misura di incertezza agli stimatori calcolati è di fondamentale interesse nel nostro lavoro, abbiamo modificato l'impostazione di Datta et al. (2010) nel modo seguente. Abbiamo sottoposto a vincolo di benchmarking ogni campione generato dall'algoritmo Markov Chain Monte Carlo per  $\xi_i$  ottenuto sulla base della (B.28). Una volta vincolato l'insieme dei valori relativi ai domini di ogni regione ne abbiamo calcolato l'esponenziale per ottenere valori dalla distribuzione a posteriori dei  $\zeta_i$  sotto il vincolo (B.29). In questo modo è possibile descrivere tutti gli aspetti della distribuzione a posteriori, inclusa la varianza. Questo modo di procedere è noto in letteratura come *transformation approach* ed è discusso in Dunson e Neelan (2003).

#### **B4.6 Specificazione delle distribuzioni a priori per gli iperparametri**

L'analisi Bayesiana di un modello prevede che a tutti i parametri incogniti sia assegnata una distribuzione a priori. Nei paragrafi precedenti, per alleggerire la presentazione non abbiamo discusso la specificazione delle distribuzioni a priori relative agli iper-parametri: coefficienti di regressione e componenti di varianza. Alla luce della complessità del modello abbiamo scelto a priori standard (dette anche “reference” in letteratura), approssimativamente non-informative.

In particolare per ognuna delle componenti dei vettori di coefficienti di regressione:  $\beta_\theta$ ,  $\beta_\pi$ ,  $\beta_\gamma$  sono state assunte distribuzioni normali diffuse, ossia caratterizzate da varianze molto elevate  $N(0,1000)$ . Benché approssimativamente uniformi, queste distribuzioni offrono il vantaggio, rispetto a distribuzioni uniformi improprie (uniformi

su tutto l'asse reale) di portare ad un calcolo più agevole delle *full conditional*, un elemento importante nella costruzione degli algoritmi di tipo Markov Chain Monte Carlo che abbiamo utilizzato per lo studio di questi modelli. Per le varianze  $\sigma_{\alpha\pi}^2$ ,  $\sigma_{\alpha\gamma}^2$ ,  $\sigma_{\alpha\theta}^2$ , delle distribuzioni condizionate delle componenti dei vettori delle intercette di tempo abbiamo proceduto alla specificazione di a priori *Inv-Gamma*(0.01,100). Anche in questo caso si tratta di distribuzioni ordinariamente utilizzate in letteratura per l'analisi di modelli gerarchici dove vengono impiegate per la trascurabile influenza che esercitano sui momenti delle distribuzioni a posteriori dei parametri di interesse e per ragioni di convenienza computazionale (costruzione delle *full conditional*).

Infine, in linea con Frühwirth-Schnatter and Wagner (2010) e Fabrizi and Trivisano (2016), abbiamo fissato a .5 i parametri  $\alpha_{t,\pi}, \alpha_{t,\theta}, \alpha_{t,\gamma}$ , e scelto una distribuzione *Gamma*(2,1) per i parametri  $\lambda_{t,\pi}, \lambda_{t,\theta}, \lambda_{t,\gamma}$ .

## B5. Selezione variabili ausiliarie

Nella descrizione dei modelli contenuta nel paragrafo (B.4) abbiamo visto come le informazioni ausiliarie vengano impiegate in tre equazioni distinte:

- L'equazione (B.21) relativa al valore atteso dei log-redditi equivalenti;
- L'equazione (B.8) relativa al valore atteso del coefficiente di concentrazione di Gini;
- L'equazione (B.3) relativa al valore atteso del tasso di deprivazione materiale.

Alcune delle 14 variabili ausiliarie descritte in Tabella B1 sono fortemente correlate tra loro, rappresentando informazione ridondante. E' stato quindi necessario effettuare una prima selezione.

Il processo di selezione è avvenuto in due passi utilizzando i dati relativi alle 110 province.

Dal primo gruppo di 14 variabili ausiliarie (regressori) si è passati ad un gruppo più ristretto di 10 eliminando, tra quelle tra loro fortemente correlate, le variabili ausiliarie che presentavano minore correlazione con le stime dirette dei parametri di interesse. Nella tabella B. 4.1 sono riportate le variabili selezionate e nella Tab B. 4.2 le correlazioni tra le covariate e le stime dirette dei parametri di interesse:



Tabella B.4.1 *Descrizione delle informazioni ausiliarie considerate nella costruzione de modelli dopo il primo passo di selezione.*

Etichetta	Descrizione	Fonte
f_dich	Numero dichiaranti/popolazione totale	Fiscale/anagrafe
rmedio_dich	Reddito totale/Numero dichiaranti	fiscale
dens	Densità di popolazione	anagrafe
dip_giov	Popolazione 0_14/popolazione 15_64	anagrafe
dip_sen	Popolazione 65 e oltre /popolazione 15_64	anagrafe
freq_stra	Popolazione straniera/popolazione totale	anagrafe
masc_stra	Numero stranieri maschi/numero stranieri femmine	anagrafe
medialog_f	Media del log-reddito	fiscale
logit_hr	Logit della variabile precedente	fiscale
logit_gini_f	Logit della variabile precedente	fiscale

Tabella B.4.2 *Coefficienti di correlazione tra le informazioni ausiliarie e le stime dirette*

stime dirette				
		tasso di povertà	gini	tasso di deprivazione
<b>Covariate</b>	f_dich	-0.851	-0.479	-0.666
	rmedio_dich	-0.588	-0.188	-0.439
	dens	-0.066	0.127	-0.036
	dip_giov	0.396	0.341	0.357
	dip_sen	-0.525	-0.364	-0.407
	freq_stra	-0.695	-0.409	-0.462
	masc_stra	-0.444	-0.277	-0.229
	medialog_f	-0.783	-0.380	-0.606
	logit_hr	0.800	0.423	0.616
	logit_gini_f	0.550	0.505	0.461

Un insieme di variabili ausiliarie ancora più ristretto è stato selezionato, per ciascuna delle equazioni sopra citate, utilizzando solo ai fini della selezione delle variabili la seguente distribuzione a priori per i parametri di regressione e stimando modelli (per il solo anno d'indagine 2010) simili a quelli descritti nelle sezioni B4.1-B4.4, tranne che per la distribuzione a priori sui parametri di regressione.

Sia  $\beta_i$  l' $i$ -esima componente del vettore dei parametri di regressione; si assume per ogni  $i$

$$\begin{aligned} \beta_i &\sim N(0, \sigma_i^2) \\ \sigma_i^2 &= 0.001(1 - \lambda_i) + c\lambda_i \\ \lambda_i &\sim \text{Bernoulli}(0.5) \end{aligned} \tag{B.31}$$

La media a posteriori del parametro  $\lambda_i$  è un indicatore dell'importanza della covariata  $i$ -esima. Il metodo di selezione delle covariate indotto dalla prior (B.31) è nota in letteratura come *Stochastic search variable selection* ed è stato introdotto in letteratura da George and McCulloch (1993). Si è scelto di operare una scelta conservativa (nel senso di non escludere troppe variabili) ponendo  $c=10$  e includendo nei modelli tutte le covariate con media a posteriori di  $\lambda_i$  maggiore di 0.1.

Di seguito i risultati per ognuno dei parametri di povertà di interesse.

**Tabella B.5** *media a posteriori del parametro  $\lambda_i$ , modello per il tasso di povertà*

Covariata	Media a posteriori	Selezionata
f_dich	1.0000	SI
rmedio_dich	0.1009	SI
dens	0.0735	
dip_giov	0.0783	
dip_sen	0.2414	SI
freq_stra	0.0737	
masc_stra	0.0728	
medialog_f	0.0944	
logit_hr	0.1045	SI
logit_gini_f	0.1449	SI

Tabella B.6 *media a posteriori del parametro  $\lambda_i$ , modello per l'indice di Gini*

Covariata	Media a posteriori	Selezionata
f_dich	0.9889	SI
rmedio_dich	0.7402	SI
dens	0.0417	
dip_giov	0.0343	
dip_sen	0.1176	SI
freq_stra	0.0417	
masc_stra	0.0314	
medialog_f	0.1885	SI
logit_hr	0.1586	SI
logit_gini_f	0.2081	SI

Tabella B.7 *media a posteriori del parametro  $\lambda_i$ , modello per la media del logaritmo del reddito equivalente*

Covariata	Media a posteriori	Selezionata
f_dich	1.0000	SI
rmedio_dich	0.2157	SI
dens	0.0459	
dip_giov	0.0190	
dip_sen	0.2482	SI
freq_stra	0.0332	
masc_stra	0.0203	
medialog_f	0.0657	
logit_hr	0.0290	
logit_gini_f	0.0289	

Tabella B.8 *media a posteriori del parametro  $\lambda_i$ , modello per il tasso di deprivazione materiale*

Covariata	Media a posteriori	Selezionata
f_dich	0.9786	SI
rmedio_dich	0.1697	SI
dens	0.1412	SI
dip_giov	0.1572	SI
dip_sen	0.1870	SI
freq_stra	0.1969	SI
masc_stra	0.2748	SI
medialog_f	0.2389	SI
logit_hr	0.2935	SI
logit_gini_f	0.1087	SI

## B6. Simulazione delle distribuzioni a posteriori

Le distribuzioni a posteriori dei parametri inclusi nei modelli considerati sono state studiate utilizzando metodi Markov Chain Monte Carlo. In particolare, abbiamo utilizzato il software OpenBugs (Thomas et al., 2006; Spiegelhalter et al., 2003). L'idea fondamentale di questi metodi è quella di generare campioni da catene di Markov la cui distribuzione stabile coincida con la distribuzione a posteriori del parametro oggetto di interesse.

Il campione generato dalla catena potrà quindi essere diviso in due parti: una prima parte (o "burn in") relativo alla transizione della catena verso la sua distribuzione stabile e la parte successiva alla convergenza, che è quello che può effettivamente essere utilizzato per descrivere la distribuzione a posteriori. Dal punto di vista tecnico occorre quindi approntare le diagnostiche necessarie a verificare l'avvenuta convergenza. Noi abbiamo scelto di generare catene lunghe di 25,000 valori per ciascuno dei parametri d'interesse e di eliminare le prime 5,000 come burn in. Che questo taglio fosse effettivamente sufficiente è stato accertato con gli strumenti dell'analisi visuale delle traiettorie delle catene, della generazione di catene parallele a partire da valori iniziali distanziati tra loro e calcolando la statistica di Gelman e Rubin.

I campioni generati dalle catene di Markov giunte a convergenza sono spesso caratterizzati da una considerevole autocorrelazione. Per ridurre l'effetto dell'autocorrelazione abbiamo fissato il parametro di *thin* a 2, prendendo cioè un valore ogni due generati.

## B7. La selezione dei modelli

Alcune delle ipotesi alla base dei modelli presentati nelle sezioni B4.1-B4.4 sono il frutto di un'accurata procedura di selezione in cui alcune delle ipotesi centrali, quali ad esempio quelle sulle intercette di tempo, sui parametri di regressione, sono state sottoposte alla verifica dei dati.

Sono state infatti considerate tre specificazioni alternative dei modelli:

- Modello 1: Intercette diverse per ogni anno, ma a priori indipendenti. Parametri di regressione diversi per ogni anno e per ogni parametro di interesse, a priori indipendenti. Effetti casuali diversi per ogni anno. Di fatto in questi modelli non c'è nessun legame tra i parametri relativi ad anni diversi ed equivalgono alla stima di modelli indipendenti per ciascuno degli anni considerati.
- Modello 2: intercette diverse per ogni anno, ma a priori indipendenti. Parametri di regressione uguali nel tempo. Effetti casuali diversi per ogni anno. Rispetto al Modello 1 è introdotta quindi l'ipotesi che i parametri di regressione siano uguali negli anni.
- Modello 3: le intercette sono ancora diverse per ogni anno; non sono però a priori indipendenti ma hanno la prior (B.4) di tipo Random walk di ordine 1. Parametri di regressione uguali nel tempo. Effetti casuali diversi per ogni anno. Si tratta quindi di modelli con la struttura descritta nelle sezioni B4.1-B4.4. Rispetto al Modello 2 viene introdotta l'ipotesi che le intercette abbiano a priori una struttura di dipendenza temporale.

La selezione tra modelli è stata basata sul Deviance Information Criterion (DIC). Tra due modelli in competizione si sceglie quello con il DIC minore. Ai fini della selezione sono stati considerati i modelli stimati sui dati delle 110 province italiane.

Di seguito sono riportati i risultati per ciascuno dei parametri di povertà di interesse:

**Tabella B.9** *Deviance Information criterion per i tre modelli considerati*

	Gini	Media log	Tasso di povertà	Tasso di DM
Modello 1	-2679	-1347	-2048	-1846
Modello 2	-2691	-1343	-2067	-1866
Modello 3	-2693	-1354	-2087	-1879

Come si può vedere per ognuno dei quattro parametri di povertà di interesse il DIC risulta minore per il Modello 3.

Tab. B.10 - Stima del tasso di povertà per i distretti socio-sanitari dell'Emilia Romagna (anni 2007-2012, riferimento al reddito)

		tasso di povertà %					
Provincia	Denominazione	2007	2008	2009	2010	2011	2012
Reggio	GUASTALLA	8.32	8.81	8.99	9.81	9.57	9.99
Reggio	CORREGGIO	7.61	7.67	8.39	9.15	9.47	8.87
Reggio	SCANDIANO	6.32	6.60	6.99	6.63	7.30	7.14
Reggio	CASTELNUOVO NE' MONTI	12.83	12.88	12.20	12.95	12.60	11.73
Modena	CARPI	7.40	7.56	7.75	7.85	8.87	8.59
Modena	MIRANDOLA	8.26	7.85	8.36	9.09	9.87	10.05
Modena	MODENA	9.34	7.85	7.76	8.72	9.25	9.03
Modena	SASSUOLO	8.22	7.69	8.45	9.04	8.85	9.47
Modena	PAVULLO NEL FRIGNANO	12.03	11.61	10.31	11.45	11.64	11.29
Modena	VIGNOLA	8.39	8.39	8.19	9.09	9.89	9.62
Bologna	CASALECCHIO DI RENO	7.21	6.45	6.23	6.99	7.22	7.25
Bologna	PORRETTA TERME	8.37	8.42	8.07	8.61	8.56	8.26
Bologna	SAN LAZZARO DI SAVENA	7.05	7.44	6.83	6.64	6.77	6.64
Bologna	IMOLA	6.94	7.02	6.52	7.31	6.86	7.58
Bologna	PIANURA EST	5.96	5.22	4.89	5.75	6.10	5.67
Bologna	PIANURA OVEST	6.12	6.26	6.14	7.00	6.43	6.52
Bologna	CITTA' DI BOLOGNA	8.23	7.99	8.20	9.06	9.74	9.32
Ferrara	OVEST	6.70	6.78	6.70	7.62	7.88	7.85
Ferrara	CENTRO-NORD	8.09	8.96	7.72	8.70	8.64	8.36
Ferrara	SUD-EST	8.83	9.36	8.83	10.29	10.47	10.59
Ravenna	RAVENNA	9.14	9.10	7.47	8.28	7.86	8.66
Ravenna	LUGO	8.31	7.82	7.57	8.72	9.28	9.75
Ravenna	FAENZA	9.77	11.01	8.44	10.62	10.70	10.85
Forli	FORLI'	9.84	9.27	8.55	9.83	9.52	10.47
Forli	CESENA - VALLE DEL SAVIO	10.98	11.10	10.90	10.13	10.53	11.33
Parma	PARMA	7.60	7.25	7.14	8.43	8.03	8.32
Rimini	RIMINI	13.73	12.96	13.21	12.71	12.49	14.13
Rimini	RICCIONE	13.92	14.65	13.11	14.15	14.25	14.66
Modena	CASTELFRANCO EMILIA	6.70	6.60	6.94	8.00	7.77	7.18
Forli	RUBICONE	10.70	12.31	11.21	13.12	12.62	12.76
Parma	FIDENZA	9.51	8.15	8.05	9.00	9.63	9.66
Piacenza	CITTA' DI PIACENZA	9.56	9.63	8.35	10.03	10.15	10.02
Piacenza	LEVANTE	9.73	10.60	9.58	10.59	10.19	9.69
Piacenza	PONENTE	10.47	10.54	9.84	11.01	10.67	10.60
Parma	VALLI TARO E CENO	13.25	13.42	12.97	14.21	12.43	12.77
Parma	SUD EST	8.56	8.48	7.75	7.82	8.08	8.12
Reggio	MONTECCHIO	8.31	8.29	8.03	8.66	8.54	8.61
Reggio	REGGIO EMILIA	8.13	6.88	7.76	9.14	9.33	9.91

Tab. B.11 - CV riferiti alla stima del tasso di povertà per i distretti socio-sanitari dell'Emilia Romagna (anni 2007-2012, riferimento al reddito)

		CV del tasso di povertà %					
Provincia	Denominazione	2007	2008	2009	2010	2011	2012
Reggio	GUASTALLA	18.40	18.99	19.55	18.27	18.37	19.04
Reggio	CORREGGIO	19.76	19.91	19.55	19.25	19.22	19.83
Reggio	SCANDIANO	19.83	19.15	20.05	18.75	19.94	19.45
Reggio	CASTELNUOVO NE' MONTI	18.95	18.75	18.83	18.31	18.52	18.17
Modena	CARPI	18.26	18.60	19.24	18.00	17.72	18.24
Modena	MIRANDOLA	19.59	19.82	19.48	19.28	19.31	18.86
Modena	MODENA	18.14	18.19	18.66	18.28	17.76	18.31
Modena	SASSUOLO	19.36	18.64	18.96	19.00	18.60	19.52
Modena	PAVULLO NEL FRIGNANO	18.40	17.65	18.07	17.61	18.11	18.08
Modena	VIGNOLA	19.37	19.22	19.04	18.15	18.44	18.83
Bologna	CASALECCHIO DI RENO	18.76	18.31	19.33	19.01	18.99	18.92
Bologna	PORRETTA TERME	19.70	19.56	20.07	19.97	19.49	19.99
Bologna	SAN LAZZARO DI SAVENA	18.26	19.12	20.18	18.79	18.45	18.22
Bologna	IMOLA	18.81	19.79	19.53	18.96	18.96	19.59
Bologna	PIANURA EST	19.50	18.46	18.51	18.66	19.16	19.08
Bologna	PIANURA OVEST	19.10	18.93	19.39	19.71	18.91	19.21
Bologna	CITTA' DI BOLOGNA	17.21	17.84	18.09	18.47	17.07	17.19
Ferrara	OVEST	20.19	20.14	20.15	19.92	20.03	19.85
Ferrara	CENTRO-NORD	16.90	18.20	18.64	17.48	17.14	17.37
Ferrara	SUD-EST	17.39	17.75	18.70	18.04	18.89	18.69
Ravenna	RAVENNA	17.60	19.26	18.34	18.45	17.53	18.19
Ravenna	LUGO	17.86	17.40	18.08	18.29	18.67	18.40
Ravenna	FAENZA	17.32	18.76	18.16	18.41	18.78	18.40
Forli	FORLI'	17.88	18.09	17.85	18.02	18.06	18.51
Forli	CESENA - VALLE DEL SAVIO	17.58	18.31	18.52	17.23	17.96	18.43
Parma	PARMA	19.28	19.08	18.68	19.50	18.52	19.00
Rimini	RIMINI	16.21	16.48	16.85	16.33	16.82	16.47
Rimini	RICCIONE	15.99	16.80	17.44	18.52	18.79	18.53
Modena	CASTELFRANCO EMILIA	20.03	19.92	19.68	20.05	19.67	19.79
Forli	RUBICONE	16.18	17.69	19.30	18.36	18.43	18.73
Parma	FIDENZA	19.56	19.77	18.78	19.04	19.51	19.49
Piacenza	CITTA' DI PIACENZA	18.77	18.52	18.13	17.47	17.72	17.72
Piacenza	LEVANTE	17.81	18.49	18.76	18.45	17.79	17.63
Piacenza	PONENTE	18.28	18.51	19.23	18.78	18.97	19.17
Parma	VALLI TARO E CENO	17.89	17.32	18.33	18.49	18.74	18.72
Parma	SUD EST	19.26	19.01	19.00	18.54	18.70	19.28
Reggio	MONTECCHIO	19.40	19.81	19.79	19.61	19.59	19.66
Reggio	REGGIO EMILIA	17.63	17.65	17.41	17.43	18.05	18.11

Tab. B.12 - Stima del gap relativo per i distretti socio-sanitari (anni 2007-2012, riferimento al reddito)

		gap relativo %					
Provincia	Denominazione	2007	2008	2009	2010	2011	2012
Reggio	GUASTALLA	24.39	21.09	20.93	19.67	22.47	20.51
Reggio	CORREGGIO	23.90	20.65	21.26	19.78	23.81	19.94
Reggio	SCANDIANO	22.77	19.92	20.01	17.60	20.25	17.80
Reggio	CASTELNUOVO NE' MONTI	26.32	23.15	22.26	20.60	23.39	21.34
Modena	CARPI	23.22	20.40	19.74	17.77	22.09	19.41
Modena	MIRANDOLA	23.32	19.77	20.73	18.42	21.67	19.58
Modena	MODENA	27.24	23.15	22.37	21.61	25.06	21.92
Modena	SASSUOLO	24.68	21.46	21.01	19.15	21.94	20.64
Modena	PAVULLO NEL FRIGNANO	24.91	22.05	20.82	19.74	22.53	20.62
Modena	VIGNOLA	24.52	20.75	19.98	19.85	22.99	20.55
Bologna	CASALECCHIO DI RENO	25.15	21.64	20.67	19.25	22.33	20.15
Bologna	PORRETTA TERME	23.99	20.55	19.95	18.78	21.63	19.41
Bologna	SAN LAZZARO DI SAVENA	25.15	22.10	21.40	19.11	21.97	20.32
Bologna	IMOLA	24.49	20.02	18.97	17.72	20.37	19.07
Bologna	PIANURA EST	22.53	18.65	17.82	16.14	20.23	17.74
Bologna	PIANURA OVEST	22.83	19.66	18.79	17.73	20.37	18.23
Bologna	CITTA' DI BOLOGNA	27.61	24.89	24.80	22.26	25.49	23.05
Ferrara	OVEST	22.33	19.13	18.46	17.34	20.76	18.59
Ferrara	CENTRO-NORD	24.77	21.11	20.14	19.21	22.40	20.46
Ferrara	SUD-EST	23.47	19.82	18.95	17.92	21.35	19.09
Ravenna	RAVENNA	24.14	20.87	19.62	17.58	21.08	19.41
Ravenna	LUGO	23.80	19.62	19.38	18.25	21.66	19.93
Ravenna	FAENZA	25.01	22.44	19.78	19.07	22.63	20.65
Forli	FORLI'	24.77	21.33	20.30	19.32	22.32	20.62
Forli	CESENA - VALLE DEL SAVIO	25.21	21.53	21.03	19.72	23.02	21.22
Parma	PARMA	25.39	23.06	22.23	21.42	23.62	21.16
Rimini	RIMINI	26.62	23.01	22.52	20.29	23.13	21.96
Rimini	RICCIONE	26.75	23.82	21.66	20.27	23.29	21.56
Modena	CASTELFRANCO EMILIA	22.61	19.04	18.89	17.48	20.84	17.73
Forli	RUBICONE	24.38	21.78	20.17	19.53	22.49	20.62
Parma	FIDENZA	25.11	20.58	20.21	18.73	22.54	20.46
Piacenza	CITTA' DI PIACENZA	26.32	22.97	22.29	21.12	24.23	22.42
Piacenza	LEVANTE	24.60	21.88	20.91	19.29	22.23	20.17
Piacenza	PONENTE	26.18	22.33	21.35	20.21	22.98	21.05
Parma	VALLI TARO E CENO	26.08	23.07	22.53	21.09	23.32	21.87
Parma	SUD EST	24.52	21.15	20.51	18.64	21.53	19.73
Reggio	MONTECCHIO	24.43	20.99	20.30	18.68	21.65	19.74
Reggio	REGGIO EMILIA	24.36	20.70	21.11	20.13	23.01	21.67



Tab. B.13 - CV riferiti alla stima del gap relativo per i distretti socio-sanitari dell'Emilia Romagna (anni 2007-2012, riferimento al reddito)

		CV del gap relativo %					
Provincia	Denominazione	2007	2008	2009	2010	2011	2012
Reggio	GUASTALLA	4.84	5.67	6.14	6.41	5.68	6.41
Reggio	CORREGGIO	4.90	5.77	5.53	6.13	5.02	6.38
Reggio	SCANDIANO	4.61	5.39	5.74	6.67	5.48	6.88
Reggio	CASTELNUOVO NE' MONTI	5.04	5.68	6.13	6.59	5.70	6.60
Modena	CARPI	4.79	5.64	6.04	6.56	5.31	6.41
Modena	MIRANDOLA	4.88	5.87	5.64	6.52	5.72	6.56
Modena	MODENA	4.17	4.87	5.41	5.61	4.76	5.65
Modena	SASSUOLO	4.61	5.36	5.89	6.35	5.41	6.03
Modena	PAVULLO NEL FRIGNANO	4.74	5.25	6.11	6.59	5.73	6.44
Modena	VIGNOLA	4.68	5.65	5.94	6.07	5.16	6.11
Bologna	CASALECCHIO DI RENO	4.22	5.32	5.69	6.35	5.50	6.26
Bologna	PORRETTA TERME	4.72	5.60	5.93	6.45	5.46	6.78
Bologna	SAN LAZZARO DI SAVENA	4.35	4.95	5.41	6.46	5.50	5.87
Bologna	IMOLA	4.45	5.67	5.98	6.51	5.73	6.14
Bologna	PIANURA EST	4.37	5.35	5.74	6.57	5.20	6.36
Bologna	PIANURA OVEST	4.79	5.67	6.24	6.47	5.69	6.47
Bologna	CITTA' DI BOLOGNA	3.88	4.33	4.61	5.07	4.31	5.04
Ferrara	OVEST	4.92	5.91	6.32	6.84	5.84	6.78
Ferrara	CENTRO-NORD	4.12	4.98	5.45	5.92	4.94	5.83
Ferrara	SUD-EST	4.54	5.46	6.30	6.78	5.63	6.83
Ravenna	RAVENNA	4.12	4.94	5.41	6.28	5.17	6.05
Ravenna	LUGO	4.67	5.48	6.13	6.72	5.49	6.40
Ravenna	FAENZA	4.65	5.10	6.05	6.46	5.33	6.24
Forli	FORLI'	4.33	5.07	5.49	5.80	4.89	5.71
Forli	CESENA - VALLE DEL SAVIO	4.41	5.19	5.53	6.04	5.15	5.95
Parma	PARMA	4.23	4.71	4.89	5.25	4.76	5.68
Rimini	RIMINI	4.23	4.87	5.22	5.76	5.23	5.72
Rimini	RICCIONE	4.32	5.08	5.71	6.69	5.67	6.46
Modena	CASTELFRANCO EMILIA	4.71	5.46	5.58	6.41	5.45	6.63
Forli	RUBICONE	4.61	5.22	6.31	6.31	5.54	6.29
Parma	FIDENZA	4.48	5.76	5.52	6.16	5.12	6.03
Piacenza	CITTA' DI PIACENZA	4.62	5.35	5.70	6.07	5.36	5.92
Piacenza	LEVANTE	4.86	5.22	5.65	6.32	5.50	6.26
Piacenza	PONENTE	4.65	5.67	6.09	6.54	5.58	6.39
Parma	VALLI TARO E CENO	4.72	5.64	5.77	6.32	5.73	6.42
Parma	SUD EST	4.83	5.53	5.69	6.42	5.61	6.40
Reggio	MONTECCHIO	4.87	5.81	6.21	6.71	5.74	6.62
Reggio	REGGIO EMILIA	4.26	5.19	5.31	5.68	4.87	5.44

Tab. B.14 - Stima dell'indice di Gini per i distretti socio-sanitari dell'Emilia Romagna (anni 2007-2012, riferimento al reddito)

		Indice di Gini %					
Provincia	Denominazione	2007	2008	2009	2010	2011	2012
Reggio	GUASTALLA	26.4	26.2	26.7	26.6	27.0	27.5
Reggio	CORREGGIO	26.5	26.2	27.2	26.7	28.5	27.2
Reggio	SCANDIANO	25.0	25.5	26.6	25.0	25.2	25.3
Reggio	CASTELNUOVO NE' MONTI	27.1	26.8	27.1	26.4	27.1	27.6
Modena	CARPI	25.7	26.2	26.2	25.2	27.3	26.8
Modena	MIRANDOLA	25.4	24.9	27.2	25.3	26.1	26.4
Modena	MODENA	31.6	30.2	30.2	30.6	32.0	30.7
Modena	SASSUOLO	27.7	27.4	27.8	26.3	26.9	28.4
Modena	PAVULLO NEL FRIGNANO	26.1	26.2	26.0	26.0	26.5	27.0
Modena	VIGNOLA	27.3	26.4	26.4	27.1	28.1	28.0
Bologna	CASALECCHIO DI RENO	29.0	28.5	28.3	27.8	28.8	29.1
Bologna	PORRETTA TERME	25.6	25.1	25.7	25.4	25.9	26.4
Bologna	SAN LAZZARO DI SAVENA	28.9	28.8	29.0	27.8	28.5	29.7
Bologna	IMOLA	27.7	25.8	25.6	25.1	25.6	26.8
Bologna	PIANURA EST	25.2	24.6	24.7	23.9	26.0	26.2
Bologna	PIANURA OVEST	25.8	25.5	25.3	25.2	25.5	25.9
Bologna	CITTA' DI BOLOGNA	32.6	33.2	34.1	31.9	32.5	32.6
Ferrara	OVEST	24.3	24.2	24.4	24.2	25.3	25.7
Ferrara	CENTRO-NORD	27.6	26.7	27.3	27.1	28.2	28.8
Ferrara	SUD-EST	24.3	23.5	23.9	23.6	24.7	25.1
Ravenna	RAVENNA	26.5	26.4	26.5	25.1	26.5	27.2
Ravenna	LUGO	25.8	24.6	25.5	25.2	26.4	27.3
Ravenna	FAENZA	27.0	27.8	25.9	25.9	27.5	28.0
Forli	FORLI'	27.0	26.5	26.6	26.6	27.0	27.7
Forli	CESENA - VALLE DEL SAVIO	26.9	26.3	27.0	26.8	27.6	28.5
Parma	PARMA	29.4	30.2	30.3	30.5	30.9	30.5
Rimini	RIMINI	28.1	27.4	27.7	26.9	27.4	28.7
Rimini	RICCIONE	27.3	27.9	26.4	26.1	26.9	27.5
Modena	CASTELFRANCO EMILIA	24.7	24.3	24.8	24.2	25.6	24.9
Forli	RUBICONE	26.2	26.4	25.3	25.2	26.0	26.6
Parma	FIDENZA	27.3	25.8	26.5	25.7	27.4	27.8
Piacenza	CITTA' DI PIACENZA	29.8	29.2	29.5	29.0	30.3	30.7
Piacenza	LEVANTE	26.4	26.7	26.8	26.0	26.8	27.3
Piacenza	PONENTE	28.9	27.4	27.4	27.1	27.8	28.5
Parma	VALLI TARO E CENO	27.5	27.6	28.0	27.2	27.5	28.5
Parma	SUD EST	27.4	26.9	27.3	26.4	26.9	28.0
Reggio	MONTECCHIO	26.6	26.2	26.4	25.7	26.4	27.0
Reggio	REGGIO EMILIA	27.1	26.5	28.0	28.0	28.3	29.6

Tab. B.15 - CV riferiti alla stima dell'indice di Gini per i distretti socio-sanitari dell'Emilia Romagna (anni 2007-2012, riferimento al reddito)

		CV dell'indice di Gini %					
Provincia	Denominazione	2007	2008	2009	2010	2011	2012
Reggio	GUASTALLA	5.66	5.72	5.77	5.37	5.71	5.60
Reggio	CORREGGIO	5.81	5.80	5.34	5.38	5.21	5.63
Reggio	SCANDIANO	5.55	5.35	5.25	5.64	5.66	5.99
Reggio	CASTELNUOVO NE' MONTI	5.84	5.71	5.75	5.75	5.74	5.67
Modena	CARPI	5.71	5.48	5.71	5.66	5.41	5.57
Modena	MIRANDOLA	5.85	5.89	5.28	5.78	5.86	5.75
Modena	MODENA	4.91	4.99	5.12	4.96	4.69	5.13
Modena	SASSUOLO	5.54	5.29	5.52	5.54	5.62	5.34
Modena	PAVULLO NEL FRIGNANO	5.46	5.25	5.59	5.59	5.79	5.68
Modena	VIGNOLA	5.45	5.53	5.61	5.31	5.31	5.40
Bologna	CASALECCHIO DI RENO	5.09	5.22	5.37	5.55	5.35	5.32
Bologna	PORRETTA TERME	5.71	5.61	5.67	5.53	5.56	5.89
Bologna	SAN LAZZARO DI SAVENA	5.14	4.92	5.14	5.40	5.48	5.06
Bologna	IMOLA	5.29	5.74	5.78	5.68	5.82	5.44
Bologna	PIANURA EST	5.42	5.36	5.48	5.70	5.32	5.44
Bologna	PIANURA OVEST	5.65	5.59	5.80	5.63	5.78	5.60
Bologna	CITTA' DI BOLOGNA	4.67	4.52	4.55	4.71	4.62	4.67
Ferrara	OVEST	5.99	5.99	5.98	5.98	5.95	5.89
Ferrara	CENTRO-NORD	5.08	5.06	5.11	5.12	5.03	5.14
Ferrara	SUD-EST	5.62	5.50	5.94	5.89	5.87	5.96
Ravenna	RAVENNA	5.11	5.19	5.26	5.52	5.34	5.24
Ravenna	LUGO	5.60	5.57	5.80	5.75	5.61	5.65
Ravenna	FAENZA	5.46	5.22	5.78	5.69	5.45	5.54
Forli	FORLI'	5.16	5.22	5.21	5.13	5.19	5.20
Forli	CESENA - VALLE DEL SAVIO	5.38	5.31	5.34	5.28	5.34	5.22
Parma	PARMA	5.22	4.94	4.73	4.76	4.90	5.12
Rimini	RIMINI	5.14	4.96	5.16	5.10	5.38	5.09
Rimini	RICCIONE	5.16	5.04	5.49	5.93	5.85	5.80
Modena	CASTELFRANCO EMILIA	5.74	5.46	5.31	5.53	5.64	5.76
Forli	RUBICONE	5.24	5.25	6.03	5.67	5.75	5.71
Parma	FIDENZA	5.34	5.79	5.27	5.35	5.26	5.37
Piacenza	CITTA' DI PIACENZA	5.35	5.33	5.34	5.31	5.16	5.26
Piacenza	LEVANTE	5.76	5.31	5.35	5.54	5.61	5.51
Piacenza	PONENTE	5.35	5.58	5.65	5.67	5.65	5.58
Parma	VALLI TARO E CENO	5.57	5.52	5.36	5.56	5.75	5.68
Parma	SUD EST	5.61	5.44	5.44	5.57	5.78	5.46
Reggio	MONTECCHIO	5.73	5.82	5.77	5.84	5.75	5.74
Reggio	REGGIO EMILIA	5.17	5.33	5.14	4.97	5.02	4.95

Tab. B.16 - Stima del tasso di deprivazione materiale per i distretti socio-sanitari dell'Emilia Romagna (anni 2007-2012, riferimento al reddito)

		Tasso di deprivazione materiale %					
Provincia	Denominazione	2007	2008	2009	2010	2011	2012
Reggio	GUASTALLA	12.4	15.4	17.4	15.7	18.5	19.5
Reggio	CORREGGIO	11.2	12.5	14.7	13.3	15.6	16.6
Reggio	SCANDIANO	12.4	15.8	16.7	16.7	18.1	17.9
Reggio	CASTELNUOVO NE' MONTI	5.6	6.5	7.2	8.0	9.5	10.8
Modena	CARPI	13.4	14.5	14.6	14.2	14.9	17.7
Modena	MIRANDOLA	10.4	11.1	13.1	14.0	15.6	16.9
Modena	MODENA	9.6	9.9	9.7	12.4	13.6	16.3
Modena	SASSUOLO	13.5	14.9	15.8	17.4	18.0	21.2
Modena	PAVULLO NEL FRIGNANO	6.0	7.3	8.1	9.4	10.8	12.4
Modena	VIGNOLA	10.1	12.8	13.7	14.0	17.5	19.2
Bologna	CASALECCHIO DI RENO	7.9	7.4	8.4	8.5	10.5	11.8
Bologna	PORRETTA TERME	8.4	10.4	11.2	14.2	12.2	14.0
Bologna	SAN LAZZARO DI SAVENA	9.0	10.6	10.2	9.4	10.5	10.6
Bologna	IMOLA	7.6	9.1	10.2	10.7	10.9	12.5
Bologna	PIANURA EST	9.1	10.3	9.7	8.4	13.8	12.1
Bologna	PIANURA OVEST	9.5	13.3	13.9	13.6	13.2	14.0
Bologna	CITTA' DI BOLOGNA	10.9	11.2	12.2	14.7	18.7	21.9
Ferrara	OVEST	11.4	13.3	13.6	13.4	14.4	15.5
Ferrara	CENTRO-NORD	8.3	7.3	8.4	9.2	10.1	10.2
Ferrara	SUD-EST	9.4	10.5	9.9	8.8	10.5	13.7
Ravenna	RAVENNA	8.6	9.8	10.7	13.3	12.8	14.1
Ravenna	LUGO	9.2	9.8	9.6	10.0	10.4	12.2
Ravenna	FAENZA	7.4	9.3	10.5	9.7	14.3	14.5
Forli	FORLI'	11.8	11.9	11.4	13.9	15.6	18.7
Forli	CESENA - VALLE DEL SAVIO	9.0	11.6	11.5	15.0	15.2	17.7
Parma	PARMA	7.6	8.4	7.4	10.5	12.1	15.2
Rimini	RIMINI	8.6	8.5	9.9	11.8	13.0	13.3
Rimini	RICCIONE	11.4	13.2	13.5	13.5	15.2	17.1
Modena	CASTELFRANCO EMILIA	12.4	13.4	14.7	17.5	16.5	19.1
Forli	RUBICONE	10.9	12.9	14.9	14.9	18.2	21.1
Parma	FIDENZA	11.2	10.4	14.6	12.9	15.4	18.1
Piacenza	CITTA' DI PIACENZA	8.9	10.3	11.1	14.9	16.0	16.8
Piacenza	LEVANTE	7.4	7.4	9.2	11.2	13.1	16.9
Piacenza	PONENTE	7.7	8.9	9.9	10.9	13.2	15.1
Parma	VALLI TARO E CENO	5.9	5.8	6.0	7.8	8.2	9.5
Parma	SUD EST	10.8	11.1	12.6	11.8	15.0	16.2
Reggio	MONTECCHIO	9.6	11.5	12.9	13.1	14.2	15.4
Reggio	REGGIO EMILIA	10.5	10.4	10.6	11.1	12.1	16.3

Tab. B.17 - CV riferiti alla stima del tasso di deprivazione materiale per i distretti socio-sanitari (anni 2007-2012, riferimento al reddito)

		CV del tasso di deprivazione materiale %					
Provincia	Denominazione	2007	2008	2009	2010	2011	2012
Reggio	GUASTALLA	26.46	25.98	27.36	22.36	23.27	22.56
Reggio	CORREGGIO	28.84	27.79	26.72	23.72	25.15	24.35
Reggio	SCANDIANO	23.92	24.62	25.53	25.58	23.44	22.04
Reggio	CASTELNUOVO NE' MONTI	39.85	37.68	36.20	32.86	32.60	28.94
Modena	CARPI	25.22	24.36	25.12	21.09	21.45	20.99
Modena	MIRANDOLA	31.68	31.30	27.66	27.75	25.46	24.80
Modena	MODENA	23.48	25.24	23.35	23.82	20.87	21.12
Modena	SASSUOLO	24.67	23.92	24.73	24.52	23.23	22.47
Modena	PAVULLO NEL FRIGNANO	27.49	26.59	25.69	24.64	25.25	25.30
Modena	VIGNOLA	25.47	25.17	23.56	22.93	21.44	20.55
Bologna	CASALECCHIO DI RENO	24.34	24.13	25.04	25.90	27.53	27.91
Bologna	PORRETTA TERME	26.61	26.91	26.45	30.80	25.53	27.85
Bologna	SAN LAZZARO DI SAVENA	25.80	27.72	27.33	26.56	27.44	23.91
Bologna	IMOLA	25.99	27.88	27.25	25.72	23.76	23.42
Bologna	PIANURA EST	23.50	22.66	21.83	24.70	23.68	22.03
Bologna	PIANURA OVEST	23.43	26.07	24.94	22.95	22.61	22.69
Bologna	CITTA' DI BOLOGNA	22.25	21.92	21.58	20.75	18.86	17.35
Ferrara	OVEST	28.44	27.70	27.07	26.41	25.71	25.24
Ferrara	CENTRO-NORD	25.73	25.04	26.29	25.87	24.06	23.06
Ferrara	SUD-EST	28.40	29.87	28.78	26.97	26.67	27.37
Ravenna	RAVENNA	23.20	24.04	24.44	28.58	22.90	21.84
Ravenna	LUGO	32.03	28.93	27.92	25.75	25.35	25.09
Ravenna	FAENZA	24.02	25.01	26.44	23.68	25.72	23.10
Forli	FORLI'	25.43	24.11	22.15	22.51	23.34	22.69
Forli	CESENA - VALLE DEL SAVIO	23.48	23.24	22.53	24.04	22.66	22.16
Parma	PARMA	24.91	25.18	24.33	25.14	23.29	24.65
Rimini	RIMINI	23.70	22.87	22.92	23.76	24.34	21.70
Rimini	RICCIONE	24.72	26.40	25.79	28.78	27.66	27.19
Modena	CASTELFRANCO EMILIA	27.65	23.26	21.44	22.95	21.61	21.84
Forli	RUBICONE	21.80	22.45	27.56	23.52	24.33	23.88
Parma	FIDENZA	28.66	26.64	25.84	21.27	21.20	21.10
Piacenza	CITTA' DI PIACENZA	26.36	25.20	25.35	25.73	23.47	22.81
Piacenza	LEVANTE	25.09	23.71	23.15	25.50	23.86	23.16
Piacenza	PONENTE	28.96	27.67	27.81	27.25	25.50	24.82
Parma	VALLI TARO E CENO	31.84	30.35	28.74	32.24	31.81	32.35
Parma	SUD EST	26.93	24.69	24.68	23.42	23.51	23.30
Reggio	MONTECCHIO	26.69	26.92	26.65	26.19	26.08	26.27
Reggio	REGGIO EMILIA	26.18	25.10	23.95	23.10	21.58	20.69

## Riferimenti bibliografici

Aaberge R., Melby I. (2005) The sensitivity of income inequality to choice of equivalence scales, *The Review of Income and Wealth*, **44**, 565-569.

Alfons A., Templ M. (2013). Estimation of social exclusion indicators from complex surveys: The R package laeken. *Journal of Statistical Software*, **54**, 15.

Atkinson A.B. (1998) Poverty in Europe, *Blackwell Publishing*, New York.

Atkinson A.B., Rainwater L.B., Smeeding T.M. (1995) Income distribution in OECD countries, Paris: OECD.

Ayala L., Jurado A., Pérez-Mayo J. (2011) Income poverty and multidimensional deprivation: lessons from cross-regional analysis, *The Review of Income and Wealth*, **57**, 40-60.

Balog M., Thornburn D. (2007) Extreme values treatment for samples from skew income distributions, *Statistics in Transition*, **8**, 139-153.

Balduzzi R., Moirano F. (2011) La rete dei distretti sanitari in Italia (editoriale), *Quaderno di Monitor n. 8*, Agenas.

Bellentani M.D., Damiani G., Ronconi A., Catania S., Bugliari Armenio L. (2011) Report dell'indagine, in *La rete dei distretti sanitari in Italia*, *Quaderno di Monitor n. 8*, Agenas.

Banca d'Italia (2006), I bilanci delle famiglie italiane nell'anno 2004, *Supplementi al Bollettino Statistico*, Nuova serie, n.7.

Bellù L.G., Liberati P. (2005), Impacts of Policies on Poverty. The Definition of Poverty, *EASYPol*, FAO, scaricabile da [www.fao.org/docs/up/easypol/321/axioms\\_pov\\_msmt\\_008en.pdf](http://www.fao.org/docs/up/easypol/321/axioms_pov_msmt_008en.pdf).

Bibby B.M., Sorensen M. (2003), Generalized hyperbolic and inverse Gaussian distributions: limiting cases and approximation of processes, in Rachev, S.T. (ed.) *Handbook of Heavy Tailed Distributions in Finance*, Elsevier Science B.V., 211-248.

Carbonaro G. (1985) Nota sulla scala di equivalenza, Presidenza del Consiglio dei Ministri, Commissione Indagine sulla Povertà "La Povertà in Italia", Studi di base.

Citro C.F., Michaels R.T. (1995) Measuring poverty: a new approach, National Academic Press, Washington D.C.

Curcetti C. (2011) Regione Emilia Romagna, in *La rete dei distretti sanitari in Italia*, *Quaderno di Monitor n. 8*, Agenas.

Datta G.S., Ghosh M., Steorts, R., Maples J. (2011). Bayesian benchmarking with applications to small area estimation, *TEST*, **20**, 574-588.

Deville J. C., Särndal C. H. (1992) Calibration estimators in survey sampling, *Journal of the American Statistical Association*, **87**, 376-382.

Dunson D.B. and Neelan B. (2003). Bayesian Inference on Order-Constrained Parameters in Generalized Linear Models, *Biometrics*, **59**, 286-295.

Eurostat (2007) Comparative EU statistics on income and living conditions: issues and challenges. Proceedings of the EU-SILC conference, Helsinki 6-8 November 2006. Eurostat Methodologies and working papers

Fabrizi E., Ferrante M.R., Pacei S. (2008) Measuring sub-national income poverty using a Small-Area multivariate approach, *The Review of Income and Wealth*, **54**, 597-615.

Fabrizi E., Trivisano C. (2016) Small area estimation of the Gini concentration coefficient, *Computational Statistics & Data Analysis*; (99): 223-234.

Fruhwirth-Schnatter S., Wagner H. (2010). Bayesian variable selection for random intercept modelling of Gaussian and non-Gaussian data. In J. Bernardo, M. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, M. West (eds.), *Bayesian Statistics*, 9, 165–200, Oxford Univ. Press.

Gelman A., Meng X., Stern H.S. (1996) Posterior predictive assessment of model fitness via realized discrepancies (with discussion), *Statistica Sinica*, **6**, 733-807.

George E. I., McCulloch, R. E. (1993). Variable selection via Gibbs sampling." *J. Am. Stat. Assoc.*, **85**: 398-409.

Isaki C.T., Fuller W.A. (1982) Survey design under the regression superpopulation model, *Journal of the American Statistical Association*, **77**, 89-96.

ISTAT (2004) La povertà assoluta: informazioni sulla metodologia di stima, documento scaricabile da:

[www.istat.it/salastampa/comunicati/non\\_calendario/20040503\\_00/poverta\\_mag04.pdf](http://www.istat.it/salastampa/comunicati/non_calendario/20040503_00/poverta_mag04.pdf)

ISTAT (2008) Reddito e condizioni di vita. Indagine sulle condizioni di vita. Anno 2005, scaricabile da <http://www.istat.it/>

ISTAT (2015), Il benessere equo e sostenibile delle province, in <http://www.besdelleprovince.it/>

ISTAT (2015), Il benessere equo e sostenibile in Italia, in [http://www.istat.it/it/files/2015/12/Rapporto BES 2015.pdf](http://www.istat.it/it/files/2015/12/Rapporto_BES_2015.pdf)

Rao J.N.K. (2003), *Small Area Estimation*, Wiley series in Survey Mehtodology, John Wiley and Sons.

Rubin D. (1984) Bayesian justifiable and relevant frequency calculation for the applied statistician, *Annals of Statistics*, **12**, 1151-1172.

Särndal C. H., Lundström S. (2005) Estimation in surveys with nonresponse, *John Wiley and Sons*, New York.

Sen A. (1983) Poor, relatively speaking, *Oxford Economic Papers*, New series, **35**, 153-169.

Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., (2003), WinBUGS User Manual Version 1.4, scaricabile al <http://www.mrc-bsu.cam.ac.uk/bugs>.

Thomas A., O' Hara B., Ligges U., Sturz S. (2006) Making BUGS Open, *R News*, **6**, 12-17.

Tillé Y., Matei A. (2008) The Sampling package. Functions for drawing and calibrating samples, scaricabile da <http://cran.r-project.org/web/packages>.

Vandewalle B., Beirlant J., Christmann A., Hubert M. (2007). A robust estimator for the tail index of Pareto-type distributions. *Computational Statistics and Data Analysis*, **51**, 6252–6268.

Van Kerm, P. (2007). *Extreme incomes and the estimation of poverty and inequality indicators from EUSILC*. Working Paper Series 2007-01. Centre d'Etudes de Populations, de Pauvrete et de Politiques Socio- Economiques International Network for Studies in Technology, Environment, Alternatives, Development.