

STIMA DELLA SERIE TEMPORALE 2009-2018 DI PARAMETRI DI  
POVERTÀ, DEPRIVAZIONE E DISUGUAGLIANZA  
PER I DISTRETTI SOCIO SANITARI  
E PER LE PROVINCE DELLA REGIONE EMILIA-ROMAGNA

effettuata sulla base dei dati tratti dalla  
Indagine su Reddito e condizioni di vita (EU-SILC)  
per gli anni di riferimento dal 2009 al 2018,  
mediante l'impiego di tecniche di  
stima per piccole aree

*Aggiornamento dei Report precedenti*

**Maria Rosaria Ferrante\*, Enrico Fabrizi\*\***

\*Dipartimento di Scienze Economiche, Università di Bologna

\*\*DISES, Università Cattolica del S. Cuore, Piacenza

Dicembre 2020

# Indice

1. Introduzione .....	3
2. Stime da modello per provincia .....	5
3. Stime da modello per distretti .....	18
Appendice 1. La metodologia di stima .....	33
1. Stima della varianza delle stime dirette.....	33
2. La stima per piccole aree .....	35
2.1. Sintesi delle principali novità.....	35
2.2. I modelli di stima per piccole aree.....	37
2.2.1 Modello univariato per i tassi.....	37
2.2.2 Modello univariato per l'indice di Gini .....	41
2.2.3 Modello univariato per la media del logaritmo del reddito .....	45
2.2.4 Un modello multivariato.....	45
2.2.5 Specificazione delle distribuzioni a priori per gli iperparametri.....	46
2.3. Selezione variabili ausiliarie.....	46
2.4. Simulazione delle distribuzioni a posteriori .....	49
2.5. La selezione dei modelli.....	50
2.6. La stima del <i>Gap</i> mediano .....	51
2.7. Benchmarking.....	53
APPENDICE 2. Il disegno campionario.....	56
Riferimenti bibliografici .....	60

## 1. Introduzione

L'obiettivo di questa ricerca è la stima di parametri di povertà, deprivazione e disuguaglianza per i distretti sociosanitari e per le province dell'Emilia-Romagna, con riferimento agli anni dal 2009 al 2018. Al fine di ottenere un quadro completo della condizione delle famiglie in Emilia-Romagna, sono stati presi in esame quattro parametri in grado di cogliere diversi aspetti del fenomeno, ossia

1. il *rischio di povertà*, talvolta denominato anche *tasso di povertà*, che misura la diffusione della povertà monetaria;
2. il *poverty Gap* mediano, che misura l'intensità della povertà;
3. l'*indice di Gini*, che misura la disuguaglianza della distribuzione dei redditi;
4. il *tasso di deprivazione materiale severa*, che è una misura non monetaria della povertà.

La scelta dei distretti sociosanitari 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. Negli anni, in Regione, 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 sociosanitaria.

Il riferimento alle province (NUTS3 nella *Nomenclatura delle unità territoriali statistiche*) è importante nell'ottica di offrire ai *policy makers* locali informazioni statistiche territorialmente disaggregate, e quindi più aderenti rispetto a quelle regionali o addirittura nazionali alle esigenze informative connesse ai processi di *governance* delle aree provinciali o metropolitane. 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 le autorità 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 ad un livello di dettaglio territoriale così "fine". Le stime qui prodotte a livello provinciale potrebbero quindi completare ed arricchire il set di indicatori del benessere a livello provinciale già reso disponibile 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.

Le stime sono state ottenute utilizzando i dati tratti dalla Indagine su reddito e condizioni di vita (Eu-Silc), con periodo di riferimento 2009-2018 e mediante l'applicazione di 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 "diretti" che l'Istituto Nazionale di Statistica (Istat) utilizza correntemente per stimare questi parametri.

Qui di seguito sono riportate le stime per piccole aree dei quattro parametri su elencati e i relativi coefficienti di variazione per provincia (Paragrafo 2) e per distretto sociosanitario (Paragrafo 3).

Con riferimento ai parametri reddituali adottiamo qui la convenzione utilizzata da Istat (Istat, 2018) in cui si adopera la dicitura "indicatore riferito all'anno  $t$ ", dove  $t$  è l'anno di riferimento dell'indagine, anche se l'indicatore è calcolato sui dati di reddito riferiti all'anno solare precedente,  $t - 1$ . (Ad esempio, il Rischio di povertà per l'anno 2018 è calcolato sulla base dei redditi riferiti all'anno 2017.)

La metodologia di stima per piccole aree e i metodi di stima della varianza degli stimatori dei parametri sono illustrati in dettaglio in APPENDICE 1.

Infine, nella APPENDICE 2 è illustrata brevemente la strategia di campionamento adottata da Istat nella Indagine EU-SILC.

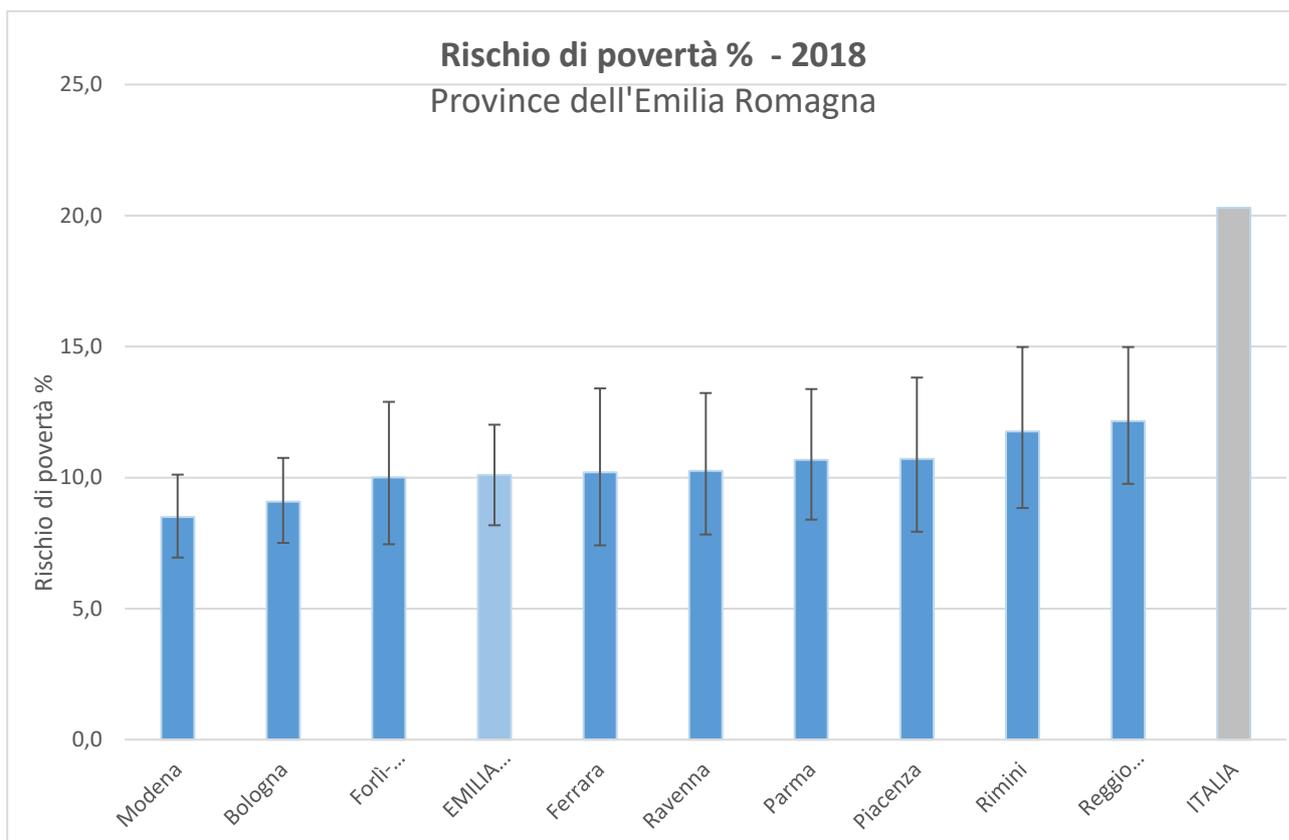
## **NOTA**

L'Istat ha di recente aggiornato i coefficienti di riporto all'universo per le indagini precedenti il 2014, rivisti sulla base dei risultati del Censimento della popolazione del 2011. Al momento della redazione del presente rapporto, i file di microdati EU-SILC aggiornati non erano stati ancora rilasciati da Istat.

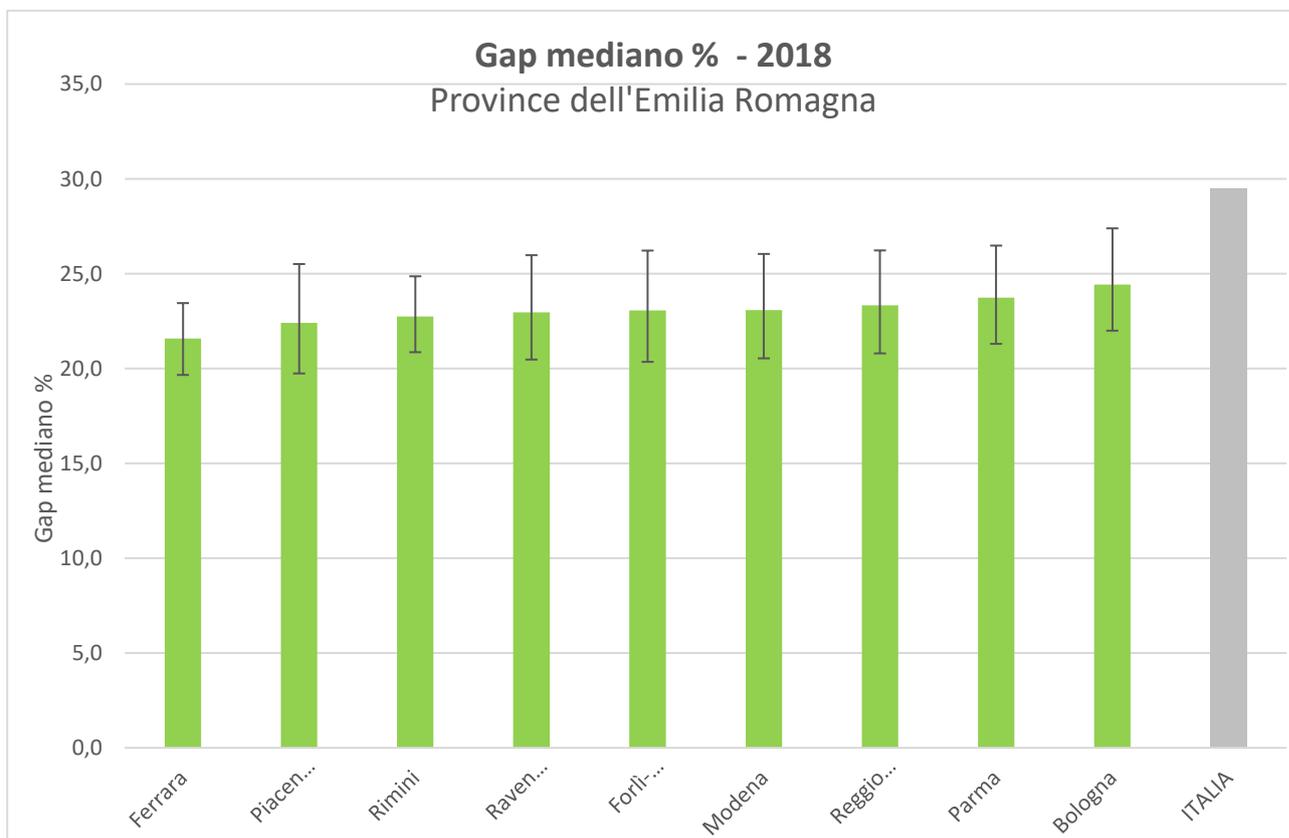
Le stime in questo Rapporto sono state ottenute sulla base della prima versione dei file di microdati EU-SILC. Si potrebbero quindi osservare delle lievi discrepanze tra le stime a livello regionale qui presentate e quelle pubblicate nei rapporti ufficiali da Istat.

## **2. Stime per piccole aree per provincia**

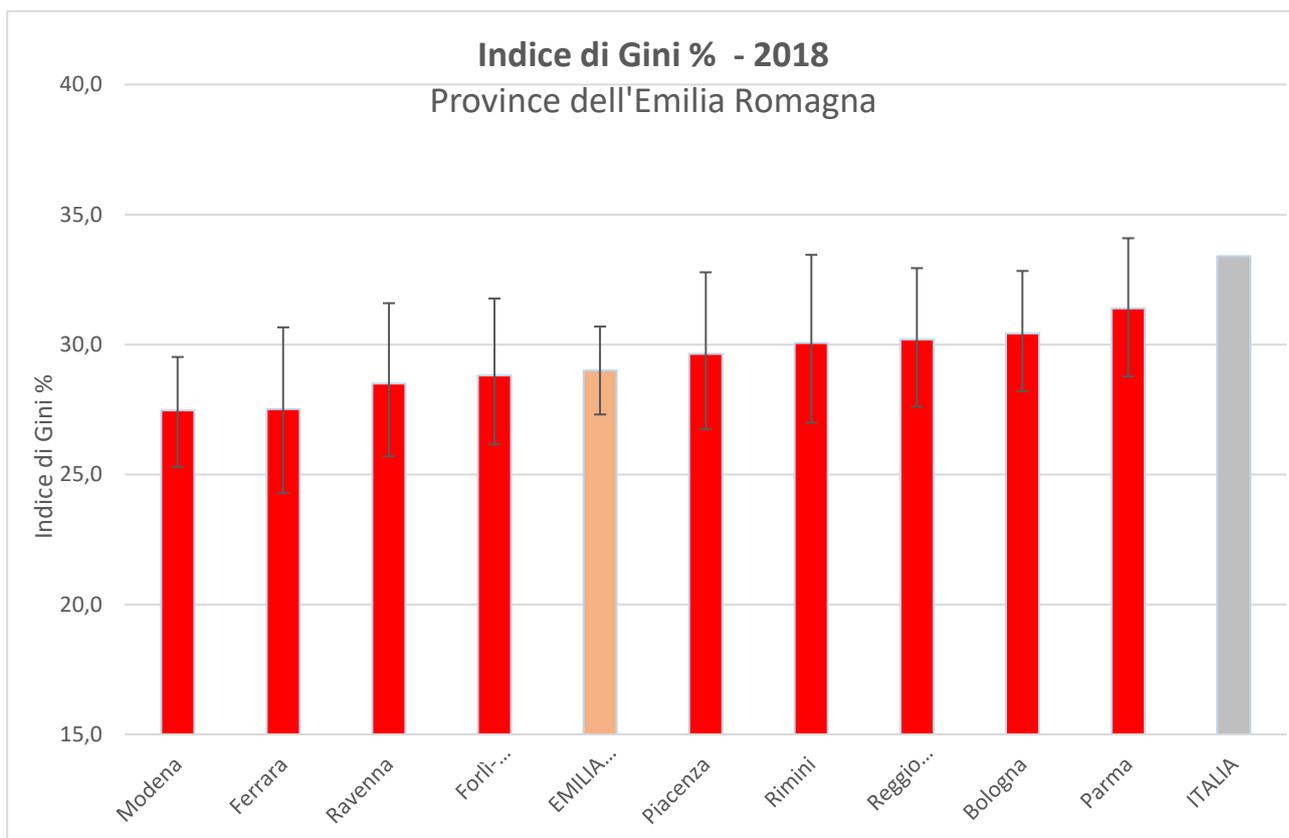
**Rischio di povertà e relativi coefficienti di variazione per provincia. Anno 2018 (valori percentuali)**



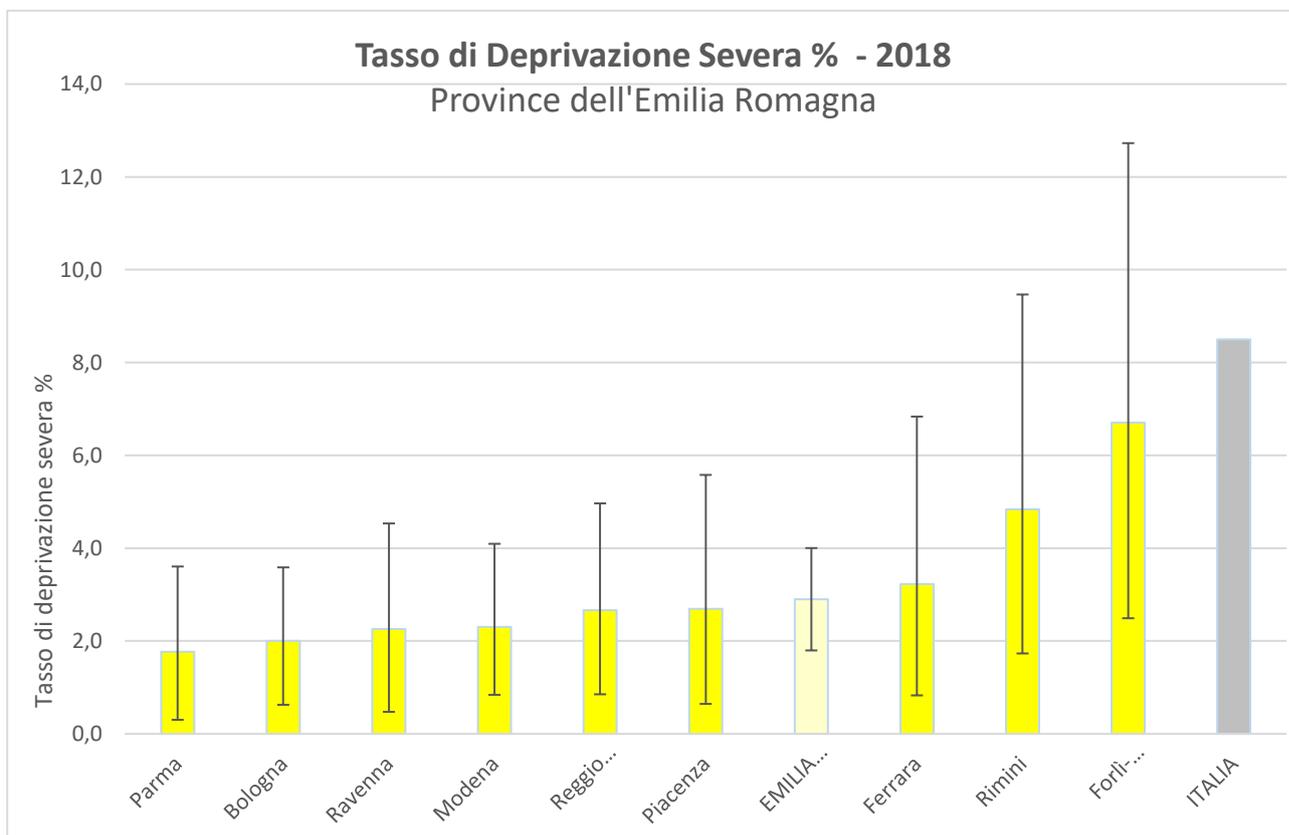
**Gap mediano e relativi coefficienti di variazione per provincia. Anno 2018 (valori percentuali)**



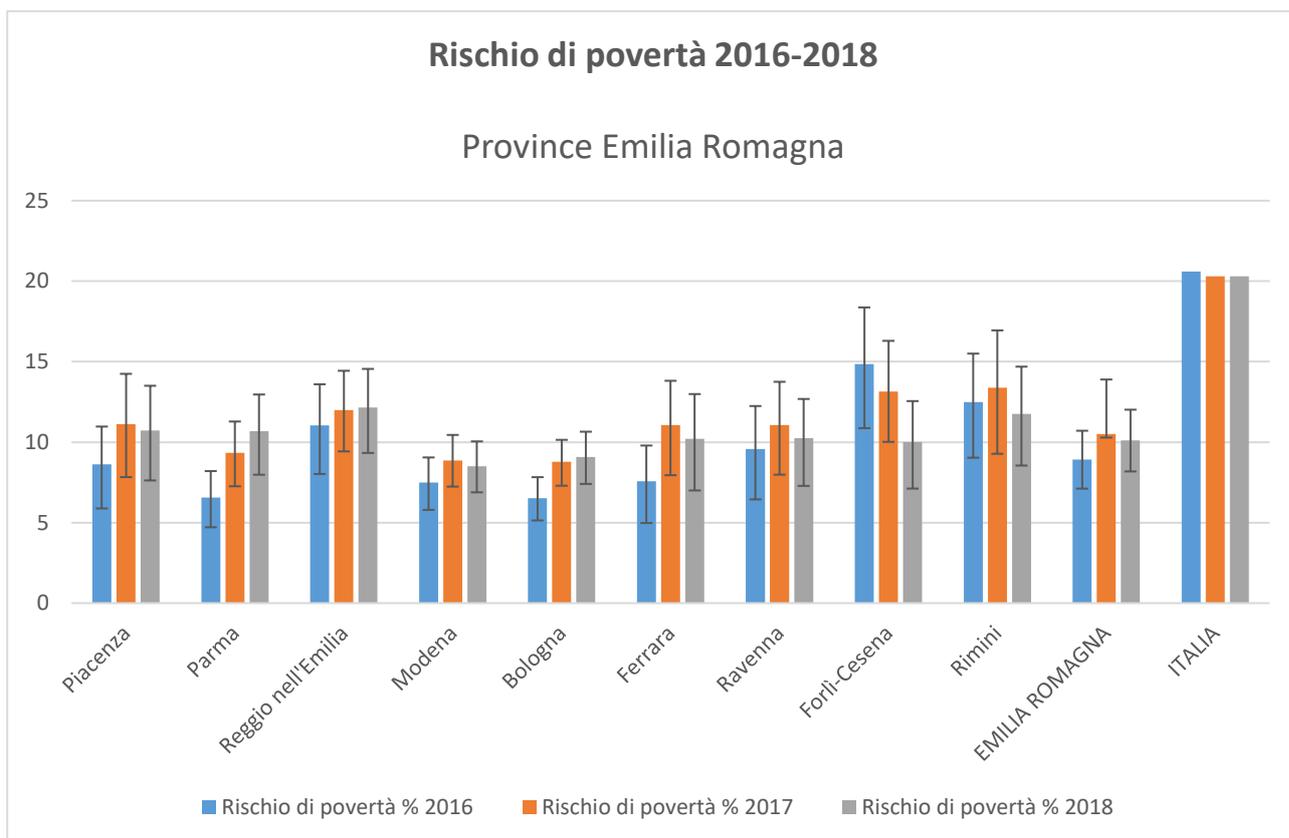
**Indice di Gini e relativi coefficienti di variazione per provincia. Anno 2018 (valori percentuali)**



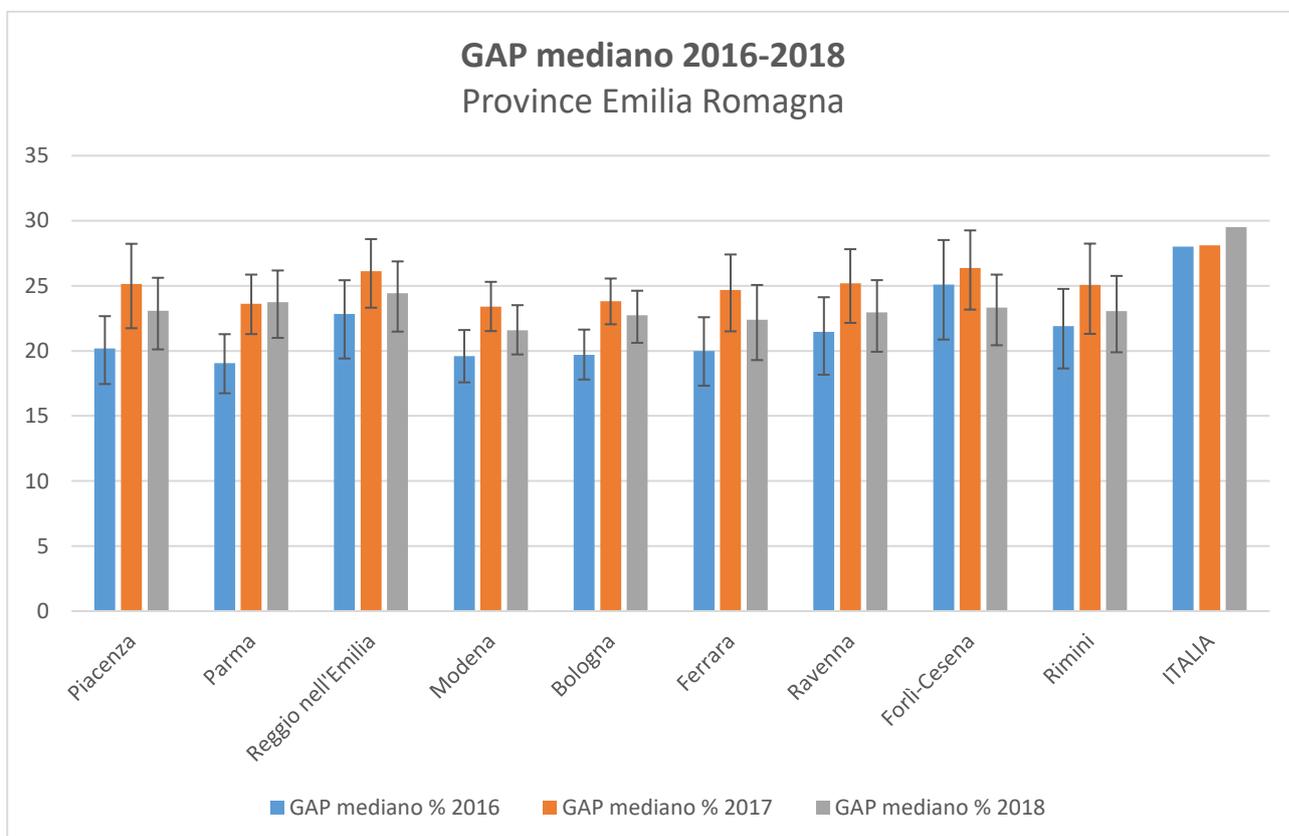
**Tasso di deprivazione severa e relativi coefficienti di variazione per provincia. Anno 2018 (valori percentuali)**



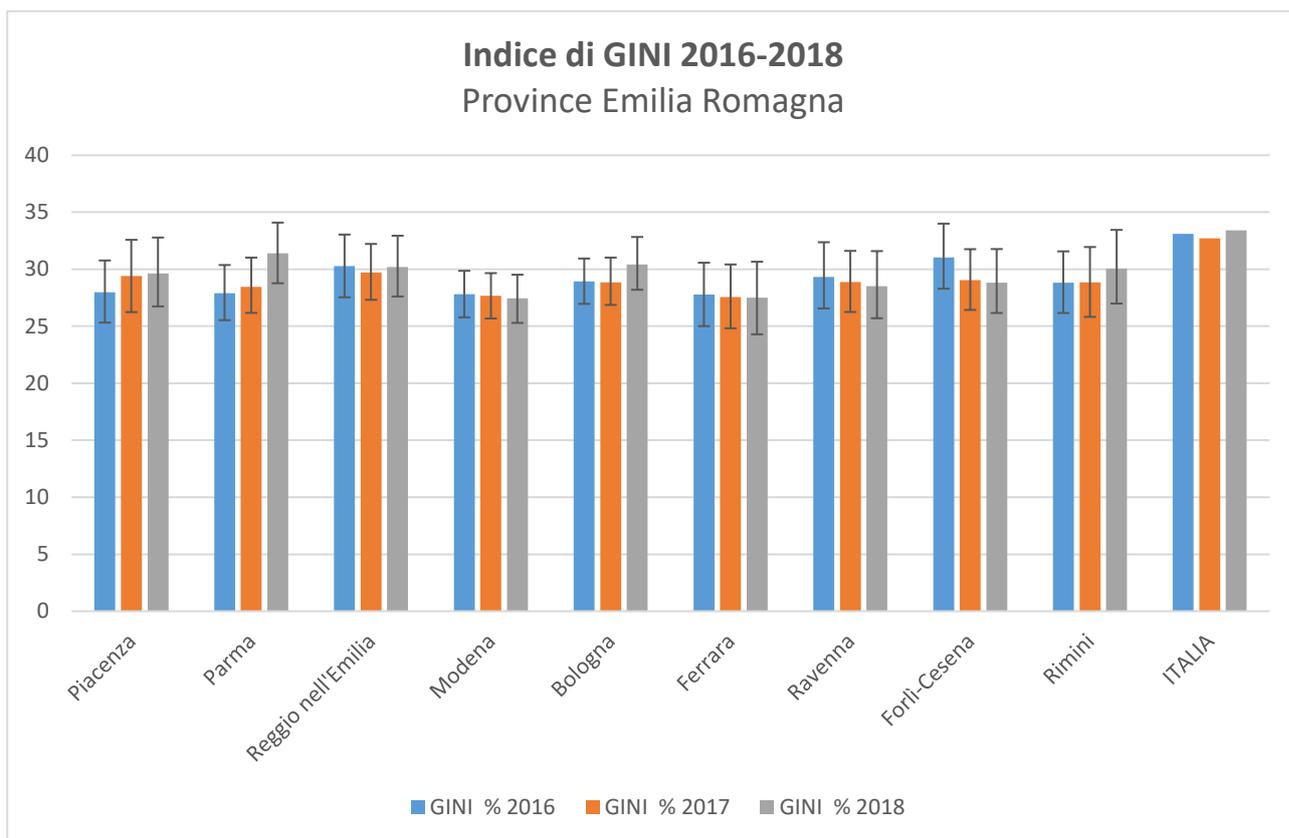
**Rischio di povertà e relativi coefficienti di variazione per provincia. Anni 2016-2018 (valori percentuali)**



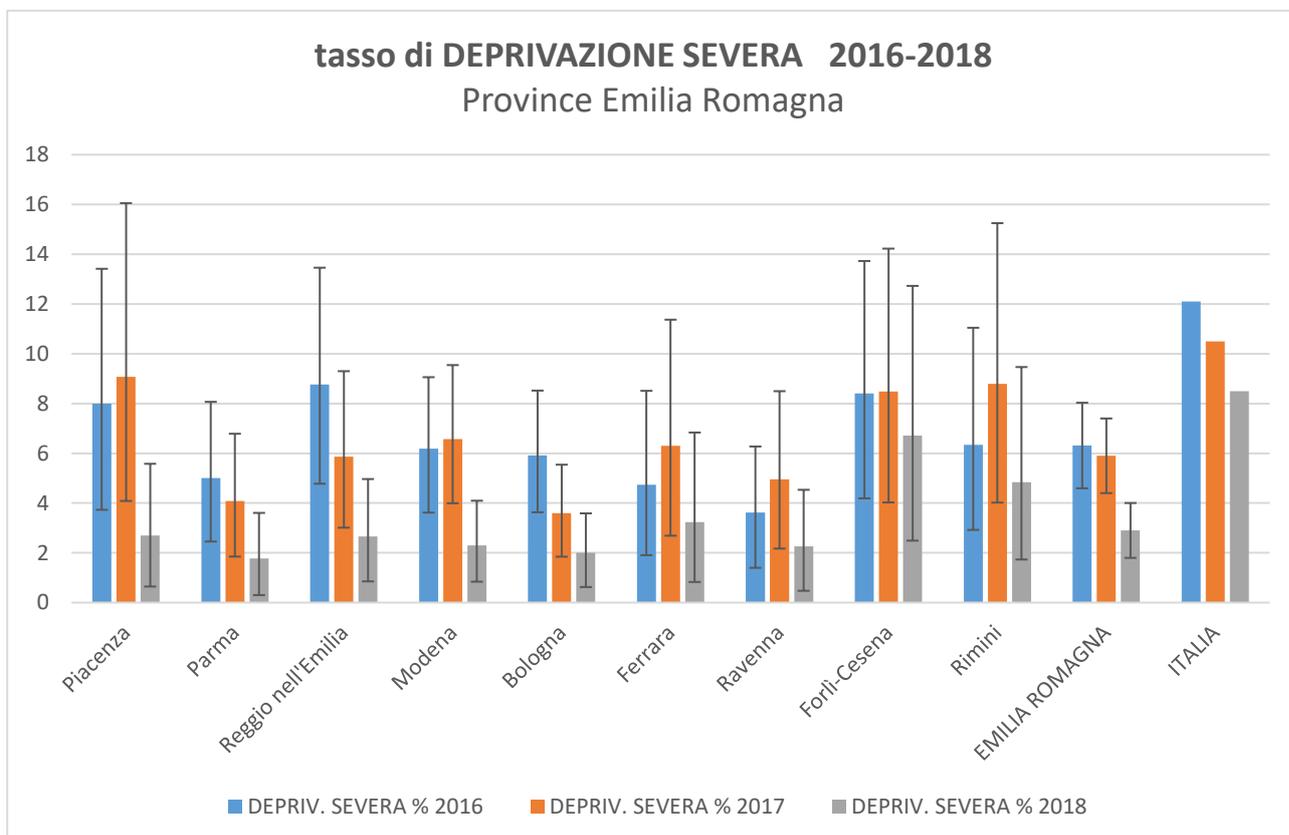
**Gap mediano e relativi coefficienti di variazione per provincia. Anni 2016-2018 (valori percentuali)**



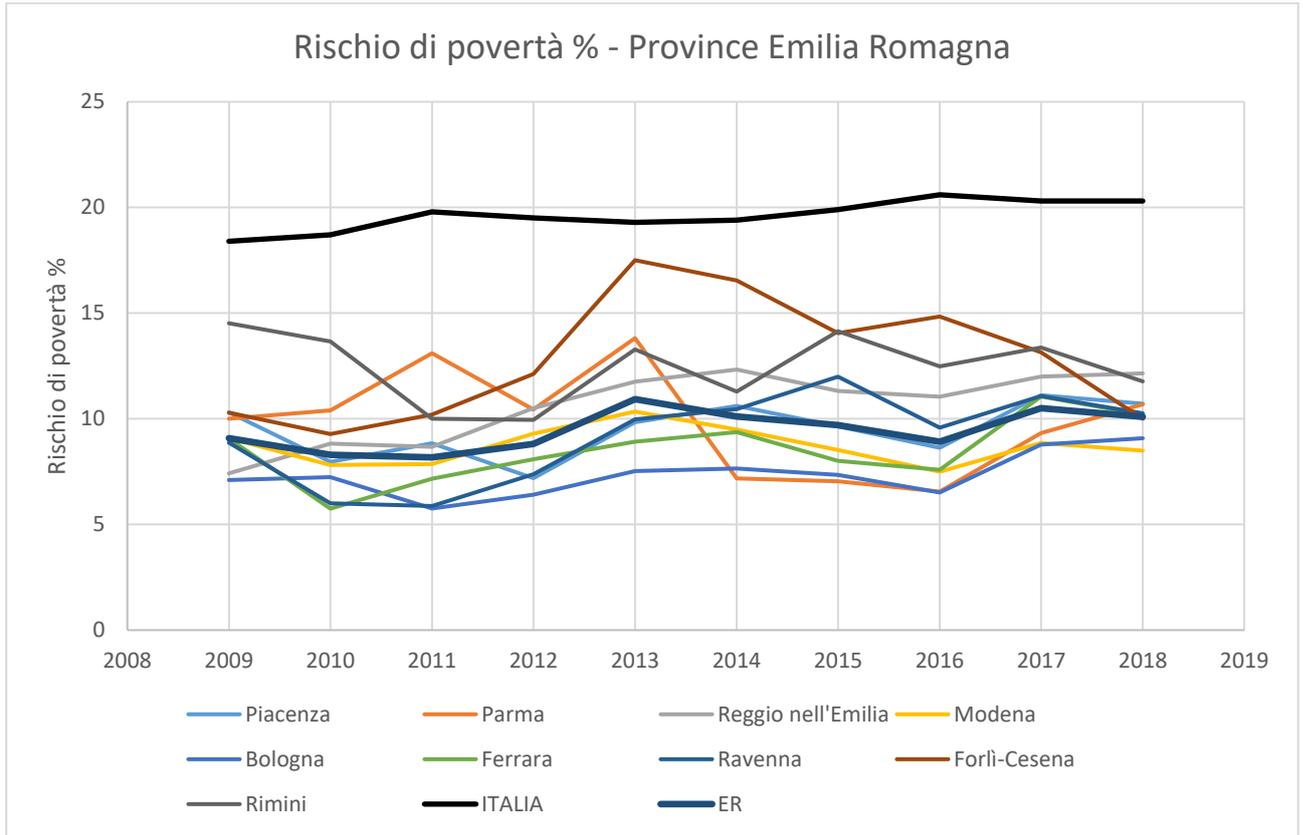
**Indice di Gini e relativi coefficienti di variazione per provincia. Anni 2016-2018 (valori percentuali)**



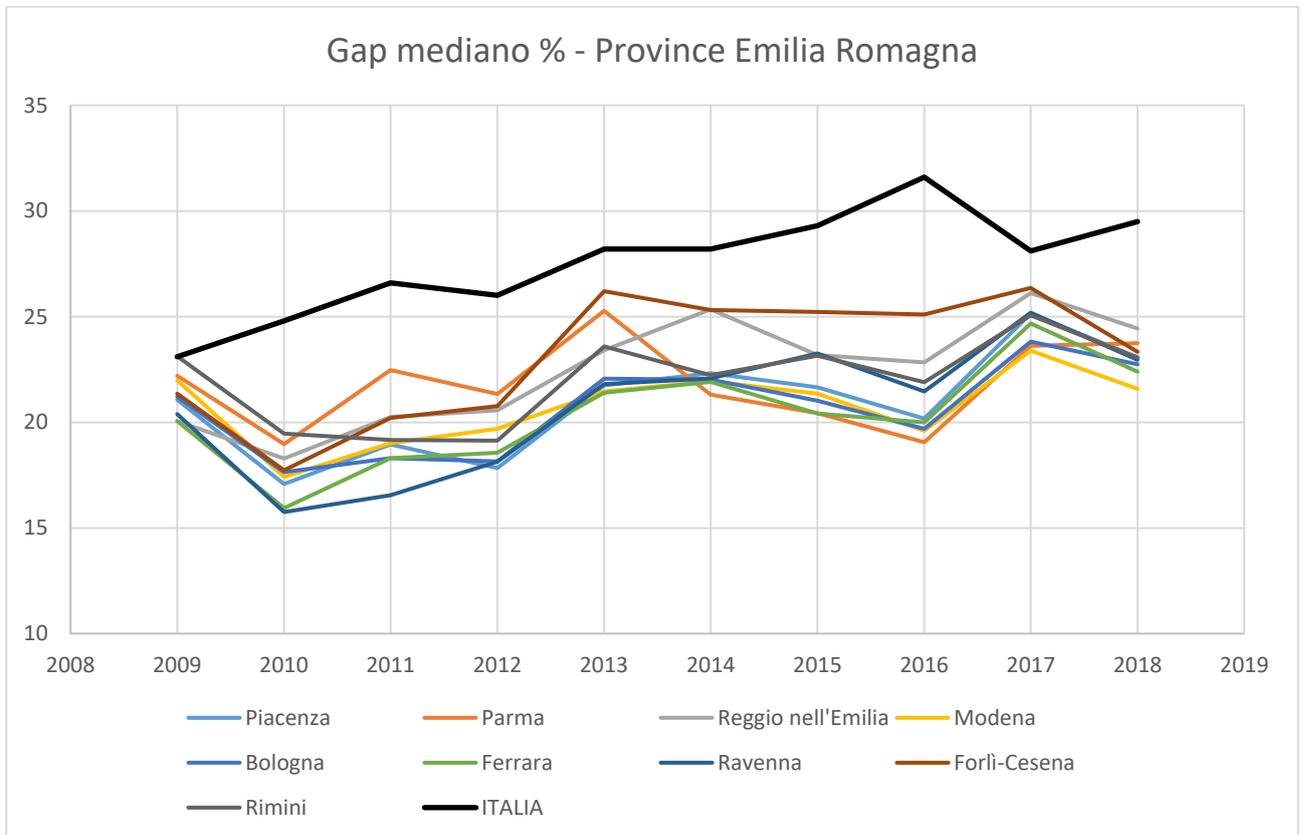
**Tasso di deprivazione e relativi coefficienti di variazione per provincia. Anni 2016-2018 (valori percentuali)**



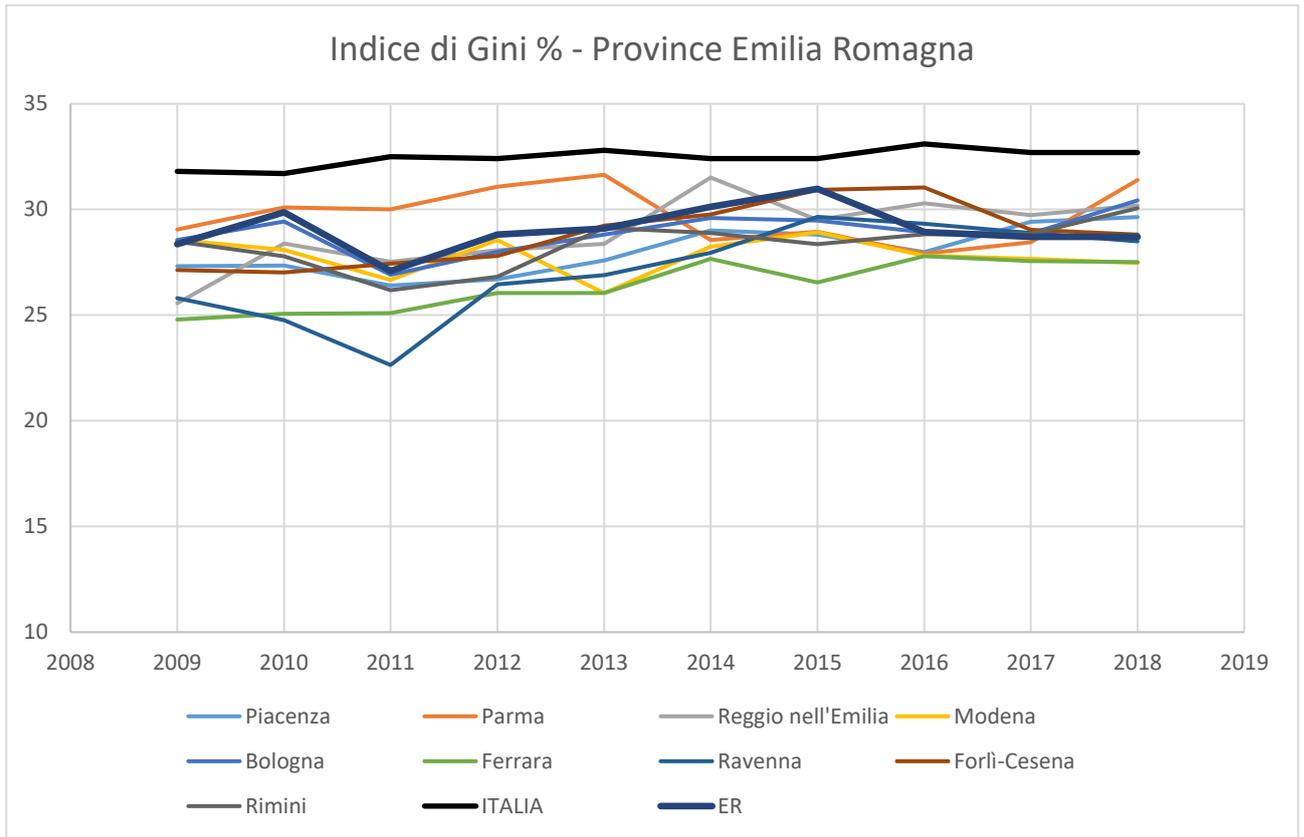
**Rischio di povertà per provincia. Anni 2009-2018 (valori percentuali)**



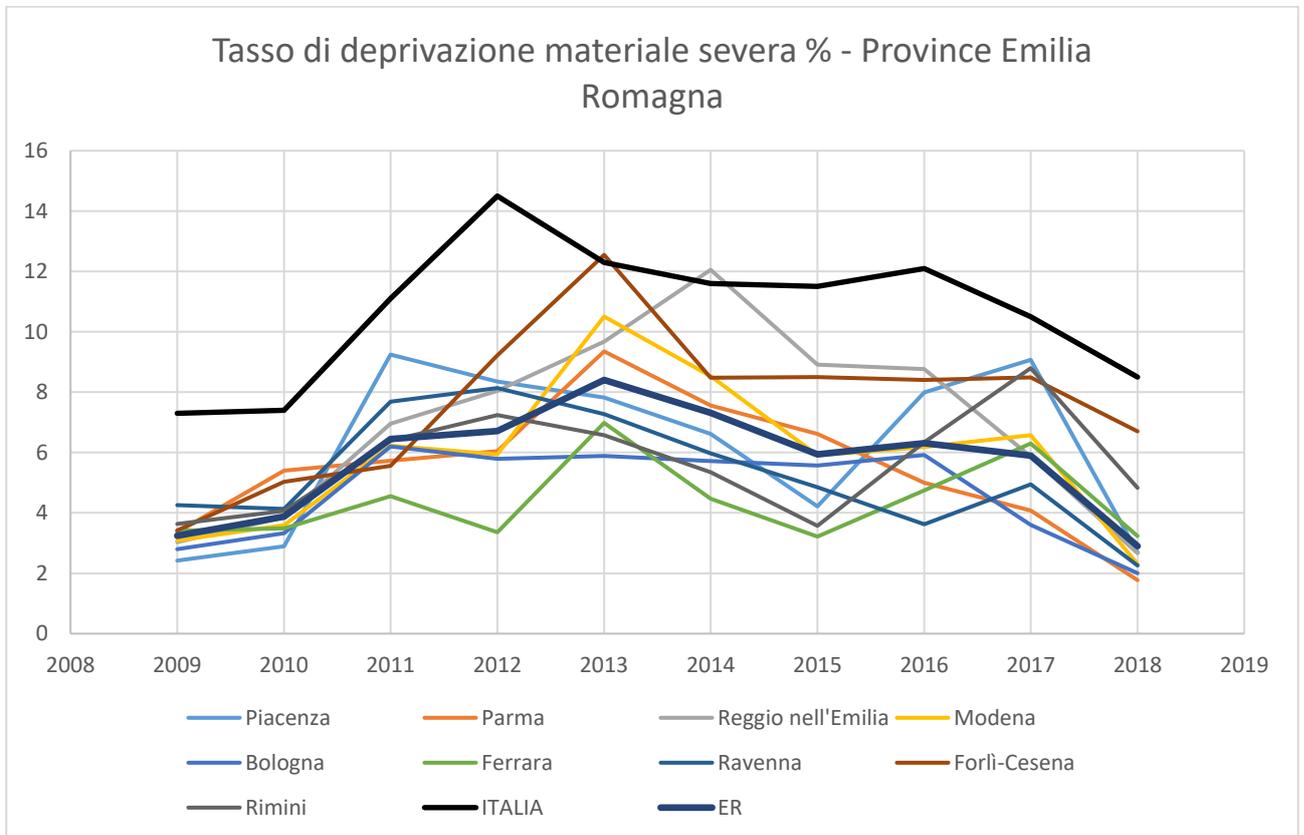
**Gap mediano per provincia. Anni 2009-2018 (valori percentuali)**



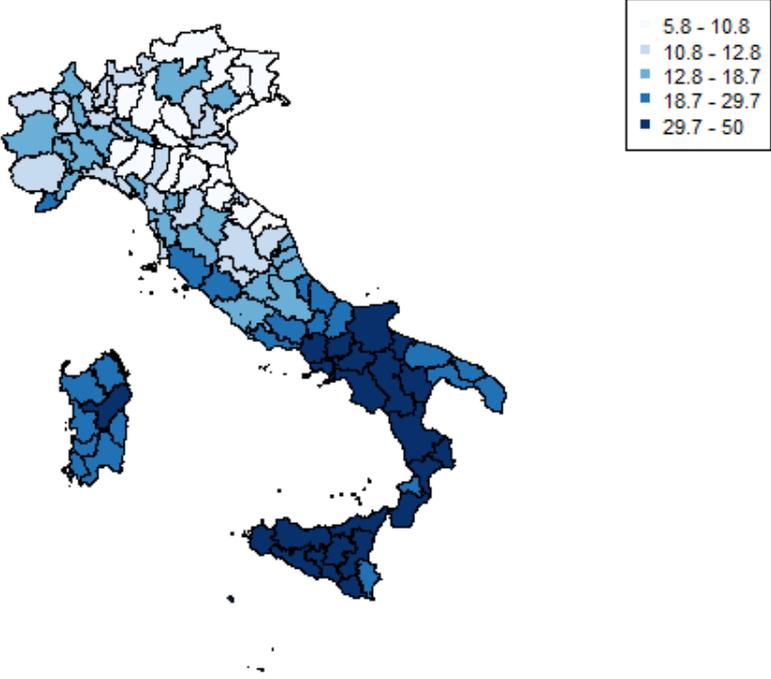
**Indice di Gini per provincia. Anni 2009-2018 (valori percentuali)**



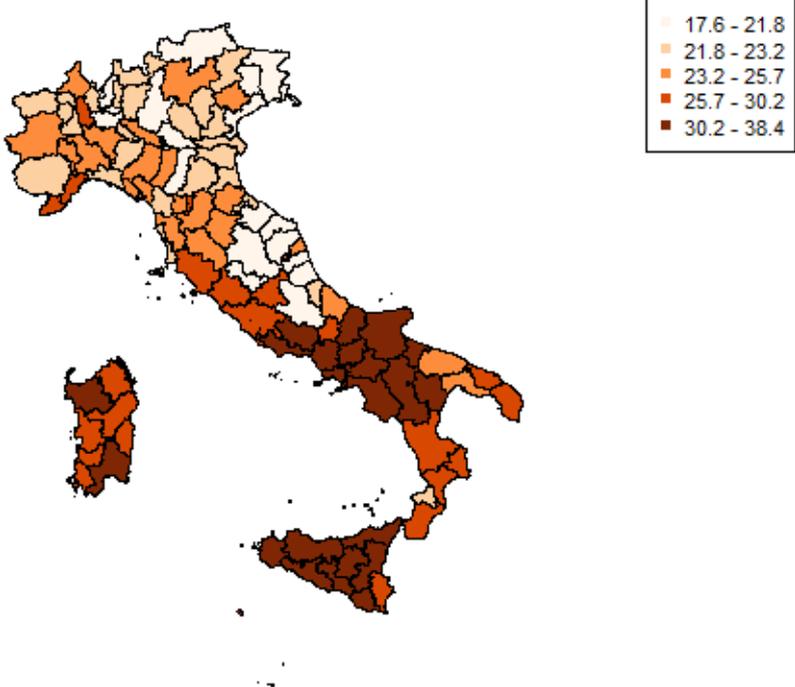
**Tasso di deprivazione severa per provincia. Anni 2009-2018 (valori percentuali)**



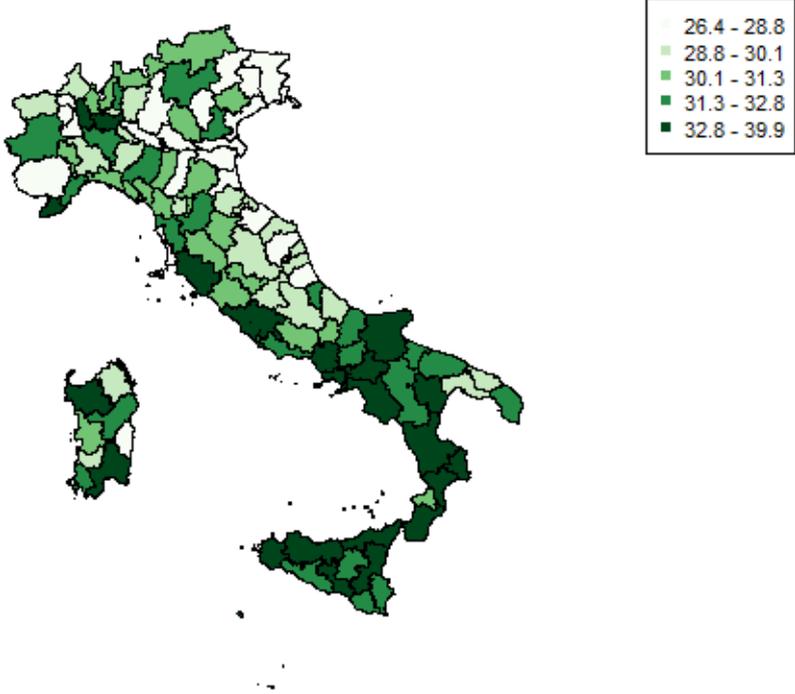
Rischio di povertà in Italia per provincia. Anno 2018 (valori percentuali)



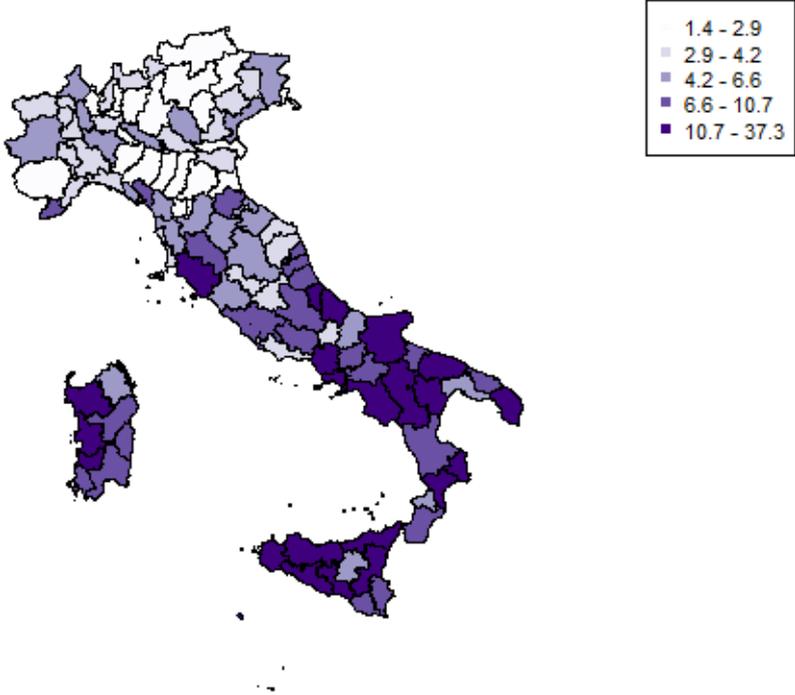
**Gap mediano in Italia per provincia. Anno 2018 (valori percentuali)**



**Indice di Gini in Italia per provincia. Anno 2018 (valori percentuali)**



Tasso di deprivazione severa in Italia per provincia. Anno 2018 (valori percentuali)



**Tab. 1 - Stime dei parametri per le province dell'Emilia-Romagna. Anni 2009-2018 (valori percentuali)**

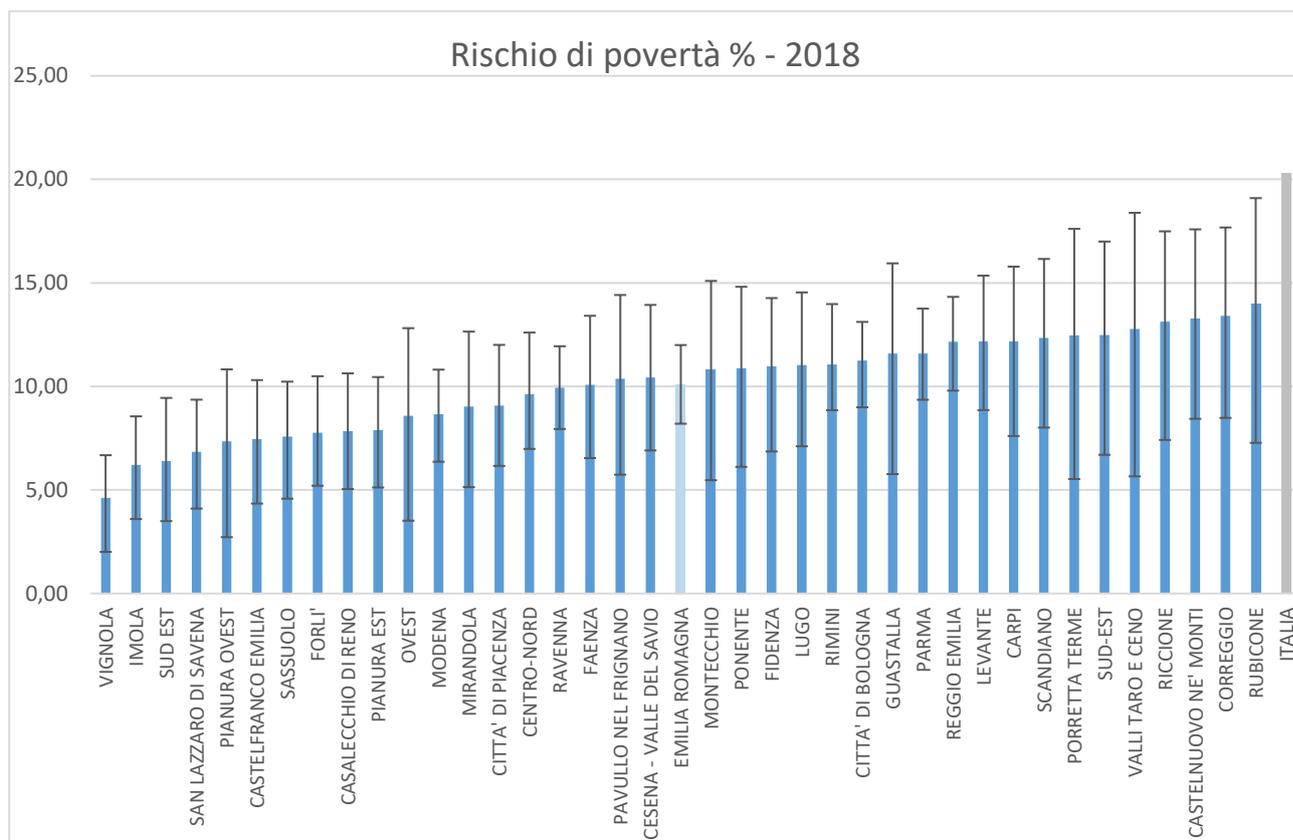
Rischio di povertà (%)											
Cod.	Prov.	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
33	PC	10,31	7,96	8,85	7,19	9,84	10,61	9,64	8,63	11,12	10,72
34	PR	10,00	10,40	13,10	10,44	13,81	7,18	7,04	6,55	9,33	10,68
35	RE	7,42	8,83	8,68	10,51	11,75	12,33	11,32	11,05	11,99	12,16
36	MO	9,09	7,81	7,86	9,29	10,34	9,49	8,53	7,50	8,85	8,50
37	BO	7,10	7,24	5,76	6,41	7,53	7,65	7,34	6,52	8,78	9,08
38	FE	9,06	5,75	7,16	8,09	8,91	9,36	8,00	7,58	11,05	10,20
39	RA	8,87	6,00	5,87	7,38	9,97	10,45	11,99	9,57	11,07	10,26
40	FC	10,29	9,28	10,21	12,12	17,50	16,54	14,05	14,85	13,15	10,00
99	RN	14,52	13,66	10,01	9,95	13,28	11,28	14,14	12,49	13,38	11,77
	Emilia-R.	9,09	8,30	8,17	8,81	10,93	10,10	9,70	8,91	10,50	10,10
	Italia	18,4	18,7	19,8	19,5	19,3	19,4	19,9	20,6	20,30	20,30
Gap mediano (%)											
Cod.	Prov.	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
33	PC	21,1	17,1	19,0	17,8	21,8	22,3	21,7	20,2	25,1	23,1
34	PR	22,2	19,0	22,5	21,3	25,3	21,3	20,4	19,1	23,6	23,7
35	RE	20,1	18,3	20,2	20,6	23,4	25,3	23,2	22,8	26,1	24,4
36	MO	22,0	17,4	19,0	19,7	21,5	21,9	21,4	19,6	23,4	21,6
37	BO	21,2	17,6	18,3	18,2	22,1	22,0	21,0	19,7	23,8	22,7
38	FE	20,1	15,9	18,3	18,6	21,4	21,9	20,4	20,0	24,7	22,4
39	RA	20,4	15,8	16,5	18,1	21,8	22,1	23,3	21,5	25,2	23,0
40	FC	21,4	17,7	20,2	20,8	26,2	25,3	25,2	25,1	26,4	23,3
99	RN	23,1	19,5	19,2	19,1	23,6	22,2	23,2	21,9	25,1	23,1
	Emilia-R.	--	--	--	--	--	--	--	--	--	--
	Italia	23,1	24,8	26,6	26	28,2	28,2	29,3	31,6	28,1	29,5
Indice di Gini (%)											
Cod.	Prov.	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
33	PC	27,3	27,4	26,4	26,7	27,6	29,0	28,8	28,0	27,2	29,6
34	PR	29,1	30,1	30,0	31,1	31,6	28,6	28,9	27,9	29,4	31,4
35	RE	25,6	28,4	27,5	28,1	28,4	31,5	29,5	30,3	28,5	30,2
36	MO	28,6	28,1	26,7	28,6	26,0	28,2	28,9	27,8	29,7	27,5
37	BO	28,6	29,4	26,9	28,0	28,8	29,6	29,5	28,9	27,7	30,4
38	FE	24,8	25,1	25,1	26,0	26,0	27,7	26,5	27,8	28,9	27,5
39	RA	25,8	24,8	22,6	26,5	26,9	28,0	29,7	29,3	27,6	28,5
40	FC	27,1	27,0	27,4	27,8	29,2	29,8	30,9	31,0	28,9	28,8
99	RN	28,5	27,8	26,2	26,8	29,1	28,9	28,4	28,8	29,0	30,1
	Emilia-R.	28,4	29,9	27,1	28,8	29,1	30,1	31,0	28,9	28,7	29,0
	Italia	31,8	31,7	32,5	32,4	32,8	32,4	32,4	33,1	32,7	33,4
Tasso di deprivazione materiale severa (%)											
Cod.	Prov.	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
33	PC	2,42	2,89	9,25	8,35	7,82	6,62	4,22	7,99	9,07	2,69
34	PR	3,36	5,40	5,73	6,05	9,35	7,56	6,61	5,00	4,08	1,77
35	RE	3,03	3,86	6,95	8,04	9,68	12,06	8,91	8,76	5,86	2,66
36	MO	3,09	3,60	6,22	5,94	10,51	8,53	5,90	6,18	6,58	2,30
37	BO	2,80	3,33	6,20	5,79	5,89	5,72	5,57	5,92	3,60	2,00
38	FE	3,42	3,49	4,56	3,35	6,98	4,47	3,22	4,74	6,30	3,23
39	RA	4,26	4,13	7,68	8,14	7,28	5,97	4,85	3,62	4,95	2,26
40	FC	3,42	5,03	5,55	9,22	12,55	8,48	8,50	8,40	8,49	6,71
99	RN	3,63	4,07	6,39	7,25	6,58	5,35	3,57	6,34	8,80	4,83
	Emilia-R.	3,25	3,88	6,44	6,71	8,41	7,31	5,94	6,31	5,90	2,90
	Italia	7,30	7,40	11,10	14,50	12,30	11,60	11,50	12,10	10,10	8,50

**Tab. 2 – Coefficienti di Variazione delle stime dei parametri, per le province dell'Emilia-Romagna. Anni 2009-2018 (valori percentuali)**

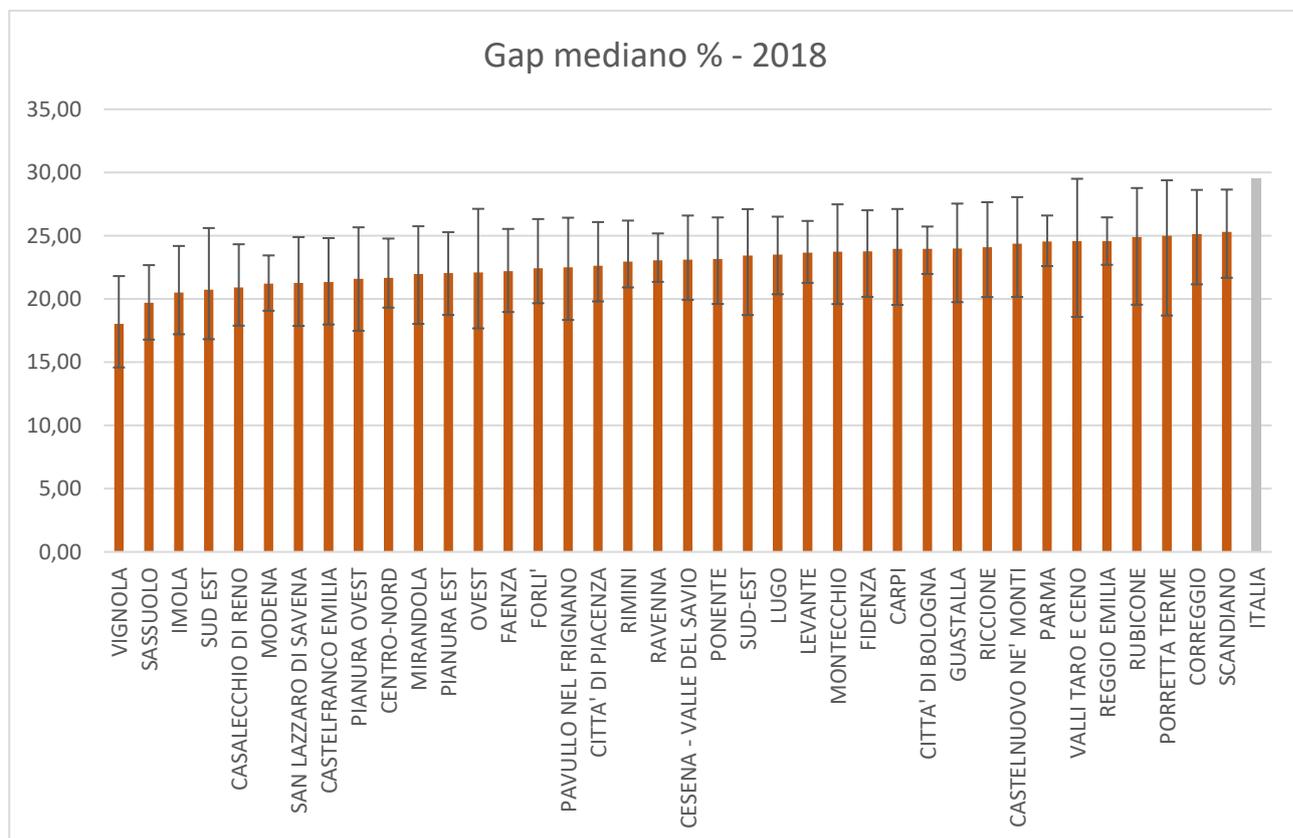
<b>Rischio di povertà (%)</b>											
<b>Cod.</b>	<b>Prov.</b>	<b>2009</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>
33	PC	14,2	16,2	15,7	16,8	13,1	13,2	15,1	15,1	15,2	14,0
34	PR	13,2	12,2	11,9	14,4	13,5	12,9	13,2	13,9	11,0	12,0
35	RE	12,7	12,8	12,9	13,6	11,2	11,7	12,8	12,8	10,5	11,0
36	MO	10,1	11,4	11,2	11,2	9,4	9,9	11,4	11,1	9,2	9,4
37	BO	8,8	10,2	11,5	10,9	9,3	9,3	9,7	10,7	8,4	9,2
38	FE	11,4	16,9	14,7	14,8	13,1	13,2	14,1	15,8	13,5	15,2
39	RA	12,6	13,6	14,7	14,9	12,6	13,1	13,6	15,3	13,0	13,7
40	FC	12,1	14,7	13,3	13,9	11,7	11,5	12,5	12,6	12,0	13,9
99	RN	11,1	12,6	14,4	17,0	13,5	14,0	14,0	12,9	14,6	13,3
<b>Gap mediano (%)</b>											
<b>Cod.</b>	<b>Prov.</b>	<b>2009</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>
33	PC	6,27	6,33	6,59	6,64	5,64	6,22	6,69	6,69	6,59	6,02
34	PR	6,20	6,04	7,15	7,24	6,91	5,80	5,96	6,16	4,92	5,60
35	RE	5,30	5,75	6,11	6,28	5,33	6,52	6,48	6,78	5,13	5,56
36	MO	5,01	5,00	5,02	5,29	4,50	4,81	5,18	5,29	4,11	4,38
37	BO	4,38	4,87	5,25	5,08	4,75	4,82	4,75	4,95	3,71	4,44
38	FE	5,17	6,54	6,15	6,29	5,61	5,80	6,29	6,66	6,01	6,52
39	RA	5,34	5,71	5,96	6,37	5,41	5,74	6,96	6,98	5,63	6,09
40	FC	5,69	6,02	6,27	6,95	7,32	7,11	7,20	7,69	5,87	5,91
99	RN	6,43	7,38	6,71	7,40	6,73	6,79	7,83	7,01	7,07	6,37
<b>Indice di Gini (%)</b>											
<b>Cod.</b>	<b>Prov.</b>	<b>2009</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>
33	PC	5,11	5,35	5,39	5,21	5,14	4,91	5,03	4,95	5,45	5,15
34	PR	4,89	4,18	4,32	4,60	4,95	4,60	4,09	4,48	4,34	4,35
35	RE	4,57	4,42	4,44	4,66	4,70	4,52	4,87	4,57	4,22	4,46
36	MO	3,87	4,06	4,15	4,09	4,25	4,12	3,73	3,71	3,64	3,97
37	BO	3,19	3,45	3,76	3,51	3,64	3,44	3,38	3,51	3,59	3,88
38	FE	4,55	5,34	5,21	4,92	5,16	4,99	4,96	5,04	5,25	6,02
39	RA	4,48	4,76	5,05	5,22	4,96	4,72	4,81	5,08	4,66	5,37
40	FC	4,45	5,09	4,52	5,08	4,92	4,59	4,41	4,73	4,69	5,01
99	RN	4,38	4,80	5,24	5,66	5,24	5,35	5,59	4,82	5,47	5,30
<b>Tasso di deprivazione materiale severa (%)</b>											
<b>Cod.</b>	<b>Prov.</b>	<b>2009</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>
33	PC	39,5	35,0	35,5	36,6	31,4	31,7	38,4	31,5	34,2	47,5
34	PR	37,2	29,3	29,7	33,8	29,0	28,0	31,2	28,9	30,9	47,6
35	RE	34,7	30,2	27,2	27,8	24,4	22,6	27,8	25,1	27,5	39,1
36	MO	31,5	27,5	25,0	27,5	20,3	21,8	25,4	22,8	21,7	36,0
37	BO	26,4	24,0	21,2	23,9	21,4	20,5	22,3	21,0	26,3	37,3
38	FE	36,6	35,6	33,6	40,9	30,1	33,3	39,9	35,9	35,5	47,5
39	RA	34,8	33,6	32,0	31,4	29,1	30,6	35,2	34,0	32,9	46,3
40	FC	34,2	32,7	30,3	31,7	26,2	26,9	31,7	29,5	30,4	39,7
99	RN	35,3	34,3	35,0	37,6	31,1	33,4	39,2	32,7	33,4	41,6

### **3. Stime per piccole aree per distretto sociosanitario**

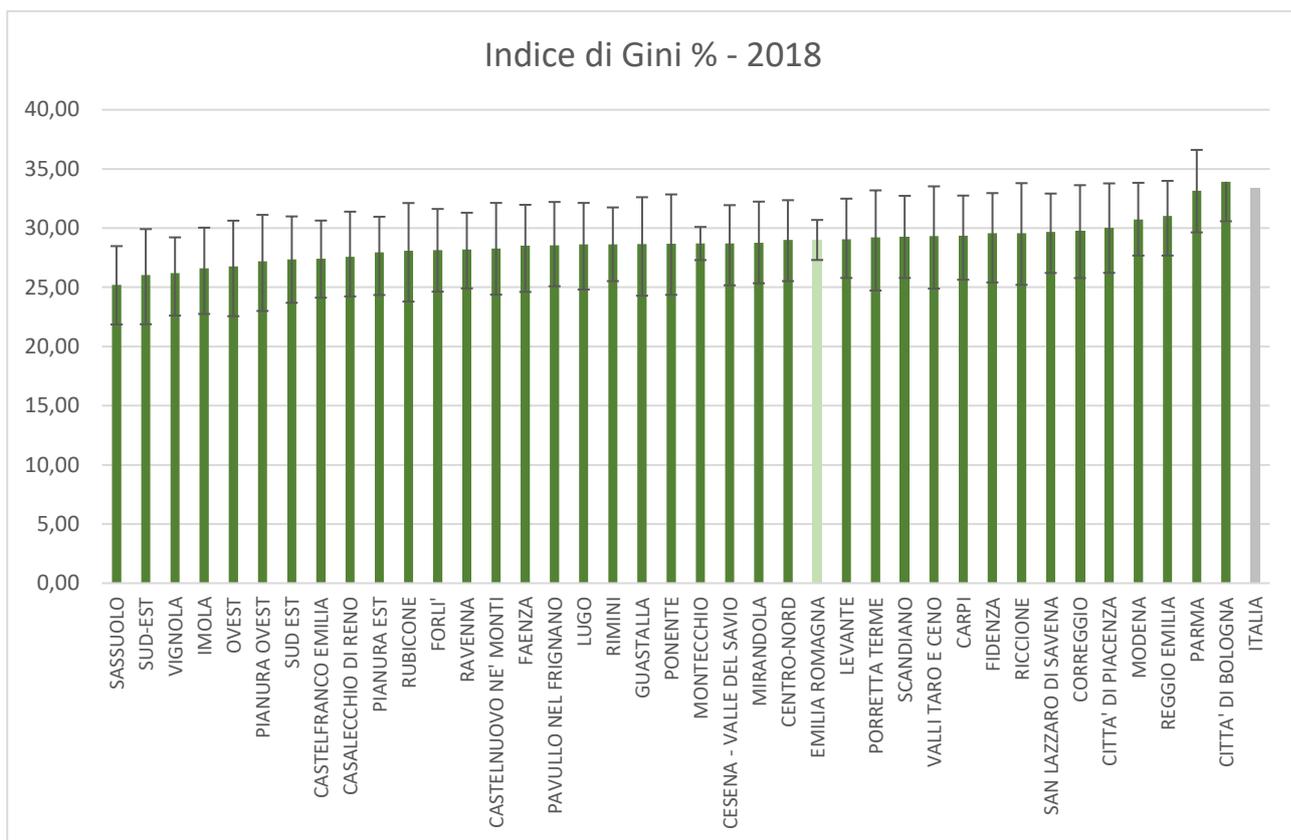
**Rischio di povertà e relativi coefficienti di variazione per distretto sociosanitario. Anno 2018 (valori percentuali)**



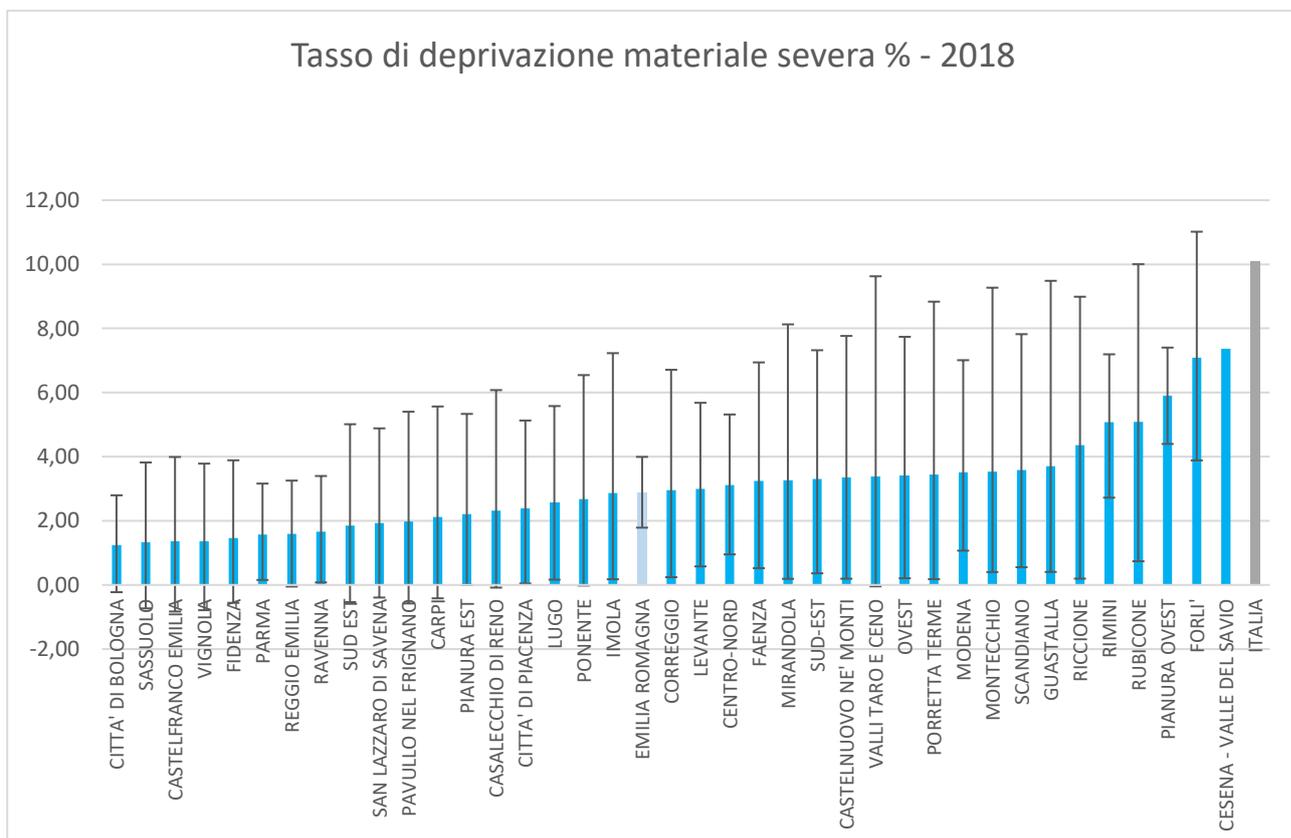
**Gap mediano e relativi coefficienti di variazione per distretto sociosanitario. Anno 2018 (valori percentuali)**



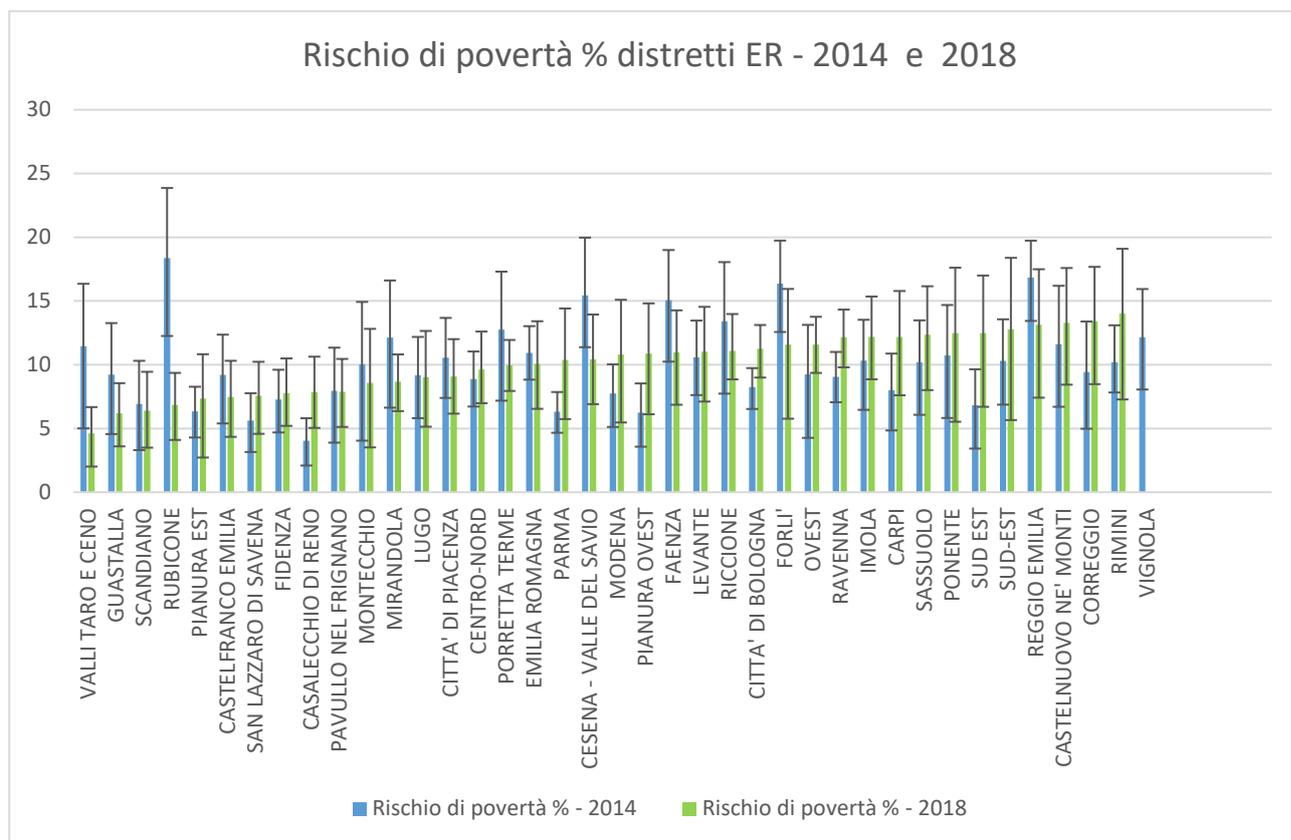
**Indice di Gini e relativi coefficienti di variazione per distretto sociosanitario. Anno 2018 (valori percentuali)**



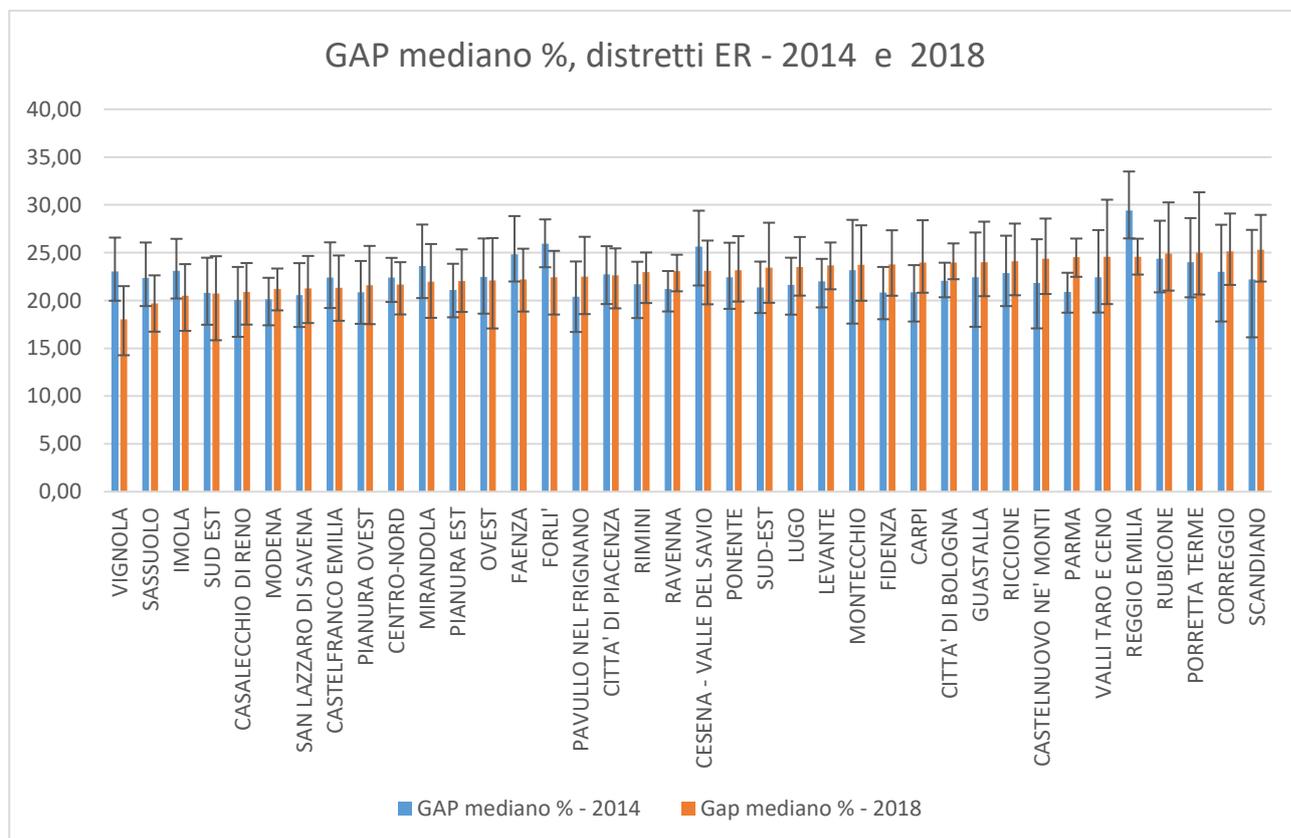
**Tasso di deprivazione materiale severa e relativi coefficienti di variazione per distretto sociosanitario. Anno 2018 (valori percentuali)**



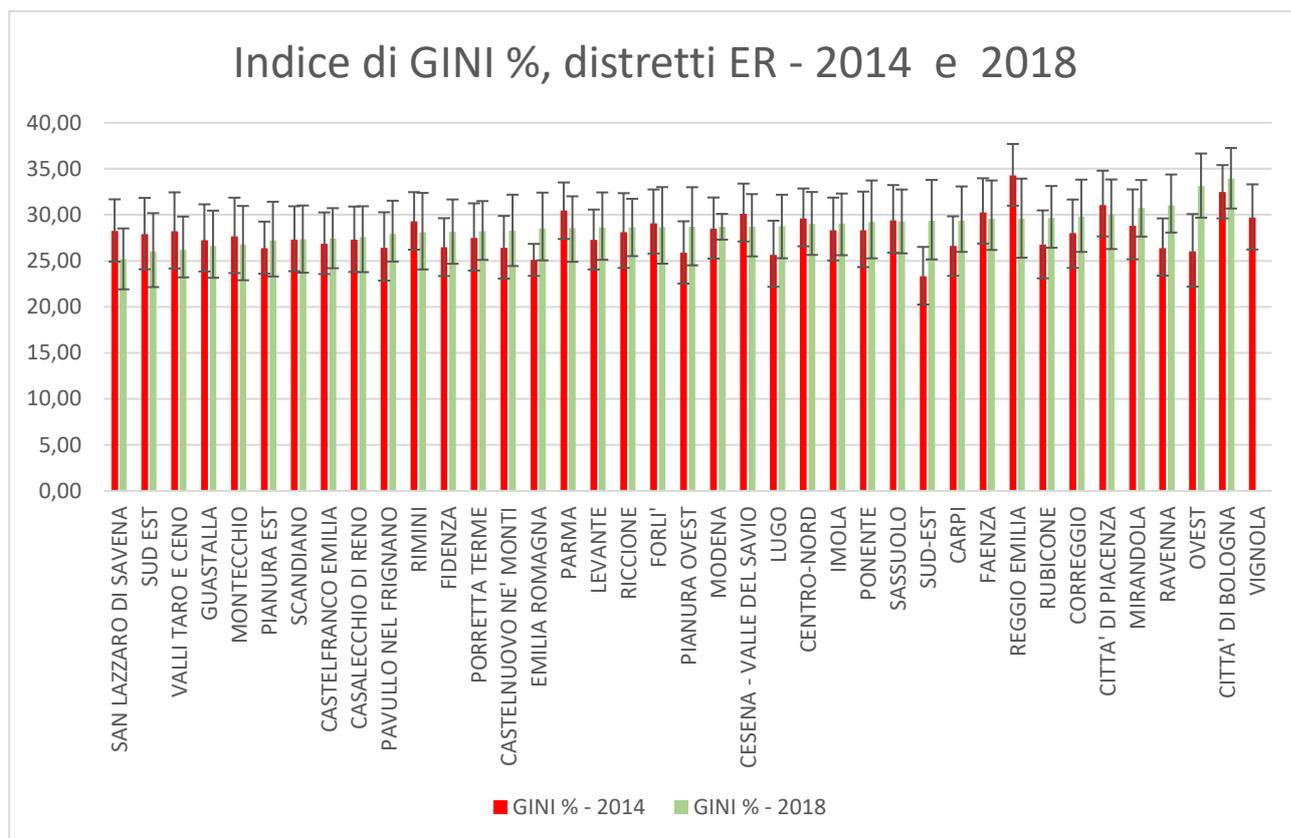
**Rischio di povertà e relativi coefficienti di variazione per distretto sociosanitario. Anni 2014 e 2018 (valori percentuali)**



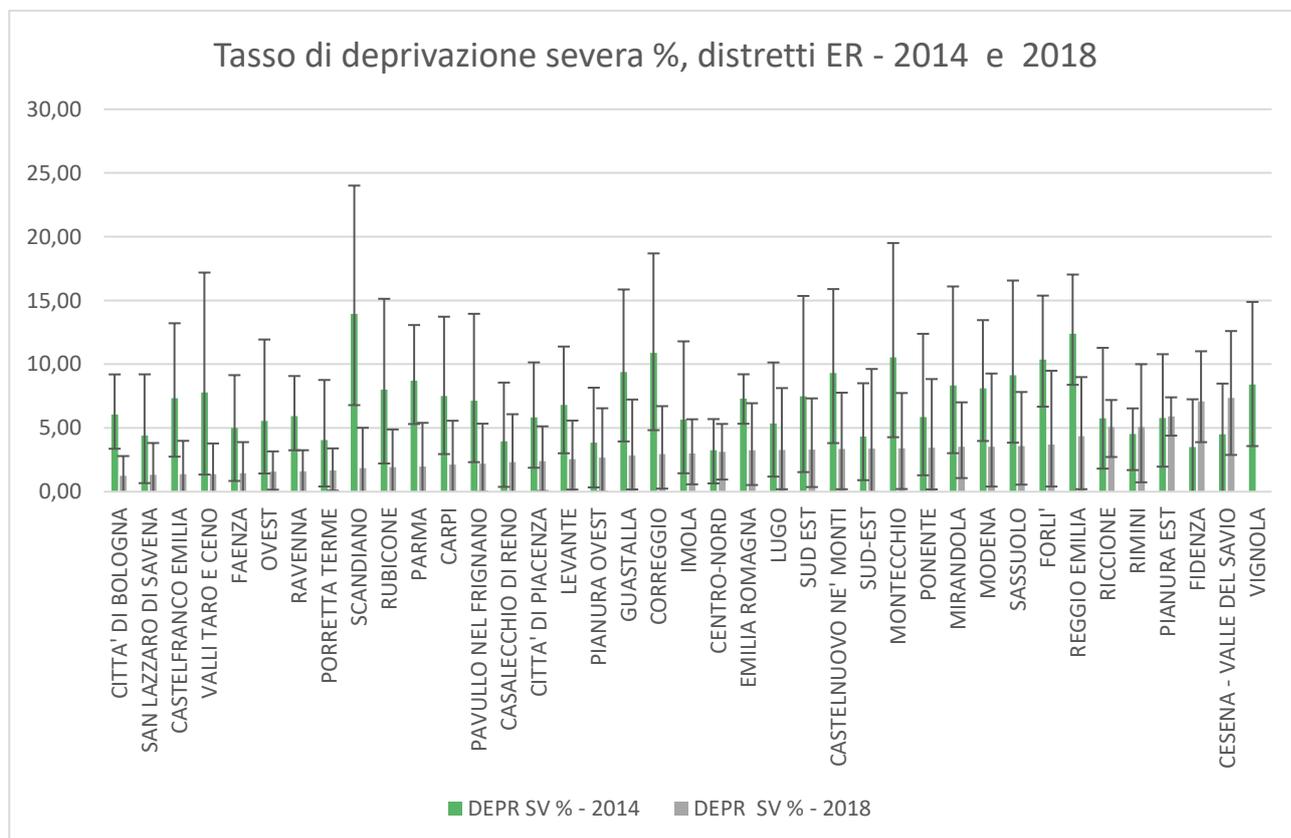
**Gap mediano e relativi coefficienti di variazione per distretto sociosanitario. Anni 2014 e 2018 (valori percentuali)**



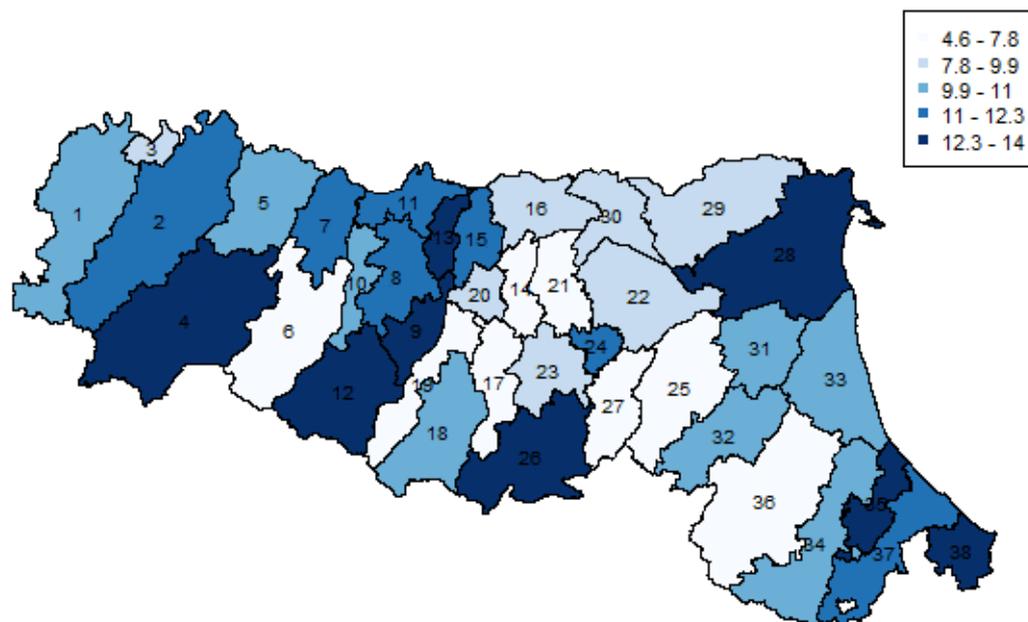
**Indice di Gini e relativi coefficienti di variazione per distretto sociosanitario. Anni 2014 e 2018 (valori percentuali)**



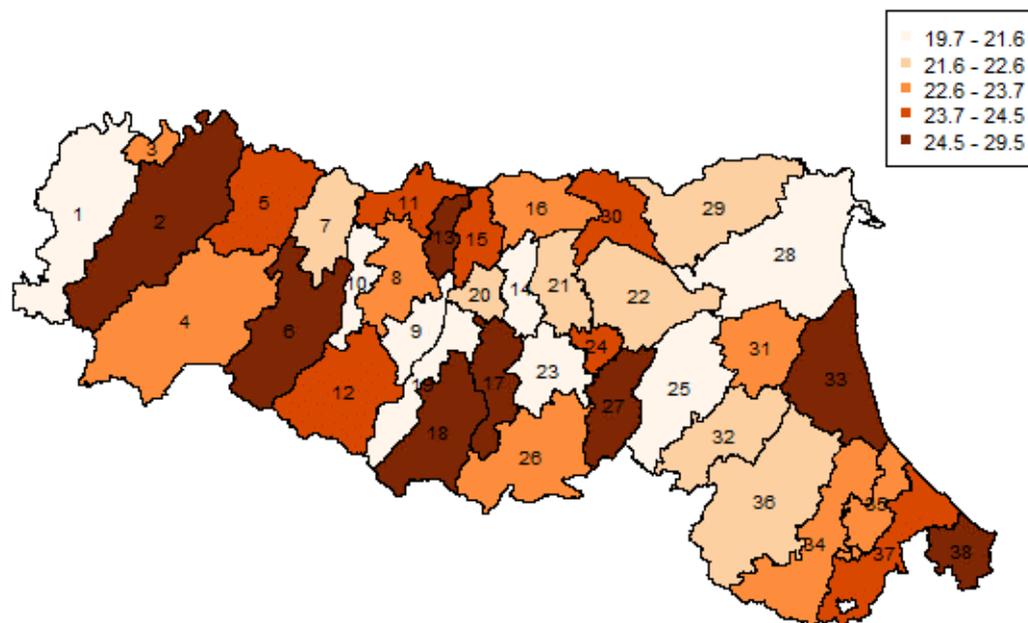
**Tasso di deprivazione materiale severa e relativi coefficienti di variazione per distretto sociosanitario. Anni 2014 e 2018 (valori percentuali)**



**Rischio di povertà in ER per distretto sociosanitario. Anno 2018 (valori percentuali)**



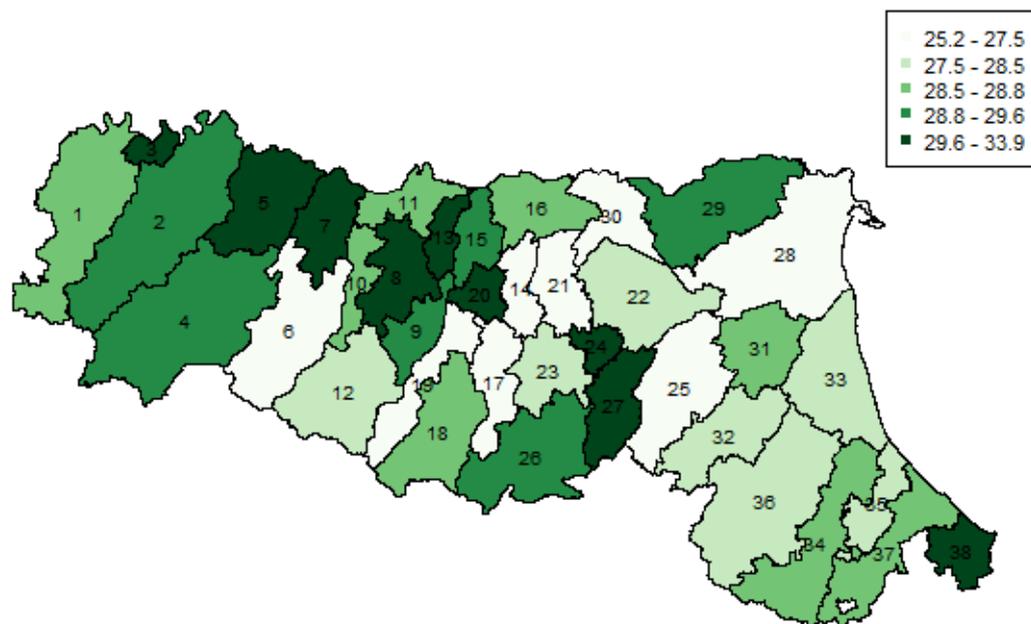
**Gap mediano in ER per distretto sociosanitario. Anno 2018 (valori percentuali)**



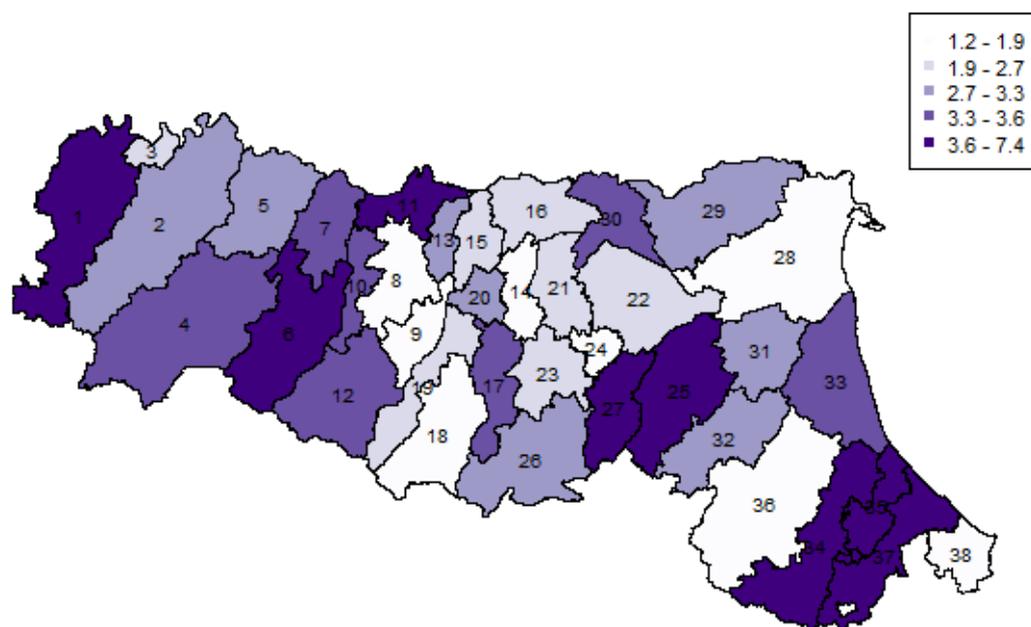
**Legenda**

1 Ponente, 2 Levante, 3 Citta Di Piacenza, 4, Valli Taro e Ceno, 5 Fidenza, 6 Sud Est, 7 Parma, 8 Reggio Emilia, 9 Scandiano, 10 Montecchio, 11 Guastalla, 12 Castelnuovo ne Monti, 13 Correggio, 14 Castelfranco Emilia, 15 Carpi, 16 Mirandola, 17 Vignola, 18 Pavullo Nel Frignano, 19 Sassuolo, 20 Modena, 21 Pianura Ovest, 22 Pianura Est, 23 Casalecchio di Reno, 24 Citta di Bologna, 25 Imola, 26 Porretta Terme, 27 San Lazzaro di Savena, 28 Sud-Est, 29 Centro-Nord, 30 Ovest, 31 Lugo, 32 Faenza, 33 Ravenna, 34 Cesena - Valle del Savio, 35 Rubicone, 36 Forlì, 37 Rimini, 38 Riccione

**Indice di Gini in ER per distretto sociosanitario. Anno 2018 (valori percentuali)**



**Tasso di deprivazione severa in ER per distretto sociosanitario. Anno 2018 (valori percentuali)**



**Legenda**

1 Ponente, 2 Levante, 3 Citta Di Piacenza, 4, Valli Taro e Ceno, 5 Fidenza, 6 Sud Est, 7 Parma, 8 Reggio Emilia, 9 Scandiano, 10 Montecchio, 11 Guastalla, 12 Castelnuovo ne Monti, 13 Correggio, 14 Castelfranco Emilia, 15 Carpi, 16 Mirandola, 17 Vignola, 18 Pavullo Nel Frignano, 19 Sassuolo, 20 Modena, 21 Pianura Ovest, 22 Pianura Est, 23 Casalecchio di Reno, 24 Citta di Bologna, 25 Imola, 26 Porretta Terme, 27 San Lazzaro di Savena, 28 Sud-Est, 29 Centro-Nord, 30 Ovest, 31 Lugo, 32 Faenza, 33 Ravenna, 34 Cesena - Valle del Savio, 35 Rubicone, 36 Forlì, 37 Rimini, 38 Riccione

**Tab. 3a. Stima del Rischio di povertà per i distretti sociosanitari dell'Emilia-Romagna e numerosità campionaria dell'ultimo anno. Anni 2009-2018 (valori percentuali e valore assoluto)**

Provincia	Denominazione	Rischio di povertà (%)										n
		2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	
Bologna	CASALECCHIO DI RENO	6,22	6,15	5,27	4,26	4,57	4,04	3,82	4,69	7,33	7,8	42
Bologna	CITTA' DI BOLOGNA	7,37	8,51	6,70	8,88	8,56	8,25	7,37	6,81	10,05	11,2	285
Bologna	IMOLA	7,65	7,17	6,42	4,65	11,34	10,33	9,61	6,92	6,91	6,2	49
Bologna	PIANURA EST	4,31	3,75	2,25	3,83	4,06	6,33	5,70	6,36	8,02	7,9	190
Bologna	PIANURA OVEST	4,89	4,45	5,88	4,81	7,49	6,25	7,71	6,85	7,89	7,3	10
Bologna	PORRETTA TERME	12,93	10,52	10,70	11,49	9,77	12,76	11,12	11,74	13,76	12,5	26
Bologna	SAN LAZZARO DI SAVENA	9,61	10,18	3,85	3,52	5,56	5,62	8,55	3,18	6,61	6,8	126
Ferrara	CENTRO-NORD	8,75	4,55	6,01	6,06	7,18	8,86	6,24	6,43	11,05	9,6	144
Ferrara	OVEST	8,07	6,46	7,57	7,93	9,22	9,26	7,29	7,68	9,89	8,6	10
Ferrara	SUD-EST	10,34	7,28	8,85	11,72	11,68	10,31	11,62	9,50	11,97	12,5	10
Forlì	CESENA - VALLE DEL SAVIO	10,56	10,16	9,10	10,51	16,88	15,43	13,73	15,20	13,29	10,4	122
Forlì	FORLÌ'	9,78	8,33	8,59	11,00	16,78	16,36	14,40	15,04	12,65	7,8	75
Forlì	RUBICONE	11,01	10,12	15,09	16,55	19,82	18,35	13,76	14,00	13,99	14,0	10
Modena	CARPI	7,05	6,25	3,22	7,13	7,48	8,00	9,98	11,50	10,60	12,2	55
Modena	CASTELFRANCO EMILIA	8,26	8,67	9,79	8,67	7,99	9,21	12,38	8,52	8,99	7,5	119
Modena	MIRANDOLA	8,89	10,70	10,05	10,65	11,16	12,14	8,49	8,22	9,45	9,0	60
Modena	MODENA	9,74	7,85	7,25	8,22	9,08	7,74	6,63	6,54	8,05	8,7	219
Modena	PAVULLO NEL FRIGNANO	13,81	7,26	5,98	6,57	11,02	7,94	7,29	7,02	9,23	10,4	65
Modena	SASSUOLO	9,27	7,53	9,68	11,24	14,11	10,21	8,83	5,26	10,30	7,6	121
Modena	VIGNOLA	8,44	6,62	9,10	11,82	12,02	12,15	7,72	6,43	5,71	4,6	142
Parma	FIDENZA	9,28	9,71	12,55	14,96	19,49	7,27	6,59	5,99	8,30	11,0	112
Parma	PARMA	10,72	10,30	13,72	8,22	12,06	6,31	7,45	7,12	10,15	11,6	227
Parma	SUD EST	8,69	10,03	8,51	9,25	9,93	6,82	4,60	2,95	7,04	6,4	57
Parma	VALLI TARO E CENO	10,49	12,90	18,82	12,20	15,32	11,46	10,21	11,12	11,50	12,8	10
Piacenza	CITTA' DI PIACENZA	8,30	5,55	8,31	5,33	9,64	10,55	9,85	10,11	13,20	9,1	57
Piacenza	LEVANTE	12,21	9,66	8,81	8,11	9,57	10,57	9,41	6,97	9,36	12,2	130
Piacenza	PONENTE	10,23	8,70	9,62	8,34	10,48	10,73	9,69	9,00	10,81	10,9	10
Ravenna	FAENZA	10,91	6,74	8,73	9,87	11,51	15,05	16,51	10,86	11,68	10,1	83
Ravenna	LUGO	5,81	5,64	5,20	8,65	10,21	9,18	8,93	7,48	12,08	11,0	61
Ravenna	RAVENNA	9,58	5,87	4,93	5,58	9,14	9,04	11,56	10,08	10,28	9,9	108
Reggio	CASTELNUOVO NÈ MONTI	13,90	12,19	12,80	13,81	12,24	11,60	8,97	11,02	13,18	13,3	53
Reggio	CORREGGIO	7,85	12,73	12,03	15,17	13,84	9,44	7,75	8,32	12,65	13,4	53
Reggio	GUASTALLA	9,13	9,53	8,44	8,20	11,36	9,22	10,12	10,91	11,98	11,6	10
Reggio	MONTECCHIO	8,59	8,80	9,00	9,83	10,82	10,03	9,25	9,95	10,85	10,8	10
Reggio	REGGIO EMILIA	5,44	7,01	8,28	10,72	13,56	16,84	14,24	12,20	12,89	12,2	173
Reggio	SCANDIANO	7,13	9,01	5,64	7,90	6,28	6,91	9,25	10,74	9,44	12,3	137
Rimini	RICCIONE	16,98	13,68	12,15	12,17	14,50	13,41	14,15	13,79	15,07	13,1	10
Rimini	RIMINI	13,26	13,65	8,90	8,80	12,65	10,20	14,13	11,82	12,51	11,1	198

**Tab. 3b. Coefficiente di variazione della stima del Rischio di povertà per i distretti sociosanitari dell'Emilia-Romagna. Anni 2009-2018 (valori percentuali)**

CV del Rischio di povertà (%)											
Provincia	Denominazione	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Bologna	CASALECCHIO DI RENO	16,2	18,7	25,6	31,0	32,0	23,5	23,6	26,5	30,8	22,3
Bologna	CITTA' DI BOLOGNA	10,3	10,0	11,0	14,2	11,2	10,0	9,7	10,9	13,3	9,8
Bologna	IMOLA	20,6	22,0	22,7	24,3	24,7	18,1	17,1	18,8	23,6	20,2
Bologna	PIANURA EST	18,6	18,6	23,8	39,7	27,1	21,8	16,0	17,6	20,0	15,4
Bologna	PIANURA OVEST	25,6	22,1	26,7	26,9	27,1	19,8	20,0	29,4	32,8	27,8
Bologna	PORRETTA TERME	20,9	19,1	23,8	24,7	22,1	26,7	20,2	21,2	24,2	19,4
Bologna	SAN LAZZARO DI SAVENA	17,7	16,9	19,9	31,7	33,1	20,1	20,8	20,6	32,4	21,5
Ferrara	CENTRO-NORD	12,3	10,5	19,2	15,8	17,6	14,8	12,3	20,8	19,1	11,5
Ferrara	OVEST	27,5	24,3	29,9	27,5	28,0	23,8	24,3	31,4	30,1	23,0
Ferrara	SUD-EST	13,9	12,7	18,5	17,3	16,4	15,6	16,3	22,5	24,8	19,7
Forlì	CESENA - VALLE DEL SAVIO	16,5	15,0	17,2	17,9	19,0	13,5	13,9	14,2	13,8	13,3
Forlì	FORLÌ'	13,3	11,6	13,7	14,6	13,5	9,7	10,9	11,9	11,5	11,6
Forlì	RUBICONE	24,1	16,4	20,4	16,5	18,7	15,9	16,3	21,8	21,2	20,4
Modena	CARPI	19,4	20,0	25,1	32,3	22,4	18,3	18,8	22,5	20,7	18,3
Modena	CASTELFRANCO EMILIA	21,4	17,4	19,5	21,4	21,9	19,8	19,3	18,0	19,9	19,9
Modena	MIRANDOLA	24,3	24,3	21,3	23,3	23,4	22,0	21,1	22,2	23,7	21,0
Modena	MODENA	14,1	13,4	17,8	18,1	17,1	14,6	16,0	16,4	17,5	14,4
Modena	PAVULLO NEL FRIGNANO	17,5	16,6	24,4	30,5	25,8	20,8	23,7	26,6	25,7	23,3
Modena	SASSUOLO	18,6	17,5	22,5	19,3	16,8	16,2	18,8	19,1	24,5	17,6
Modena	VIGNOLA	22,0	20,9	23,2	20,6	18,9	17,8	16,5	19,0	23,3	21,5
Parma	FIDENZA	16,7	22,2	16,7	16,0	15,8	13,1	17,1	18,1	20,1	16,7
Parma	PARMA	12,2	10,8	10,4	9,8	13,6	10,1	12,8	11,7	11,7	9,4
Parma	SUD EST	23,9	19,9	19,5	20,7	18,5	17,0	22,9	29,7	37,1	21,0
Parma	VALLI TARO E CENO	19,7	19,1	20,4	17,5	24,4	19,6	25,4	27,6	26,3	25,3
Piacenza	CITTA' DI PIACENZA	20,0	18,2	25,7	18,8	26,8	15,4	15,0	15,8	16,4	13,6
Piacenza	LEVANTE	19,9	12,7	16,5	18,2	18,9	14,4	14,0	15,3	20,4	15,7
Piacenza	PONENTE	25,0	18,8	25,9	24,5	27,7	21,4	21,2	23,0	25,4	20,5
Ravenna	FAENZA	19,9	17,8	18,0	17,6	19,1	16,1	14,9	15,0	18,3	16,1
Ravenna	LUGO	20,0	22,2	20,0	22,7	18,6	17,7	17,4	17,8	22,9	17,7
Ravenna	RAVENNA	10,1	11,2	12,3	15,4	16,5	11,4	11,0	10,3	11,2	11,0
Reggio	CASTELNUOVO NÈ MONTI	29,2	21,9	24,0	23,4	23,7	18,8	20,5	25,4	24,2	20,3
Reggio	CORREGGIO	31,4	29,2	21,0	21,3	20,1	19,8	22,9	25,7	23,2	17,5
Reggio	GUASTALLA	23,0	23,7	26,9	21,0	22,0	18,9	23,9	28,3	26,5	23,2
Reggio	MONTECCHIO	22,2	28,5	27,8	28,6	26,9	24,4	27,6	30,1	28,1	24,5
Reggio	REGGIO EMILIA	15,2	15,7	14,8	12,9	10,6	10,0	9,6	11,2	11,5	10,8
Reggio	SCANDIANO	21,5	19,9	18,8	24,3	21,3	21,8	25,7	22,5	20,2	19,5
Rimini	RICCIONE	12,5	12,0	15,3	21,2	21,6	18,6	19,5	20,3	20,3	18,9
Rimini	RIMINI	7,1	7,9	7,9	15,0	15,5	11,0	13,1	10,4	12,1	11,6

**Tab. 4a. Stima del *Gap* mediano per i distretti sociosanitari dell'Emilia-Romagna e numerosità campionaria dell'ultimo anno. Anni 2009-2018 (valori percentuali e valore assoluto)**

Gap mediano (%)												
Provincia	Denominazione	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	n
Bologna	CASALECCHIO DI RENO	20,5	17,2	17,3	16,1	20,3	20,0	18,2	17,7	22,3	20,9	42
Bologna	CITTA' DI BOLOGNA	21,8	18,6	18,6	18,7	21,8	22,0	20,6	19,5	24,0	24,0	285
Bologna	IMOLA	21,2	17,2	17,8	16,0	23,4	23,1	22,4	19,7	22,2	20,5	49
Bologna	PIANURA EST	19,2	15,4	13,5	16,1	20,3	21,1	20,3	19,3	23,6	22,0	190
Bologna	PIANURA OVEST	20,0	16,0	18,3	16,3	21,5	20,9	21,3	19,9	23,2	21,6	10
Bologna	PORRETTA TERME	23,0	19,6	21,3	21,3	22,9	24,0	23,3	23,2	26,8	25,0	26
Bologna	SAN LAZZARO DI SAVENA	22,1	18,7	15,5	14,9	21,4	20,6	21,9	15,9	22,1	21,3	126
Ferrara	CENTRO-NORD	19,8	15,2	17,8	17,1	20,5	22,4	18,6	19,1	24,9	21,7	144
Ferrara	OVEST	20,3	16,9	19,1	19,0	22,2	22,5	20,2	20,5	25,0	22,1	10
Ferrara	SUD-EST	20,2	16,7	18,6	20,4	22,1	21,4	22,7	21,0	24,5	23,4	10
Forlì	CESENA - VALLE DEL SAVIO	21,1	18,1	19,1	19,7	25,8	25,6	26,8	25,5	26,3	23,1	122
Forlì	FORLÌ'	21,7	17,6	20,3	20,5	26,3	25,9	26,1	26,5	27,2	22,4	75
Forlì	RUBICONE	22,2	18,0	21,2	22,7	26,4	24,4	24,4	24,1	25,7	24,9	10
Modena	CARPI	21,6	16,6	14,6	18,5	19,2	20,9	22,6	22,1	24,2	24,0	55
Modena	CASTELFRANCO EMILIA	22,7	18,5	20,5	20,4	20,1	22,4	24,8	20,4	24,0	21,3	119
Modena	MIRANDOLA	21,6	19,5	19,9	20,3	21,9	23,6	21,4	20,1	24,1	22,0	60
Modena	MODENA	22,7	17,2	18,7	19,6	20,0	20,1	19,5	18,1	21,9	21,2	219
Modena	PAVULLO NEL FRIGNANO	23,6	16,4	16,9	17,5	21,3	20,4	20,1	18,6	23,1	22,5	65
Modena	SASSUOLO	23,1	17,4	19,4	20,1	23,6	22,4	22,0	17,4	24,2	19,7	121
Modena	VIGNOLA	22,0	16,6	19,4	21,2	22,2	23,0	20,8	18,7	21,2	18,0	142
Parma	FIDENZA	22,0	19,1	22,2	24,5	28,2	20,8	19,4	17,9	22,6	23,8	112
Parma	PARMA	23,1	19,2	23,0	21,3	25,5	20,9	21,0	19,1	24,0	24,5	227
Parma	SUD EST	21,5	19,2	20,6	20,3	25,2	20,8	18,6	14,5	21,4	20,7	57
Parma	VALLI TARO E CENO	21,6	19,8	23,5	21,5	25,1	22,4	21,8	21,6	24,1	24,6	10
Piacenza	CITTA' DI PIACENZA	19,8	15,1	18,1	16,2	21,5	22,7	22,0	20,9	27,0	22,6	57
Piacenza	LEVANTE	21,9	18,0	18,9	18,1	21,4	22,0	21,4	18,4	23,9	23,7	130
Piacenza	PONENTE	20,9	17,4	19,3	18,7	22,0	22,4	21,7	20,3	24,7	23,2	10
Ravenna	FAENZA	21,3	15,7	17,8	19,5	22,4	24,8	26,4	23,4	25,9	22,2	83
Ravenna	LUGO	17,4	15,3	15,6	18,9	21,9	21,6	21,6	20,9	26,2	23,5	61
Ravenna	RAVENNA	20,7	15,7	15,4	16,7	21,3	21,2	23,1	21,7	25,0	23,1	108
Reggio	CASTELNUOVO NÈ MONTI	22,3	18,9	20,8	21,5	21,9	21,8	20,2	21,9	25,6	24,4	53
Reggio	CORREGGIO	20,1	20,5	22,0	23,7	24,3	23,0	20,3	20,8	26,0	25,1	53
Reggio	GUASTALLA	20,4	18,6	20,0	19,1	22,8	22,4	21,8	22,5	25,5	24,0	10
Reggio	MONTECCHIO	20,5	18,4	20,2	20,3	22,7	23,2	21,4	22,1	25,0	23,7	10
Reggio	REGGIO EMILIA	18,1	17,4	20,2	20,4	24,3	29,4	24,9	23,6	27,2	24,6	173
Reggio	SCANDIANO	20,4	19,8	18,8	19,4	20,3	22,2	21,5	23,4	24,7	25,3	137
Rimini	RICCIONE	24,2	19,5	20,2	20,5	23,9	22,9	23,4	22,8	25,8	24,1	10
Rimini	RIMINI	22,5	19,5	18,3	18,1	23,5	21,7	23,0	21,3	24,7	23,0	198

**Tab. 4b. Coefficiente di variazione della stima del *Gap* mediano per distretti sociosanitari dell'Emilia-Romagna. Anni 2009-2018 (valori percentuali)**

CV del <i>Gap</i> mediano (%)											
Provincia	Denominazione	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Bologna	CASALECCHIO DI RENO	7,2	9,0	11,8	13,0	8,7	9,3	10,1	11,6	8,4	7,7
Bologna	CITTA' DI BOLOGNA	4,5	5,0	5,7	4,6	4,0	4,2	4,4	5,4	4,1	4,0
Bologna	IMOLA	7,2	7,6	9,4	10,1	7,6	7,0	7,6	9,2	7,7	8,5
Bologna	PIANURA EST	7,6	10,4	18,2	11,8	8,8	6,7	7,1	8,1	6,7	7,6
Bologna	PIANURA OVEST	8,2	10,3	10,8	10,8	7,3	7,9	9,8	11,8	9,8	9,5
Bologna	PORRETTA TERME	8,2	9,0	10,5	10,1	9,1	8,7	9,1	11,2	9,2	10,6
Bologna	SAN LAZZARO DI SAVENA	7,4	8,2	13,1	13,7	8,2	8,2	8,3	14,2	8,7	8,3
Ferrara	CENTRO-NORD	4,2	7,6	6,8	8,0	6,4	5,2	9,6	8,6	5,1	6,3
Ferrara	OVEST	7,9	9,7	9,5	10,3	8,3	8,8	12,5	11,8	8,8	10,8
Ferrara	SUD-EST	5,2	6,7	6,5	7,5	6,7	6,4	9,5	9,4	7,2	9,0
Forlì	CESENA - VALLE DEL SAVIO	6,2	6,3	7,8	8,3	6,2	7,6	7,8	7,3	6,8	7,2
Forlì	FORLÌ'	4,7	5,4	6,9	6,3	4,8	4,9	5,1	5,5	5,3	7,5
Forlì	RUBICONE	7,9	7,7	7,7	10,2	8,4	7,7	8,9	9,1	8,7	9,5
Modena	CARPI	7,4	8,5	14,3	8,8	7,3	7,1	8,3	9,4	7,7	8,3
Modena	CASTELFRANCO EMILIA	7,4	7,6	9,0	9,1	7,9	7,8	8,5	8,6	8,0	8,2
Modena	MIRANDOLA	7,4	8,4	8,7	9,3	9,1	8,2	8,1	9,4	8,1	8,7
Modena	MODENA	5,7	6,4	7,6	7,4	5,8	6,3	6,7	7,1	6,1	5,2
Modena	PAVULLO NEL FRIGNANO	7,8	8,6	11,6	10,4	8,7	9,0	9,6	10,2	9,5	9,1
Modena	SASSUOLO	7,1	7,9	7,7	7,7	8,6	7,3	7,4	9,7	7,1	7,5
Modena	VIGNOLA	7,6	8,1	8,2	8,2	7,9	7,2	7,3	9,4	8,6	10,1
Parma	FIDENZA	7,6	6,8	7,3	7,3	7,4	6,5	7,5	8,5	7,3	7,4
Parma	PARMA	4,5	4,3	4,8	6,7	5,1	5,1	4,3	4,6	3,8	4,1
Parma	SUD EST	8,4	7,6	10,4	7,8	7,8	8,3	11,3	17,2	8,6	10,9
Parma	VALLI TARO E CENO	8,3	8,9	9,0	9,6	8,2	9,7	11,0	11,9	10,5	11,2
Piacenza	CITTA' DI PIACENZA	7,8	11,2	8,1	12,1	6,1	6,7	6,5	6,8	5,9	7,0
Piacenza	LEVANTE	4,9	6,0	6,9	7,7	5,6	5,7	6,2	9,0	7,1	5,2
Piacenza	PONENTE	7,0	9,0	8,7	10,5	7,4	7,6	8,4	9,7	7,8	7,5
Ravenna	FAENZA	7,5	6,4	6,6	7,5	6,2	7,0	7,6	8,1	7,0	7,3
Ravenna	LUGO	10,2	7,1	8,3	7,1	6,6	6,8	8,0	9,3	7,3	6,7
Ravenna	RAVENNA	4,2	4,3	5,8	6,9	4,4	5,0	4,5	4,4	4,8	4,2
Reggio	CASTELNUOVO NÈ MONTI	10,0	9,4	9,8	10,8	8,7	10,8	11,6	10,9	9,7	8,1
Reggio	CORREGGIO	9,8	8,7	9,3	9,8	8,9	11,1	11,9	10,5	8,6	7,7
Reggio	GUASTALLA	8,3	8,7	8,7	8,8	7,6	11,3	11,8	10,7	9,6	8,1
Reggio	MONTECCHIO	9,6	9,0	10,3	9,7	9,1	11,6	12,3	11,2	9,8	8,1
Reggio	REGGIO EMILIA	6,2	5,8	5,1	4,2	4,1	6,0	5,1	4,7	4,7	3,8
Reggio	SCANDIANO	8,3	8,1	10,8	9,0	10,2	12,8	10,2	9,6	8,8	6,9
Rimini	RICCIONE	5,7	8,2	8,1	9,1	8,0	8,1	9,5	9,3	8,5	7,8
Rimini	RIMINI	4,0	3,8	7,5	7,5	5,5	7,0	5,0	6,3	6,0	5,8

**Tab. 5a. Stima dell'indice di Gini per i distretti sociosanitari dell'Emilia-Romagna e numerosità campionaria dell'ultimo anno. Anni 2009-2018 (valori percentuali e valore assoluto)**

Indice di Gini (%)												
Provincia	Denominazione	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	n
Bologna	CASALECCHIO DI RENO	27,0	28,0	27,3	27,7	26,9	27,3	26,6	28,2	28,2	27,6	42
Bologna	CITTA' DI BOLOGNA	31,6	33,1	31,4	32,5	31,4	32,5	31,6	32,3	31,8	33,9	285
Bologna	IMOLA	26,1	25,3	25,0	24,3	27,3	28,3	27,7	27,5	26,1	26,6	49
Bologna	PIANURA EST	23,1	23,4	22,4	25,6	25,0	26,4	26,6	26,8	27,3	27,9	190
Bologna	PIANURA OVEST	24,1	24,0	25,7	24,5	25,5	25,9	26,9	27,5	26,7	27,2	10
Bologna	PORRETTA TERME	25,4	26,8	26,3	26,5	25,8	27,5	27,2	28,2	28,1	29,2	26
Bologna	SAN LAZZARO DI SAVENA	29,3	29,9	26,2	26,8	28,8	28,2	30,4	27,6	28,3	29,7	126
Ferrara	CENTRO-NORD	26,3	26,4	26,5	27,0	27,1	29,6	27,1	28,9	28,5	29,0	144
Ferrara	OVEST	23,8	24,1	23,9	24,8	25,2	26,0	25,8	26,7	26,0	26,8	10
Ferrara	SUD-EST	22,5	22,3	22,0	23,4	23,2	23,3	25,1	25,1	24,6	26,0	10
Forlì	CESENA - VALLE DEL SAVIO	25,9	27,5	26,1	27,4	29,0	30,1	30,8	29,9	28,3	28,7	122
Forlì	FORLÌ'	26,5	26,5	27,4	27,6	28,5	29,1	28,7	30,6	29,0	28,2	75
Forlì	RUBICONE	27,1	25,5	26,0	26,6	27,1	26,7	26,2	26,7	26,5	28,1	10
Modena	CARPI	25,4	25,8	22,3	26,4	24,6	26,6	28,4	28,5	27,5	29,4	55
Modena	CASTELFRANCO EMILIA	25,2	25,9	25,5	26,7	24,7	26,9	29,0	27,0	26,8	27,4	119
Modena	MIRANDOLA	24,6	28,4	25,8	26,4	26,4	28,8	27,0	28,3	28,0	28,8	60
Modena	MODENA	29,9	30,1	29,7	31,3	29,4	28,5	28,4	29,6	28,5	30,7	219
Modena	PAVULLO NEL FRIGNANO	27,5	24,9	24,4	25,1	26,9	26,4	26,2	26,9	27,0	28,6	65
Modena	SASSUOLO	28,1	27,9	26,6	26,7	28,2	29,4	29,3	27,3	28,5	25,2	121
Modena	VIGNOLA	26,7	26,2	27,3	29,0	28,2	29,7	27,8	28,6	26,9	26,2	142
Parma	FIDENZA	25,9	27,2	27,0	29,8	29,6	26,5	25,5	26,2	27,0	29,6	112
Parma	PARMA	30,9	31,2	31,6	30,7	31,0	30,5	31,6	31,6	30,9	33,2	227
Parma	SUD EST	26,9	28,6	26,5	26,6	27,3	27,9	27,2	25,5	27,1	27,4	57
Parma	VALLI TARO E CENO	26,8	28,6	28,7	27,2	27,7	28,2	27,8	29,0	28,1	29,3	10
Piacenza	CITTA' DI PIACENZA	27,9	27,7	27,9	28,8	29,7	31,0	30,9	31,2	31,7	30,0	57
Piacenza	LEVANTE	27,4	27,5	26,0	25,5	26,7	27,3	27,0	26,4	27,2	29,1	130
Piacenza	PONENTE	27,2	27,2	26,9	27,6	27,8	28,3	28,0	28,5	27,9	28,7	10
Ravenna	FAENZA	28,3	25,0	25,2	27,7	27,6	30,3	31,3	30,1	28,6	28,5	83
Ravenna	LUGO	22,4	23,7	22,5	25,8	26,2	25,6	24,9	26,6	27,6	28,6	61
Ravenna	RAVENNA	27,5	26,6	24,0	26,0	27,2	26,4	28,4	27,9	27,8	28,2	108
Reggio	CASTELNUOVO NÈ MONTI	26,7	26,8	26,3	27,3	25,8	26,4	26,6	28,1	28,1	28,3	53
Reggio	CORREGGIO	25,9	28,3	27,2	29,3	27,8	28,0	27,3	28,0	28,6	29,8	53
Reggio	GUASTALLA	25,5	26,3	26,0	25,7	27,4	27,2	27,8	28,8	28,1	28,7	10
Reggio	MONTECCHIO	26,3	26,4	25,9	26,7	27,0	27,7	27,6	28,3	28,0	28,4	10
Reggio	REGGIO EMILIA	25,4	27,2	28,0	28,8	31,0	34,3	32,6	31,7	31,5	31,0	173
Reggio	SCANDIANO	26,5	28,3	25,1	26,1	24,2	27,3	27,8	29,9	27,8	29,3	137
Rimini	RICCIONE	29,1	26,2	26,7	27,5	27,7	28,1	28,1	28,8	28,4	29,6	10
Rimini	RIMINI	27,9	29,0	26,3	27,0	29,6	29,3	28,7	29,0	28,6	28,6	198

**Tab. 5b. Coefficiente di variazione della stima dell'indice di Gini per i distretti sociosanitari dell'Emilia-Romagna. Anni 2009-2018 (valori percentuali)**

CV dell'indice di Gini (%)											
Provincia	Denominazione	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Bologna	CASALECCHIO DI RENO	5,7	6,3	6,6	6,5	6,5	6,6	6,4	7,0	6,9	6,6
Bologna	CITTA' DI BOLOGNA	4,4	4,7	4,9	4,7	4,7	4,6	4,4	4,8	4,7	4,9
Bologna	IMOLA	7,0	6,8	6,6	6,7	6,4	6,2	6,4	6,4	6,5	7,1
Bologna	PIANURA EST	5,9	6,0	6,2	5,7	6,1	5,5	5,3	5,8	5,2	5,9
Bologna	PIANURA OVEST	6,9	7,1	6,4	6,6	6,5	6,6	7,6	7,6	7,5	7,6
Bologna	PORRETTA TERME	6,7	6,9	6,8	6,7	7,6	6,8	6,7	7,1	6,4	7,2
Bologna	SAN LAZZARO DI SAVENA	5,3	5,8	6,3	6,5	5,6	6,2	5,9	5,8	5,8	5,7
Ferrara	CENTRO-NORD	5,3	5,7	5,8	5,4	5,4	5,4	5,1	5,5	5,1	5,9
Ferrara	OVEST	7,9	7,9	7,9	7,8	7,7	7,7	7,7	7,5	7,5	7,6
Ferrara	SUD-EST	6,2	6,8	6,8	6,9	7,4	7,0	7,9	7,8	7,6	7,8
Forlì	CESENA - VALLE DEL SAVIO	5,9	6,0	6,1	6,3	6,0	5,4	4,8	5,2	5,5	6,0
Forlì	FORLÌ'	6,0	5,9	5,7	5,9	5,8	6,1	6,2	6,0	6,1	6,3
Forlì	RUBICONE	6,0	7,1	6,8	7,4	7,1	7,0	7,8	7,8	7,5	7,6
Modena	CARPI	6,7	7,0	6,5	6,2	6,2	6,2	6,5	6,7	6,5	6,3
Modena	CASTELFRANCO EMILIA	6,1	5,9	6,6	6,4	6,5	6,3	5,8	5,7	6,0	6,0
Modena	MIRANDOLA	7,8	6,3	6,7	7,2	7,2	6,6	6,7	6,3	6,4	6,3
Modena	MODENA	5,5	5,6	5,4	5,3	5,6	5,9	5,7	5,5	5,1	5,1
Modena	PAVULLO NEL FRIGNANO	5,9	6,9	7,1	7,2	6,7	7,2	7,2	6,8	7,1	6,5
Modena	SASSUOLO	6,4	6,7	6,4	6,3	6,2	6,5	6,1	6,3	6,3	6,7
Modena	VIGNOLA	7,1	6,6	5,9	6,1	6,1	6,1	5,8	5,8	5,4	6,0
Parma	FIDENZA	7,6	5,9	6,2	6,4	6,2	6,0	5,6	5,6	5,6	6,3
Parma	PARMA	5,3	5,2	5,3	5,6	5,6	5,1	5,1	5,1	4,7	5,3
Parma	SUD EST	6,5	6,5	6,5	6,5	6,8	7,1	6,8	6,4	6,5	6,8
Parma	VALLI TARO E CENO	6,7	6,6	6,7	7,6	7,4	7,4	7,6	7,1	7,2	7,5
Piacenza	CITTA' DI PIACENZA	6,6	6,4	5,9	6,2	5,8	5,9	5,8	6,0	5,8	6,3
Piacenza	LEVANTE	6,0	6,4	6,8	6,6	6,2	6,2	6,0	5,9	5,9	5,9
Piacenza	PONENTE	6,9	7,5	7,5	7,6	7,4	7,3	7,5	7,4	7,4	7,6
Ravenna	FAENZA	6,0	6,3	6,4	6,4	6,3	6,1	5,9	6,1	5,6	6,5
Ravenna	LUGO	6,1	6,3	7,0	6,7	6,9	7,1	6,6	7,0	6,6	6,5
Ravenna	RAVENNA	5,8	5,8	6,2	5,8	6,0	5,9	6,0	5,9	5,9	6,0
Reggio	CASTELNUOVO NÈ MONTI	7,0	7,2	7,3	7,4	6,7	6,7	6,7	6,6	6,3	6,9
Reggio	CORREGGIO	7,5	6,5	6,4	6,8	6,8	6,8	6,6	6,2	6,4	6,7
Reggio	GUASTALLA	7,0	7,6	6,3	6,9	6,8	6,9	7,5	7,3	7,1	7,4
Reggio	MONTECCHIO	7,7	7,6	7,8	7,6	7,5	7,5	7,5	7,5	7,4	7,6
Reggio	REGGIO EMILIA	5,8	5,5	5,5	5,4	5,6	5,0	5,5	5,6	5,3	5,0
Reggio	SCANDIANO	6,1	5,9	6,5	6,4	7,0	6,6	6,5	6,3	6,3	6,2
Rimini	RICCIONE	5,5	6,1	7,6	7,5	7,5	7,4	7,5	7,3	7,2	7,5
Rimini	RIMINI	5,4	5,5	5,6	6,0	5,7	5,5	5,9	5,3	5,6	5,7

**Tab. 6a. Stima del Tasso di deprivazione severa per i distretti sociosanitari dell'Emilia-Romagna e numerosità campionaria dell'ultimo anno. Anni 2009-2018 (valori percentuali e valore assoluto)**

Tassodi deprivazione severa (%)												
Provincia	Denominazione	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	n
Bologna	CASALECCHIO DI RENO	2,6	2,9	5,1	5,6	4,2	4,0	4,2	5,0	4,5	3,7	42
Bologna	CITTA' DI BOLOGNA	2,2	2,5	5,8	6,0	5,9	6,1	6,0	7,3	2,2	3,0	285
Bologna	IMOLA	3,2	3,7	7,2	5,2	3,4	5,7	6,7	4,8	3,9	3,6	49
Bologna	PIANURA EST	2,3	2,1	3,5	7,0	4,2	5,8	4,9	3,7	3,9	3,4	190
Bologna	PIANURA OVEST	4,7	5,1	6,4	4,7	5,6	3,8	6,0	5,3	5,0	2,1	10
Bologna	PORRETTA TERME	3,6	4,1	9,7	7,0	5,8	4,0	6,5	6,9	6,0	3,3	26
Bologna	SAN LAZZARO DI SAVENA	3,7	2,3	6,2	6,7	3,3	4,4	5,5	4,1	4,8	3,5	126
Ferrara	CENTRO-NORD	3,1	2,6	4,1	2,4	5,1	3,2	1,8	3,8	6,6	1,3	144
Ferrara	OVEST	3,5	3,4	5,2	6,1	6,5	5,5	5,0	5,6	6,3	2,0	10
Ferrara	SUD-EST	3,9	3,6	3,6	3,6	7,0	4,3	4,7	5,0	5,8	1,4	10
Forlì	CESENA - VALLE DEL SAVIO	3,2	5,1	6,8	7,9	5,8	4,5	5,6	8,5	8,1	2,3	122
Forlì	FORLÌ'	3,6	4,5	4,7	10,8	16,3	10,4	11,2	8,7	9,7	3,4	75
Forlì	RUBICONE	3,2	4,2	4,2	8,5	9,2	8,0	7,2	6,7	6,6	1,9	10
Modena	CARPI	2,9	3,8	3,7	3,7	7,9	7,5	6,8	7,9	5,3	2,9	55
Modena	CASTELFRANCO EMILIA	4,1	3,3	7,9	5,7	10,5	7,3	3,9	3,5	6,9	2,2	119
Modena	MIRANDOLA	2,6	3,6	6,4	7,2	9,1	8,3	7,0	4,5	4,6	2,5	60
Modena	MODENA	2,5	2,3	5,6	6,0	9,6	8,1	4,9	4,6	7,8	1,2	219
Modena	PAVULLO NEL FRIGNANO	2,5	3,3	5,5	6,5	8,0	7,1	5,6	8,0	6,2	3,4	65
Modena	SASSUOLO	3,8	3,8	6,9	7,1	11,6	9,1	5,5	6,3	6,3	3,1	121
Modena	VIGNOLA	3,5	2,8	5,5	7,2	8,6	8,4	9,4	8,6	7,7	3,3	142
Parma	FIDENZA	3,4	8,0	5,3	7,2	6,7	3,5	6,7	5,3	4,9	1,7	112
Parma	PARMA	2,9	3,1	5,7	4,3	9,5	8,7	6,8	4,4	3,3	2,6	227
Parma	SUD EST	4,2	6,3	4,5	9,0	7,3	7,5	6,6	3,7	4,1	3,2	57
Parma	VALLI TARO E CENO	3,8	4,8	5,8	8,3	8,3	7,8	7,0	7,2	6,3	7,1	10
Piacenza	CITTA' DI PIACENZA	2,1	2,2	11,2	9,8	6,2	5,8	4,1	6,2	8,3	7,4	57
Piacenza	LEVANTE	2,2	2,5	7,7	7,9	7,7	6,8	4,3	10,1	11,3	1,6	130
Piacenza	PONENTE	3,1	2,9	7,6	7,9	6,4	5,9	4,7	6,6	7,0	5,1	10
Ravenna	FAENZA	3,5	4,6	5,5	10,9	5,2	5,0	6,8	3,6	4,6	4,4	83
Ravenna	LUGO	5,4	3,3	5,8	7,4	5,9	5,4	4,7	4,8	6,2	1,4	61
Ravenna	RAVENNA	4,0	3,6	9,0	7,7	6,9	5,9	4,3	2,7	4,5	5,1	108
Reggio	CASTELNUOVO NÈ MONTI	4,2	5,0	8,3	11,0	9,3	9,3	7,5	6,8	6,2	1,5	53
Reggio	CORREGGIO	3,0	4,3	8,1	12,1	9,7	10,9	7,4	6,8	7,4	2,4	53
Reggio	GUASTALLA	3,4	3,9	6,7	8,1	7,1	9,4	9,0	7,7	6,4	3,0	10
Reggio	MONTECCHIO	3,3	3,5	6,8	9,1	8,4	10,5	8,8	7,7	6,1	2,7	10
Reggio	REGGIO EMILIA	2,9	3,3	5,2	6,2	8,9	12,4	8,9	9,7	6,0	3,4	173
Reggio	SCANDIANO	2,2	2,2	8,5	9,6	8,7	13,9	11,7	8,8	3,6	1,8	137
Rimini	RICCIONE	5,4	3,1	5,7	7,4	6,3	5,8	5,3	6,2	6,8	3,5	10
Rimini	RIMINI	2,7	3,9	6,2	7,5	5,2	4,5	2,9	6,1	9,8	1,6	198

**Tab. 5b. Coefficiente di variazione della stima del Tasso di deprivazione severa per i distretti sociosanitari dell'Emilia-Romagna. Anni 2009-2018 (valori percentuali)**

CV del Tasso di deprivazione severa (%)											
Provincia	Denominazione	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Bologna	CASALECCHIO DI RENO	45,6	50,6	43,2	46,4	53,4	52,5	45,3	50,5	45,1	62,6
Bologna	CITTA' DI BOLOGNA	32,6	34,5	26,0	27,8	27,2	24,4	24,2	24,0	36,9	56,9
Bologna	IMOLA	46,1	50,1	40,6	46,7	56,2	46,7	44,0	51,4	43,3	52,0
Bologna	PIANURA EST	44,3	51,7	43,6	38,6	48,0	39,0	37,4	49,3	38,7	58,3
Bologna	PIANURA OVEST	48,9	50,4	40,7	48,5	49,4	52,9	51,6	58,2	54,5	71,0
Bologna	PORRETTA TERME	48,8	52,0	46,5	47,8	59,5	52,3	46,4	52,7	48,0	62,2
Bologna	SAN LAZZARO DI SAVENA	46,4	55,0	43,2	48,5	57,0	49,4	43,0	49,7	44,0	43,2
Ferrara	CENTRO-NORD	29,7	32,5	29,1	52,9	31,3	38,7	67,3	36,2	22,6	86,0
Ferrara	OVEST	47,6	48,4	48,0	51,2	45,9	49,8	54,2	49,4	40,5	77,2
Ferrara	SUD-EST	36,5	39,0	44,2	50,4	38,5	44,8	47,0	46,1	39,0	85,3
Forlì	CESENA - VALLE DEL SAVIO	43,2	34,7	32,4	38,5	50,2	50,1	45,0	30,8	32,4	67,4
Forlì	FORLÌ'	29,4	27,3	28,8	22,7	18,5	21,5	22,4	23,0	22,2	66,4
Forlì	RUBICONE	48,1	44,6	44,8	41,3	41,9	40,2	44,1	43,2	41,9	69,8
Modena	CARPI	48,7	51,5	46,7	57,9	33,9	36,1	40,5	39,6	35,7	63,2
Modena	CASTELFRANCO EMILIA	49,2	48,1	42,4	50,2	34,8	36,7	42,0	44,2	35,7	60,3
Modena	MIRANDOLA	58,3	52,0	46,2	53,1	39,2	40,3	44,1	42,5	38,1	80,0
Modena	MODENA	43,1	46,2	36,1	39,7	28,6	30,0	33,9	33,7	26,7	61,3
Modena	PAVULLO NEL FRIGNANO	49,6	51,3	46,1	54,1	38,2	41,9	45,0	45,5	41,6	55,6
Modena	SASSUOLO	44,7	45,7	39,9	44,9	33,0	36,7	37,3	36,0	33,7	34,6
Modena	VIGNOLA	49,1	49,9	42,4	45,6	33,7	34,6	36,4	35,0	33,5	52,6
Parma	FIDENZA	46,6	30,8	40,3	36,3	43,3	59,5	35,2	37,2	36,8	49,7
Parma	PARMA	32,4	31,8	26,2	34,4	23,2	22,7	22,5	27,0	29,3	53,8
Parma	SUD EST	43,5	36,4	47,0	39,4	45,1	45,8	40,7	50,7	44,6	50,5
Parma	VALLI TARO E CENO	45,2	34,5	45,9	49,5	51,2	51,2	49,5	48,7	50,7	25,3
Piacenza	CITTA' DI PIACENZA	49,9	52,9	27,3	30,7	36,8	35,3	41,4	34,2	29,2	32,5
Piacenza	LEVANTE	47,4	45,7	31,1	34,1	30,8	30,6	37,4	24,2	24,0	48,0
Piacenza	PONENTE	49,4	56,7	38,1	42,8	46,6	46,6	50,0	41,2	37,5	22,2
Ravenna	FAENZA	40,9	39,2	42,2	34,3	44,6	42,4	37,0	49,8	44,0	50,9
Ravenna	LUGO	34,4	40,3	40,1	38,2	41,5	41,7	39,0	44,8	40,2	88,5
Ravenna	RAVENNA	26,4	26,8	21,1	24,8	24,1	24,9	27,6	41,4	30,1	45,3
Reggio	CASTELNUOVO NÈ MONTI	54,0	54,2	42,9	40,8	42,4	33,1	35,1	45,2	46,9	77,7
Reggio	CORREGGIO	64,0	51,7	39,3	38,1	40,5	32,3	33,2	41,8	43,0	52,6
Reggio	GUASTALLA	52,7	57,5	37,5	35,8	39,7	32,0	38,4	46,1	50,5	42,3
Reggio	MONTECCHIO	61,4	60,4	46,7	42,3	45,6	36,8	39,3	47,6	52,2	62,2
Reggio	REGGIO EMILIA	36,9	33,5	27,6	25,5	24,0	18,0	20,4	22,7	26,3	73,2
Reggio	SCANDIANO	60,9	60,6	38,1	34,8	38,7	31,6	33,1	38,2	51,6	76,1
Rimini	RICCIONE	31,3	47,6	43,3	41,8	41,2	42,0	40,6	41,0	42,0	63,7
Rimini	RIMINI	31,6	19,5	20,5	21,5	26,0	27,4	38,0	21,5	14,9	52,1

## Appendice 1. La metodologia di stima

### 1. Stima della varianza delle stime dirette

Il disegno campionario di EU-SILC, descritto nella Appendice 2, è complesso e non riconducibile a nessun disegno “notevole” a cui poter applicare stimatori delle varianze (siano essi stimatori diretti, per linearizzazione o basati su metodi di ri-campionamento) noti in letteratura. Inoltre, la natura complessa delle operazioni di calibrazione e di inseguimento delle unità *panel* rende problematica la ricostruzione esatta delle informazioni relative al disegno e alla costruzione dei pesi effettuate da Istat.

Non è possibile calcolare le probabilità di inclusione del secondo ordine che sarebbero necessarie per calcolare gli stimatori “diretti” delle varianze. Le operazioni di calibrazione effettuate da Istat sul campione originario che, in molti algoritmi di ri-campionamento devono essere ri-effettuate per ciascun campione estratto nella simulazione, non sono ricostruibili con esattezza. Inoltre, per effetto della natura panel del campione, non è possibile ricostruire, sulla base delle informazioni disponibili, quali unità di primo livello, ossia i comuni, siano incluse nel campione per esser state effettivamente campionate e quali vi compaiano per esclusivo effetto delle regole di inseguimento.

Nella letteratura che si è occupata della stima della varianza di stimatori calcolati utilizzando i dati EU-SILC è emerso come l’applicazione di stimatori delle varianze che tenessero conto di alcuni elementi essenziali del disegno, ancorché non di tutti, porti a stime sufficientemente affidabili (Biewen and Jenkins, 2006; Goedemé, 2013). La nostra procedura di stima della varianza segue questo tipo di impostazione.

Una prima semplificazione del disegno consiste nel considerare l’intero campione EU-SILC come se fosse estratto simultaneamente, trascurando la natura di *panel* ruotato del campione.

Gli elementi che consideriamo essenziali nel disegno e che includiamo nella stima della varianza sono elencati di seguito.

- Il raggrupparsi degli individui all’interno di *cluster* definiti dalle famiglie. È un aspetto molto importante poiché, nonostante tutti gli indicatori siano definiti in termini di individui, il reddito equivalente (o lo stato di deprivazione materiale) è lo stesso per tutti gli individui all’interno dello stesso *cluster*.

- La stratificazione iniziale delle unità.
- La variabilità dei pesi. Considerare questa variabilità è importante in quanto, in primo luogo, essa è notevole; inoltre, poiché i pesi sono sostanzialmente indipendenti dal reddito equivalente, questa variabilità rappresenta un fattore di inflazione della varianza.

Una volta adottate queste semplificazioni, è stato utilizzato uno stimatore delle varianze per linearizzazione di tipo standard, così come è implementato nel pacchetto *Survey* del software *R* (Lumley, 2004, 2017) per tutti gli stimatori in esame. Fa eccezione l'indice di concentrazione di Gini, per cui è stato utilizzato un algoritmo diverso, basato su una approssimazione lineare dell'indice, implementato nel pacchetto *Convey* di *R* (Pessoa et al., 2018). In quest'ultimo caso, il risultato ottenuto utilizzando il pacchetto *Convey* è stato successivamente modificato alla luce della correzione della distorsione che abbiamo utilizzato.

Nella edizione 2015 del Rapporto veniva usato un algoritmo di tipo *bootstrap* basato su alcune semplificazioni del disegno in linea con quelle che abbiamo appena descritto. Gli algoritmi *bootstrap* permettono di ottenere le varianze per qualsiasi stimatore (anche non lineare) al prezzo di un notevole "sforzo" computazionale. L'utilizzo di stimatori *bootstrap* permette, inoltre, la stima della correlazione tra stimatori diretti di parametri diversi. Queste correlazioni possono essere usate in un certo tipo di modelli multivariati che sono stati considerati in passato come opzione possibile, ma non nel modello che è stato utilizzato nel Rapporto ediz. 2015 e che è stato mantenuto, seppur con qualche modifica ed estensione, in questa edizione.

Per due *wave* consecutive dell'indagine, abbiamo confrontato le stime *bootstrap* con quelle ottenute utilizzando gli stimatori per linearizzazione. I risultati sono molto simili. Ciò conferma le aspettative teoriche: a parità di disegno, non vi sono ragioni per ritenere che stimatori *bootstrap* e per linearizzazione si comportino in modo sostanzialmente diverso. L'unico dubbio riguardava la stima della varianza dell'indice di Gini, che è una funzione non lineare delle osservazioni e per cui era possibile che la linearizzazione operasse una semplificazione eccessiva; anche in questo caso le stime sono risultate sostanzialmente coincidenti. Abbiamo deciso di utilizzare gli stimatori per linearizzazione perché sono più "leggeri" dal punto di vista computazionale.

Infine, nella presentazione dei risultati, le stime regionali sono prodotte con la nostra metodologia di produzione delle stime dirette, quelle Italia sono fonte Eurostat.

## 2. La stima per piccole aree

### 2.1. Sintesi delle principali novità

La stima del parametro di deprivazione severa è stata effettuata mediante un modello che ha la medesima struttura del modello usato per la deprivazione nel Rapporto ediz. 2015. Per la stima del rischio di povertà e dell'indice di Gini, invece, è stato utilizzato un approccio diverso in alcuni punti.

Al fine di aumentare la quantità di informazione utilizzata nella stima SAE (*small area estimation*) di questi parametri (oltre che del *Gap* mediano), sono stati utilizzati nella modellazione tre ulteriori parametri: la media del logaritmo dei redditi (già utilizzata nel 2015), il rischio di povertà mediano (frazione di individui con reddito inferiore alla mediana nazionale), l'*affluence rate* (frazione di individui con reddito superiore a due volte la mediana nazionale).

Sono stati quindi modellati complessivamente cinque parametri, che riassumono informazioni su tutta la distribuzione del reddito.

Sono state confrontate due categorie di modelli rispondenti a due ipotesi fondamentali:

hp1: indipendenza tra i cinque parametri;

hp2: dipendenza tra i cinque parametri.

Le due categorie di modelli hanno in comune alcune caratteristiche:

1. legame tra covariate e parametro di interesse specificato attraverso una funzione *logit* (tranne che per la media dei logaritmi, in questo caso la funzione link è la funzione identità);
2. *deff* (*design effect*) noto e stimato attraverso funzioni di *smoothing* della varianza (tranne che per la media dei logaritmi, in questo caso la varianza è assunta nota e pari alla stima diretta della varianza);
3. intercette specifiche per ogni anno di indagine; le intercette non sono indipendenti ma seguono un processo RW (*random walk*) gaussiano di ordine 1; l'obiettivo è quello di ottenere *smoothing* temporale;
4. Effetti casuali generati da un'unica distribuzione (di tipo Variance-Gamma) per tutti gli anni di indagine considerati.

La dipendenza tra i cinque parametri nell'hp2 è stata ottenuta specificando per gli effetti casuali una distribuzione di tipo Variance-Gamma multivariata (anche in questo caso la distribuzione è la stessa per tutti gli anni di indagine).

Le caratteristiche 2 e 3 sono comuni ai modelli utilizzati nel 2015.

È stato verificato, attraverso uno studio preliminare di sensibilità, che la specificazione della funzione link logistico (anziché la probit, utilizzata nel 2015) migliora notevolmente il mixing dei modelli e la velocità computazionale senza produrre cambiamenti apprezzabili nei risultati.

L'uso di cinque parametri, anziché 3, e l'aggiunta di tre anni di indagine aumenta notevolmente i tempi di calcolo dei modelli, soprattutto nel caso multivariato (circa 8 ore per ottenere catene MCMC di lunghezza pari a 20000).

Sono quindi stati stimati i seguenti tre modelli.

Modello 1: ipotesi di indipendenza tra i cinque parametri; parametri di regressione uguali per tutti gli anni di indagine

Modello 2: ipotesi di dipendenza tra i cinque parametri; parametri di regressione uguali per tutti gli anni di indagine

Modello 3: ipotesi di dipendenza tra i cinque parametri; parametri di regressione diversi per ogni anno di indagine

Con riferimento ai 3 modelli stimati sono stati ottenuti i seguenti valori del *Deviance information criterion*:

Modello 1: -16240

Modello 2: -18490

Modello 3: -18390

Per la stima del *Gap* mediano è stato seguito un procedimento simile a quello utilizzato nel 2015 solo nei seguenti punti:

1. non sono state utilizzate le stime dirette in quanto altamente imprecise e distorte;
2. una stima del *Gap* è ricavata a partire dai parametri stimati nel modello di stima SAE utilizzando un'ipotesi distributiva del reddito equivalente nei domini di studio.

Vi sono tuttavia notevoli elementi di novità. In sintesi:

1. è stata cambiata l'ipotesi distributiva sottostante; è stata utilizzata una distribuzione di Dagum anziché una Log-Normale;
2. l'informazione utilizzata è più ricca (cinque parametri anziché tre) e consente di caratterizzare in modo più completo la distribuzione del reddito nei domini.

Gli elementi di novità introdotti sono stati testati su un ampio studio di simulazione presentato in Fabrizi, Ferrante and Trivisano (2019).

## 2.2. I modelli di stima per piccole aree

Indichiamo per semplicità con  $\hat{t}_{itk}$ ,  $i = 1, \dots, m$ ,  $t = 1, \dots, T$ ,  $k = 1, 2, \dots, 5$  lo stimatore diretto del  $k$ -esimo parametro di interesse nell'area  $i$  nell'anno  $t$ ;  $m$  è il numero delle aree considerate,  $T = 9$  il numero degli anni;  $k = 1$  identifica il tasso di povertà,  $k = 2$  identifica la frazione di famiglie con reddito inferiore al reddito mediano nazionale,  $k = 3$  identifica la frazione di famiglie con reddito superiore al doppio del reddito mediano nazionale,  $k = 4$  identifica l'indice di Gini,  $k = 5$  identifica la media del logaritmo del reddito. L'indice  $i$  identifica aree che ricoprono l'intero territorio nazionale. Il motivo per cui sono stati considerati ulteriori parametri oltre ai due parametri target (tasso di povertà, indice di Gini) è chiarito più avanti.

### 2.2.1 Modello univariato per i tassi

Poiché per  $k = 1, 2, 3$   $\hat{t}_{itk} \in (0, 1)$  possiamo assumere:

$$\hat{t}_{itk} \sim \text{Beta}(\theta_{itk}(f_{itk} - 1), (1 - \theta_{itk})(f_{itk} - 1)) \quad (1)$$

che implica  $E(\hat{t}_{itk}) = \theta_{itk}$ ,  $V(\hat{t}_{itk}) = \frac{\theta_{itk}(1-\theta_{itk})}{f_{itk}}$ .  $\theta_{itk}$  è il tasso per la popolazione del dominio  $i$  e l'anno  $t$ .

Poiché  $\frac{V_{CCS}(\hat{t}_{itk})}{V(\hat{t}_{itk})} = \frac{f_{itk}}{n_{it}^*}$ ,  $f_{itk}$  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  $\kappa_{tk, des}$  nell'equazione

$$\frac{\hat{\epsilon}_{itk}(1-\hat{\epsilon}_{itk})}{\hat{V}_{lin}(\hat{\epsilon}_{itk})} = \kappa_{tk,des}n_{it}^* + e_{itk} \quad (2)$$

utilizzando il metodo dei minimi quadrati. Potremo allora porre  $f_{itk} = \hat{\kappa}_{tk,des}n_{it}^*$  con  $\hat{\kappa}_{tk,des}$  interpretabile direttamente come misura di efficienza del piano di meno efficiente di quello casuale semplice). Le stime  $\hat{\kappa}_{tk,des}$  sono riportate di seguito per  $k=1,2,3$ .

Il modello 1 e l'equazione 2 sono utilizzati anche per il tasso di deprivazione severa; nel seguito sono riportate le stime di  $\hat{\kappa}_{t,des}$  per questo parametro.

**Tabella 1. Valori stimati del parametro  $\kappa_{t1,des}$  (Rischio di povertà) e relativo  $R^2$**

Anno	Stima parametro	$R^2$
2008	0.265	0.723
2009	0.259	0.784
2010	0.267	0.861
2011	0.235	0.823
2012	0.218	0.878
2013	0.221	0.734
2014	0.233	0.892
2015	0.243	0.882
2016	0.216	0.803
2017	0.239	0.874
2018	0.227	0.860

**Tabella 2. Valori stimati del parametro  $\kappa_{t2,des}$  (Frazione di famiglie con reddito inferiore al reddito mediano nazionale) e relativo  $R^2$**

Anno	Stima parametro	$R^2$
2008	0.247	0.956
2009	0.249	0.938
2010	0.251	0.944
2011	0.232	0.952
2012	0.220	0.960
2013	0.209	0.946
2014	0.257	0.970
2015	0.262	0.971
2016	0.239	0.962
2017	0.250	0.968
2018	0.239	0.851

**Tabella 3. Valori stimati del parametro  $\kappa_{t3,des}$  (Frazione di famiglie con reddito superiore al doppio del reddito mediano nazionale) e relativo  $R^2$**

Anno	Stima parametro	$R^2$
2008	0.286	0.807
2009	0.307	0.821
2010	0.310	0.908
2011	0.336	0.902
2012	0.335	0.839
2013	0.346	0.671
2014	0.381	0.874
2015	0.395	0.858
2016	0.336	0.867
2017	0.342	0.572
2018	0.311	0.805

Per modellare  $\theta_{itk}$  specifichiamo il seguente modello:

$$\text{logit}(\theta_{itk}) = \alpha_{tk} + \mathbf{x}_{itk}^t \beta_k + v_{itk} \quad (3)$$

dove  $\mathbf{x}_{itk}$  è un vettore di variabili esplicative dell'area  $i$  al tempo  $t$  per il parametro  $k$ ,  $v_{itk}$  è un effetto casuale,  $\alpha_{tk}$  è un'intercetta specifica di anno. Il modello assume inoltre che il parametro di regressione  $\beta_k$  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 2.7).

Riguardo al vettore  $\alpha_\pi$  delle intercette specifiche di anno, si assume che:

$$\begin{aligned} \alpha_{tk} | \alpha_{\setminus tk} &\sim N(\alpha_{t+1,k}, \sigma_{\alpha k}^2) & t = 1 \\ \alpha_{tk} | \alpha_{\setminus tk} &\sim N\left(\frac{\alpha_{t-1,k} + \alpha_{t+1,k}}{2}, \frac{\sigma_{\alpha k}^2}{2}\right) & t = 2, \dots, T-1 \\ \alpha_{tk} | \alpha_{\setminus tk} &\sim N(\alpha_{t-1,k}, \sigma_{\alpha k}^2) & t = T \end{aligned} \quad (4)$$

dove  $\alpha_{\setminus tk}$  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 che:

$$v_{itk} | v\sigma_k^2 \sim N(0, v\sigma_k^2) \text{ e } v\sigma_k \sim \text{half-t}(v_k, A_k).$$

Tale specificazione è equivalente ad assumere:

$v_{itk} = v_{itk}^* \sqrt{\xi_k}$ , con  $v_{itk}^* \sim N(0, \eta_k^2)$ ,  $\xi_k \sim \text{Gamma}(.5, A_k^{-2})$ ,  $\eta_k^{-2} \sim \text{Gamma}(v_k/2, v_k)$ ; si ha inoltre che

$$v_{itk} | \eta_k \sim VG\left(.5, \sqrt{\frac{2}{\eta_k^2 A_k}}, 0, 1\right) \quad (5)$$

(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  $E(v_{itk} | \eta_k) = 0$  e  $V(v_{itk} | \eta_k) = \frac{\eta_k^2 A_k^2}{2}$ . L'uso di tale distribuzione presenta notevoli vantaggi rispetto alla usuale 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.

## 2.2.2 Modello univariato per l'indice di Gini

Poiché  $\hat{t}_{it4} \in (0,1)$ , possiamo specificare un modello di tipo Beta anche per questo parametro:

$$\hat{t}_{it4} \sim \text{Beta}\left(\frac{2f_{it4}}{1-\theta_{it4}} - \theta_{it4}, \frac{2f_{it4} - \theta_{it4}(1+\theta_{it4})}{1+\theta_{it4}} \frac{(1-\theta_{it4})}{\theta_{it4}}\right) \quad (6)$$

dove  $\theta_{it4}$  indica il parametro oggetto principale di stima, ovvero l'indice di concentrazione in popolazione. Notiamo che, pur essendo basato sulla distribuzione Beta, il modello (6) è diverso da (1) in quanto, nonostante il valore atteso coincida con il parametro oggetto di stima, la varianza ha una struttura diversa (Fabrizi e Trivisano, 2016). Infatti, il modello (6) implica  $E(\hat{t}_{it4}) = \theta_{it4}$ ,

$$V(\theta_{it4}) = \theta_{it4}^2(1 - \theta_{it4}^2)(2f_{it4})^{-1}. \quad (7)$$

La struttura di varianza (7) è motivata di seguito.

I parametri  $f_{it4}$  sono assunti noti nel modello; essi sono stimati esogenamente con un modello di "variance smoothing".

Assumiamo, inoltre, analogamente alla (3), che:

$$\text{logit}(\theta_{it4}) = \alpha_{t4} + \mathbf{x}_{it4}^t \beta_4 + v_{it4} \quad (8)$$

dove  $\mathbf{x}_{it4}$  è un vettore di variabili esplicative dell'area  $i$  al tempo  $t$ ,  $v_{it4}$  è un effetto casuale,  $\alpha_{t4}$  è un'intercetta specifica di anno. Il modello assume inoltre che il parametro di regressione  $\beta_4$  sia lo stesso per ogni anno. Per i parametri  $\alpha_{t4}$  e  $v_{it4}$  si è scelta una specificazione uguale a quella utilizzata per gli analoghi parametri nel modello per i tassi, quindi rispettivamente le specificazioni (4) e (5).

### 2.2.2.1 Lo “smoothing” della varianza dell’indice di concentrazione di Gini

Per semplicità espositiva verrà illustrato il caso della stima dei parametri  $f_{it4}$  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:

$$y_{ij} \sim N(\mu_i, \sigma_i^2) \quad (9)$$

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

$${}_{LN}\theta_{i4} = 2\Phi\left(\frac{\sigma_i}{\sqrt{2}}\right) - 1 = \operatorname{erf}\left(\frac{\sigma_i}{\sqrt{2}}\right) \quad (10)$$

Se assumiamo il campionamento casuale semplice una stima diretta (consistente) di  ${}_{LN}\theta_{i4}$  è data da

$$g_{2i} = \operatorname{erf}\left(\frac{s_i}{2}\right) \quad (11)$$

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

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

È quindi possibile dimostrare che

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

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

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

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

Dalla (10) è facile ottenere che

$$\sigma_i = \sqrt{2}\Phi^{-1}\left(\frac{LN\theta_{i4}+1}{2}\right) \quad (14)$$

Per praticità possiamo riscrivere (B.14) come  $\sigma_i \cong \frac{\sqrt{\pi}}{2} \text{logit}\left(\frac{LN\theta_{i4}+1}{2}\right)$ , sfruttando la relazione

$\text{logit}(x) \cong \sqrt{\frac{3}{\pi}}\Phi^{-1}(x)$ . La relazione può essere ulteriormente semplificata notando che, poiché in

un intorno di 1  $\log(1+x) \simeq 1+x$ ,  $\text{logit}\left(\frac{LN\theta_{i4}+1}{2}\right) \simeq 2LN\theta_{i4}$  e quindi

$$\sigma_i \cong LN\theta_{i4}\sqrt{\pi}. \quad (15)$$

Dalle formule (15) e (13) si ottiene

$$V(g_{2i}) \simeq (2n_i^*)^{-1}LN\theta_{i4}^2(1 - 0.5\pi LN\theta_{i4}^2). \quad (16)$$

Poiché la relazione (16) può portare, per valori elevati di  $LN\theta_{i4}$ , a valori negativi di  $V(g_{2i})$ , essa è sostituita da

$$V(g_{2i}) \simeq (2n_i^*)^{-1}LN\theta_{i4}^2(1 - LN\theta_{i4}^2) \quad (17)$$

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

$$\hat{V}_{lin}(\hat{t}_{i4}) = \kappa_{4,des} (2n_i^*)^{-1}\hat{t}_{i4}^2(1 - \hat{t}_{i4}^2) + e_{i4} \quad (18)$$

Abbiamo stimato la relazione (18) utilizzando il metodo dei minimi quadrati ordinari ottenendo per ciascun anno le stime del parametro  $\kappa_{4,des}$  riportate nella Tabella 4.

Infine, per i parametri  $f_{it4}$  nella (7), alla luce della (18), si assume  $f_{it4} = \frac{\hat{\kappa}_{t4,des}}{n_{it}^*}$ .

**Tabella 4 . Valori stimati del parametro  $K_{4,des}$  (Indice di Gini) e relativo  $R^2$**

Anno	Stima parametro	$R^2$
2008	0.153	0.762
2009	0.227	0.911
2010	0.183	0.780
2011	0.200	0.909
2012	0.164	0.770
2013	0.156	0.731
2014	0.156	0.784
2015	0.184	0.794
2016	0.157	0.746
2017	0.171	0.805
2018	0.146	0.793

### 2.2.3 Modello univariato per la media del logaritmo del reddito

In questa sezione presentiamo un modello per la media del logaritmo del reddito equivalente.

Al primo livello assumiamo che

$$\hat{t}_{it5} \sim N(\theta_{it5}, \delta_{it}^2) \quad (19)$$

I parametri  $\delta_{it}^2$  sono assunti noti e uguali alla loro stima diretta.

Si assume inoltre che

$$\theta_{it5} = \alpha_{t5} + \mathbf{x}_{it5}^t \beta_5 + v_{it5}, \quad (20)$$

dove  $\mathbf{x}_{it5}$  è un vettore di variabili esplicative dell'area  $i$  al tempo  $t$ ,  $v_{it5}$  è un effetto casuale,  $\alpha_{t5}$  è un'intercetta specifica di anno. Il modello assume inoltre che il parametro di regressione  $\beta_5$  sia lo stesso per ogni anno. Per i parametri  $\alpha_{t5}$  e  $v_{it5}$  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 (4) e (5).

### 2.2.4 Un modello multivariato

Un modello multivariato, alternativo ai cinque modelli univariati specificati nei paragrafi precedenti, è stato specificato assumendo una distribuzione multivariata per gli effetti casuali; in particolare:

$$\mathbf{v}_{it\bullet} \sim NMV(\mathbf{0}, \Sigma) \quad (21)$$

$$\Sigma | a_1, \dots, a_5 \sim Inv - Wishart(v + 1, 2v \text{diag}(a_1^{-1}, \dots, a_5^{-1})) \quad (22)$$

$$a_k \sim Inv - Gamma(.5, 1/A_k)$$

Questa scelta a priori comporta che marginalmente  $v\sigma_k \sim half - t(v, A_k)$ , un'ipotesi analoga a quella vista per i modelli univariati. Inoltre, per  $v = 2$  essa induce marginalmente una distribuzione a priori uniforme sulle correlazioni tra effetti casuali.

### 2.2.5 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:  $\beta_k$  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  $\sigma_{\alpha k}^2$  delle distribuzioni condizionate delle componenti dei vettori delle intercette di tempo abbiamo proceduto alla specificazione a priori di *Inv - Gamma*(0.1,0.1). 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) e dopo opportuna analisi di sensitività, sia nel caso di modelli univariati che nel caso del modello univariato è stato scelto  $\nu_k = 2$  e  $A_k = 1$

### 2.3. Selezione variabili ausiliarie

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

- l'equazione (20) relativa al valore atteso dei log-redatti equivalenti;
- l'equazione (8) relativa al valore atteso del coefficiente di concentrazione di Gini;
- l'equazione (3) relativa al valore atteso dei tre tassi.

Alcune delle variabili ausiliarie descritte in Tabella 1 sono fortemente correlate tra loro, rappresentando informazione ridondante. È 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 variabili ausiliarie (regressori) si è passati ad un gruppo più ristretto di 11 eliminando, tra quelle tra loro fortemente correlate, le variabili ausiliarie che presentavano minore correlazione con le stime dirette dei parametri di interesse. Nella tabella seguente sono riportate le variabili selezionate:

**Tabella 5. 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
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_f	Logit del tasso di povertà	fiscale
logit_gini_f	Logit dell'indice di Gini	fiscale
logit_rich_f	Logit del tasso di dichiaranti con reddito superiore al doppio della mediana nazionale	Fiscale
Logit_hrm_f	Logit del tasso di dichiaranti con reddito inferiore al reddito mediano nazionale	fiscale

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 simili a quelli descritti nelle sezioni 2.2.1-2.2.4, tranne che per la distribuzione a priori sui parametri di regressione e per il fatto che sono stati stimati modelli separati per ogni anno d'indagine.

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

$$\beta_{ik} \sim N(0, \sigma_{ik}^2)$$

$$\sigma_{ik}^2 = 0.001(1 - \lambda_{ik}) + c_k \lambda_{ik}$$

$$\lambda_{ik} \sim \text{Bernoulli}(.5)$$

La media a posteriori del parametro  $\lambda_{ik}$  è un indicatore dell'importanza della covariata  $i$ -esima per il  $k$ -esimo parametro di interesse. Il metodo di selezione delle covariate indotto dalla prior (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  $c_k = 10$ . Per ogni parametro di interesse sono state incluse nei modelli tutte le covariate per le quali la media sugli anni delle medie a posteriori di  $\lambda_{ik}$  sia risultata maggiore di 0.2.

Di seguito sono riportate per ognuno dei parametri di interesse le covariate inserite nei modelli.

**Tabella 6. Media a posteriori del parametro  $\lambda_i$ , modello per il tasso di povertà**

Covariata	hr	Gini	hr mediano	affluence	media dei log	deprivaz. severa
f_dich	x	x	X	x	x	x
dens						
dip_giov						x
dip_sen	x	x	X			x
freq_stra						x
masc_stra						x
medialog_f					x	
logit_hr_f	x			x		x
logit_gini_f		x		x		
logit_rich_f		x	X	x		
Logit_hrm_f			X			

## 2.4. 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 10, prendendo cioè un valore ogni dieci generati.

## 2.5. La selezione dei modelli

Alcune delle ipotesi alla base dei modelli presentati nelle sezioni 2.2.1-2.2.4 sono il frutto di un'accurata procedura di selezione in cui alcune delle ipotesi centrali sono state sottoposte alla verifica dei dati.

Sono state considerate tre specificazioni alternative dei modelli:

- Modello 1: insieme dei 5 modelli univariati delle sezioni 2.2.1, 2.2.2, 2.2.3.
- Modello 2: modello multivariato (si veda par. 2.2.4).
- Modello 3: come modello 2 ma assumendo che i parametri di regressione siano diversi per ogni anno di indagine.

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:

**Tabella 7. *Deviance Information criterion* per i tre modelli considerati**

	DIC
Modello 1	-16240
Modello 2	-18490
Modello 3	-18390

Viene selezionato il Modello 2.

Al fine di ottenere stime per i 38 distretti della regione Emilia-Romagna il modello 2 è stato stimato su un insieme di 139 aree costituito dall'unione delle province italiane, escluse quelle dell'Emilia-Romagna, e dei distretti dell'Emilia-Romagna.

## 2.6. La stima del 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. 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 su un'ipotesi sulla distribuzione del reddito equivalente nelle aree di studio. È 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.

La metodologia utilizzata per la stima del *Gap* mediano è illustrata in Fabrizi, Ferrante and Trivisano (2019). In questo lavoro viene presentata un approccio che, a partire da un'ipotesi distributiva sul reddito equivalente nelle aree, utilizza le catene MCMC generate dalle distribuzioni a posteriori dei parametri  $\theta_{itk}$  per generare delle catene MCMC per il *Gap* mediano. Mediante un ampio studio di simulazione nello stesso lavoro viene dimostrato che:

1. la media della catena MCMC generata per il *Gap* è uno stimatore con buone proprietà frequentiste (approssimativamente non distorto e con MSE comparabile con quello degli stimatori bayesiani dei parametri  $\theta_{itk}$ );
2. l'ipotesi distributiva di tipo Dagum è per il caso italiano preferibile ad altre ipotesi (in particolare alla distribuzione Beta prime, e quella di Singh-Maddala, che, insieme alla Dagum, rappresentano tre casi particolari della distribuzione Beta generalizzata del secondo tipo).

Nel seguito si descrive sinteticamente la metodologia. Per semplicità di esposizione tratteremo il problema con riferimento ad un solo anno di calendario). Sia  $M$  la mediana nazionale del reddito equivalente.

Sotto l'ipotesi distributiva di tipo Dagum, i cinque parametri su cui si basano i modelli descritti nei paragrafi precedenti possono essere espressi come segue:

$$Dag\theta_1 = F(0.6M, a, b, p) \quad (23)$$

$$Dag\theta_2 = F(M, a, b, p) \quad (24)$$

$$Dag\theta_3 = 1 - F(2M, a, b, p) \quad (25)$$

$$Dag\theta_{i4} = \frac{\Gamma(p)\Gamma(2p+1/a)}{\Gamma(2p)\Gamma(p+1/a)} \quad (26)$$

$$Dag\theta_5 = \frac{\{\psi(p)-\psi(1)\}}{a} + \log(b) \quad (27)$$

dove  $a, b, p$  sono i parametri che indicizzano la distribuzione,  $F$  è la funzione di ripartizione,  $\Gamma$  rappresenta la funzione gamma,  $\psi$  rappresenta la funzione digamma.

Inoltre, sotto l'ipotesi distributiva Dagum il *Gap* mediano può essere espresso come:

$$1 - \frac{F^{-1}(Dag\theta_1/2, a, b, p)}{F^{-1}(Dag\theta_1, a, b, p)} \quad (27)$$

Sia ora  $MC\theta_{sik}$  il valore  $s$ -esimo,  $s = 1, \dots, S$ , della catena MCMC generata dalla distribuzione a posteriori del parametro  $\theta_{ik}$  relativo all'area  $i$ -esima,  $k = 1, 2, \dots, 5$ .

Per ogni  $i, s$  viene minimizzata rispetto ai parametri  $a, b, p$  la funzione:

$$\sum_{k=1}^5 \left( \frac{MC\theta_{sik} - Dag\theta_{sik}}{MC\theta_{sik}} \right)^2 \quad (28)$$

I valori  $\hat{a}_{is}, \hat{b}_{is}, \hat{p}_{is}$  che minimizzano tale funzione indicizzano una distribuzione di Dagum con parametri  $Dag\hat{\theta}_{sik}$  approssimativamente uguali a quelli relativi ai valori  $s$  della catena MCMC ( $MC\theta_{sik}$ ).

Sulla base della relazione (27), viene generata una catena MCMC per il *Gap* mediano. Il valore  $s$ -esimo di tale catena per l'area  $i$  sarà dato da:

$$1 - \frac{F^{-1}(\widehat{Dag\theta_{1is}}/2, \hat{a}_{is}, \hat{b}_{is}, \hat{p}_{is})}{F^{-1}(\widehat{Dag\theta_{1is}}, \hat{a}_{is}, \hat{b}_{is}, \hat{p}_{is})}$$

La media di tali catene MCMC è uno stimatore bayesiano del *Gap* mediano. Tale stimatore non è funzione delle stime dirette del *Gap* mediano. Esso è funzione, attraverso una trasformazione basata

sull'ipotesi distributiva Dagum del reddito equivalente e il modello SAE utilizzato, delle stime dirette dei cinque parametri su cui si basano i modelli descritti nei paragrafi precedenti. La modellazione congiunta di più parametri, oltre ai due parametri target (tasso di povertà e indice di Gini) ha due effetti positivi: in un modello multivariato la struttura di somiglianza tra i cinque parametri considerati migliora in termini di efficienza la stima dei due parametri target; l'utilizzo di più parametri che descrivono aspetti diversi della distribuzione del reddito in un'area rende più precisa la scelta di una particolare distribuzione teorica utilizzata per generare il *Gap* mediano.

Si noti che con un procedimento simile è possibile stimare altri parametri per i quali non si hanno stime dirette o per i quali si ha poca informazione (stime dirette imprecise e assenza di covariate). Ad esempio, il quintile share ratio può essere stimato prendendo la media a posteriori di una catena MCMC il cui generico valore  $s$ , sotto l'ipotesi Dagum, è espresso come

$$\frac{1-F_{(1)}(x_{80}, \hat{a}_{is}, \hat{b}_{is}, \hat{p}_{is})}{F_{(1)}(x_{20}, \hat{a}_{is}, \hat{b}_{is}, \hat{p}_{is})},$$

dove  $F_{(1)}^{-1}(x_{80}, \dots)$  è il momento incompleto di ordine 1 di una distribuzione troncata al 80° percentile. In modo analogo è definito  $F_{(1)}^{-1}(x_{20}, \dots)$ .

## 2.7. Benchmarking

Al fine di ottenere stime per provincia e per distretto con un grado elevato di coerenza, si è proceduto a sottoporre a *benchmarking* la stima di alcuni parametri.

In particolare, è stata utilizzata una procedura di *benchmarking* per i parametri il cui valore ad un livello di aggregazione superiore (ad esempio la regione) a quello considerato (ad esempio le provincie) può essere ottenuto come opportuna media aritmetica delle stime al livello meno aggregato (ciò vale per i tre tassi e la media dei logaritmi).

Anziché sottoporre a *benchmarking* le medie a posteriori dei parametri di interesse, si è preferito agire al livello di singolo valore della catena MCMC per due motivi principali:

1) la generazione di una catena MCMC consente di ottenere facilmente una valutazione del livello di precisione della stima *benchmarked*;

2) a partire dalle catene MCMC sottoposte a vincolo di *benchmarking* è possibile derivare stime del *Gap* mediano con un maggiore livello di coerenza secondo la procedura descritta nel paragrafo (2.6).

Consideriamo dapprima il caso delle province.

Sia  $B\theta_{sitk}$  il valore  $s$ -esimo,  $s = 1, \dots, S$ , della catena MCMC del parametro  $\theta_{itk}$  sottoposta a vincolo di *benchmarking* per la provincia  $i$ -esima e l'anno  $t$ ;  $k = 1,2,3,5$ .

Il vincolo di *benchmarking* assicura che la somma pesata di queste quantità riproduca la stima diretta ottenuta per l'intera regione alla quale appartiene il dominio (provincia). Ovvero:

$$\sum_{i:i \in R} w_i B\theta_{sitk} =_R \hat{t}_{itk} \quad (29)$$

dove  $R\hat{t}_{itk}$  è la stima diretta del parametro  $k$ -esimo nell'anno  $t$  nella regione  $R$  cui appartiene la provincia  $i$ . Il sistema dei pesi nella (29) dev'essere tale che

$$w_i > 0, \sum_{i:i \in R} w_i = 1.$$

In questa applicazione essi sono stati scelti pari a

$$w_i = \frac{P_i}{\sum_{i:i \in R} P_i},$$

dove  $P_i$  è la popolazione anagrafica della provincia  $i$ . Esistono impostazioni alternative per ottenere stimatori soddisfino un vincolo del tipo (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.

Nel caso dei distretti si è proceduto in modo analogo. Questa volta però il vincolo (29) è espresso come segue:

$$\sum_{i:i \in P} w_i B\theta_{sitk} =_{P,B} \hat{\theta}_{itk},$$

dove  $B\theta_{sitk}$  rappresenta il valore  $s$ -esimo,  $s = 1, \dots, S$ , della catena MCMC del parametro  $\theta_{itk}$  sottoposta a vincolo di *benchmarking* per il distretto  $i$ -esimo e l'anno  $t$ ;  $k = 1,2,3,5$

$$P, B\hat{\theta}_{itk}$$

è la media a posteriori sottoposta a vincolo di *benchmarking* del  $k$ -esimo nell'anno  $t$  per la provincia  $P$  cui appartiene il dominio  $i$ .

La procedura è dunque gerarchica: dapprima (*step 1*) vengono sottoposte a *benchmarking* le stime provinciali; il vincolo è costituito dalla stima diretta della regione cui le province appartengono. In seguito (*step 2*) vengono sottoposte a *benchmarking* le stime dei distretti; in questo secondo caso il vincolo è costituito dalla stima provinciale (in questo caso è una stima da modello) ottenuta allo *step 1*.

**Tabella 8. Valori stimati del parametro  $\kappa_{t,des}$  (tasso di deprivazione materiale severa) e relativo  $R^2$**

Anno	Stima parametro	$R^2$
2008	0.285	0.676
2009	0.249	0.567
2010	0.268	0.648
2011	0.224	0.675
2012	0.196	0.659
2013	0.202	0.641
2014	0.246	0.794
2015	0.241	0.702
2016	0.234	0.736
2017	0.234	0.730
2018	0.267	0.577

## APPENDICE 2. Il disegno campionario

Il campione annuale EU-SILC per l'Italia è composto dall'unione di 4 campioni longitudinali. Ogni anno il 25% delle famiglie campionate esce dalla rilevazione e un numero corrispondente viene campionato dalla popolazione per rimpiazzarle. Il piano di campionamento può quindi essere descritto come un panel ruotato, secondo uno schema che possiamo riassumere in questo modo:

	<i>wave x</i>	<i>wave x+1</i>	<i>wave x+2</i>	<i>wave x+3</i>	<i>wave x+4</i>
<i>wave x</i>	ABCD				
<i>wave x+1</i>		BCDE			
<i>wave x+2</i>			CDEF		
<i>wave x+3</i>				DEFG	
<i>wave x+4</i>					EFGH

Il campione longitudinale delle famiglie "nuove" è selezionato secondo un piano di campionamento stratificato a due stadi, 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. Le soglie di dimensione demografica che definiscono gli strati variano da regione a regione (per maggiori dettagli, si veda Istat, 2008). Le unità di secondo stadio, le famiglie, sono campionate in modo casuale dalle liste anagrafiche.

Nel file dei dati relativi ad EU-SILC rilasciato da Istat sono pubblicati i coefficienti di riporto all'universo (pesi campionari).

I pesi campionari sono costruiti a partire 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. Tuttavia, non solo le probabilità di inclusione non sono costanti, ma la probabilità di effettiva inclusione nel campione è

modificata per effetto della non risposta e dell'attrito. Per questa ragione i pesi "base" ottenuti come inverso delle probabilità di inclusione vengono calibrati in modo tale che le stime del totale per variabili il cui totale a livello di popolazione è noto riproducano esattamente questo valore.

I pesi rilasciati da Istat per IT-SILC 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. Un secondo passo di calibrazione viene poi effettuato con l'obiettivo di migliorare l'efficienza degli stimatori.

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 Cutillo, 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 840 celle è stato valutato il numero di rispondenti effettivo  $m_z^r$  e il numero di rispondenti teorico,  $m_z$  ( $z = 1, \dots, 840$ ), 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 collapsamento 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”:

$$w_i = \frac{1}{\pi_i} \frac{m_z}{m_z^r}, i \in s$$

dove  $s$  indica l’intero campione trasversale.

Riguardo alla calibrazione a totali noti, i pesi campionari 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.

I pesi calibrati ottenuti da Istat secondo la procedura ora illustrata sono pensati per la produzione di stime affidabili a livello regionale. La somma dei pesi, che possiamo leggere come coefficienti di riporto all’universo, all’interno del campione di una certa regione, riproduce la dimensione della popolazione della regione stessa, ottenuta dalle fonti anagrafiche. Questa identità non è più

verificata se si considerano sottocampioni più piccoli, quali quelli specifici di provincia e di distretto sanitario. La calibrazione adottata impone infatti che le stime regionali delle variabili rispetto a cui si è calibrato (classi d'età, sesso) riproducano esattamente i totali regionali noti (da fonte anagrafica); ma non garantiscano che questi totali siano rispettati a livello di sottopopolazioni. Questa caratteristica dei pesi campionari ha un ruolo importante nella scelta degli stimatori diretti per le quantità di interesse.

Notiamo che, 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).

## Riferimenti bibliografici

- Aitchison, J., Brown, J.A.C. (1963) The Lognormal distribution with special reference to its uses in economics. Cambridge University Press.
- 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.
- Biewen M., Jenkins S.P. (2006), Variance estimation for generalized entropy and Atkinson inequality indices: the complex survey data case, Oxford Bulletin of Economics and Statistics, 68, 371-383.
- Datta G.S., Ghosh M., Steorts, R., Maples J. (2011). Bayesian *benchmarking* with applications to small area estimation, TEST, 20, 574-588.
- Fabrizi E., Ferrante M.R., Trivisano C. (2019). Small area estimation of the relative median poverty *Gap*, sottoposto per pubblicazione alla rivista Journal of the Royal Statistical Society, series A, *in fase di seconda revisione*.
- Fabrizi E., Ferrante M.R. Pacei S. (2015), Stima della serie temporale 2007-2012 dei parametri di povertà e deprivazione per i distretti sociosanitari e per le province della regione Emilia-Romagna, Rapporto prodotto per la Regione Emilia-Romagna.
- Fabrizi E., Trivisano C. (2016) Small area estimation of the Gini concentration coefficient, Computational Statistics & Data Analysis; (99): 223-234.
- Fruhworth-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.
- Goedemé T. (2013), How much confidence can we have in EU-SILC? Complex sample designs and the standard Error of the Europe 2020 poverty indicators, *Social Indicators Research*, 110, 89-110.
- George E. I., McCulloch, R. E. (1993). Variable selection via Gibbs sampling." J. Am. Stat. Assoc., 85: 398-409.
- Istat (2018), Statistiche Report, Anno 2017, Condizioni di vita, reddito e carico fiscale delle famiglie, 6 dicembre 2018.

Lumley T. (2004) Analysis of complex survey samples. *Journal of Statistical Software* 9(1): 1-19

Lumley T. (2017) survey: analysis of complex survey samples. R package version 3.32.

Pessoa D., Damico A. and Jacob G. (2018). convey: Income Concentration Analysis with complex survey samples. R package version 0.2.1.

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.