A functional approach to small area estimation of the relative median poverty gap

Published Version:
Fabrizi, E., Ferrante, M., Trivisano, C. (2020). A functional approach to small area estimation of the relative median poverty gap. JOURNAL OF THE ROYAL STATISTICAL SOCIETY. SERIES A. STATISTICS IN SOCIETY, 183(3 (June)), 1273-1291 [10.1111/rssa.12562].

Availability:
This version is available at: https://hdl.handle.net/11585/762944 since: 2020-06-24

Published:
DOI: http://doi.org/10.1111/rssa.12562

A functional approach to small area estimation of the relative median poverty gap

Enrico Fabrizi*
DISES, Università Cattolica del S. Cuore, Piacenza, Italy

Maria Rosaria Ferrante,
Dipartimento di Scienze Economiche, Università di Bologna, Italy

Carlo Trivisano,
Dipartimento di Scienze Statistiche 'P. Fortunati',
Università di Bologna, Italy

January 11, 2020

Abstract

We consider the estimation of the relative median poverty gap (RMPG) at the level of Italian provinces using data from the European Survey on Income and Living Conditions. The overall sample size does not allow reliable estimation of income distribution related parameters at the provincial level; therefore, small area estimation techniques has to be used. The specific challenge in estimating the RMPG is that, as it summarizes the income distribution of the poor, samples for estimating it for small subpopulations are even smaller than those available in other parameters. We propose a Bayesian strategy where various parameters summarizing the distribution of income at the provincial level are modelled by means of a multivariate small area model. To estimate the RMPG, we relate these parameters to a distribution describing income and namely the Generalized Beta of the second kind (GB2). Posterior draws from the multivariate model are then used to generate draws for the GB2 area-specific parameters and then of the RMPG defined as their functional.

Keywords: GB2 distribution; hierarchical Bayes; income inequality; poverty; complex sample surveys.

1 Introduction

The relative median at-risk-of-poverty gap is one of the indicators endorsed by the European Union for the assessment of social cohesion (European Commis-

*Corresponding author's address: Enrico Fabrizi, DISES, Università Cattolica del S. Cuore, Via Emilia Parmense 84, 29122 Piacenza, Italy, e-mail:enrico.fabrizi@unicatt.it

sion, 2004). It is defined as the median distance of the individual poor equivalized income from a threshold defined as the 60% of national median, relative to this threshold. The relative median at-risk-of-poverty gap (from now on, RMPG) is an important complement to the information provided by the head-count ratio measure of poverty (at-risk-of-poverty rate) as it offers an insight on how deep is the poverty experienced by the median poor, regardless of how many live below the poverty line.

At-risk-of-poverty rates, RMPGs, as well as many other poverty and income inequality measures are annually calculated by EUROSTAT for most EU-member states using data from the European Survey on Income and Living Conditions (EU-SILC), conducted under harmonized guidelines (see Atkinson and Marlier, 2010, for a general introduction). Estimates of these parameters are published also for large regions or social groups within countries. This paper is about estimating RPMG in small areas, that is for a collection of population subsets ('areas') for which the subset-specific sample sizes are not large enough to obtain decent precision from ordinary survey-weighted estimators (that are labelled as *direct estimators* in the small area literature).

We note that the problem of sample sizes *not large enough* is more severe for the RMPG than for other summaries of the income distribution as it is a (scaled) quantile of the poor income distribution whose direct estimation is based only on those who are poor, usually a minority of the sample units. For instance, if the prevalence of the poor ranges from 5% to 33% the expected area-specific sample sizes available to estimate the sample mean will be from 3 to 20 times larger than those available for the estimation of the RMPG.

Specifically, we consider the problem of estimating the RMPG for Italian administrative provinces using data from the Italian section of the EU-SILC survey. In Italy there are 110 provinces corresponding to the NUTS 3 level according to Eurostat nomenclature of territorial units for statistics (Eurostat, 2019). Provincial administrations play an important role in implementing policies decided at higher levels (national or regional) and in co-ordinating the activities of lower administrative levels (municipalities and health districts). We consider data from the 2013 wave of the EU-SILC survey and auxiliary information known at the provincial level obtained from various sources, including fiscal archives of the Italian Ministry of finance and population registers.

Small area estimation is about complementing the insufficient information provided by area-specific samples with auxiliary information known from external sources (Censuses, administrative archives,...). The complementing is typically achieved by using models that can be specified at either the area or the unit level (Pfeffermann, 2013).

In this paper we consider area-level models (Rao and Molina, 2015, chapter 5). These models are less demanding in terms of required information as only direct estimates, associated measures of uncertainty and summaries at the area level of the auxiliary variables are needed. They can represent the only viable strategy to the secondary data analysis that does not have access to the details of the sampling design and relevant unit-level information. Moreover, some typical problems met when using unit-level models, such as possible inconsistencies in definitions and measurement techniques for auxiliary variables between the sample survey and the auxiliary source, are sidestepped. See Tarozzi and Deaton (2009) and Tzavidis et al. (2018) for more general discussions of these topics. In our application, we have limited access to some information on the sampling design and dispose only of area-level summary statistics for the auxiliary information we consider in the models.

As it relies on area-level models, this research is different from previous literature on small area estimation of the RMPG (Molina and Rao, 2010; Molina et al., 2014) that focuses on unit-level modelling.

The inputs of an effective area-level model are: *i*) a set of area-level approximately unbiased estimates endowed with reliable sampling variability measures; *ii*) a vector of area-level auxiliary information with good predictive power for the parameter in question. If we denote η_d the RPMG in area d, $\hat{\eta}_d$ its direct estimate, $\mathbf{x}_d$ a vector of area level auxiliary information, a typical area level model is not a viable strategy as direct estimators of RPMG are biased (as the median is) and very imprecise in small samples (see results in Appendix 1); moreover auxiliary variables with good predictive power are difficult to find for η_d.

Our alternative strategy can be summarized as follows. We consider $\boldsymbol{\theta}_d$, a vector of additional small area parameters for which approximately unbiased direct estimators and predictive auxiliary information is available. As they are not of direct interest, we label $\boldsymbol{\theta}_d$ as *nuisance* small area parameters. We specify a small area model for $\boldsymbol{\theta}_d$. The components of $\boldsymbol{\theta}_d$ can be related functionally to each other via $\boldsymbol{\xi}_d$, a vector of parameters characterizing a distribution we assume for income in area d, so that $\boldsymbol{\theta}_d = \boldsymbol{\theta}(\boldsymbol{\xi}_d)$. The solution in $\boldsymbol{\xi}_d$ of this system of equations can then be used to functionally estimate $\eta_d = \eta(\boldsymbol{\xi}_d)$ under the distribution assumed to describe income.

A few technical comments are in order: *i*) we consider five *nuisance* small area parameters θ_{kd} so that $\boldsymbol{\theta}_d = \{\theta_{kd}\}$, $k = 1, \ldots, 5$; they include three head-count ratios based on different thresholds, a concentration index and the mean of the log-income; their choice is aimed at providing a description of the whole income distribution at the area level. More details will be given in section 2.2; *ii*) we specify a multivariate small area model for $\boldsymbol{\theta}_d$. Multivariate models have a long tradition in small area estimation dating back at least to Ghosh et al. (1996) and they usually lead to more efficient estimators as they exploit the correlation between parameters; *iii*) the parametric distribution we consider for income is the GB2 (Generalized Beta of the second kind McDonald, 1984) that is widely used in the literature. We also consider three distribution that are special cases of the GB2 distribution (Dagum, Singh-Maddala, Beta of the second kind) that depend on three parameters. The recourse to these special cases is motivated by computational sustainability; more details on this point will be given in sections 4.2 and 5; *iv*) the number of *nuisance* parameters is larger than the size of $\boldsymbol{\xi}$ characterizing the GB2 distribution: this entails a solution of the system $\boldsymbol{\theta}_d = \boldsymbol{\theta}(\boldsymbol{\xi}_d)$ based on the minimization of a loss function that allows more flexible and numerically stable solutions.

The core of this methodology, that is the estimation of $\boldsymbol{\xi}$ by solving $\boldsymbol{\theta}_d = \boldsymbol{\theta}(\boldsymbol{\xi}_d)$, was introduced in Graf and Nedyalkova (2014). Here we apply it to a small area estimation problem in the framework of a hierarchical Bayesian model. Specifically, we approximate posterior distributions of $\boldsymbol{\theta}_d$ by means of Markov Chains Monte Carlo (MCMC) algorithms. By solving $\boldsymbol{\theta}_d = \boldsymbol{\theta}(\boldsymbol{\xi}_d)$ for each MCMC draw we obtain Markov chains for the parameters characterizing the assumed income distribution at the area level. The $\eta_d = \eta(\boldsymbol{\xi}_d)$ can be exploited to generate a Markov Chain converging to the posterior of the target parameter η_d.

Predictors of nuisance parameters are design-consistent (see section 3), i.e. their point predictors converge to area-specific population descriptive quantities regardless of misspecifications of the multivariate model. Asymptotically the estimator of η_d converges to the functional of these population quantities that depends on the assumption of GB2 distributed income in the area. As a consequence, the dependence on the assumption of these distribution remains, but the estimator is robust with respect to misspecifications of the multivariate small area model.

The rest of the paper is organized as follows. Section 2 introduces the data set we consider in this application and direct estimation of the *small area parameters* involved in the study. In section 3 we introduce the multivariate small area estimation model that provides the basis for the estimation of the RPMG. Section 4 includes a short review of the Generalized Beta of the second kind distribution and its special cases and the illustration of our functional estimation methodology. The estimation of RMPG at the level of Italian provinces is illustrated in section 5, with some discussion. As the method is rather complex, we explore the frequentist properties of the proposed estimators by means of a simulation exercise, based on the same sample data (section 6). Concluding remarks are provided in section 7.

2 The data and direct estimation of small area parameters

2.1 The data

We analyze data from the 2013 wave of the EU-SILC. The survey is conducted in many countries across the European Union by the relevant National Institutes of Statistics using harmonized questionnaires and survey methodologies. Although following common guidelines, sampling designs can differ from country to country. In Italy, the EU-SILC is a rotating panel survey with 75% overlap of samples in successive years. The fresh part of the sample is drawn according to a stratified two-stage sample design, where municipalities (LAU 2 level, see Eurostat, 2019) are the primary sampling units (PSUs), while households are the secondary sampling units (SSUs). The PSUs are divided into strata according to their population size and the SSUs are selected by systematic sampling in each PSU.

We target administrative provinces. The 110 Italian provinces have largely different populations ranging from the 4.3 million inhabitants of Rome, down to less 0.1 million (Medio Campidano, Isernia, Ogliastra). Provinces are unplanned domains for the EU-SILC survey. For the 2013 wave that we consider in this article, province-specific sample sizes range from 6 up to 882 in terms of households and from 10 to 2018 in terms of individuals. The median province-specific sample size is 115 households (274 individuals).

2.2 Direct estimation

Let's consider a population P of size N and a partition of it into D small areas $\{P_1, \ldots, P_d, \ldots, P_D\}$ of size N_d, $\sum_{d=1}^{D} N_d = N$. A sample of overall size n is drawn from the population according to a complex design such as the stratified multi-stage design with a rotating panel component used in EU-SILC.

Area-specific samples sizes are denoted n_d so that $\sum_{d=1}^{D} n_d = n$. A survey weight w_{dj} is associated to each unit in the sample ($j = 1, \ldots, n_d$, $d = 1, \ldots, D$) reflecting both inclusion probabilities and non-response corrections. We target a variable y, the equivalized disposable income, defined as the total disposable household income divided by the equivalized household size calculated according to the modified OECD scale (see Fusco et al., 2010).

Although our ultimate focus is the estimation of the RPMG, we consider several population descriptive quantities at the area level that we label *small area parameters*. To avoid confusion, we denote the RPMG at the area level with η_d and the vector of *nuisance* small area parameters as $\boldsymbol{\theta}_d = \{\theta_{kd}\}$ with $k = 1, \ldots, 5$. Whenever $n_d > 0$ these parameters can be estimated using area-specific samples using Hàjek type (Hàjek, 1958) or other design based estimators that we can assume approximately unbiased. We label these estimators as direct and denote them $\hat{\eta}_d, \hat{\theta}_{kd}$.

The RMPG is defined as $\eta = \{pt_1 - Me_p(y)\}/pt_1$, where $Me_p(y)$ is the median income of the poor, i.e. $Me_p(y) = Me(y|y \leq pt_1)$ and pt_1 is the national poverty threshold, defined, in the EU-SILC framework as 60% of the national median of equivalized income. A survey weighted estimator of η_d is given by

$$\hat{\eta}_d = \frac{pt_1 - \widehat{mp}_d}{pt_1} \tag{1}$$

where

$$\widehat{mp}_d = \begin{cases} \frac{1}{2}(y_{(jd)} + y_{(j+1,d)}) & \texttt{if } \sum_{i=1}^{j} w_{(i)} = 0.5 \sum_{i=1}^{n_{dp}} w_{(i)} \\ \\ y_{(j+1,d)} & \texttt{if } \sum_{i=1}^{j} w_{(i)} < 0.5 \sum_{i=1}^{n_{dp}} w_{(i)} < \sum_{i=1}^{j+1} w_{(i)} \end{cases}$$

$n_{dp} \leq n_d$, is the number of poor in the sample specific to domain d, $y_{(i)} \leq y_{(i+1)}$, $i = 1, \ldots, n_{dp}$ is the non decreasing sequence of poor incomes. $\hat{\eta}_d$ is likely to be more imprecise than $\hat{\theta}_{kd}$ as it based only on the income of those below pt_1 in the sample, typically a minority. Moreover, in very small samples it can be substantially biased. A small design-based simulation exercise, based on EU-

SILC data and reported in Appendix 1, explores the size of bias and variance of this estimator in small samples.

The *nuisance* parameters we consider in this application are: i) the at-risk-of-poverty rate, $\theta_1 = E\{\mathbf{1}(y \leq pt_1)\}$, a poverty count based on the threshold pt_1 and that represent the most popular poverty measure in the EU; ii) the proportion of people living with an equivalized income below the national median: $\theta_2 = E\{\mathbf{1}(y \leq Me(y))\}$; ii) an affluence rate defined as the proportion of individuals for which $y > pt_3$ where pt_3 is some high threshold, that we fix at twice the national sample median in line with Peichl et al. (2010): $\theta_3 = E\{\mathbf{1}(y > pt_3)\}$. Affluence rates are useful to describe the right tail of the y distribution at the area level; iv) the Gini concentration index, can be defined as $\theta_4 = \Delta(2E(y))^{-1}$ where $\Delta = E\{|y_s - y_t|\}$ with y_s, y_t identically distributed as y; v) the mean of the log-income, i.e. $\theta_5 = E\{log(y)\}$.

We now present direct estimators for the *nuisance* parameters θ_{kd}. For $k = 1, 2$ they can be written as:

$$\hat{\theta}_{kd} = \frac{\sum_{j=1}^{n_d} w_{dj} \mathbf{1}(y_{dj} < pt_k)}{\sum_{j=1}^{n_d} w_{dj}} \tag{2}$$

When $k = 1$, we have the at-risk-of-poverty rate while for $k = 2$ we define $pt_2 = Me(y)$, i.e. $pt_1 = 0.6pt_2$. We note that, when estimated at the whole population level $\hat{\theta}_{2.} = 0.5$; in specific domains it can be read as a departure of the local median from that of entire population. The direct estimator of θ_{3d} is defined as

$$\hat{\theta}_{3d} = \frac{\sum_{j=1}^{n_d} w_{dj} \mathbf{1}(y_{dj} > pt_3)}{\sum_{j=1}^{n_d} w_{dj}} \tag{3}$$

We note that pt_1, pt_2, and pt_3 rely on the estimated national median of the equivalized income. As this estimate is based on a very large national sample, we will overlook the uncertainty associated to these thresholds and threat them as fixed constants.

The most popular direct estimators θ_4, for instance the one considered in Alfons and Templ (2013), are biased in small samples. In line with Fabrizi and

Trivisano (2016) we consider a nearly unbiased direct estimator that accounts also for the fact that individuals in the same household share the same income:

$$\hat{\theta}_{4d} = \frac{1}{2\hat{\bar{Y}}_d} \frac{\sum_{j=1}^{n_d} \sum_{k=1}^{n_d} w_{dj} w_{dk} |y_{dj} - y_{dk}|}{\hat{N}_d^2 - \sum_{h=1}^{m_d} \tilde{w}_{dh}^2}. \tag{4}$$

where $\hat{\bar{Y}}_d = \hat{N}_d^{-1} \sum_{j=1}^{n_d} w_{dj} y_{dj}$, $\hat{N}_d = \sum_{j=1}^{n_d} w_{dj}$ is the Horwitz-Thompson estimator of the domain size; moreover, m_d is the number of households sampled in domain d and $\tilde{w}_{dh} = \sum_{j=1}^{n_h} w_{dj}$ is the sum of weights associated to the n_h individuals living in household h ($h = 1, \ldots, m_d$).

An approximately unbiased estimator of θ_5 can be defined as

$$\hat{\theta}_{5d} = \frac{\sum_{j=1}^{n_d} w_{dj} \log y_{dj}}{\sum_{j=1}^{n_d} w_{dj}} \tag{5}$$

The direct estimators $\hat{\theta}_{kd}$ are nearly unbiased but their variance can be large when n_d is small. In the case of of the EU-SILC survey, their variances will be larger than those we would have obtained with simple random samples of the same number of individuals. In the first place, the same equivalized income is shrared by all individuals in the same household (perfect intra-cluster correlation). Moreover, the design effect of the EU-SILC survey for Italy is larger than 1 even considering variables at the household level; although the design is stratified at the first stage, clustering of households within municipalities, unequal selection probabilities and weighting corrections to counter non response cause efficiency losses (see Clemenceau and Museux, 2007; Goedemé, 2013, for more details).

To estimate the variances of $\hat{\theta}_{kd}$ we consider a two steps approach: first a bootstrap algorithm, described in Fabrizi et al. (2011) is used to obtain preliminary variance estimates. These *raw* variances are then used to estimate design effects and other parameters of variance smoothing models that will be described in section 5. We note that the bootstrap algorithm does not incorporate all details of the EU-SILC sample design for Italy because of limited access to municipality level clustering and longitudinal tracking information; based on previous literature (see Goedemé, 2013; Biewen and Jenskins, 2006) we assume that once essential features of the designs are accounted for (stratification, clustering at the household level, unequal selection probabilities and weighting), good approximations to actual sampling variances can be obtained. As pointed out in Tzavidis et al. (2018), variance smoothing is a delicate step in building an area-level model, so special attention will be devoted to the assessment and fit quality of these smoothing models in section 5.

3 A multivariate small area model for parameters related to equivalized income distribution

In this section we describe a multivariate model for θ_{kd}, $k = 1, \ldots, 5$. In line with the typical specification of small area models, ours has two levels: i) a sampling model that provides a likelihood for the direct estimators and relates them to the underlying population parameters; ii) a linking model that relates the small area parameters to auxiliary information and to each other by means of exchangeable random effects according to the principle of *borrowing strength.*

The recourse to a multivariate model is motivated by the fact that the five parameters represent different aspects of the area-level distribution of the target variable y. The estimates $\hat{\theta}_{kd}$, represent summaries of the same area-specific samples, so it is natural to assume they are correlated, and to specify a multivariate sampling model. We do this by means of a gaussian copula function in line with Fabrizi et al. (2016). See Souza and Moura (2016) for other applications of copula functions in the small area context. We present the sampling model in two steps: first, we introduce the marginal sampling models, then the copula function is used to account for their dependence structure.

For the rates θ_{kd}, $k = 1, 2, 3$, in line with Fabrizi et al. (2016), we speficy a zero-inflated Beta sampling model to account for the fact that rates range in the $(0, 1)$ interval and that when m_d is small, the direct estimate can be zero, i.e. $\hat{\theta}_{kd} = 0$ even if it is assumed, as we do $\theta_{kd} > 0$:

$$
\begin{aligned}
f\big(\hat{\theta}_{kd}|\theta^{\star}_{kd}, \hat{\phi}_{kd}\big) &= \big(1-\theta^{\star}_{kd}\big)^{m_d}\mathbf{1}(\hat{\theta}_{kd}=0) \\
&+ \big\{1-\big(1-\theta^{\star}_{kd}\big)^{m_d}\big\} dBeta(A_{kd}, B_{kd})\mathbf{1}(\hat{\theta}_{kd}>0)
\end{aligned} \tag{6}
$$

where $A_{kd} = \theta^{\star}_{kd}\big(\hat{\phi}_{kd}-1\big)$, $B_{kd} = \big(1-\theta^{\star}_{kd}\big)\big(\hat{\phi}_{kd}-1\big)$. See Ospina and Ferrari (2012), Wieczorek and Hawala (2011) for alternative specifications of zero-inflated beta regression allowing also for $\theta_{kd} = 0$.

The quantities $\hat{\phi}_{kd}$ can be interpreted as an effective sample size in terms of individuals and are estimated using variance smoothing models. See section 5 for more details on these models and estimation leading to $\hat{\phi}_{kd}$. The parameter $\theta^{\star}_{kd}$ is defined as $\theta^{\star}_{kd} = E\big(\hat{\theta}_{kd}|\hat{\theta}_{kd} > 0, \theta_{kd}, \hat{\phi}_{kd}\big)$ so the parameter we are actually interested in is given by

$$
\theta_{kd} = \theta^{\star}_{kd}\big\{1-\big(1-\theta^{\star}_{kd}\big)^{m_d}\big\} = E\big(\hat{\theta}_{kd}|\theta^{\star}_{kd}, \hat{\phi}_{kd}\big)
$$

Note that in (6) we assume that $P(\hat{\theta}_{kd} = 0)$ depends explicitly on the underlying rate $\theta^{\star}_{kd}$ and the number m_d of households sampled from domain d.

The sampling model for the Gini concentration coefficient is based on a Beta likelihood, with a parameterization we take from Fabrizi and Trivisano (2016):

$$
\hat{\theta}_{4d} \sim Beta\left(\frac{2\hat{\phi}_{4d}}{1+\theta_{4d}} - \theta_{4d}, \frac{2\hat{\phi}_{4d} - \theta_{4d}(1+\theta_{4d})}{1+\theta_{4d}}\frac{1-\theta_{4d}}{\theta_{4d}}\right) \tag{7}
$$

As a consequence $E(\hat{\theta}_{4d}|\hat{\phi}_{4d}) = \theta_{4d}$, $V(\hat{\theta}_{4d}|\hat{\phi}_{4d}) = \theta^2_{4d}\big(1-\theta^2_{4d}\big)\big(2\hat{\phi}^{-1}_{4d}\big)$. See 5 for details on variance model used to obtain the quantities $\hat{\phi}_{4d}$, that will be treated as known.

The sampling model for the mean of the log-incomes $\hat{\theta}_{5d}$ is a normal Fay-Herriot model:

$$
\hat{\theta}_{5d} \sim N\Big(\theta_{5d}, \hat{\phi}^{-1}_{5d}\Big) \tag{8}
$$

Variances $\hat{\phi}_{5d}^{-1}$ are estimated using the bootstrap algorithm discussed in Fabrizi et al. (2016). The assumption of known variances for normal small area models is in line with most literature (see Rao and Molina, 2015, chapter 5). It is also consistent with (6) and (7) as we consider a two parameter distribution where one of the two parameters is assumed known.

The Gaussian copula (Clemen and Reilly, 1999) used to model the direct estimators' dependence structure is parametrized in terms of the correlation matrix $\mathbf{R}$ of a Gaussian multivariate distribution. In detail, we assume that:

$$f(\hat{\theta}_{1d}, \ldots, \hat{\theta}_{kd}) = \frac{g_1(\hat{\theta}_{1d}) \times \cdots \times g_k(\hat{\theta}_{kd})}{|\mathbf{R}|^{1/2}} = \exp\left\{ -\frac{1}{2}\mathbf{z}_k^T(\mathbf{R}^{-1} - \mathbf{I}_k)\mathbf{z}_k \right\} \quad (9)$$

with $\mathbf{z}_k^T = \big(\Phi^{-1}\{F_1(\hat{\theta}_{1d})\}, \ldots, \Phi^{-1}\{F_5(\hat{\theta}_{kd})\}\big)$; the marginal densities $f_k(\hat{\theta}_{kd})$, $k = 1, \ldots, 5$ are defined in (6)-(8) while $F_k(\hat{\theta}_{kd})$ are the associated cumulative distribution functions. The matrix $\mathbf{R}$ is to be estimated from the data. For the specific application we consider in this paper, the estimation procedure will be outlined in section 5.

The linking models for the three rates and the Gini coefficients are based on a logit link

$$logit(\theta_{kd}) = \mathbf{x}_{kd}^t \beta_k + v_{kd} \quad (10)$$

$(k = 1, \ldots, 4)$, while an identity link is considered for θ_{5d}:

$$\theta_{5d} = \mathbf{x}_{5d}^t \beta_k + v_{5d} \quad (11)$$

The vector $\mathbf{x}_{kd}$ contains for each parameter and each area auxiliary information known at the area level. Note that x_{kd} and β_k may vary with k; but the first element of x_{kd} is 1 in all cases.

The multivariate relationship among the population parameters θ_{kd} is incorporated in the distributional assumption for $\mathbf{v}_d = (v_{kd})$, $k = 1, \ldots, 5$:

$$\mathbf{v}_d \sim MVN\big(\mathbf{0}, \mathbf{\Sigma}_v\big) \quad (12)$$

where MVN denotes the multivariate normal distribution. For Σ_v we specify a prior within the family proposed by Huang and Wand (2013) with the purpose of keeping the analytical and computational tractability of the inverse Wishart but improving the non-informativity properties:

$$\begin{aligned} \mathbf{\Sigma}_v | a_1, \ldots, a_k &\sim \text{Inv-Wishart}\big(\nu + 1, 2\nu \text{diag}(a_1^{-1}, \ldots, a_k^{-1})\big) \\ a_k &\sim \text{Inv-Gamma}\Big(\frac{1}{2}, \frac{1}{A_k}\Big), k = 1, \ldots, 5. \end{aligned} \tag{13}$$

This prior marginally induces $\sigma_k \sim \text{half} - t(\nu, A_k)$. The choice $\nu = 2$ allows for a diffuse prior, close to the popular half-Cauchy ($\nu = 1$); moreover it induces a marginal uniform prior on the correlations between the random effects. We choose $A_k = 1$ after careful consideration of the scale of the parameters' distribution and some sensitivity analysis.

For all parameters the point predictor of the small area mean is obtained summarizing the posterior distribution of θ_{kd} using quadratic loss, so that $\tilde{\theta}_{kd} = E(\theta_{kd}|\mathbf{d})$, $k = 1, \ldots, 5$ and $\mathbf{d}$ where shortcut notation for the data.

It can be shown that conditionally on $\mathbf{\Sigma}_v$, $\tilde{\theta}_{kd}$, $k = 1, \ldots, 5$ is design consistent provided that $\hat{\theta}_{kd}$ are. For the definition of design consistency we refer to Fuller (2009), p. 41. For a proof of this design consistency property see Appendix 2.

4 The proposed estimation strategy for the relative median poverty gap

4.1 The generalized Beta of the second kind distribution and its special cases

The generalized beta distribution of the second kind (GB2; McDonald, 1984) is a four parameter distribution which is acknowledged as an excellent descriptor of income distributions (Dastrup et al., 2007; Jenkins, 2009; Graf and Nedyalkova, 2011). The GB2 density can be written as:

$$f(x; a, b, p, q) = \frac{a}{bB(p,q)} \frac{(x/b)^{ap-1}}{\big(1 + (x/b)^a\big)^{p+q}} \mathbf{1}(x > 0) \tag{14}$$

where $a, b, p, q > 0$ and $B(p, q)$ is the Beta function. With the exception of b which is a scale parameter, the other three parameters are all shape parameters: a can be interpreted as an overall shape parameter, p rules the right tale, while q the left one. For a general description of the properties of the GB2 distribution see Kleiber and Kotz (2003, chapter 6.1), Graf et al. (2011a).

In the economy of this study we are interested in the expression of the *small area parameters* $\eta_d, \boldsymbol{\theta}_d$ introduced in Section 2.2 when the equivalized income variable is assumed to be GB2 distributed. We use the notation $\theta_{kd|GB2}$, $\eta_{d|GB2}$ to the denote the expression of θ_{kd} under the GB2 assumption:

$$\begin{aligned}
\theta_{1d|GB2} &= F\big(pt_1, a_d, b_d, p_d, q_d\big) && (15)\\
\theta_{2d|GB2} &= F\big(pt_2, a_d, b_d, p_d, q_d\big) && (16)\\
\theta_{3d|GB2} &= 1 - F\big(pt_3, a_d, b_d, p_d, q_d\big) && (17)\\
\theta_{4d|GB2} &= \frac{B(2p_d + 1/a_d, 2q_d - 1/a_d)}{B(p_d + 1/a_d, 2q_d - 1/a_d)} && (18)\\
&\times \big\{p_d^{-1} G_1(a_d, p_d, q_d) + (p_d + 1/a_d)^{-1} G_2(a_d, p_d, q_d)\big\} && (19)\\
\theta_{5d|GB2} &= \frac{\big\{\psi(p_d) - \psi(q_d)\big\}}{a_d} + \log(b_d) && (20)\\
\eta_{d|GB2} &= 1 - \frac{F^{-1}\big(\theta_{1d|GB2}/2, a_d, b_d, p_d, q_d\big)}{F^{-1}\big(\theta_{1d|GB2}, a_d, b_d, p_d, q_d\big)} && (21)
\end{aligned}$$

Note that F in (15)-(17) is the cumulative distribution function while in (19) $G_1(.)$ and $G_2(.)$ are generalized hypergeometric series (see McDonald, 1984, for a detailed definition) depending on all the distribution parameters except the scale b_d while $\psi(.)$ in (20) is the di-gamma function.

The GB2 distribution encompasses several special cases. In this research we consider the Beta of the second kind (B2) distribution ($a = 1$) the Dagum distribution ($q = 1$) and the Singh-Maddala distribution ($p = 1$). For these special cases the expressions (15) - (21) are simpler and notably so for the Gini coefficient (19) that reduces to:

$$
\begin{aligned}
\theta_{4d|B2} &= \frac{B\big(2p_d, 2q_d - 1\big)}{2pB^2(p_d, q_d)} &\quad (22)\\
\theta_{4d|Dagum} &= \frac{\Gamma(p_d)\Gamma(2p_d + 1/a_d)}{\Gamma(2p_d)\Gamma(p_d + 1/a_d)} &\quad (23)\\
\theta_{4d|SM} &= 1 - \frac{\Gamma(q_d)\Gamma(2q_d - 1/a_d)}{\Gamma(2q_d)\Gamma(q_d - 1/a_d)} &\quad (24)
\end{aligned}
$$

where $\Gamma(.)$ is the Gamma function. The considered special cases of the GB2 are also those identified by McDonald et al. (2013) as the ones characterized by skewness-kurtosis spaces encompassing the largest portion of income data set in their cross-country analysis of the Luxembourg Income Study database. Kakamu (2016), using a simulation study based on data generated from GB2 distributions, characterizes parameters regions in which the fit of the Dagum distribution is superior to that of the SM distribution and vice-versa. Intuitively, data with a heavy right tail should be better fit by SM and those with a more moderate skewness by the Dagum distribution. Kleiber (1996) expects the Dagum distribution to fit better than the SM in most real data set; actually its skweness-kurtosis space includes that of the SM in the direction of more moderate and even negative skewness. The B2 distribution is considered especially for its popularity in the literature (Chotikapanich et al., 2012).

4.2 Indirect estimation of the RMPG

Let $\boldsymbol{\xi}_d = (a_d, b_d, p_d, q_d)$ denote the parameters of the GB2 distribution we assume to describe the income distribution in area d. As areas are many, this description would imply a very large set of parameters to be estimated; this cannot be done using area-specific samples, as they are typically small. We use the multivariate model to accomplish this task. Under this GB2 assumption:

$$\boldsymbol{\theta}_d = \boldsymbol{\theta}\left(\boldsymbol{\xi}_d\right)$$

according to formulas (15) - (20). Using the multivariate model of section 3 we can draw from $p(\boldsymbol{\theta}_d|\mathbf{d})$. For each draw $\boldsymbol{\theta}_{rd}$, $r = 1, \ldots, R$ we can solve $\boldsymbol{\theta}_{rd} = \boldsymbol{\theta}(\boldsymbol{\xi}_{rd})$ in $\boldsymbol{\xi}_{rd}$ thus obtaining a draw from $p(\boldsymbol{\xi}_d|\mathbf{d})$. We can then use

$$\eta_d = \eta(\boldsymbol{\xi}_d)$$

defined according to (21) to simulate from $p(\eta_d = \eta(\boldsymbol{\xi}_d)|\mathbf{d})$, by drawing $\eta_{rd} = \eta(\boldsymbol{\xi}_{rd})$.

Several technical details about the implementation of this approach now follow. We note that $p(\theta_{kd}|\mathbf{d})$ depends on the way we modelled the direct estimators $\hat{\theta}_{kd}$ but not on the GB2 we assume for the income distribution in the areas. If the size of $\boldsymbol{\theta}_d$ and $\boldsymbol{\xi}_d$ were the same, a solution to the system $\boldsymbol{\theta}_d = \boldsymbol{\theta}(\xi_d)$ can be slow or even impossible to find with numeric methods. In line with Graf and Nedyalkova (2014), section 5, we use a vector $\boldsymbol{\theta}_d$ of five elements to solve for the four parameters characterizing the GB2 distribution by minimizing a relative quadratic loss function:

$$L(\boldsymbol{\theta}_{rd}, \boldsymbol{\xi}_{rd}) = \sum_{k=1}^{5} \left\{ \frac{\theta_{krd} - \theta_{krd|GB2}(\boldsymbol{\xi}_{rd})}{\theta_{krd}} \right\}^2 \tag{25}$$

With respect to Graf and Nedyalkova (2014) we select a different set of *nuisance* parameters and namely the θ_{kd}, $k = 1, \ldots, 5$ discussed in section 3. Except for θ_{5d} all parameters have approximately the same scale (as they range between 0 and 1), while the latter is much bigger in scale. For this reason when solving the system we consider the scaled values $\theta^{\star}_{r5d} = \theta_{r5d} - \log(K)$ where K is a suitably chosen constant that makes scales of all parameters more homogeneous. The solution of the system with the original set of parameters $\boldsymbol{\xi}_{rd} = (a_{rd}, b_{rd}, p_{rd}, q_{rd})$ can be obtained from $\boldsymbol{\xi}^{\star}_{rd} = (a_{rd}, b^{\star}_{rd}, p_{rd}, q_{rd})$ using a property of the GB2 distribution as $b_{rd} = Kb^{\star}_{rd}$. In line with Graf et al. (2011a) and Graf and Nedyalkova (2014) we set the constraints $a_{rd}p_{rd} > 1$ and $a_{rd}q_{rd} > 2$ which ensure that the implicitly defined $X_{rd} \sim GB2(a_{rd}, b_{rd}, p_{rd}, q_{rd})$

are such that $E(X_{rd}^{-1}) < +\infty$ and $E(X_{rd}^2) < +\infty$.

The minimum is searched using numerical methods and namely the popular Levenberg-Marquardt algorithm. Theoretical properties and efficient implementations of this algorithm have been studied in many papers (e.g. Moré, 1978). Kanzow et al. (2004) show global convergence properties of the algorithm when the constraints set is a convex set as in our problem.

Because of the mathematical complexity of (19) the solution leading to the indirect estimation of the GB2 parameters can be slow to find, making the whole method impractical. For this reason we consider three special cases of the GB2: Beta of the second kind, Dagum and Singh-Maddala distributions, characterized by three parameters and much simpler formulas for the Gini coefficient (see 22, 23, 24). We keep the same set of five *small area parameters* and a loss function analogous to (25), i.e. $L^{(i)}(\boldsymbol{\theta}_{rd}, \boldsymbol{\xi}_{rd})$, $i = 1, 2, 3$ for the indirect estimation of the three distribution parameters.

For each draw θ_{rkd}, $r = 1, \ldots, R$, we estimate three parallel non-linear systems: one for each of the three special cases of the the GB2, thus generating separate chains for the three set of distribution parameters. Although the three systems are solved instead of one, this strategy is computationally much more efficient than the one based on the GB2 distribution. If we denote with $\hat{\boldsymbol{\xi}}_{rd}$ a solution to (25) the distribution that minimizes $\sum_{r=1}^{R} L^{(i)}(\boldsymbol{\theta}_{rd}, \hat{\boldsymbol{\xi}}_{rd})$ in i is chosen, separately for each area, as the income distribution model. As a consequence, we adapt possibly different models to the data from different areas.

A point predictor for η_d can be obtained summarizing the posterior distribution $p(\eta_d|\mathbf{d})$; if quadratic loss is adopted it will be given by the posterior mean $\tilde{\eta}_d = E(\eta_d|\mathbf{d})$.

The small area estimator obtained in this way is not design-consistent as it depends on assuming the GB2 as a description of income within the areas even in large samples. Nonetheless it is robust with respect to misspecifications of the small area model as $\tilde{\boldsymbol{\theta}}_d$ is design consistent and thus converging to $\boldsymbol{\theta}_d$ regardless of model misspecifications. Asymptotically the posterior distribution

$p(\eta_d|\mathbf{d})$ will collapse on the solution of $\eta_d = \eta(\boldsymbol{\xi}_d)$: the dependence on the GB2 does remain, but that on the multivariate model does not.

5 An application to Italian EU-SILC data: estimation of RMPG in Italian provinces

In this section we illustrate the estimation of the RMPG η_d and the *nuisance* parameters θ_{kd} for the Italian administrative provinces. Input data come from the 2013 EU-SILC survey sample for Italy and consist of $(\hat{\theta}_{kd}, \hat{\phi}_{kd}, \mathbf{R})$, $k = 1, \ldots, 5$, $d = 1, \ldots, D$. We obtain an estimate of $\mathbf{R}$ starting from Spearman correlations $\rho_r(.,.)$ among the $\hat{\theta}_{kd}$. Rough estimates of $\rho_r(\hat{\theta}_{kd}, \hat{\theta}_{k'd})$ can be obtained using the bootstrap algorithm output (see section 2.2). We denote these estimates as $cor_{boot}(\hat{\theta}_{kd}, \hat{\theta}_{k'd})$. As most of the areas are small, to get stable estimates, we first assume that correlations $\rho_r(\hat{\theta}_{kd}, \hat{\theta}_{k'd})$ are constant across areas i.e. $\rho_r(\hat{\theta}_{kd}, \hat{\theta}_{k'd}) = \rho_r(\hat{\theta}_k, \hat{\theta}_{k'})$ and propose averaged estimates $\hat{\rho}_r(\hat{\theta}_k, \hat{\theta}_{k'}) = (\sum_{d=1}^{D} w_d)^{-1} \sum_{d=1}^{D} w_d cor_{boot}(\hat{\theta}_{kd}, \hat{\theta}_{k'd})$ with $w_d = n_d$. To obtain even more stable results, we then restrict the average to the set of the largest areas and namely to those with a sample size above the median, thus assuming $w_d = n_d \mathbf{1}\{n_d > Me(n_d)\}$. As the matrix $\mathbf{R}$ describes the dependence structure of $\hat{\theta}_{kd}$ on a transformed scale, we finally exploit the invariance of Spearman correlation under non decreasing monotone transformations and the sin transformation to switch from Spearman to Pearson correlations (see Elfadaly and Garthwaite, 2013, for details).

The parameters $\hat{\phi}_{kd}$ are estimated using variance smoothing models. Specifically, for the rates $\hat{\theta}_{kd}$, $k = 1, 2, 3$ the variances estimated using the bootstrap algorithm $v_{boot}(\hat{\theta}_{kd})$ are smoothed using the models:

$$\frac{\hat{\theta}_{kd}(1 - \hat{\theta}_{kd})}{v_{boot}(\hat{\theta}_{kd})} = \nu_k n_d + e_{kd}$$

where, for the residuals e_{kd} we assume $E(e_{kd}) = 0$ and $V(e_{kd}) = \varrho_k$. For the

Gini concentration coefficient, a different smoothing model is adopted:

$$\frac{\hat{\theta}_{4d}^2(1-\hat{\theta}_{4d}^2)}{v_{boot}(\hat{\theta}_{4d})} = \nu_4 n_d + e_{4d}$$

See Fabrizi and Trivisano (2016) for a motivation of this model. The least squares estimators $\hat{\nu}_k$ are then used to compute $\hat{\phi}_{kd} = \nu_k n_d, = 1, \ldots, 4$. For our data the squared correlations describing the fit of these models equal 0.82, 0.95, 0.78, 0.78 for $k = 1, \ldots, 4$ respectively.

These data are complemented by auxiliary information from administrative archives. A description of auxiliary variables, defined at the provincial level can be found in Appendix 3. The candidate auxiliary variables are many, some are highly correlated with each other, so selection is needed. Although the model is multivariate, we selected covariates to be used in equations (10) and (11) from the univariate models. Auxiliary variable selection is based on the methodology introduced in George and McCullogh (1993). Details on the variable selection process can be found in Appendix 3 as well.

All codes used in the estimation exercise are written in `R`. Posterior distributions for the multivariate model are based on Metropolis-Hastings type of MCMC algorithms. Specifically we used the software `jags` called through the R package `rjags` (Plummer et al., 2016). For all parameters single Markov Chains of length 50,000 are run. To assess the convergence of each chain, beside visual inspection of the chains, we use the Heidelberg-Welch diagnostics (Heidelberg and Welch, 1983; Carlin and Cowles, 1996) that reduces to testing the null hypothesis of a stationary path using the Cramer-von-Mises statistic. A conservative burn-in of 10,000 is used before calculating these statistics. The Heidelberg-Welch diagnostics are based on a single chain; a multichain approach was not advisable in our problem as a careful setting of the initial value is needed to speed up the convergence. In the overwhelming majority of chains the p-value associated to the Heidelberg-Welch diagnostics is above 0.05; for the chains of the parameters θ_{1d}, θ_{2d}, θ_{4d}, θ_{5d} in more than 98% of the cases, for θ_{3d} slightly more than 95% of the cases. In calculating posterior summaries, one every 30^{th} draw is kept. This severe *thinning* of the chains is partly motivated by their relatively poor mixing; this depends on the fact that *nuisance* parameters are strongly correlated, as they are all summaries of the same distributions. Moreover, we want to keep the posterior sample size small as its size defines the number of times the non-linear system discussed in section 4.2 needs to be solved. The overall sample from the posterior is of size $R = 3,000$.

Each draw from the posterior distribution of θ_{kd}, $k = 1, \ldots, 5$ is used to solve the constrained non-linear system discussed in section 4.2. Specifically we work with the Levenberg-Marquardt nonlinear least-squares algorithm as implemented in the `nlsLM` function of the `R` package `minpack.lm` (Elzhov et al., 2016). Initial values are set solving the system on the ensemble of the posterior means $E(\theta_{kd}|\mathbf{d})$ with a precision $1.0 \times 10E - 10$, while a precision $1.0 \times 10E - 5$ is used to assess convergence of solutions for the systems based on individual draws.

The application run in about 2 hours using a 4 cores 5500u processor (2.44GhZ, 8GB ram memory). We tried to run the same application using the GB2 instead of its special cases as the reference distribution: the computing times rise to about 40 hours. This motivates our choice of considering a solution based on the three parameters special cases of the GB2.

A special case of the GB2 distribution is chosen separately for each area according to the methodology illustrated in section 4.2. The Dagum distribution is chosen in the large majority of areas (95 times), the Singh-Maddala for 14 areas and the B2 only in one area. This result is in line with expectations from the literature (Kleiber, 1996; McDonald et al., 2013) as discussed in section 4.1. For the purposes of the analysis of this data set the methodology could then be simplified and the only Dagum distribution considered. Nonetheless this may depend on specific features of our data and it is not necessarily a general result (see Kakamu, 2016).

Markov chains for η_d (RMPG) are generated from those of the parameters of the chosen distributions. The Heidelbergt-Welch diagnostics computed for the chains η_d result in p-value greater than 0.05 in 96% of the cases. As this percentage are in line with the type-I error of the test, we can conclude that the convergence is satisfying also for these chains.

As a further check we apply the functional approach used to generate posterior chains for η_d to the *nuisance* parameters θ_{kd} and compare the posterior obtained in this way to those directly obtained from the multivariate model described in section 3. We focus our comparisons on posterior means and standard deviations calculating ratios of the posterior summaries obtained according to the two methods. These ratios show some variation across areas. For posterior means we have that for all parameters and all areas the difference is less than 5% with the exception of θ_4 (Gini concentration coefficient) for which the difference is between 5% and 10% in 20% of the areas; posterior means obtained with the functional being slightly smaller (3% on average). For all parameters, posterior standard deviations are very close on average (less than 2%) with the exception of θ_4 and θ_5 for which the posterior standard deviations based on the functional approach are 5% larger on average. In the large majority of areas the difference is less than 10% and for θ_1, θ_2 and θ_3 less than 5%.

In table 1 we present how efficient is our approach in reducing the standard errors associated to the estimators. We define

$$ser(\eta_d) = \frac{sd(\eta_d|\mathbf{d})}{se(\hat{\eta}_d)} \tag{26}$$

where $se(\hat{\eta}_d)$ is computed according to the bootstrap algorithm of Fabrizi et al. (2011). We calculate also $ser(\theta_{kd})$ that are defined similarly; $se(\hat{\theta}_{kd})$ is calculated according to the methodology illustrated in section 2.2. We recognize that this comparison involve two quantities that are logically different as the numerator is a posterior *sd* and the denominator a *se* with respect to the randomization distribution induced by sampling. Nonetheless this type of comparisons are common in small area literature.

The improvement in precision allowed by $\tilde{\eta}_d$ with respect to $\hat{\eta}_d$ is dramatic; on average the posterior standard deviation is slightly more than one quarter of that of the direct estimator. Only in large areas, and especially so if located in the Sourth of the country where poverty prevalence is higher $sd(\eta_d|\mathbf{d})$ is more than one half of $se(\hat{\eta}_d)$. The posterior standard deviations $sd(\theta_{kd}|\mathbf{d})$ are on average half the size of the standard error $se(\hat{\theta}_{kd})$ of direct estimators; different reduction levels in different areas can be explained by different area-specific sample sizes.

Parameter	η	$\theta_{1.}$	$\theta_{2.}$	$\theta_{3.}$	$\theta_{4.}$	$\theta_{5.}$
Min.	0.064	0.102	0.113	0.122	0.078	0.169
1st Qu.	0.168	0.380	0.413	0.303	0.303	0.549
Median	0.265	0.482	0.493	0.398	0.362	0.627
Mean	0.284	0.483	0.511	0.414	0.383	0.627
3rd Qu.	0.358	0.586	0.601	0.506	0.467	0.745
Max.	0.711	0.904	0.93	0.885	0.831	0.926

Table 1: Distribution of the standard error reduction (ser_{kd}) defined in equation (26) across the 110 provinces (areas); η = RMPG, θ_1 = at-risk-of-poverty rate, θ_2 = share of population with income below the median, θ_3 = affluence rate, θ_4 = Gini concentration coefficient, θ_5 = mean of log-income.

Statistics Canada (2007) suggests that estimates whose associate coefficient of variation (CV) is less than 16.6% are reliable enough for general use, those with a CV between 16.6% and 33.3% can be published but accompanied by a warning to users while those with even larger CV should be deemed as completely unreliable and not published. In figure 1 we plot the histograms of $CV(\eta_{kd}|\mathbf{d})$, $CV(\theta_{kd}|\mathbf{d})$, using the thresholds suggested by Statistics Canada (2007). We note that, although popular, these criteria can be too exigent for the estimation of small proportions when a high coefficient of variation can be the effect of a small estimate; in this case, that encompasses our θ_1 and θ_3, alternative criteria in terms of standard errors can be used (see European Commission, 2013, page 13). We keep the Statistics Canada criteria as, from figure 1 it is apparent that for all parameters the small area estimates we produce are suitable for publication with few problematic cases for the affluence rate θ_3, attributable the low point estimates. Notably the posterior coefficients of variation are acceptable in all cases for the RMPG.

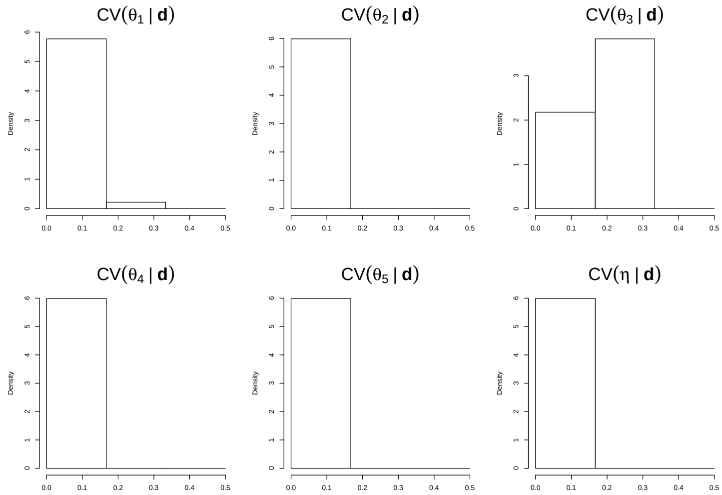

Figure 1: Histograms of the posterior coefficient of variations over the 110 provinces. The breaks in the histograms plot coincide with those suggested by Statistics Canada (2007)

6 A simulation exercise

The methodology we presented for the estimation of the RMPG is complex as it involves a multivariate hierarchical Bayesian model and, for each MCMC draw, the solution of a non-linear system based on a parametric assumption on the distribution of equivalized income in the areas. The good performances in terms of posterior coefficient of variation that appears in figure 1 can be misleading if the point estimates were heavily biased. In this section, we introduce a simulation study to assess the frequentist properties of the RMPG predictor. Specifically we focus on bias, mean square error and the frequentist coverage of probability intervals based on posterior quantiles. These properties will be evaluated also for the predictors of *nuisance* parameters θ_{kd}.

The simulation exercise is based on the same EU-SILC sample considered in our application. We assume it as a synthetic population, from which we repeatedly draw stratified samples and estimate the small area parameters for areas larger than those considered in the application. As the synthetic population is held fixed, the simulation can be labeled as design based.

We target administrative regions as areas of interest, an higher level administrative body with respect to the provinces considered in the application; each region includes several provinces; the two exceptions, Valle d'Aosta and Molise, that include only 1 and 2 provinces respectively, are excluded from the syntethic population. Administrative regions are planned domain of the EU-SILC survey in Italy. We draw stratified samples from the synthetic population with strata defined by these regions. The size of the 18 administrative regions in the synthetic population ranges, in terms of households from 386 to 1846 with a median size of 998. Stratified samples, drawn without replacement, are allocated proportionally with a sampling rate of 0.115, chosen so that the median size of region-specific samples in the simulation matches the median of province-specific samples in the application. With respect to the application, sample sizes are less variable as they range from 44 to 212 (and not from 6 to 882 as in the case of province-specific samples in the application).

For each of the $S = 1000$ samples drawn from the synthetic population we replicate the methodology illustrated in section 5; also the details related to MCMC computation and the non-linear system remain the same.

Let's denote with ${}_P\boldsymbol{\theta}_d$, ${}_P\eta_d$ the syntethic population target parameters, where ${}_P\boldsymbol{\theta}_d = \{{}_P\theta_{kd}\}$ $k = 1, \dots, 5$, while the Bayes estimators based on quadratic loss are denoted as ${}_s\tilde{\boldsymbol{\theta}}_d = E\big({}_P\boldsymbol{\theta}_d|\mathbf{d}_s\big)$, ${}_s\tilde{\eta}_d = E\big({}_P\eta_d|\mathbf{d}_s\big)$ where $\mathbf{d}_s$ denotes the data from the $s-th$ replicated sample. If we use the shortcut $.\tilde{\theta}_{kd}$ to denote the

Bayes estimator for θ_{kd} when averaged over the S replications we can define:

$$RRMSE(\tilde{\theta}_{kd}) = \frac{1}{S}\sum_{s=1}^{S}\frac{\sqrt{\left({}_s\tilde{\theta}_{kd} - {}_P\,\theta_{kd}\right)^2}}{{}_P\theta_{kd}} \quad (27)$$

$$RBIAS(\tilde{\theta}_{kd}) = \frac{1}{S}\sum_{s=1}^{S}\frac{\left({}_s\tilde{\theta}_{kd} - {}_P\,\theta_{kd}\right)}{{}_P\theta_{kd}} \quad (28)$$

$$COV(\tilde{\theta}_{kd}; 1-\alpha) = \frac{1}{S}\sum_{s=1}^{S}\mathbf{1}\left({}_s q_{\alpha/2} \leq_s \theta_{kd} \leq_s q_{1-\alpha/2}\right) \quad (29)$$

where ${}_s q_{\alpha/2}$, ${}_s q_{1-\alpha/2}$ are the α and $1-\alpha$ quantiles of $p({}_P\theta_{kd}|\mathbf{d}_s)$. Specifically we consider $\alpha = 0.05$. Definitions for $RRMSE(\tilde{\eta}_d)$, $RBIAS(\tilde{\eta}_d)$, $COV(\tilde{\eta}_d, 1-\alpha)$ follow accordingly.

In Table 2 we present results for the indicators (27)-(29): we show the three quartiles (Q_1, Me, Q_3) of the distribution of these three indicators across the 18 regions considered in the simulation.

Direct estimators		θ_1	θ_2	θ_3	θ_4	θ_5	η
RBIAS	Q_1	-0.005	-0.002	-0.006	-0.005	0.000	0.016
	Me	-0.002	0.000	-0.001	-0.003	0.000	0.035
	Q_3	0.003	0.002	0.007	-0.002	0.000	0.123
RRMSE	Q_1	0.205	0.090	0.250	0.072	0.005	0.320
	Me	0.257	0.116	0.329	0.081	0.006	0.426
	Q_3	0.283	0.124	0.486	0.092	0.009	0.466
Bayesi estimators		θ_1	θ_2	θ_3	θ_4	θ_5	η
RBIAS	Q_1	-0.057	-0.023	-0.046	-0.029	-0.002	-0.054
	Me	0.019	0.003	0.073	-0.004	0.000	0.012
	Q_3	0.101	0.027	0.108	0.028	0.002	0.101
RRMSE	Q_1	0.093	0.043	0.139	0.034	0.002	0.108
	Me	0.115	0.055	0.156	0.041	0.003	0.141
	Q_3	0.160	0.074	0.241	0.066	0.006	0.205
COV (., 0.95)	Q_1	0.904	0.880	0.933	0.871	0.904	0.911
	Me	0.977	0.983	0.975	0.985	0.955	0.937
	Q_3	0.987	0.985	0.986	0.995	0.979	0.953

Table 2: First, third quartiles and median of $RRMSE$, $RBIAS$, $COV(., 0.95)$ with respect to the 18 regions considered in the simulation. θ_1 = at-risk-of-poverty-rate, θ_2 = share of population with income below the median, θ_3 = affluence rate, θ_4 = Gini concentration coefficient, θ_5 = mean of log-income, η = RMPG.

The RRMSE associated to RMPG has the same magnitude of those of the at-risk-of-poverty rate ($\tilde{\theta}_{1d}$) and affluence rate ($\tilde{\theta}_{3d}$), a good result if we read it considering the little information the direct estimation of the RMPG provides. Smaller RRMSE can be either attributed to a size effect ($\tilde{\theta}_{2d}$ has an MSE similar to that of $\tilde{\theta}_{1d}$ but a larger denominator) or to the more power auxiliary variables have for some parameters (specifically this is the case of the mean of the log-incomes, $\tilde{\theta}_{5d}$). The relative bias is, in all cases, when averaged across areas, close to 0, that is the shrinkage does not imply a systematic tendency to over- or under-estimate the corresponding population parameters. As far as RMPG is concerned, the relative bias is, despite their indirect estimation, small in most of the areas. Negative or positive biases on individual areas is due to a shrinkage effect that is more pronounced when the sample size is small.

Interval estimates based on posterior quantiles ($q_{\alpha/2}$, $q_{1-\alpha/2}$) usually have an approximate $1 - \alpha$ frequentist coverage if the bias of the posterior mean is small and posterior standard deviation is close to the frequentist standard error. Table 2 shows that in some cases the coverage is below the frequentist nominal level; these cases are those characterized by relatively higher bias levels. In some other cases we have a coverage above the nominal (frequentist) level; this is due to a tendency of posterior standard deviations to be slightly larger than the frequentist standard errors (we can estimate from MC replications).

To complete the comparison, for η_d, we simulated also an estimator associated to a *standard* Fay-Herriot type of model assuming approximate normality of $\hat{\eta}_d$, $var(\hat{\eta}_d)$ as known and set equal to their actual values resulting from MC replications. We selected auxiliary variables from those described in Appendix 3 and namely the variables x_1 , the anti-logit of x_6 and x_9 that proved to be those providing the best fit. The $ARRMSE$ results equal to 0.249 and the $ARBIAS$ to 0.059. $ACOV(0.95)$ is very close (slightly above) the nominal level; nonetheless some of the intervals are so wide that the lower bound is negative. This estimator is therefore effective in improving the efficiency of the direct estimator but clearly inferior to $\tilde{\eta}_d$; this finding is in line with our expectation: not only the $\hat{\eta}_d$ are very unreliable but it is difficult to auxiliary variables with a good predictive power.

7 Conclusions

In this research we focused on the estimation of the relative median poverty gap (RMPG), a popular measure of poverty severity, motivated by the need to estimate it at the small area level using Italian data from the EU-SILC survey.

We present a small area estimation method based on area-level modelling, that requires only survey based direct estimators and area-level summaries from auxiliary sources. Area-level modelling is therefore less data demanding with respect ot unit-level models that, when applied to the estimation of non-linear functional of the target variable population values, require knowledge of individual level values of the auxiliary variables, a requirement that implies non trivial data quality and disclosure problems.

The specific nature of the RMPG, for which direct estimators are in most cases completely unreliable, led us to consider a functional estimation method. We build on a method of using summary statistics to estimate parameters of an underlying income distribution due to Graf and Nedyalkova (2014), apply it within the framework of MCMC based Bayesian inference, and we use it in the opposite direction to estimate the RMPG (i.e. using estimated income distribution parameter to obtain an estimate of a population descriptive quantity).

Our methodology implies a number of choices, some of them driven by computational reasons. Specifically we propose to use three-parameters special cases of the GB2 to describe income distribution in the small area as this choice reduced computational times by a factor of 20. This computational gain was crucial, especially in view of the simulation exercise we introduced in section 6, to assess frequentist properties of the introduced Bayesian predictors.

Simulation results confirm that the method we propose can produce reliable small area estimates of the RMPG. The proposed methodology can be applied to the estimation of other parameters with problems similar to those of the RMPG, such as the quintile share ratio. More details on the estimation of this parameter can be found in Appendix 4.

References

Alfons A. and Templ M. (2013), Estimation of Social Exclusion Indicators from Complex Surveys: The R Package `laeken`, *Journal of Statstical Software*, **54**, 15.

Atkinson A.B., Marlier E. (2010), *Income and living conditions in Europe*, Eurostat Statistical books, Publication Office of the European Union, Luxembourg.

Biewen M., Jenkins S.P. (2006), Variance estimation for generalized entropy and Atkinson inequality indices: the complex survey data case, *Oxford Bullettin of Economics and Statistics*, **68**, 371–383.

Chotikapanich D., Griffiths W.E., Rao D.S.P., Valencia V. (2012), Global income distributions and inequality, 1993 and 2000: incorporating country-level inequality modeled with Beta distributions, *Review of Economics and Statistics*, **94**, 52–73.

Clemenceau A., Museux J.P. (2007), EU-SILC (community statistics on income and living conditions: general presentation of the instrument), in *Comparative EU statistics on Income and Living Conditions: Issues and Challenges*, Proceedings of the EU-SILC conference (Helsinki, 6-8 November 2006), Eurostat Methodologies and Working papers, Publication Office of the European Union, Luxembourg.

Cowles M.K., Carlin B.P. (1996), Markov Chain Monte Carlo convergence diagnostics: a comparative review, *Journal of the American Statistical Association*, **91**, 883–904.

Clemen R.C., Reilly T. (1999), Correlations and copulas for decision and risk analysis, *Management Science*, **45**, 228–224.

Datrup S.R., Hartshorn R., McDonald J.B. (2007), The impact of taxes and transfer payments on the distribution of income: a parametric comparison, *Journal of Economic Inequality*, **5**, 353–359.

Elfadaly F.G., Garthwaite P.H. (2013), Elicing Dirichlet and Gaussian copulas prior distributions for multinomial models, unpublished manuscript downlodadble at http://mcs-brains.open.ac.uk/elicitation/Copula

Elzhov T.V., Mullen K.M., Spiess A.N., Bolker B. (2016), `minpack.lm`: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds, url = `https://CRAN.R-project.org/package=minpack.lm`.

European Commission (2004), A New Partnership for Cohesion: Convergence, Competitiveness, Cooperation. Third report on economic and social cohesion, Office for the Official Publications of the European Communities, Luxembourg.

European Commission (2010), Macro determinants of individual income poverty in 93 regions of Europe Luxembourg: Publications Office of the European Union, `doi:10.2785/53721`

European Commission (2013), Handbook on precision requirements and variance estimation for ESS household survey, Luxembourg: Publications Office of the European Union, `doi:10.2785/1357`

Eurostat (2019), Methodological manual on territorial typologies, 2018 edition, Publications Office of the European Union, Luxembourg, `doi:10.2785/930137`

Fabrizi E., Ferrante M.R. and Pacei S. (2008), Measuring sub-national income poverty by using a small area multivariate approach, *The Review of Income and Wealth*, **54**, 597–5.

Fabrizi E., Ferrante M.R., Pacei S. and Trivisano C. (2011), Hierarchical Bayes multivariate estimation of poverty rates based on increasing thresholds for small domains, *Computational Statistics and Data Analysis*, **55**, 1736–47.

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

Fabrizi E., Ferrante M.R., Trivisano C. (2016), Hierarchical Beta regression models for the estimation of poverty and inequality parameters in small areas, in: Pratesi, Monica (Ed.): Analysis of poverty data by small area methods, John Wiley and Sons, 299-314.

Fuller W.A. (2009), Sampling Statistics, Wiley Series in Survey Methodology, John Wiley and Sons, New York.

Fusco A., Guio A.C., Marlier E. (2010), Chracterizing the income poor and the materially deprived in European countries, in *Eurostat Statistics books: Income and living conditions in Europe,* (Atkinson A.B. and Marlier E. (eds.), Publication Office of the European Union, Luxembourg.

George E.I., McCulloch R.E. (1993), Variable selection via Gibbs sampling, *Journal of the American Statistical Association*, **88**, 881–9.

Ghosh M., Nangia N., Kim D. (1996), Estimation of Median Income of Four-Person Families: A Bayesian Time Series Approach, *Journal of the American Statistical Association*, **91**, 1423–31.

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.

Gil A., Segura J., Temme N.M. (2007), Numerical methods for special functions, SIAM, Philadelphia.

Giorgi G.M., Gigliarano C. (2017), The Gini concentration index: a review of the inference literature, *Journal of Economic Surveys*, **31**, 1130-1148.

Graf M., Nedyalkova D., Muennich R., Seger J., Zins S. (2011), Parametric estimation of income distributions and indicators of poverty and social exclusion, Deliverable 2.1 of the AMELI project.

Graf M., Nedyalkova D. (2011), Parametric Estimation of Income Distributions and Derived Indicators Using the GB2 Distribution, in B. Hulliger (ed.), Report on the Simulation Results, Deliverable 7.1 of the AMELI project, chapter 7.1.

Graf M., Nedyalkova D. (2014), Modeling of income and indicators of poverty and social exclusion using the Generalized Beta Distribution of the second kind, *Review of Income and Wealth*, **60**, 821–832.

Hàjek J. (1958), On the theory of ratio estimates, *Aplikace Mathematiki*, **3**, 384–298.

Heidelberger P., Welch P.D. (1983), Simulation Run Length Control in the Presence of an Initial Transient, *Operations Research*, **31**, 1109–1144.

Huang A., Wand M.P. (2013), Simple marginally noninformative prior distributions for covariance matrices, *Bayesian Analysis*, **8**, 439–452.

Jenkins S.P. (2009), Distributionally-Sensitive Inequality Indices and the GB2 Income Distribution, *The Review of Income and Wealth*, **55**, 392–398.

Kakamu K.(2016), Simulation Studies Comparing Dagum and Singh-Maddala Income Distributions, *Computational Economics*, **48**, 593–605.

Kanzow C., Yamashita N., Fukushima M. (2004), Levenberg-Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints, *Journal of Computational and Applied Mathematics*, **172**, 375–397.

Kleiber C.(1996), Dagum vs. Singh-Maddala income distributions, *Economics Letters*, **53**, 265–268.

Kleiber C., Kotz S. (2003), Statitical size distributions in economics and actuarial sciences, Wiley Series in Probability and Statistics, New York.

Langel M., Tillé Y. (2011), Statistical inference for the quintile share ratio, *Journal of Statistical Planning and Inference*, **141**, 2976–2985.

McDonald J.B. (1984), Some generalized functions for the size distribution of income, *Econometrica*, **52**, 647–3.

McDonald J.B., Sorensen J., Turley P.A. (2013), Skewness and kurtosis properties of income distribution models, *The Review of Income and Wealth*, **59**, 360–4.

Molina I., Rao J.N.K. (2010), Small area estimators of poverty indicators, *The Canadian Journal of Statistics*, **38**, 369–385.

Molina I., Nandram B., Rao J.N.K. (2014), Small area estimation of general parameters with application to poverty indicators: a hierchical Bayes approach, *The Annals of Applied Statistics*, **8**, 852–885.

Moré, J.J. (1978), The Levenberg-Marquardt algorithm: implementation and theory, in Watson G.A. (ed.) *Numerical Analysis, Lecture Notes in Mathematics*, **630**, 105–116.

Ospina R., Ferrari S.L.P. (2012), A general class of zero-or-one inflated beta regression models, *Computational Statistics and Data Analysis*, **56**, 1609-1623.

Pfeffermann D. (2013), New important developments in small area estimation, *Statistical Science*, **28**, 40-68.

Peichl A., Schaefer T., Scheicher C. (2010), Measuring richness and poverty: a micro data application to Europe and Germany, *The Review of Income and Wealth*, **56**, 597–9.

Perugini C., Martino G. (2008), Income inequality within European regions: determinants and effects on growth, *Review of Income and Wealth*, 54, 373–406.

Plummer M., Stukalov A., Denwood M. (2016). `rjags` (R package version 4-6, pp. 10). Vienna, Austria: The Comprehensive R Archive Network

Rao J.N.K., Molina I. (2015), Small area estimation, Wiley Series in Survey Methodology, New York.

Statistics Canada (2007) 2005 Survey of Financial Security - Public Use Microdata File, User Guide. Published by authority of the Minister responsible for Statistics Canada, `http://www.statcan.gc.ca/pub/13f0026m/13f0026m2007001-eng.htm`.

Souza D.B., Moura F.A.S. (2016), Multivariate Beta regression with application in small area estimation, *Journal of Official Statistics*, **32**, 747–768.

Tarozzi A., Deaton A. (2009), Using census and survey data to estimate poverty and inequality for small areas, *Review of Economics and Statistics*, **91**, 773–792.

Tzavidis N., Zhang L.C., Luna A., Schmid T., Rojas-Perilla N. (2018), From start to finish: a framework for the production of small area official statistics, *Journal of the Royal Statistical Society, ser. A*, **181**, 927-972.

van der Vaart A.W., Wellner J.A. (1996), Weak convergence and empirical processes with applications to statistics, Springer Series in Statistics, New York.

Wieczorek J., Hawala S. (2011), A Bayesian zero-one inflated Beta model for estimating poverty in U.S. counties, JSM Proceedings, Section on Survey Research Methods, 2812-2822.

Zellner A. (1971), Bayesian and non-Bayesian analysis of the log-normal distribution and log-normal regression, *Journal of the American Statistical Association*, **66**, 327–330.

Supplementary material

Appendix 1: small sample properties of the RMPG direct estimator

To assess the bias of the relative median poverty gap (RMPG) in small sample we run a design based simulation based on the 2013 EU-SILC sample we considered in section 5. We use the sample as synthetic population and we use the 21 NUTS2 adiminstrative regions of Italy as domains. The Monte Carlo experiment consist in drawing repeatedly stratified samples with proportional allocation and a 5% sampling rate. We consider households as the sampling units; in line with the definitions of the EU-SILC survey all individuals in the same household share the same income and the RMPG is defined at the individual level. We obtain very small samples (the sample household range from 3 to 18) similar in size to the poor household sub-samples that we meet in our application. Results, summarizing $5,000$ Monte Carlo replication are reported in table 3.

Sample size (m_d)	Rel. Bias	CV
$3 \leq m_d \leq 5$	23.12	69.33
$6 \leq m_d \leq 10$	13.60	55.09
$11 \leq m_d \leq 18$	3.88	36.78

Table 3: Average relative bias and average coefficient of variation (in percentage) in the estimation of RMPG

When the poor households in the sample is less than 10 the bias is large and cannot be overlooked if the estimate is going to be used as an input for a small area estimation model. A large portion of the province-specific sample sizes we deal with in our application are below this threshold, especially in view of an overall poverty rate of 18% at the national level.

Appendix 2: Robustness of the proposed small area estimator

Let's first consider the rates θ_{kd}, $k = 1, 2, 3$. We note that for large m_d, $\theta^{\star}_{kd} \cong \theta_{kd}$ so that

$$f(\hat{\theta}_{kd}|\theta_{kd}) = Beta\Big(\theta_{kd}(\hat{\phi}_{kd} - 1), (1 - \theta_{kd})(\hat{\phi}_{kd} - 1)\Big)$$

This Beta likelihood can be approximated by a Normal, as the conditions stated in Gil et al. (2007), section 10.5, for this approximation are satisfied provided we assume $\theta_{kd}/(1 - \theta_{kd})$ is bounded away from 0, consistently with (6). Consequently

$$f(\hat{\theta}_{kd}|\theta_{kd}) \cong N\left(\theta_{kd}, \frac{\theta_{kd}(1 - \theta_{kd})}{\hat{\phi}_{kd}}\right)$$

We now study the posterior distribution of θ_{kd}, $k = 1, 2, 3$ conditional on $\mathbf{\Sigma}_v$ and the rest of the parameters assuming, without loss of generality that $\mathbf{x}^t_{kd}\beta_k = 1$ and setting to 1 also the relevant element of $\mathbf{\Sigma}_v$

$$\begin{aligned} g(\theta_{kd}|\hat{t}_{dk}, \hat{\phi}_{dk}) &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{2\hat{\phi}_{kd}}{\theta_{kd}(1-\theta_{kd})}}\exp\left\{-\frac{\hat{\phi}_{kd}}{2\theta_{kd}(1-\theta_{kd})}(\hat{\theta}_{kd} - \theta_{kd})^2\right\} \times \\ &\times \exp\left\{-\frac{1}{2}\Big(\log\frac{\theta_{kd}}{1-\theta_{kd}} - \mu\Big)^2\right\} \end{aligned}$$

For all $x \leq \hat{\theta}_{kd}$ we have that

$$\begin{aligned} \int_0^x g(\theta_{kd}|\hat{\theta}_{dk}, \hat{\phi}_{dk})d\theta_{kd} &\leq \sqrt{\frac{\hat{\phi}_{kd}}{\pi}}\exp\left\{-\frac{\hat{\phi}_{kd}}{2x(1-x)}(\hat{\theta}_{kd} - x)^2\right\} \times \\ &\times \int_0^x \{\theta_{kd}(1-\theta_{kd})\}^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\Big(\log\frac{\theta_{kd}}{1-\theta_{kd}} - \mu\Big)^2\right\}d\theta_{kd} \end{aligned} \quad (30)$$

as $\frac{1}{\sqrt{2\pi}}\sqrt{\frac{2\hat{\phi}_{kd}}{\theta_{kd}(1-\theta_{kd})}}\exp\left\{-\frac{\hat{\phi}_{kd}}{2\theta_{kd}(1-\theta_{kd})}(\hat{\theta}_{kd} - \theta_{kd})^2\right\}$ is monotonically increasing in θ_{kd} on $(0, x)$. Since the integral appearing in (31) is finite and $\sqrt{\frac{\hat{\phi}_{kd}}{\pi}}\exp\left\{-\frac{\hat{\phi}_{kd}}{2x(1-x)}(\hat{\theta}_{kd} - x)^2\right\} \to 0$ as $\hat{\phi}_{kd} \to +\infty$ we have that $\int_0^x g(\theta_{kd}|\hat{\theta}_{dk}, \hat{\phi}_{dk})d\theta_{kd} \to 0$ when $\hat{\phi}_{kd} \to +\infty$, a condition that is equivalent to $m_d \to +\infty$.

Similarly, for all $x \geq \hat{\theta}_{kd}$ we have that

$$\begin{aligned} \int_x^1 g(\theta_{kd}|\hat{\theta}_{dk}, \hat{\phi}_{dk}) d\theta_{kd} \quad &\leq \quad \sqrt{\frac{\hat{\phi}_{kd}}{\pi}} \exp\left\{ -\frac{\hat{\phi}_{kd}}{2x(1-x)}(\hat{\theta}_{kd} - x)^2 \right\} \times \\ &\times \quad \int_x^1 \{\theta_{kd}(1-\theta_{kd})\}^{\frac{1}{2}} \exp\left\{ -\frac{1}{2}\Big(\log \frac{\theta_{kd}}{1-\theta_{kd}} - \mu\Big)^2 \right\} d\theta_{kd} \end{aligned} \tag{31}$$

as $\frac{1}{\sqrt{2\pi}}\sqrt{\frac{2\hat{\phi}_{kd}}{\theta_{kd}(1-\theta_{kd})}} \exp\left\{ -\frac{\hat{\phi}_{kd}}{2\theta_{kd}(1-\theta_{kd})}(\hat{\theta}_{kd} - \theta_{kd})^2 \right\}$ is monotonically decreasing in θ_{kd} on $(x, 1)$.

It easily follows that

$$\int_0^1 g(\theta_{kd}|\hat{\theta}_{dk}, \hat{\phi}_{dk}) d\theta_{kd} \to 0 \tag{32}$$

as the sample size grows large, and $E(\theta_{kd}|\mathbf{d}, \Sigma_v) \to \hat{\theta}_{kd}$ from which design consistency follows.

A parallel argument follows the small area estimator of the Gini coefficient, i.e. θ_{4d}. In this case as well, using the general results from Gil et al. (2007), section 10.5 we can approximate the Beta likelihood:

$$f(\hat{\theta}_{4d}|\theta_{4d}) \cong N\left(\theta_{4d}, \frac{\theta_{4d}^2(1-\theta_{4d})^2}{2\hat{\phi}_{kd}}\right)$$

Proof of desing consistency follows along the same lines we have seen for θ_{kd}, $k = 1, 2, 3$. The parameter θ_{5d} is modelled using a Normal likelihood for $\hat{t}_{5d}$ and the proof is even more simple.

The posterior distribution involved in the minimization (25) converges to the design-consistent direct estimators $\hat{\theta}_{kd}$ $k = 1, \ldots, 5$ as the sample size grows large. It is easy to note that estimators $\hat{\theta}_{kd}$, $k = 1, 2, 3, 5$ are in fact methods of moments estimators; $\hat{\theta}_{4d}$ can be also seen as an estimator in the same class (see Giorgi and Gigliarano, 2016). Thereby, in large samples, (25) converges to a function of $\hat{\theta}_{kd}$, $k = 1, \ldots, 5$ that can be viewed as a generalized method of moments criterion function. Assuming the GB2 is an adequate description of the income distribution in the area, consistency of $\tilde{\eta}_d$ follows from the argmax (arg-min) continous mapping theorem (van der Vaart and Wellner, 1996, chapter 3).

This result implies that $\tilde{\eta}_d$ enjoys design-consistency type of robustness with respect to mis-specifications of the multivariate small area model discussed in section 3. Nonetheless we cannot talk of design-consistency as the assmption of GB2 distribution for income is still playing a role.

A design-consistent estimator for η_d can be obtained using composite estimation

$$\tilde{\eta}_d^{dc} = \gamma_d \hat{\eta}_d + (1 - \gamma_d)\tilde{\eta}_d \tag{33}$$

where $\gamma_d \in (0, 1)$ is some weight going to 0 when $var(\hat{\eta}_d) \to 0$ and to 1 when the information provided by the direct estimator is much larger with respect to that proposed by the model. We propose

$$\gamma_d = \frac{|\tilde{\mathbf{\Sigma}}_d|^{1/5}}{|\tilde{\mathbf{\Sigma}}_d|^{1/5} + var(\hat{\eta}_d)} \tag{34}$$

where $\tilde{\mathbf{\Sigma}}_d = E(\mathbf{\Sigma}|\mathbf{d})$ and $\mathbf{\Sigma}_d$, the random effects covariance matrix is defined (12). $|\tilde{\mathbf{\Sigma}}_d|^{1/5}$ summarize the information provided by the multivariate model and generalizes the variance of the random effects ordinarily used in Fay-Herriot model. An hierarchical Bayes version of (33) can be obtained by drawing samples from its posterior distribution, that can be easily expressed as a function of that of $|\tilde{\mathbf{\Sigma}}_d|^{1/5}$. In principle we can replace $var(\hat{\eta}_d)$ with $|\hat{\mathbf{V}}_d|^{1/5}$, where $\hat{\mathbf{V}}_d$ is the covariance matrix of the *nuisance* parameters variance estimators, as $|\tilde{\mathbf{\Sigma}}_d|^{1/5}$ and $|\hat{\mathbf{V}}_d|^{1/5}$ are more directly comparable; nonetheless this would lead to an unjustified large γ_d as $var(\hat{\eta}_d)$ is much larger than $|\hat{\mathbf{V}}_d|^{1/5}$ in practical situations. We do not insist on (33) for two reasons: first, we think that assuming the GB2 for income is not a particularly strong assumption, especially as the left tail, the one involved in the definition of the RMPG is concerned; secondly the low efficiency of $\hat{\eta}_d$ leads in practice to composite estimators dominated by $\tilde{\eta}_d$, i.e. the estimator we proposed.

Appendix 3: Auxiliary information used in the estimation

Auxiliary information is obtained from publicly available archives at the municipal level, and then aggregated to obtain province level variables. Literature on poverty and income inequality determinants within regional communities is vast; a review of it is out of the scope of this paper. See European Commission (2010), Perugini and Martino (2008) among other references. In small area estimation, we do not aim to obtain an explanatory model for the target variable, rather, we use auxiliary information as a tool to improve the precision of estimators. Since auxiliary information should be accurately known at the area level, the choice is severely limited by this requirement.

A preliminary selection of variables was based on results from previous studies (Fabrizi et al., 2016; Fabrizi and Trivisano, 2016). Although several sources were initially considered the most powerful auxiliary variables are obtained from the fiscal archives held by the Italian Ministry of Finance. The variables we directly consider in this study are: percentage of residents aged more than 15 filling tax forms (x_1), total taxable income claimed by private residents divided by the overall population size (x_2), the share of population aged 65 or more (x_3), the mean log income (x_4), the logit transform of the Gini index (x_5), headcount ratio poverty rate (x_6), share of people with income below the median (x_8) and affluence rate (x_7). Variables x_4-x_8 are approximations calculated from fiscal income distributions published at the municipal level by the Ministry of Fincance. The rates are not only approximated but also based on approximated thresholds.

In variable selection we consider univariate models. Specifically sampling models are those described in (6), (7) and (8). Also linking models are the same, i.e., (10) and (11), but we assume independent random effects: $v_{kd} \sim N(0, \tau_k^2)$, $\tau_k \sim Unif(0, C_k)$ for some large C_k instead of (12).

For the β_k in (10) and (11), in line with George and McCullogh (1993) we assume a *spike and slab* prior on the coefficients associated to candidate

Parameter	θ_1	θ_2	θ_3	θ_4	θ_5
x_1	✓	✓	✓	✓	✓
x_2	✓	✓	✓	✓	✓
x_3	✓			✓	✓
x_4		✓		✓	
x_5	✓	✓	✓	✓	
x_6	✓	✓		✓	
x_7	✓	✓	✓		
x_8	✓	✓	✓	✓	

Table 4: Summary of the variable selection procedure. Checkmark is used to indicate when a variable is selected into a model

auxiliary variables:

$$\begin{aligned} \beta_{kj} &\sim N(0, \zeta_{kj}), j = 1, \ldots, p = 8 \\ \zeta_{kj} &= (1 - \gamma_{kj}) \times 0.001 + \gamma_{kj} \times M \\ \gamma_{kj} &\sim Ber(0.5) \end{aligned}$$

We set $M = 10$ after a careful sensitivity analysis. This value is conservative in allowing the selection of a relatively large number of regressors in the models. The resuls of variable selection are summarized in table 4.

We also consider more severe M, leading to more parsimonious models, but the effect on posterior distribution of θ_d is negligibile.

Appendix 4: estimation of the quintile share ratio

The quintile share ratio is defined as the sum of incomes in first quintile divided by the sum of incomes in the last. This measure of income inequality is not of direct interest in this research, but it is considered as it offers the opportunity to illustrate how the indirect methodology introduced to estimate the RMPG can be applied to estimate other summaries of the equivalized income distribution. A direct estimator of the quintile share ratio can be defined as follows:

$$\hat{\kappa}_d = \frac{\sum_{j=1}^{n_d} w_{dj} y_{dj} \mathbf{1}\{y_{dj} \geq \hat{q}_{0.8}(d)\}}{\sum_{j=1}^{n_d} w_{dj} y_{dj} \mathbf{1}\{y_{dj} \leq \hat{q}_{0.2}(d)\}} \tag{35}$$

where $\hat{q}_{0.2}$, $\hat{q}_{0.8}$ are the 20^{th} and 80^{th} percentiles of the equivalized income distribution estimated from the $d-th$ area-specific sample. See Langel and Tillé (2011) for more details.

We note that when the sample size is small, $\hat{q}_{0.2}$, $\hat{q}_{0.8}$ can be substantitally biased and $\hat{\kappa}_d$ as well. Moreover summations in (35) involve only 40% of the sample observations n_d, so the estimator $\hat{\kappa}_d$ is very likely to be very imprecise in small samples.

The quintile share ratio (κ_d) under the GB2 assumption is given by:

$$\kappa_{d|GB2} = \frac{1 - F_{(1)}\big(x_{80}, a_d, b_d, p_d, q_d\big)}{F_{(1)}\big(x_{20}, a_d, b_d, p_d, q_d\big)} \tag{36}$$

where $F_{(1)}(x_{80}, \dots) = E(X|X \leq x_{80})/E(X)$ is the incomplete moment of order 1 for the distribution truncated in the 80^{th} percentile and $F_{(1)}(x_{20}, \dots)$ is defined analoguously for the 20^{th} percentile.

An indirect estimator of κ_d can be obtained in the line illustrated in section 4.2.