Mapping non-monetary poverty at multiple geographical scales

Mapping non-monetary poverty at multiple geographical scales

Silvia De Nicolò,[1] **Enrico Fabrizi**[2] **and Aldo Gardini**[3,*]

[1, 3]Department of Statistical Sciences, University of Bologna, Italy and [2]DISES, Catholic University of the Sacred Heart, Italy

[*]Corresponding author. aldo.gardini@unibo.it

Abstract

Poverty mapping is a powerful tool to study the geography of poverty. The choice of the spatial resolution is central as poverty measures defined at a coarser level may mask their heterogeneity at finer levels. We introduce a small area multi-scale approach integrating survey and remote sensing data that leverages information at different spatial resolutions and accounts for hierarchical dependencies, preserving estimates coherence. We map poverty rates by proposing a Bayesian Beta-based model equipped with a new benchmarking algorithm accounting for the double-bounded support. A simulation study shows the effectiveness of our proposal and an application on Bangladesh is discussed.

Key words: Beta regression, Benchmarking, Demographic and Health Survey, Development economics, Small area estimation.

1. Introduction

The first goal of the 2030 Agenda for Sustainable Development of the United Nations is to eradicate poverty in all its forms. Given that, current research trends are focusing on the complexity, spatial heterogeneity, and geographical roots of poverty to properly design out-of-poverty paths (Allard and Allard, 2017; Fan and Cho, 2021; Zhou and Liu, 2022). Poverty has a multifaceted nature compounded by its spatial dimension (Gauci, 2005): as an example, we can mention that widely used classifications divide it into urban and rural poverty (Christiaensen and Todo, 2014).

For these reasons, poverty mapping is sparking interest in welfare economics and geography studies (Jean et al., 2016; Hall et al., 2023), to support the investigation of the spatial distribution and regional characteristics of poverty. In addition, if performed at a granular level, poverty mapping serves as a useful preliminary tool to identify poor households and to guide the placement and evaluation of geographically targeted interventions (Aiken et al., 2022). It enables the diversification of policy instruments employed from area to area, it can be easily combined with individual targeting or other antipoverty measures and it is recognized as an effective and cost-saving tool for poverty reduction (Bigman and Fofack, 2000; Galasso and Ravallion, 2005).

When performing poverty mapping, a natural problem relates to the choice of the spatial scale which strictly depends on the purpose of the analysis. This adds up to a more technical problem that arises when grouping areal data from neighboring zones into increasingly larger spatial units. Indeed, such an aggregation induces a smoothing effect that can mask or distort information on spatial heterogeneity (patchiness) of

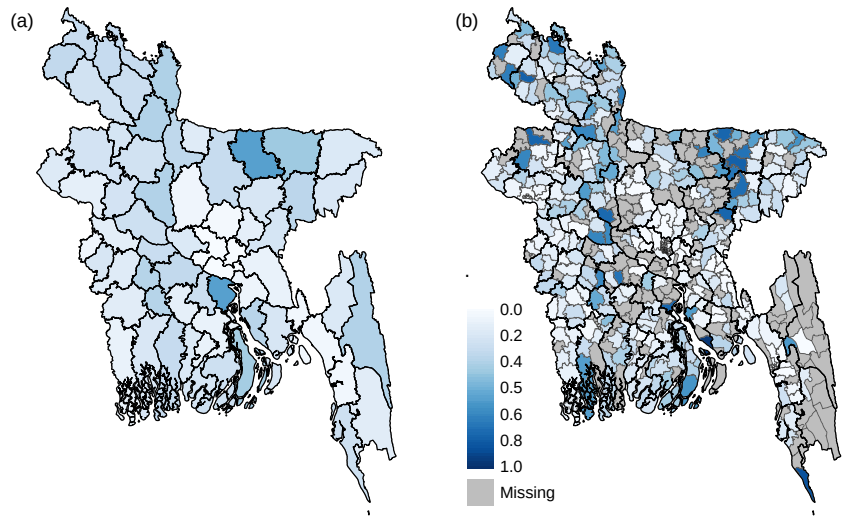

Fig. 1: Poverty rate estimates at zila level (district) on the left-hand side and at upazila level (sub-district) on the right-hand side. Note that upazila level heterogeneity is masked when aggregating at the zila level.

the target poverty measure (e.g., see Figure 1). This is known in geography as the scaling problem and represents one of the aspects of the Modifiable Area Unit Problem (MAUP, Waller and Gotway, 2004; Bradley et al., 2017). This effect strongly depends on the heterogeneity of the aggregated areas and might lead to misleading conclusions. A simple strategy for mitigating the impact of MAUP is to undertake the analysis by using multiple scales at once (Kolaczyk and Huang, 2001).

The choice of spatial scale is also strongly connected with the availability of data sources. In particular, surveys on income, consumption, and/or living conditions are commonly used in poverty mapping but typically provide reliable estimates at high levels of spatial aggregation (e.g., regional or national). Survey estimates are increasingly unreliable when considering finer spatial levels, translating into smaller and smaller sample sizes (Pratesi and Salvati, 2016). In addition, remote sensing (RS) or mobile phone usage data have recently been employed in poverty mapping, especially in developing countries (Jean et al., 2016; Schmid et al., 2017). As opposed to survey data, such alternative sources are highly informative at fine spatial levels, while losing power when aggregated.

A good practice in poverty mapping is to promote the efficient integration of multiple sources by means of statistical models (Steele et al., 2017; Zhao et al., 2019). To do so, we have to optimize the informative power that distinct sources have at different scales while accounting for hierarchical dependencies among levels. In this sense, a multi-scale procedure enables such an integration, while overcoming the aforementioned scaling problem. To the best of our knowledge, no explicit formulation of a

multi-scale modeling approach has been proposed in the poverty mapping literature. Some proposals exploit geo-statistical models (Puurbalanta, 2021; Sohnesen et al., 2022) or machine learning methods (Yeh et al., 2020; Lee and Braithwaite, 2022; Chi et al., 2022) to obtain a continuous surface estimation over the study area, avoiding the scaling problem. These methods rely on the assumption of infinite population, delivering an estimate of poverty risk. For policy targeting purposes, the focus is on poverty prevalence in administrative areas, i.e. taking into account also the distribution of the population over areas. Switching from risk to prevalence requires the population density at a very granular level, that is available only with uncertainty. Moreover, these methods also require an accurate localization of respondents units or clusters, information that is not always available to the analyst.

In contrast, our methodology sets in the framework of the Small Area Estimation (SAE) literature, aiming at estimating population quantities in areas (or domains) not originally planned by the survey design (Tzavidis et al., 2018; Rao and Molina, 2015, Ch. 4). Such techniques rely on the assumption of finite population and account for complex survey design. SAE methods have been widely used in poverty mapping (Pratesi, 2015; Corral et al., 2022); they either rely on models defined at the unit-level (Elbers et al., 2003; Molina and Rao, 2010; Molina et al., 2014) or at the area-level (Casas-Cordero Valencia et al., 2016). The latter ones are adopted by the US Census Bureau to map poverty through the SAIPE programme (Bell et al., 2016). Such programme is particularly relevant as it estimates poverty at multiple spatial scales: state, counties, and school districts, determining annual allocation of federal funds.

We use data from the Bangladesh Demographic and Health Survey (DHS; Fabic et al., 2012) by considering as non-monetary poverty measure the proportion of people in the area falling in the first quintile of the national distribution of the Wealth Index (WI; Rutstein and Johnson, 2004; Pirani, 2014). WI is a composite score that summarizes the living conditions of a household and that can be read as a measure of socioeconomic status (Poirier et al., 2020). Our goal is to create poverty maps at two distinct spatial levels: zilas, corresponding to administrative districts, and upazilas, equivalent to sub-districts comparable to counties or boroughs. While achieving this, it is essential that the poverty estimates maintain consistency across these layers and align with the overall national poverty level.

We adopt an area-level approach that we believe is reasonable to model the data available in our specific application, as clarified in what follows. Firstly, we do not have access to exact geographical locations of sampled households or clusters, which are released with addition of a significant spatial noise that does not guarantee their correct assignment to the target administrative unit. Secondly, individual level covariates for the entire population are not available, being crucial for standard unit-level strategies in SAE literature. On the other hand, we dispose of census population counts at the same administrative unit level, particularly useful to reach the desired consistency across layers through benchmarking.

The contribution of this paper to the literature is twofold. Firstly, we propose a Bayesian multi-scale model for poverty mapping that combines survey and alternative sources, available at both finer and coarser spatial levels, and does not require the exact location of survey respondents. Previous proposals that deal with multiple levels in SAE convey information in a single direction: from finer to coarser, as the so-called

sub-area or two-fold models (Torabi and Rao, 2014; Erciulescu et al., 2019; Krenzke et al., 2020), or from coarser to finer, through geo-statistical models (Benedetti et al., 2022). Our proposal, instead, leverages the information streams in both directions by considering multiple predictors with shared random effects, introducing in the poverty mapping framework the approach proposed by Aregay et al. (2016, 2017) in the disease mapping literature. The main adjustment with respect to the latter method relates to the use of survey data instead of population counts, involving a careful handling of sampling error and out-of-sample domains. To account for these features, we develop ad-hoc estimators for the target quantities at multiple levels.

The second contribution of the paper is proposing a new benchmarking algorithm for proportions, aimed at ensuring the consistency of small area estimates with those obtained for large domains or the whole population (Bell et al., 2013). Conventional benchmarking methods do not account for the bounded support of poverty rates and may return invalid outcomes, possibly lying below zero or above one. For instance, we mention the popular raking procedure employed, among the others, in the SAIPE programme (Bell et al., 2016). This issue is addressed in Okonek and Wakefield (2022), focusing on the case of inexact benchmarking. However, the sampling methods they propose are not efficient when exact benchmarking is required, as in our applied problem. Conversely, our procedure restricts the parameter space to the unit interval by adopting a different approach based on a suitable loss function. To obtain benchmarked estimates endowed with uncertainty measures, we rely on the idea of posterior projection (Dunson and Neelon, 2003; Sen et al., 2018).

The insufficient survey sample sizes in most zilas and upazilas (one-third of the upazilas are out-of-sample) drive us to integrate survey with RS data at the upazila level. Open access RS data encompasses night light radiance, population structure and population density, along with geographical characteristics, land use, infrastructures, and other social and economic features known to be poverty predictors at granular levels (Zhou and Liu, 2022).

Since our target is the estimation of poverty rates, we resort to Beta-based models (Liu et al., 2014; Fabrizi et al., 2016b; Janicki, 2020). More in detail, we consider the Extended Beta distribution proposed in De Nicolò et al. (2022) to account for some data characteristics. Indeed, our data presents challenging features such as the excess of zero/one values, the intra-cluster correlation induced by the survey design and the presence of out-of-sample areas.

The paper is organized as follows. Section 2 introduces the notation; Section 3 is devoted to a detailed description of the proposed small area model defined at multiple levels, prior specification, and posterior inference algorithms, along with a review of other standard models. The benchmarking proposal is illustrated in Section 4. Section 5 presents two simulation studies: a fully model-based simulation exercise with the twofold purpose of testing different estimation strategies within our model and assessing the impact of benchmarking on estimators properties and a second model-based simulation exercise that exploits a pseudo-population to test the performance of our model in comparison with alternatives. An application of poverty mapping in Bangladesh using DHS and RS data is illustrated in Section 6. Conclusions are drawn in Section 7.

2. Notation and survey-based estimators

The statistical models discussed afterward primarily rely on survey data. In this section, we introduce the notation and basic survey quantities, such as the poverty rate estimators and related uncertainty measures, which constitute the starting point of our modeling framework. Let $\mathcal{U}$ denotes a finite population of N individuals, on which we define two nested partitions, although the following setting can be easily generalized to multiple partitions. The first partition divides $\mathcal{U}$ into D non-overlapping domains $\mathcal{U}_d$ at a coarser level, labeled as *areas*, each constituted by N_d units. A second partition at a finer level divides each $\mathcal{U}_d$, $d = 1, \dots, D$, into M_d *sub-areas* $\mathcal{U}_{dj}$ of size N_{dj}, so that $M = \sum_{d=1}^{D} M_d$ is the overall number of sub-areas.

An overall sample of h households and n individuals is drawn from $\mathcal{U}$ according to a possible complex design. This sample can be partitioned into area and sub-area sub-samples of h_d and h_{dj} households and n_d and n_{dj} individuals, with $d = 1, \dots, D$ and $j = 1, \dots, \widetilde{M}_d$. $\widetilde{M}_d$ denotes the number of sub-areas in area d with a positive sample size, whereas sub-areas $j = \widetilde{M}_d + 1, \dots, M_d$ are those without units in the sample. The notations $\hat{\bar{Y}}_d$ and $\hat{\bar{Y}}_{dj}$ denote the estimators based on survey data (direct estimators) of a poverty rate for areas $d = 1, \dots, D$ and the nested sub-areas $j = 1, \dots, M_d$, respectively. Let w_{dji} denotes the individual sampling weight of household $i = 1, \dots, h_{dj}$ pertaining to area d and sub-area j, n_{dji} the number of household components and y_{dji} the indicator variable denoting its poverty status. As our target domains are not planned in the survey design, we employ an Hájek-type estimator (Hájek, 1971) defined as

$$\hat{\bar{Y}}_{dj} = \frac{\sum_{i=1}^{h_{dj}} w_{dji} n_{dji} y_{dji}}{\sum_{h=1}^{h_{dj}} w_{dji} n_{dji}}, \quad j = 1, \dots, \widetilde{M}_d, \ d = 1, \dots, D \tag{1}$$

$$\hat{\bar{Y}}_{d} = \frac{\sum_{j=1}^{\widetilde{M}_d} \sum_{i=1}^{h_{dj}} w_{dji} n_{dji} y_{dji}}{\sum_{j=1}^{\widetilde{M}_d} \sum_{i=1}^{n_{dj}} w_{dji} n_{dji}} = \frac{\sum_{j=1}^{\widetilde{M}_d} \left(\sum_{i=1}^{h_{dj}} w_{dji} n_{dji} \right) \hat{\bar{Y}}_{dj}}{\sum_{j=1}^{\widetilde{M}_d} \sum_{i=1}^{h_{dj}} w_{dji} n_{dji}}. \tag{2}$$

Those estimators are approximately unbiased for the unknown population proportions denoted with θ_d and θ_{dj}, assuming that $0 < \theta_d < 1$ and $0 < \theta_{dj} < 1$. Defining as t the national poverty rate and as $s_d = N_d/N$ the known population share, it holds that $\sum_{d=1}^{D} s_d \theta_d = t$. Furthermore, the sub-area partition implies

$$\sum_{j=1}^{M_d} s_{dj} \theta_{dj} = \theta_d, \quad d = 1, \dots, D, \tag{3}$$

where $s_{dj} = N_{dj}/N_d$ with $\sum_d s_d = \sum_j s_{dj} = 1$.

Survey estimates from (1) and (2) may be highly imprecise due to the small sample sizes. To assess the reliability of estimates, we need to introduce a measure of uncertainty, included also in the modeling framework of Section 3. To account for the complex design, we focus on the effective sample sizes $\tilde{n}_d$ and $\tilde{n}_{dj}$, i.e. the sample sizes that would be required by a simple random sample to retrieve estimators with standard errors equal to those of $\hat{\bar{Y}}_d$ and $\hat{\bar{Y}}_{dj}$ obtained under complex design. Focusing on the area case, we can express the sampling variance of $\hat{\bar{Y}}_d$ as $\theta_d(1-\theta_d)\tilde{n}_d^{-1}$, being the estimator of a proportion. The equivalent sample sizes incorporate the gains

and losses in efficiency due to stratification, weighting, and the possibly strong intra-cluster correlation. They can be expressed in terms of the design effect DEff_d, i.e. the ratio between the variance of the direct estimator and its variance under simple random sampling, as $\tilde{n}_d = n_d/\text{DEff}_d$. The same reasoning applies for $\hat{\bar{Y}}_{dj}$. We remark that, given the different correlation structures occurring within areas and sub-areas, generally $\tilde{n}_d \neq \sum_{j=1}^{M_d} \tilde{n}_{dj}$.

The estimated uncertainty measures allow the detection of spatial levels with unreliable estimates, thus requiring SAE techniques. In our framework, two different spatial levels are considered and we assume that they both present unreliable estimates as surveys are generally planned for national (or regional) aggregates.

3. Small Area Models

As motivated in Section 1, we focus on models defined at the area level. In particular, we consider a hierarchical Bayesian model for proportions based on an extended Beta distribution, introduced in De Nicolò et al. (2022), by extending it for multiple spatial scales. We recall its single-level specification in Section 3.1, whereas multi-level specifications follow in Section 3.2.

3.1. The extended Beta model

The extended Beta (EB) model is defined as a mixture of a Beta distribution and two Dirac components defined on the boundaries 0 and 1 (Fabrizi et al., 2016a; Janicki, 2020). Let us consider the mean-precision parametrization of the Beta random variable (Ferrari and Cribari-Neto, 2004) namely, if $Y|\mu,\phi \sim \text{Beta}(\mu\phi,(1-\mu)\phi)$, then its density function can be defined as

$$f_B(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\,\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \quad y \in (0,1).$$

Under such parametrization, $\mu \in (0,1)$ is the expectation and $\phi \in (0,+\infty)$ is a dispersion parameter as $\mathbb{V}[y|\mu,\phi] = \mu(1-\mu)(\phi+1)^{-1}$. This parametrization provides a first intuitive justification for the use of the Beta as a likelihood for proportions, as the first two moments match those of the proportion estimator under simple random sampling.

We use the notation $\hat{\bar{Y}}_{dj}|\mu_{dj},\phi_{dj},\pi_{0dj},\pi_{1dj} \sim EB(\mu_{dj},\phi_{dj},\pi_{0dj},\pi_{1dj})$ to define the extended Beta model for sub-area dj, where the mixture is made explicit as follows

$$\begin{aligned}\hat{\bar{Y}}_{dj}|\mu_{dj},\phi_{dj},\pi_{0dj},\pi_{1dj} \sim \pi_{0dj} \times \mathbf{1}\{\hat{\bar{Y}}_{dj}=0\} + \pi_{1dj} \times \mathbf{1}\{\hat{\bar{Y}}_{dj}=1\} + \\ + (1-\pi_{0dj}-\pi_{1dj}) \times \text{Beta}\left(\mu_{dj}\phi_{dj},(1-\mu_{dj})\phi_{dj}\right) \mathbf{1}\{\hat{\bar{Y}}_{dj} \in (0,1)\},\end{aligned} \tag{4}$$

with $\mathbf{1}\{A\}$ denoting an indicator function assuming value one if A occurs, and zero otherwise. The probabilities of observing 0 and 1 values are denoted by π_{0dj} and π_{1dj}, respectively. As previously mentioned, μ_{dj} is the location parameter that can be further modeled specifying a linear predictor with possible covariates and random effects. Lastly, in line with most literature on area-level models, the dispersion parameter ϕ_{dj} is considered to be known to allow identifiability and replaced by $\tilde{n}_{dj}-1$ (Janicki, 2020). We assume that direct estimates equal to 0 or 1 are due to a censoring

process, as $\theta_{dj} \in (0,1)$ and the observation of boundary values 0 or 1 occurs as a result of scarce sample sizes or strong intra-cluster correlations.

We model π_{0dj} and π_{1dj} in a parsimonious way, accounting for sample features and probabilistic assumptions. Firstly, μ_{dj} is interpreted as the expectation of non-censored observations. Secondly, the observed individual poverty status has a dependency structure which is simplified by assuming an exchangeable dependency among each pair of observations within areas. Under this assumption, the two probabilities are defined as $\pi_{1dj} = \mu_{dj}\lambda_s^{h_{dj}-1}$ and $\pi_{0dj} = [1+\mu_{dj}(\lambda_s-2)]^{h_{dj}-1}/(1-\mu_{dj})^{h_{dj}-2}$. The correlation between observations is modeled by λ_s which has bounded support: $\max\left\{0, \max_d \frac{2\mu_{dj}-1}{\mu_{dj}}\right\} \leq \lambda_s \leq 1$ (De Nicolò et al., 2022).

Under the EB model, the population proportion θ_{dj} is defined by its expectation

$$\mathbb{E}\left[\hat{\bar{Y}}_{dj}|\mu_{dj},\lambda_s\right] = \left[1-\mu_{dj}\lambda_s^{h_{dj}-1} - \frac{[1+\mu_{dj}(\lambda_s-2)]^{h_{dj}-1}}{(1-\mu_{dj})^{h_{dj}-2}}\right]\mu_{dj} + \mu_{dj}\lambda_s^{h_{dj}-1}. \quad (5)$$

In the following subsections, the EB is assumed for area-specific direct estimates, coherently adjusting the notation as $\hat{\bar{Y}}_d|\mu_d,\phi_d,\pi_{0d},\pi_{1d} \sim EB(\mu_d,\phi_d,\pi_{0d},\pi_{1d})$ and with correlation parameter λ_a.

3.2. Small area models on multiple spatial scales

As mentioned before, poverty has a geographical pattern that varies with respect to the spatial scale of measurement. This feature can be modeled through proper tools known in spatial statistics as multi-scale models. A widespread approach is to account for the scaling effect by exploiting a likelihood factorization holding for Gaussian and Poisson models (Kolaczyk and Huang, 2001). This is popular in the disease mapping framework, where responses are counts of rare events occurrences (Louie and Kolaczyk, 2006), which is not our case. Another way to implement multi-scale models is to follow the proposal by Aregay et al. (2016, 2017), building models with distinct linear predictors for each spatial scale that share common random effects. It represents an interesting procedure as it is applicable to any distributional assumption for the response. This is particularly useful since we need Beta-based models to correctly account for the bounded support of poverty rates. For this reason, we extend the approach of Aregay et al. (2016) to deal with poverty rate estimation through the EB likelihood.

In the area-level models literature, the sole proposal that handles multiple spatial scales is the sub-area (or two-fold) model, usually outlined under Gaussian assumptions (Torabi and Rao, 2014; Erciulescu et al., 2019). However, it cannot be considered as a proper multi-scale model since only sub-area direct estimates and associated uncertainty measures are employed. In this way, the survey information available at the coarser level is discarded, even if more reliable. In contrast, our modeling proposal, labeled as shared multi-scale model, employs survey data at both levels and is presented in Section 3.2.1. Sub-area model with EB likelihood is outlined in Section 3.2.2 and, in Section 3.2.3 we define as a benchmark a multi-scale model without shared random effects, labeled as independent multi-scale model.

3.2.1. Shared multi-scale model

The proposed shared multi-scale (S-MS) model is defined on two distinct spatial levels:

$$\hat{\bar{Y}}_d|\mu_d,\phi_d,\pi_{0d},\pi_{1d} \overset{ind}{\sim} EB(\mu_d,\phi_d,\pi_{0d},\pi_{1d}), \quad d=1,\dots,D,$$
$$\text{logit}\,(\mu_d) = \alpha_a + \mathbf{x}_d^T\boldsymbol{\beta}_a + u_d; \tag{6}$$
$$\hat{\bar{Y}}_{dj}|\mu_{dj},\phi_{dj},\pi_{0dj},\pi_{1dj} \overset{ind}{\sim} EB(\mu_{dj},\phi_{dj},\pi_{0dj},\pi_{1dj}), \quad j=1,\dots\widetilde{M}_d,\ d=1,\dots,D,$$
$$\text{logit}\,(\mu_{dj}) = \alpha_s + \mathbf{x}_{dj}^T\boldsymbol{\beta}_s + v_{dj} + u_d; \tag{7}$$

where $\mathbf{x}_d$ and $\mathbf{x}_{dj}$ are vectors of area and sub-area covariates. The corresponding coefficients $\boldsymbol{\beta}_a$ and $\boldsymbol{\beta}_s$ differ between predictors (6) and (7) to account for changes in the functional relationship induced by the spatial scale (ecological bias). Note that such a model promotes an exchange of information between levels through the use of the shared area-specific random effect u_d. In such a way, the information nested at finer levels directly contributes to poverty estimation at coarser levels, enabling the communication flow to be two-way. Moreover, u_d accounts for the correlation between levels.

The poverty rate θ_{dj} related to in-sample sub-areas is assumed to be equal to the model expectation in (5), i.e.

$$\theta_{dj} = \mathbb{E}\left[\hat{\bar{Y}}_{dj}|\mu_{dj},\lambda_s\right],\ j=1,\dots,\widetilde{M}_d,\ \forall d. \tag{8}$$

Due to the granularity of the finer administrative level, several sub-areas can be out-of-sample (OOS), while having the associated auxiliary information. The prediction of related poverty rates is crucial and we define their predictor as follows

$$\theta_{dj}^{OOS} = \text{logit}^{-1}\left(\alpha_s + \mathbf{x}_{dj}^T\boldsymbol{\beta}_s + v_{dj} + u_d\right), j=\widetilde{M}_d+1,\dots,M_d,\forall d. \tag{9}$$

Such a predictor is basically equal to μ_{dj}, and does not include sampling-related quantities such as h_{dj} and λ_s. This has a simple motivation: in the case of OOS domains, the sampling process does not exist by definition and the sampling model, as well as its expectation in (5), is not defined. By considering the degenerate case in which a sole household is drawn for sub-area dj, i.e. $h_{dj}=1$, the direct estimate would be only 0 or 1 and, by equations (5) and (8), θ_{dj} boils down to μ_{dj}. Such behaviour, as a heuristic solution, is considered also for the degenerate case $h_{dj}=0$.

Consequently, the predictor for area poverty rates is obtained by combining (8) and (9) as:

$$\theta_d = \sum_{j=1}^{\widetilde{M}_d} s_{dj}\theta_{dj} + \sum_{j=\widetilde{M}_d+1}^{M_d} s_{dj}\theta_{dj}^{OOS},\ d=1,\dots,D; \tag{10}$$

considering as known the population shares s_{dj}. Such a predictor presents some appealing features: firstly, it is able to exploit all the available auxiliary information at the finer spatial level; secondly, it contains also valuable information coming from the coarser level through the shared random effect u_d. In addition, it automatically preserves the coherence between poverty rates defined at different levels.

Even if the weighted mean that characterizes predictor (10) is able to carry on information from different spatial layers at once, we are aware that another specification for the area-level predictor is available, i.e. the basic model expectation

$$\theta_d^{\bullet} = \mathbb{E}\left[\hat{\bar{Y}}_d|\mu_d, \lambda_a\right], \; d = 1, \dots, D. \tag{11}$$

While it may seem to be a natural predictor for the proposed model, it presents some drawbacks in the multi-scale problem we face. Indeed, if the areas are constituted by heterogeneous sub-areas, the scaling problem at the coarser spatial level induces covariates to be less powerful in predicting the rate. Conversely, if the auxiliary information at the finer level is not relevant or it is homogeneous across sub-areas and/or there are no OOS domains, predictor (11) can be a viable alternative to (10), as observed in the epidemiological mapping problem discussed by Aregay et al. (2016). To better clarify these aspects, we compare the behaviours of the two alternative predictors by means of a simulation study detailed in Section 5.1.

3.2.2. Sub-area models

The sub-area (SA) model defined with the EB likelihood is defined as:

$$\begin{aligned}
&\hat{\bar{Y}}_{dj}|\mu_{dj}, \phi_{dj}, \pi_{0dj}, \pi_{1dj} \overset{ind}{\sim} EB(\mu_{dj}, \phi_{dj}, \pi_{0dj}, \pi_{1dj}), \quad j = 1, \dots \widetilde{M}_d, \; d = 1, \dots, D; \\
&\text{logit}\,(\mu_{dj}) = \alpha_s + \mathbf{x}_{dj}^T \boldsymbol{\beta}_s + \mathbf{x}_d^T \boldsymbol{\beta}_a + v_{dj} + u_d.
\end{aligned} \tag{12}$$

Such a model accounts for correlation within areas through u_d, but it ignores the multiple scales of the problem since u_d is not informed by any quantity defined at the coarser level. Similarly to the S-MS model, we assume $\theta_{dj} = \mathbb{E}\left[\hat{\bar{Y}}_{dj}|\mu_{dj}, \lambda_s\right]$; whereas, to obtain estimates at the area level, θ_d is defined exploiting the hierarchical partition as in (10). Note that, in contrast with the other models, the regression defined at the sub-area level uses both area and sub-area covariates (Erciulescu et al., 2019).

3.2.3. Independent multi-scale model

A naive approach to produce estimates at distinct levels of disaggregation would be fitting a model with two independent layers. In this way, the auxiliary information at a given level is combined with corresponding direct estimates, determining a multi-scale modeling procedure. The independent multi-scale (I-MS) model is defined as follows:

$$\begin{aligned}
&\hat{\bar{Y}}_d|\mu_d, \phi_d, \pi_{0d}, \pi_{1d} \overset{ind}{\sim} EB(\mu_d, \phi_d, \pi_{0d}, \pi_{1d}), \quad d = 1, \dots, D; \\
&\text{logit}\,(\mu_d) = \alpha_a + \mathbf{x}_d^T \boldsymbol{\beta}_a + u_d \\
&\hat{\bar{Y}}_{dj}|\mu_{dj}, \phi_{dj}, \pi_{0dj}, \pi_{1dj} \overset{ind}{\sim} EB(\mu_{dj}, \phi_{dj}, \pi_{0dj}, \pi_{1dj}), \quad j = 1, \dots \widetilde{M}_d, \; d = 1, \dots, D; \\
&\text{logit}\,(\mu_{dj}) = \alpha_s + \mathbf{x}_{dj}^T \boldsymbol{\beta}_s + v_{dj}.
\end{aligned} \tag{13}$$

The target proportions are defined as $\theta_d = \mathbb{E}\left[\hat{\bar{Y}}_d|\mu_d, \lambda_a\right]$ and $\theta_{dj} = \mathbb{E}\left[\hat{\bar{Y}}_{dj}|\mu_{dj}, \lambda_s\right]$, with the model expectations defined in (5). As a consequence, the coherency between rates at multiple levels is not preserved. In addition, the linear predictor related to the finer level in (13) does not present an area-specific random effect to model the hierarchical dependence between layers.

3.3. Prior distributions

The three defined models receive the same priors for the corresponding parameters. Concerning the intercepts α_a and α_s, a zero-mean Gaussian prior with a standard deviation of 2.5 is set. A regularized horseshoe prior (Piironen and Vehtari, 2017) is adopted for the vectors of regression coefficients $\boldsymbol{\beta}_a$ and $\boldsymbol{\beta}_s$, in order to automatically incorporate the variable selection step, mimicking the behavior of a spike-and-slab prior. Such a shrinkage prior is particularly appealing since it is defined as a mixture of continuous distributions, being implementable in `Stan`, but forcing to zero the coefficients related to negligible covariates. In the case of P area-specific covariates, it is defined at the area layer as follows:

$$\beta_{ai}|\zeta_i, \tau, \iota \overset{ind}{\sim} \mathcal{N}\left(0, \tau^2 \frac{\iota^2 \zeta_i^2}{\iota^2 + \tau^2 \zeta_j^2}\right), \; i = 1, \ldots, P; \tag{14}$$

with $\zeta_i \overset{ind}{\sim}$ Half-Cauchy$(0, 1)$, $\forall i$, $\iota^2 \sim$ Inverse-Gamma$\,(2.5, 2.5)$ and $\tau \sim$ Half-Cauchy$(0, \tau_0)$. τ_0 represents an important hyperparameter to set, defined as $\tau_0 = \frac{P_0 \tilde{s}}{(P - P_0)\sqrt{D}}$, where P_0 is an initial guess of the number of non-zero coefficients and $\tilde{s}^2$ is the pseudo-variance of a generic observation under the assumed model. The same setting is applied to $\boldsymbol{\beta}_s$. Both the sets of area and sub-area specific random effects present a variance-gamma shrinkage prior (Tang et al., 2018). In the case of area-specific random effects, the hierarchy that defines the prior is:

$$\begin{aligned} u_d|\xi_d, \sigma_u &\overset{ind}{\sim} \mathcal{N}\left(0, \xi_d \sigma_u^2\right), \; d = 1, \ldots, D; \\ \xi_d &\overset{ind}{\sim} \text{Gamma}(0.5, 1), \; d = 1, \ldots, D; \\ \sigma_u &\sim \text{Half-}\mathcal{N}(0, 1); \end{aligned} \tag{15}$$

where the set of ξ_d, $\forall d$, denotes local scales, whereas σ_u is a global scale. A parallel structure is build for the sub-area random effect v_{dj}, $\forall j, d$, introducing the global scale σ_v. To complete the model, a uniform prior is specified for the correlation parameter $\lambda_a|\mu_1, \ldots, \mu_D \sim \text{Unif}\left(\max\{0, \max_d(2\mu_d - 1)/\mu_d\}; 1\right)$, and the same for λ_s with μ_d replaced by μ_{dj}.

3.4. Posterior Inference

Posterior inference has been performed via Markov Chain Monte Carlo (MCMC) techniques. This has been implemented through Hamiltonian Monte Carlo (HMC) sampling algorithm via the `Stan` language and the `rstan` package (Carpenter et al., 2017). The estimation has been performed with 4 chains, each including 4,000 iterations, where the first 2,000 have been considered as warm-up iterations and discarded. The convergence of the algorithm is monitored through the Gelman-Rubin statistic, and the chain autocorrelation is checked through the effective sample size. The expectations of $\theta_d|\mathbf{y}$ and $\theta_{dj}|\mathbf{y}$ are adopted as point estimators, labeled as model-based or hierarchical Bayes (HB) estimators: $\hat{\theta}_d = \mathbb{E}[\theta_d|\mathbf{y}]$ and $\hat{\theta}_{dj} = \mathbb{E}[\theta_{dj}|\mathbf{y}]$. Their estimates, together with other posterior summaries such as credible intervals and posterior variances, can be easily approximated exploiting MCMC draws. Note that $\theta_d|\mathbf{y}$, if defined through (10), can be retrieved by combining draws from $\theta_{dj}|\mathbf{y}$ and

$\theta_{dj}^{OOS}|\mathbf{y}$. Posterior samples from the latter quantity are obtained exploiting draws from $\boldsymbol{\beta}_s|\mathbf{y}$ and $u_d|\mathbf{y}$. As v_{dj} constitutes a random effect from an unobserved sub-area, we propagate the uncertainty by drawing samples from the variance-gamma shrinkage priors (De Nicolò et al., 2022). The HB estimators under the EB model enjoy the property of design-consistency, namely conditioning on higher level parameters $\hat{\theta}_d \xrightarrow{p} \hat{\bar{Y}}_d$ and $\hat{\theta}_{dj} \xrightarrow{p} \hat{\bar{Y}}_{dj}$ (see Section S1 in the Supplementary Material). Lastly, as a measure of model goodness-of-fit, we adopt the leave-one-out information criterion (LOOIC, Vehtari et al., 2017), particularly useful in the model comparison step. Lastly, the posterior predictive distribution is exploited to assess model assumptions implementing posterior predictive checks (Gabry et al., 2019).

4. The Bayesian benchmarking proposal

The definition of area level estimates through (10) preserves the coherency of poverty rates between the two levels. However, the coherency with respect to higher levels of aggregation (e.g., national) is not guaranteed. Indeed, the poverty rate at higher levels may be known in the population or reliably estimated through surveys; such value, hereafter called benchmark, might be used to constrain the linear combination of sub-area estimates. In this section, we consider the problem of finding a set of constrained estimators $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_{11}, \dots \tilde{\theta}_{DM_d})^T$, for the target parameters $\boldsymbol{\theta} = (\theta_{11}, \dots \theta_{DM_d})^T$. The linear constraint is imposed by the poverty rate at the finest spatial scales and it is defined by the combination of both equations in (3) as

$$\sum_{d=1}^{D} \sum_{j=1}^{M_d} q_{dj} \tilde{\theta}_{dj} \quad = \quad t, \tag{16}$$

where $q_{dj} = s_{dj} s_d$ denotes the population share of sub-area dj at national level. Note that $\sum_d \sum_j q_{dj} = 1$, and t is, in our case, the poverty rate at the national level.

The problem of benchmarking has been tackled from different perspectives in the Bayesian literature. The most popular strategy follows a decision-theoretic framework: starting from a loss function $L(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$, the benchmarked estimators are obtained by minimizing the posterior risk $\mathbb{E}[L(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})|\mathbf{y}]$ under linear constraint (Datta et al., 2011). Despite its simple implementation, this method is applied to point estimators obtaining only a set of benchmarked point estimators. Thus, it represents a non-fully Bayesian strategy, not delivering posterior measures of uncertainty. To solve this problem, fully Bayesian methods have been developed. Among the others, Zhang and Bryant (2020) include the constraint through a suitable prior distribution; Janicki and Vesper (2017) search for a constrained joint posterior distribution with minimum Kullback-Leibler distance to the unconstrained one, whereas Okonek and Wakefield (2022) opt for an importance sampling approach.

Our approach relies on the concept of posterior projection (Dunson and Neelon, 2003; Sen et al., 2018; Nandram et al., 2022) that has already been considered by Patra (2019) within the SAE framework. In this case, the posterior samples from $\theta_{dj}|\mathbf{y}$ are projected into the feasible set defined by the benchmarking constraint, minimizing the distance from the original ones. Specifically, this distance is defined by means of a loss function. In contrast to other fully Bayesian methods, such an approach is more

computationally efficient, particularly when the fitted model is not trivial as in our case. The choice of a suitable loss function for our inferential problem is discussed in Section 4.1 and the main results are illustrated in Section 4.2.

4.1. Choice of the loss function

The mapping required for posterior projection can be represented by a loss function $L(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$. Defining with ψ_{dj} a generic weight, the most common loss function is the weighted quadratic one, i.e. $L(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) = \sum_{d=1}^{D} \sum_{j=1}^{M_d} \psi_{dj}(\theta_{dj} - \tilde{\theta}_{dj})^2$. This function is suitable when each θ_{dj} is defined on the real line but not appropriate when it is bounded, as in our case where $\boldsymbol{\theta} \in \Theta = (0,1)^M$. Coherently, the constrained parameter space has to be bounded too, being defined as $\tilde{\Theta} = \{\tilde{\boldsymbol{\theta}} \in (0,1)^M | \sum_d \sum_j q_{dj} \tilde{\theta}_{dj} = t\}$. A similar issue is considered by Ghosh et al. (2015) which propose a generalized Kullback-Leibler loss function that restricts the parameter space to positive-defined values and avoids estimates falling below 0. A possible adaptation to the case in $(0,1)$ has been proposed by Aitchison (1992) contemplating the simple logit transformation, i.e. $\sum_{j=1}^{M_d} \psi_{dj}(\text{logit}(\theta_{dj}) - \text{logit}(\tilde{\theta}_{dj}))^2$. Such an option, while being intuitively simple, leads to a non-convex function.

We propose a weighted loss function pertaining to the Bregman family (Banerjee et al., 2005) that suitably restricts both $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$ to lie on the unit interval, defined as follows

$$L(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) = \sum_{d=1}^{D} \sum_{j=1}^{M_d} \psi_{dj} \left[\theta_{dj} \log \left(\frac{\theta_{dj}}{\tilde{\theta}_{dj}} \right) + (1 - \theta_{dj}) \log \left(\frac{1 - \theta_{dj}}{1 - \tilde{\theta}_{dj}} \right) \right]. \tag{17}$$

In this way, we target all the solutions in $\tilde{\Theta}$ which minimize $L(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})|\mathbf{y}$. The class of Bregman loss functions includes a number of functions defined over different domains, including the quadratic and the Kullback-Leibler loss functions considered by Ghosh et al. (2015). Such a class of functions is appealing since it ensures that the posterior risk $\mathbb{E}[L(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})|\mathbf{y}]$ is minimized by the posterior expectation $\mathbb{E}[\boldsymbol{\theta}|\mathbf{y}]$, a popular Bayes estimator in the small area literature. Figure S1 in the Supplementary Material displays the Bregman loss function in comparison with the quadratic loss function. Note that the proposed function is strictly defined in $(0,1)$, being symmetric only when $\theta = 0.50$.

To fully define a weighted loss function, its weights must be specified by the user. Popular choices for ψ_{dj} are generally q_{dj}, $q_{dj}/\hat{\theta}_{dj}$, or an inversely proportional function of the sampling variances and/or the posterior variances of target parameters (Rao and Molina, 2015, Ch. 6). In our case, we opt for the conservative option $\psi_{dj} = q_{dj}$, so that all estimates are adjusted depending only on their distance to the boundaries of the support.

4.2. Benchmarked posterior projection

The benchmarking strategy relying on the posterior projection approach is based on the idea of drawing samples from the unconstrained posterior $p(\boldsymbol{\theta}|\mathbf{y})$, defined on Θ, and projecting them on the constrained parameter space $\tilde{\Theta}$, inducing a mapping function $f : \Theta \to \tilde{\Theta}$. The following theorem determines the projection mapping that must be applied to the unconstrained estimators.

Theorem 1 *The projection of* $\boldsymbol{\theta}|\mathbf{y}$ *on* $\tilde{\Theta} \subseteq (0,1)^M$ *induced by the minimum distance mapping based on* (17) *is a random variable denoted with* $\tilde{\boldsymbol{\theta}}|\mathbf{y}$*. The induced transformation is unique and given by:*

$$\tilde{\theta}_{dj}|\mathbf{y} = \frac{1}{2} + \frac{\psi_{dj}}{2\gamma q_{dj}} \left\{ \sqrt{\left(1 - \gamma \frac{q_{dj}}{\psi_{dj}}\right)^2 + 4\theta_{dj}\gamma \frac{q_{dj}}{\psi_{dj}}} - 1 \right\} \Big| \mathbf{y} \tag{18}$$

where γ *is given as the solution of the non-linear equation*

$$\frac{1}{2\gamma} \left[\sum_{d=1}^{D} \sum_{j=1}^{M_d} \psi_{dj} \left\{ \sqrt{\left(1 - \frac{\gamma q_{dj}}{\psi_{dj}}\right)^2 + 4\gamma\theta_{dj} \frac{q_{dj}}{\psi_{dj}}} - 1 \right\} + \gamma q_{dj} \right] = t. \tag{19}$$

Proof See Section S2 in the Supplementary Material. □

When inference is carried out via MCMC methods, draws from $p(\tilde{\boldsymbol{\theta}}|\mathbf{y})$, i.e. the projected posterior, can be easily obtained by applying the transformation defined in Theorem 1 to posterior samples from $p(\boldsymbol{\theta}|\mathbf{y})$. The related posterior summaries can be computed as described in Section 3.4, with point estimators denoted with $\hat{\tilde{\theta}}_d$ and $\hat{\tilde{\theta}}_{dj}$.

As mentioned before, non-fully benchmarking methods target the constrained minimization of a posterior risk, i.e. $\mathbb{E}[L(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})|\mathbf{y}]$. Note that our projection approach targets the constrained minimization of $L(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})|\mathbf{y}$, instead. One could argue that the conceptual approach is distinct and may lead to different results with respect to the standard posterior risk minimization, inducing non-comparability issues. Due to the linearity property of the derivative of the Lagrangian function with respect to θ_{dj}, the constrained minimization of the posterior risk leads to a solution that is identical to (18) and (19) but with $\theta_{dj}|\mathbf{y}$ replaced by $\mathbb{E}[\theta_{dj}|\mathbf{y}]$. Thus, Jensen's inequality explains the discrepancy between the two results. When $\psi_{dj} = q_{dj}$, the mapping function $f : \Theta \to \tilde{\Theta}$ may be either slightly concave or slightly convex depending on the sign and magnitude of $t - \sum_d \sum_j q_{dj}\theta_{dj}$ as visually illustrated in Figure S2 of the Supplementary Material: in this case, the discrepancy is negligible. In contrast, if $\psi_{dj} \neq q_{dj}$, the mapping function may be highly non-linear and results could be incomparable.

5. Simulation studies

In this section, we study how statistical models introduced in Section 3 deal with a multi-scale data problem by evaluating the frequentist properties of resulting estimators. To delve into various aspects that deserve investigation, we present two distinct studies. Firstly, we compare different estimation strategies within S-MS model framework, as discussed in Section 3.2.1, to identify the most suitable one and we examine the impact of benchmarking on frequentist properties through a fully model-based study (Section 5.1). This involves generating direct estimates from a small area model. However, such a setting is not suitable for comparing the performances of different models. To this aim, we perform a second study involving a model-based simulation, that generates a pseudo-population from which we iteratively draw samples mimicking the application setting, discussed in Section 5.2.

		ARAB			ARRMSE			25% Poorest areas		MAE
Sc.	Est.	IS	OOS	Total	IS	OOS	Total	AAB_{25}	$ARMSE_{25}$	Total
1	$\hat{\theta}_i^{\bullet}$	-	-	0.108	-	-	0.217	0.020	0.054	0.037
	$\hat{\tilde{\theta}}_i^{\bullet}$	-	-	0.108	-	-	0.209	0.020	0.052	0.036
	$\hat{\theta}_i$	-	-	0.081	-	-	0.209	0.017	0.050	0.036
	$\hat{\tilde{\theta}}_i$	-	-	0.086	-	-	0.206	0.016	0.048	0.035
2	$\hat{\theta}_i^{\bullet}$	0.130	0.241	0.186	0.228	0.349	0.288	0.069	0.092	0.056
	$\hat{\tilde{\theta}}_i^{\bullet}$	0.125	0.159	0.142	0.242	0.311	0.276	0.033	0.071	0.044
	$\hat{\theta}_i$	0.079	0.135	0.107	0.227	0.289	0.258	0.030	0.065	0.048
	$\hat{\tilde{\theta}}_i$	0.101	0.142	0.122	0.233	0.284	0.259	0.027	0.062	0.044

Table 1. Summary measures concerning the estimators at the area level for the simulation study of Section 5.1. ARAB, ARRMSE, and MAE are computed on all the areas; AAB_{25} and $ARMSE_{25}$ concerns the 25% poorer areas. Concerning ARAB and ARRMSE, values obtained for the areas without missing sub-areas (IS), with missing sub-areas (OOS) and overall (Total) are reported for Scenario 2.

5.1. Fully model-based simulation

The primary goal of this simulation exercise is to assess two predictors under the S-MS model: θ_d as defined in equation (10), and $\theta_d^{\bullet}$ as presented in equation (11). In addition, the comparison of predictors before and after the benchmarking step, the latter ones being denoted by $\tilde{\theta}_d$ and $\tilde{\theta}_d^{\bullet}$, is a focal point of discussion. To achieve these objectives, a fully model-based simulation study is chosen. A pro of such an approach is to better control the impact of a covariate on the generative process, as well as to design proper out-of-sample schemes that in turn depend on the covariate. This, indeed, is in line with the conditions outlined in the motivating application, which will be discussed in Section 6.

The study assumes a population subdivided into $D = 20$ areas, each containing $M_d = 10$ sub-areas. Within each area, the number of sub-areas with households $h_{dj} = 30$, $h_{dj} = 60$, or $h_{dj} = 90$ is distributed as 5, 3, and 2, respectively. Additionally, a sub-area covariate x_{dj} is generated from a Gaussian distribution with a mean of zero and a standard deviation of 0.8. A total of $B = 1,000$ direct estimates associated with sub-areas are randomly generated under the sub-area model (12), with parameter values set to $\alpha_s = -1.4$, $\beta = -1$ (representing the coefficient related to x_{dj}), $\lambda = 0.8$, $\sigma_u = 0.2$, and $\sigma_v = 0.2$. Subsequently, area-level estimates are obtained through aggregation.

Let us indicate with θ_i as the true value for the generic area or sub-area i and $\hat{E}_i^{(b)}$ as the generic estimates obtained at iteration b, with $E_i \in \left\{\theta_i, \tilde{\theta}_i, \theta_i^{\bullet}, \tilde{\theta}_i^{\bullet}\right\}$. To determine the 90% credible intervals we consider 5^{th} and 95^{th} posterior percentiles, labeled as $E_{i,L}^{(b)}$ and $E_{i,U}^{(b)}$. We define bias, root mean squared error (RMSE), and

coverage for 90% credible intervals for each area or sub-area as

$$\text{Bias}\left[\hat{E}_i\right] = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{E}_i^{(b)} - \theta_i\right), \quad \text{RMSE}\left[\hat{E}_i\right] = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\hat{E}_i^{(b)} - \theta_i\right)^2},$$

$$\text{Cov}\left[E_i\right] = \frac{1}{B}\sum_{b=1}^{B}\mathbf{1}\left\{\theta_i \in \left[E_{i,L}^{(b)}; E_{i,U}^{(b)}\right]\right\}.$$

As summary measures of bias and errors over the areas, we define the average relative absolute bias (ARAB), the average relative root mean squared error (ARRMSE), and the mean absolute error (MAE) as

$$\text{ARAB} = \frac{1}{I}\sum_{i=1}^{I}\left|\frac{\text{Bias}\left[\hat{E}_i\right]}{\theta_i}\right|, \quad \text{ARRMSE} = \frac{1}{I}\sum_{i=1}^{I}\frac{\text{RMSE}\left[\hat{E}_i\right]}{\theta_i},$$

$$\text{MAE} = \frac{1}{IB}\sum_{i=1}^{I}\sum_{b=1}^{B}\left|\hat{E}_i^{(b)} - \theta_i\right|.$$

Since identifying areas characterized by extreme poverty is a primary objective in poverty mapping, it becomes crucial to assess the performance of small area models under these circumstances. To achieve this, the effectiveness of the estimators in the subset comprising the poorer 25% of areas is further evaluated using absolute indicators: $\text{AAB}_{25} = I_{25}^{-1}\sum_{i=1_{25}}^{I_{25}}\left|\text{Bias}\left[\hat{E}_i\right]\right|$ and $\text{ARRMSE}_{25} = I_{25}^{-1}\sum_{i=1}^{I_{25}}\text{RMSE}\left[\hat{E}_i\right]$.

The estimators undergo evaluation in two distinct scenarios: 1) where all sub-areas are sampled, and 2) where, in half of the areas, 40% of sub-areas are OOS. Note that the process that determines OOS sub-areas depends on the covariate x_{dj}. The latter scenario is designed to replicate a situation of selection bias for sampled sub-areas, mirroring the motivating application.

The comprehensive summaries of the results are presented in Table 1, while Figure S3 in the Supplementary Materials illustrates the distribution of area-specific measures. In the first scenario, the various estimation strategies perform comparably. However, the proposed method, i.e., $\hat{\theta}_i$ and $\hat{\tilde{\theta}}_i$, before and after benchmarking, respectively, exhibits lower bias and errors (RMSE and MAE), either overall either in the poorest areas. The estimates obtained in this scenario appear consistent with the overall level set by the benchmark; for this reason, the impact of benchmarking is reasonably negligible. In contrast, clearer differences emerge in scenario 2, where the gap between the alternative predictors widens. The proposed method demonstrates significantly better performance, particularly in areas with OOS sub-areas. Indeed, the predictor θ_i, which leverages auxiliary information at a finer level of aggregation, yields estimates less biased and in accordance with the benchmark. On the other hand, the estimator relying solely on coarser level auxiliary information exhibits a severe underestimation (see Figure S3), a condition mitigated by the impact of benchmarking. This mitigation results in both bias and error reduction, while also improving coverage (refer to Figure S3). We remark that the objective of benchmarking procedure is not to improve the frequentist performances of estimates, but rather

to constrain the results to respect important administrative vincula. For this reason, a possible slight worsening of such properties after benchmarking is not a concern.

5.2. Model-based simulation with pseudo-population

The attention now shifts to a comparative analysis of estimators derived from various models. In this context, we opt for a model-based simulation study that uses a pseudo-population, particularly advantageous when the objective is to compare different models while accurately capturing the intricacies of a multi-scale problem within a survey design setting. To achieve this, we explore two aggregation levels, i.e., areas and sub-areas, and examine two distinct scenarios that account for varying ratios of between and within-area variability.

For each scenario, a population of $N = 180,000$ individuals is organized into $C = 3,600$ clusters, each containing 50 units, thereby emulating the two-stage structure common in various survey schemes. These units are further classified into $D = 30$ areas and $M = 150$ sub-areas, where each area consists of $M_d = 5$ sub-areas with distinct numbers of clusters $C_{dj} \in \{10, 10, 20, 30, 50\}$. A continuous variable is generated for each unit using a three-fold Gaussian model with an individual standard deviation of $\sigma_\varepsilon = 0.2$, cluster effect scale of $\sigma_c = 0.2$, area effect scale of σ_a, and sub-area effect scale of σ_s. In scenario 1, the sub-area variation ($\sigma_s = 0.13$) dominates the area variation ($\sigma_a = 0.08$), resulting in pronounced heterogeneity within the areas. In scenario 2, the area variation ($\sigma_a = 0.13$) prevails over sub-area variation ($\sigma_s = 0.08$), leading to homogeneity within the areas. Such values for the scale parameters are retrieved from the variance decomposition related to real data discussed in Section 6. Lastly, the dichotomous response is obtained by assigning a value of 1 to individuals below the first quintile of the generated continuous variable, and 0 otherwise.

From each synthetic population, $B = 1,000$ samples are drawn through a two-stage sampling procedure with an approximate sampling rate of 2%. To account for the presence of out-of-sample sub-areas, each drawn sample intentionally excludes units from 37 sub-areas across 15 areas. Consequently, half of the areas include individuals from all sub-areas, while in the remaining half, units from two or three sub-areas are not observed. The selection process that determines OOS sub-areas involves two distinct steps: (a.1) 15 out of 30 areas are randomly selected, (a.2) in each selected area, two or three sub-areas are randomly determined as missing and remain constant across iterations. Subsequently, a two-stage sampling scheme is set up to mimic the DHS one. At each iteration, (b.1) 10% of clusters are drawn using stratified simple random sampling in each sub-area; (b.2) within each cluster, 50% of the cluster units are randomly sampled. Refer to Section S4 of the Supplementary Material for a formal and detailed overview of the population generation and sampling processes.

The frequentist properties defined in the previous section are now applied to assess the benchmarked estimators, i.e., $E_i = \tilde{\theta}_i$. These estimators pertain to both area and sub-area levels and are derived from the models discussed in Section 3, relying on the EB distribution: S-MS, SA, and I-MS. As additional baselines, we include in the simulation study the estimators related to two popular models, (a) the model by Fay and Herriot (I-FH, 1979) independently specified at both layers, similarly to the

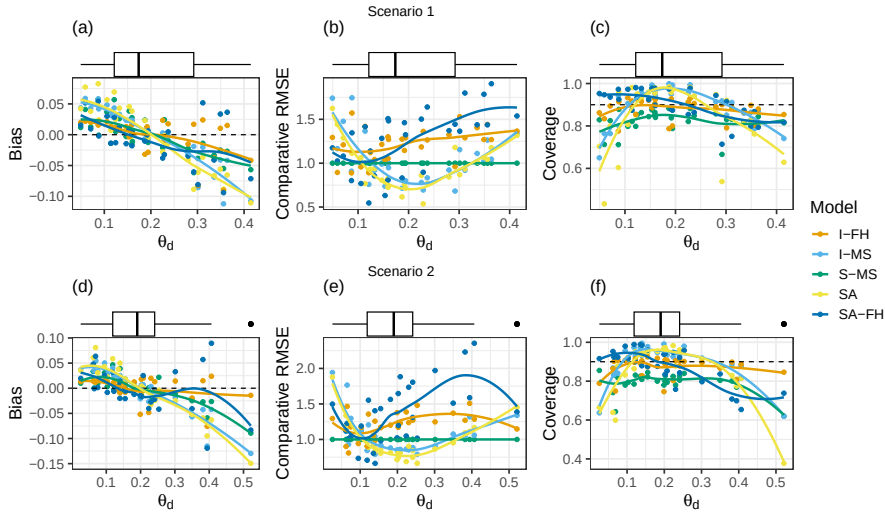

Fig. 2: Ensamble distribution of biases, comparative RMSEs (i.e. divided by RMSEs of S-MS) and frequentist coverages with respect to the population poverty rate θ_d.

SAIPE approach, and (b) the sub-area model by Erciulescu et al. (SA-FH, 2019). To constrain predictions on the unit interval, we employ a logit transformation at the linking level (You and Rao, 2002; Liu et al., 2014). In this study, LOOIC is introduced as an additional goodness-of-fit measure, useful in comparing models assuming the EB distribution. Additionally, to monitor if ranking across the areas is correctly preserved, we also study the behavior of the rank correlation coefficient. Indeed, the ability to accurately rank the poorest areas might be of interest in poverty mapping (Aiken et al., 2022).

Interesting cues about the performances of the compared area-specific estimators can be deduced from Figure 2. It reports the comparative RMSEs, obtained by dividing each RMSE by the corresponding S-MS RMSE, jointly with biases and frequentist coverages versus the population values. A first outcome that can be detected is the superiority of models based on the EB distribution on FH-type models in terms of RMSE and MAE. However, FH-type models exhibit reduced bias and good coverage. These visual findings are further supported by the summary measures presented in Table 2.

Let us now focus on models based on the EB distribution. The S-MS estimators evidently have lower biases, if compared to I-MS and SA cases: such differences become relevant for areas whose true values are far from their average (about 50% of them). This trend is particularly noticeable in areas with out-of-sample sub-areas, resulting in a correspondingly lower ARAB (Table 2). Such a decrease in the bias leads to some differences also in the behaviour of RMSE and MAE too: S-MS produces more efficient estimators in the tails of the *ensemble* distribution of population poverty rates $\{\theta_1, \ldots, \theta_D\}$, becoming less efficient in the central part of the distribution. This

		ARAB			ARRMSE			25% Poorer areas		MAE
Sc	Model	IS	OOS	Total	IS	OOS	Total	AAB_{25}	$ARMSE_{25}$	Total
	I-MS	0.221	0.288	0.225	0.378	0.440	0.409	0.059	0.085	0.053
	S-MS	0.087	0.246	0.167	0.323	0.468	0.395	0.037	0.080	0.051
1	SA	0.201	0.363	0.282	0.334	0.463	0.398	0.068	0.084	0.052
	I-FH	0.127	0.140	0.134	0.405	0.510	0.458	0.037	0.106	0.062
	SA-FH	0.113	0.224	0.169	0.334	0.463	0.446	0.056	0.124	0.065
	I-MS	0.239	0.263	0.251	0.426	0.454	0.440	0.053	0.088	0.052
	S-MS	0.072	0.206	0.139	0.347	0.478	0.413	0.035	0.082	0.051
2	SA	0.219	0.310	0.264	0.386	0.473	0.429	0.059	0.085	0.051
	I-FH	0.120	0.113	0.117	0.436	0.514	0.475	0.023	0.109	0.061
	SA-FH	0.139	0.238	0.189	0.534	0.477	0.506	0.058	0.138	0.069

Table 2. Summary measures concerning the estimators at the area level for the simulation study of Section 5.2. ARAB, ARRMSE and MAE are computed on all the areas; AAB_{25} and $ARMSE_{25}$ concerns the 25% poorer areas. Concerning ARAB and ARRMSE, values obtained for the areas without missing sub-areas (IS), with missing sub-areas (OOS) and overall (Total) are distinctly reported.

	Model				
Sc	I-MS	S-MS	SA	I-FH	SA-FH
1	0.7504	0.7687	0.7565	0.7381	0.7226
2	0.7898	0.8025	0.7982	0.7745	0.7472

Table 3. Summary measures concerning the estimators at the area level for the simulation study of Section 5.2: average Spearman's rank correlation values on overall areas.

is in line also with results concerning AAB_{25} and $ARMSE_{25}$ in the context of the poorest areas. By focusing on the 90% credible intervals, we observe that S-MS is affected by a slight under-coverage (median: 0.83 in scenario 1 and 0.81 in scenario 2). However, the performances deteriorate more slowly if compared to the other methods in the non-central parts of the ensemble distribution. We can point out that the I-MS model, characterized by the absence of an area-specific random effect, leads to poorer performances if compared to SA and S-MS models. This is more evident in Scenario 2 where the data-generating process mimics the case $\sigma_a > \sigma_s$ that characterizes the application.

Figure S4 and Table S1 in Supplementary Material report the results concerning the estimation of sub-area parameters. It can be noted that the relative indicators (ARAB and ARRMSE) seem to indicate that FH-type models have good behaviour when targeting sub-areas. However, looking at the absolute indicators AAB_{25}, $ARMSE_{25}$, and MAE, it can be noted that the performances of such models are sensibly worse, as, for example, the $ARMSE_{25}$ basically doubles those of EB-based models. In fact, the nice results reported by relative indicators are due to better performances in areas with small θ_d, as can be detected by the overall trend of Figure S4. In addition, the three EB-based models behave quite similarly when targeting

sub-areas, with the exception of I-MS, which still shows lower performances in Scenario 2. Lastly, we highlight that the average LOOIC of the S-MS model is widely lower both for sub-area (Table S1) and area (-17.3 and -17.1 for I-MS compared to -29.5 and -33.5 for S-MS, for scenarios 1 and 2 respectively) levels in both scenarios, denoting a remarkable improvement in the predictive abilities of this model.

As mentioned, an interesting property that estimators should have is the preservation of the rank across the areas and sub-areas. To check it, the average rank correlation coefficient is set out for areas (Table 3) and sub-areas (Table S2 in the Supplementary Material). From this perspective, we notice a distinction in outcomes between EB-based models and FH-type models; observing that the S-MS (for areas) and SA (for sub-areas) better preserve the rank across domains.

Focusing on the different studied properties, the estimators under the S-MS model represent a good trade-off at the area level; whereas, at the sub-area levels, differences across methods are more diluted. Indeed, S-MS estimators are characterized by a more moderate shrinkage and, for this reason, they are able to reduce the bias that affects other EB-based models, approaching the levels of FH-type models. However, when focusing on the error and the ranking ability, it emerges that EB-based models have greater performances. We also note that S-MS model behaves better in the tails of the poverty rates ensemble distribution. Small and, especially, large poverty rates are relevant for the implementation of geographically targeted alleviation policies as they can identify poverty hotspots. Therefore, an improvement in the estimation of such values may be of great importance in this context.

Nonetheless, we advise against generalizing these findings, as, akin to all model-based simulations, they depend on generated data and specific assumptions that inherently limit their scope. As differing assumptions about population parameters may potentially affect the validity of results, we stress the importance of cautiously interpreting and evaluating their applicability.

6. Poverty mapping in Bangladesh

The aim of this section is to map poverty in Bangladesh at multiple levels of administrative divisions: the Administrative Level AL-2, comprising $D = 64$ zilas as areas, and the AL-3 with $M = 544$ upazilas as sub-areas. The target poverty measure is the proportion of people below the 20th percentile of the WI national distribution. We consider DHS survey data complemented with RS and geographical data available from multiple open sources. Data and sources are shortly illustrated in Section 6.1, whereas results are discussed in Section 6.2.

6.1. Data

The Bangladesh DHS, of which we consider the 2014 wave, targets the population residing in non-institutional dwelling units by providing a two-stage stratified sample with 17,300 households and $n = 81,624$ individuals overall. The sample sizes in terms of individuals range from 135 to 4,730 (median 986) for zilas and from 16 to 1,884 (median: 160) for upazilas. All the zilas are in-sample, whereas about one third of upazilas (179) are not included in the sample.

	LOOIC (Standard Error)		
Models	SA	I-MS	S-MS
Zila level	- (-)	-138.5 (14.7)	-152.6 (12.7)
Upazila level	-118.8 (33.8)	-119.6 (34.1)	-126.0 (33.3)

Table 4. LOOIC values and their standard errors for all the considered models.

For each household, a WI score is computed by combining a set of answers on the availability of durable assets and housing welfare characteristics (Fabic et al., 2012). Households lying in the first quintile of the national distribution of WI are labeled as *poor*. Since poverty is usually investigated individually, the analysis is carried out at the individual level by assuming that all components of the same household share the same WI score. A sampling weight is associated with each household, accounting for unequal probabilities of selection and non-responses. The survey estimates, as defined in Section 2, range from 0 to 0.53 (median: 0.21) with just one 0 value for zilas; while they span from 0 to 0.96 (median: 0.16) with 66 zero values for upazilas.

To retrieve the effective sample sizes, we opt to estimate the design effect starting from the strata level as in Schmid et al. (2017). Such uncertainty estimates confirm the unreliability of survey estimates both at zila and upazila levels motivating us to employ SAE: we observe coefficients of variation higher than 0.20 in more than 80% of overall in-sample estimates. Note that $\tilde{n}_d$ and $\tilde{n}_{dj}$ are much lower than n_d and n_{dj} respectively since the poverty status is defined at the household level and the correlation within clusters (i.e. Census enumeration areas) is very strong.

We integrate 46 RS variables as covariates computed both at the area and sub-area levels by cutting rasters through shapefiles (for more details, see De Nicolò et al., 2022, and Section S5 in the Supplementary Material). Specifically, we consider the demographic composition of areas and sub-areas by including the population density and its disaggregation by age and sex classes. The nighttime light radiance and the distances to main facilities and infrastructures (e.g. nearest city or healthcare site) are included as indicators of area urbanization and economic development. As the agricultural sector is a driving force for Bangladeshi economy, its territorial and climatic characterization may be proxies of productivity. The first one has been considered by including land-use variables that specify the distance from nearest areas with specific use classification (e.g. cultivated, woody-tree, artificial surface), the elevation above sea level, and topographic slope. The second one is fulfilled through bio-climatic variables such as the overall and seasonal variations of temperature and rainfall (e.g., annual mean, standard deviation, and temperature diurnal range).

6.2. Results

The alternative models SA, I-MS, and S-MS are fitted on Bangladeshi data using the HMC algorithm: the chains converged to the stationary distribution (all the Gelman-Rubin statistics are equal to one) and all the effective sample sizes are greater than 1000. Table 4 provides details about the model comparison in terms of LOOIC, for both spatial levels. According to it, the S-MS model shows better goodness-of-fit

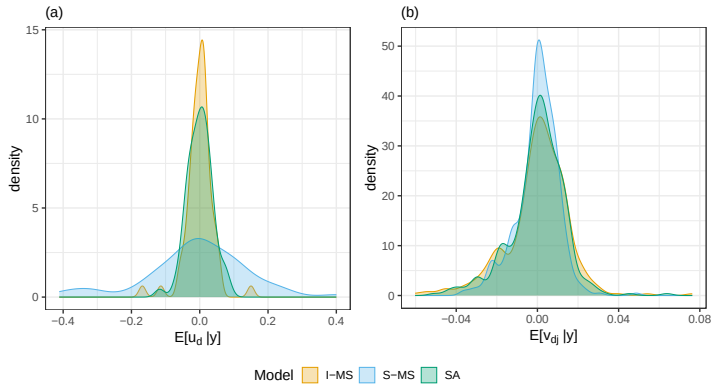

Fig. 3: Densities of posterior means of u_d and v_{dj}, $\forall d$, $\forall j$, for each model.

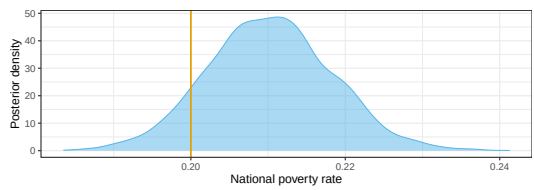

Fig. 4: Posterior distribution of the unbenchmarked national poverty rate (blue density) compared to the benchmark (vertical orange line).

with respect to the other proposals, while SA and I-MS perform similarly. The main motivation for this finding is that the S-MS model has the area-specific random effect u_d, $\forall d$, characterizing the distributions at both levels, thus promoting an exchange of information between areas and sub-areas. Conversely, the other models define u_d at only one level, thereby being informed only by observations at the corresponding level. This translates into a better ability of S-MS model to capture specific area features through random effects. To elaborate more on this point, the ensemble distributions of $\mathbb{E}[u_d|\mathbf{y}]$ are depicted in the left panel of Figure 3: note that under SA and I-MS models such distributions are markedly more concentrated around zero, not capturing area-specific peculiarities. Looking at the ensemble distribution of $\mathbb{E}[v_{dj}|\mathbf{y}]$, shown in the right panel of Figure 3, we observe that upazila effects seem partially absorbed by zila effects for the S-MS model. However, no relevant differences among models are recorded in this case.

To simplify the discussion, from now on, we focus on S-MS model estimates due to the aforementioned reasons. Figure S5 in the Supplementary Material presents the posterior predictive check based on the density function. It illustrates that the kernel densities of pseudo-samples generated from the model do not exhibit systematic deviations from the kernel density of the direct estimates, suggesting that the distributional assumptions of the models are met. To obtain benchmarked estimates

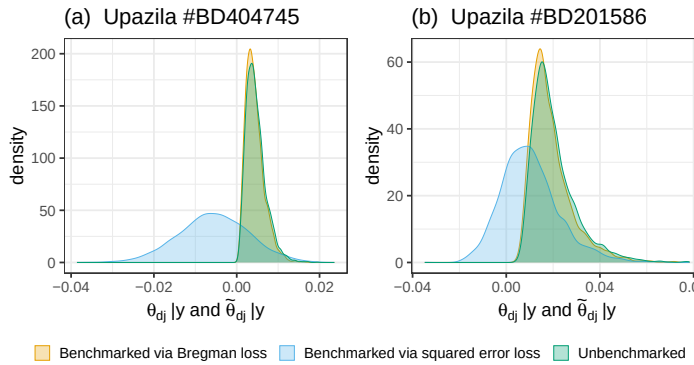

Fig. 5: Posterior distributions of benchmarked and unbenchmarked estimates in two upazilas.

with respect to the national level, we project the posterior distributions of θ_{dj}, $\forall d$, $\forall j$ exploiting the result of Theorem 1. Exploiting the definition of the response, we set $t = 0.2$; the posterior distribution of the unbenchmarked national poverty rate is visually compared to t in Figure 4. Furthermore, we evaluate the original posterior distributions with their projections under the proposed Bregman and weighted squared error loss functions. In particular, Figure 5 shows it for two upazilas with poverty rates close to zero. Note that, under the weighted squared error loss, projections can be markedly outside the support taking negative values. In contrast, under the Bregman loss function, benchmarked posteriors are correctly defined in $(0, 1)$.

The shrinking process induced by the model at both spatial levels is depicted in Figure 6, by comparing survey estimates with model-based estimates. As expected, the shrinkage is stronger at the upazila level, given the downward precision of survey estimates. The differences between survey and model-based estimates may be high in case of very low effective sample sizes. Moreover, the right panel of Figure 6 shows that discrepancies at the area level may be also due to a high percentage of out-of-sample sub-areas. In Figure 7, ensemble densities of sub-area estimates are split up between in-sample and out-of-sample ones, together with a collection of the most relevant covariates. This illustrates that the composition of out-of-sample sub-areas is quite heterogeneous, incorporating both remote sub-areas (e.g. a large number of sub-areas on Chittagong Hill tracts) and urban or suburban ones. As a consequence, the out-of-sample poverty rates show a higher polarization between a set of urban and less poor sub-areas and a set of very poor rural ones. The latter is characterized by greater poverty levels than those fitted for in-sample sub-areas. This confirms that the sample selection process determining out-of-sample sub-areas can be affected by specific features.

The spatial distribution of model-based estimates is displayed in Figure 8, enabling us to highlight the heterogeneity of poverty rates inside each zila. This is not clear by the zila map that averages out such estimates; whereas, at a finer level, the upazila map bears additional and valuable information. The reduction of standard deviation

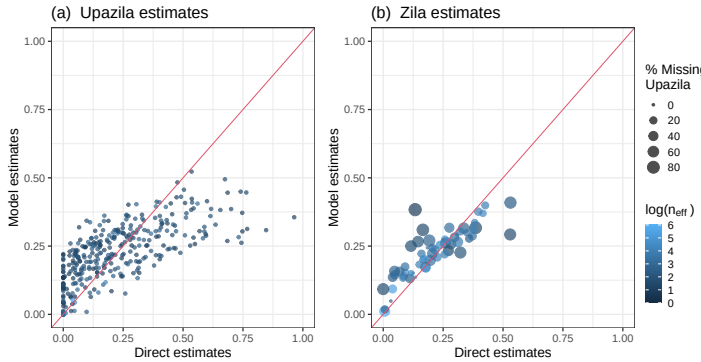

Fig. 6: Shrinkage process at both spatial levels, bisector line in red.

of model-based estimates with respect to survey-based ones is of 51% on average for zilas and 77% for upazilas.

At a first glance, the poverty patterns are consistent with the literature (Kam et al., 2005; Imam et al., 2019), capturing consolidated dynamics. The metropolitan regions of Dhaka and Chittagong retain the lowest poverty levels, in contrast with peripheral areas. High poverty incidence domains overlap with ecologically poor zones (Kam et al., 2005). We can mention the territories more severely exposed to climate change effects and floodings such as the Haor depression (Sylhet basin lowlands) in the north-east, some areas at the edge of major rivers (Haque and Jahan, 2015), those exposed to droughts in the Rangpur division (north-west) and the Chittagong Hill Tract (south-east). On the other hand, the drought-prone southern Rajshahi and Khulna divisions experience lower poverty levels due to the good irrigation coverage (Kam et al., 2005). However, when disaggregating estimates at the upazila level, the picture becomes more clear. The poorest regions present a remarkable variety and it is possible to detect few upazilas having lower poverty levels than the neighboring ones. They correspond to main urban centers (indicated by red dots in Figure 8) such as Rangpur, Dinajpur, and Saidpur cities in Rangpur division; Sylhet, Kishoreganj, and Mymensingh in the north-east side and Raozan in the south-east side.

7. Conclusions

In this paper, we introduce a multi-scale modeling framework for poverty mapping at multiple spatial resolutions to avoid the scaling problem. The multi-scale estimation of poverty rates constitutes as an application-relevant problem in Official Statistics, as testifies the aforementioned SAIPE programme. The main aim is to produce reliable poverty maps that are coherent at different geographical layers. Such tools are valuable for poverty evaluation, poverty targeting and to develop place-based policies. On this line, we complement the proposal with a novel benchmarking algorithm

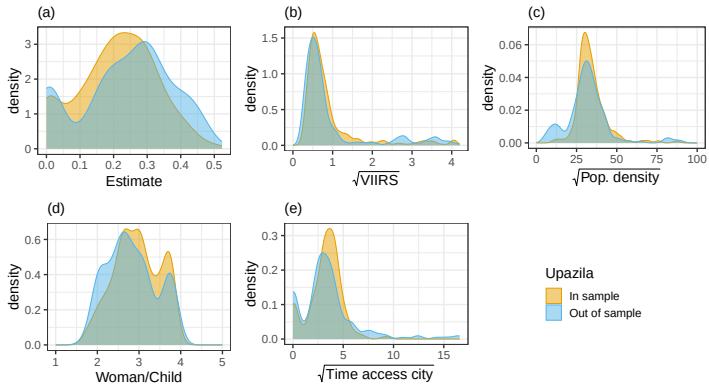

Fig. 7: Densities of model-based estimates and the most relevant covariates compared between in-sample and out-of-sample sub-areas.

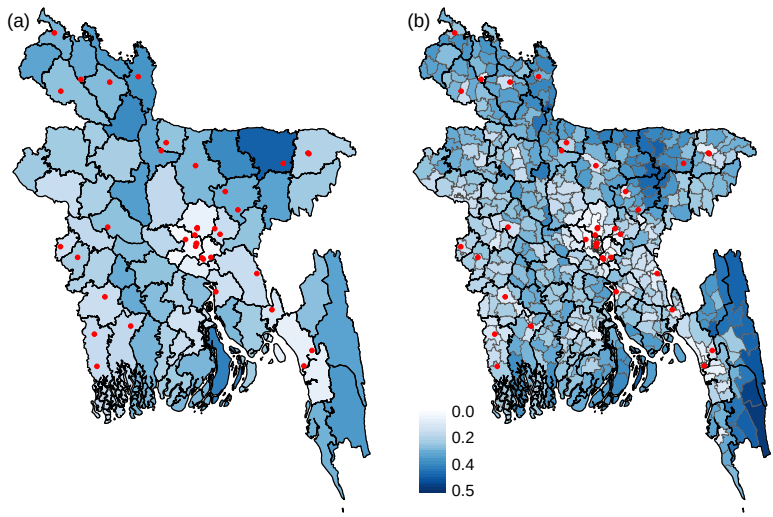

Fig. 8: Model-based estimates at the zila level (left-hand side) and at the upazila level (right-hand side); red points indicate the 37 major cities.

that ensures the accordance of poverty estimates across levels. This allows to preserve poverty rate estimates and associated credible intervals within their support, restricted to $(0, 1)$.

The simulation study we carried out highlights the effectiveness of our proposal, showing better performances in terms of error and rank with respect to existing alternatives. In particular, this is evident for poverty rates lying in the tails of their distribution, which are less shrunken towards the average. In this way, our modeling

strategy is able to retain domains with extremely high poverty rates: this has relevant practical implications since poverty hotspots need to be identified as being the target of poverty relief actions.

In our application on Bangladeshi data, rural and remote regions that mostly overlap with droughts- and floods-prone zones with high exposure to climate change effects, are classified as very poor. In this sense, geo-targeted policies may foster risk mitigation strategies (e.g., insurance) as well as further expand irrigation coverages via specific water management policies (Pal et al., 2011; Hossain et al., 2022). Furthermore, our multi-scale poverty mapping approach permits us to spot critical zones located within larger administrative units.

A possible limitation of our model concerns the fact that the sampling dependency arising from the overlap between area and sub-area samples is ignored. Nonetheless, the incorporation of such dependency within a multi-scale model is challenging as it does not resemble a conventional bivariate structure. A possible solution is to adopt copula functions (Souza and Moura, 2016). A proposal to extend our model in this direction is discussed in Section S7 of the Supplementary Material. We note that the resulting estimates are very close to those obtained under the share multi-scale model, supporting its robustness. Such extension seems promising, nonetheless, it involves further methodological challenges that are out of the scope and worth to be addressed in future work.

It is possible to extend our proposal to explicitly account for the spatial structure in the statistical model, similarly to Aregay et al. (2016). However, results of our application do not provide evidence of a residual spatial trend. This can be due to the spatial informative power of remote sensing covariates employed in the model. In this sense, an interesting extension would be to consider additional and more granular levels, such as the cluster one, in our multi-level strategy. This would enable us to fully exploit the grainy nature of RS data. Lastly, as possible future directions of research, we mention the use of multiple quintiles of the WI index national distribution through a multivariate framework, and, moving beyond poverty measurement, the mapping of other well-being measures such as inequality indicators.

Supplementary Material

A pdf file that contains the proof of Theorem 1, additional information about the simulation study, and the dataset discussed in the application. Furthermore, a zipped folder with the Stan and R codes to run the model on generated data is available.

Acknowledgments

Work supported by the Data and Evidence to End Extreme Poverty (DEEP) research programme. DEEP is a consortium of the Universities of Cornell, Copenhagen, and Southampton led by Oxford Policy Management, in partnership with the World Bank - Development Data Group and funded by the UK Foreign, Commonwealth & Development Office.

Data availability

The complete dataset from DHS cannot be released due to data disclosure agreement.

Funding

The work of Silvia De Nicolò was partially funded by the ALMA IDEA 2022 grant (CUP J45F21002000001) supported by the European Union - NextGenerationEU and PNRR funds, PE10 project – ONFOOD, "Research and innovation network on food and nutrition. Sustainability, Safety and Security – Working ON Foods" (code PE0000003, CUP J33C22002860001). The work of Aldo Gardini was partially supported by MUR on funds FSE REACT EU - PON R&I 2014-2020 and PNR (D.M. 737/2021) for the RTDA_GREEN project under Grant J41B21012140007.

References

E. Aiken, S. Bellue, D. Karlan, C. Udry, and J. E. Blumenstock. Machine learning and phone data can improve targeting of humanitarian aid. *Nature*, 603(7903): 864–870, 2022.

J. Aitchison. On criteria for measures of compositional difference. *Mathematical Geology*, 24(4):365–379, 1992.

S. W. Allard and S. Allard. *Places in need: The changing geography of poverty*. Russell Sage Foundation, 2017.

M. Aregay, A. B. Lawson, C. Faes, R. S. Kirby, R. Carroll, and K. Watjou. Multiscale measurement error models for aggregated small area health data. *Statistical methods in medical research*, 25(4):1201–1223, 2016.

M. Aregay, A. B. Lawson, C. Faes, R. S. Kirby, R. Carroll, and K. Watjou. Comparing multilevel and multiscale convolution models for small area aggregated health data. *Spatial and Spatio-temporal Epidemiology*, 22:39–49, 2017.

A. Banerjee, X. Guo, and H. Wang. On the optimality of conditional expectation as a Bregman predictor. *IEEE Transactions on Information Theory*, 51(7):2664–2669, 2005.

W. Bell, G. S. Datta, and M. Ghosh. Benchmarking small area estimators. *Biometrika*, 100(1):189–202, 2013.

W. R. Bell, W. W. Basel, and J. J. Maples. An overview of the US census bureau's small area income and poverty estimates program. *Analysis of poverty data by small area estimation*, pages 349–378, 2016.

M. H. Benedetti, V. J. Berrocal, and R. J. Little. Accounting for survey design in Bayesian disaggregation of survey-based areal estimates of proportions: An application to the American Community Survey. *The Annals of Applied Statistics*, 16 (4):2201–2230, 2022.

D. Bigman and H. Fofack. Geographical targeting for poverty alleviation: An introduction to the special issue. *The World Bank Economic Review*, 14(1):129–145, 2000.

J. R. Bradley, C. K. Wikle, and S. H. Holan. Regionalization of multiscale spatial processes by using a criterion for spatial aggregation error. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(3):815–832, 2017.

B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 2017.

C. Casas-Cordero Valencia, J. Encina, and P. Lahiri. Poverty mapping for the Chilean comunas. *Analysis of Poverty Data by Small Area Estimation*, pages 379–404, 2016.

G. Chi, H. Fang, S. Chatterjee, and J. E. Blumenstock. Microestimates of wealth for all low-and middle-income countries. *Proceedings of the National Academy of Sciences*, 119(3):e2113658119, 2022.

L. Christiaensen and Y. Todo. Poverty reduction during the rural–urban transformation– The role of the missing middle. *World Development*, 63:43–58, 2014.

P. Corral, I. Molina, A. Cojocaru, and S. Segovia. Guidelines to small area estimation for poverty mapping. Technical report, World Bank, 2022.

G. Datta, M. Ghosh, R. Steorts, and J. Maples. Bayesian benchmarking with applications to small area estimation. *Test*, 20(3):574–588, 2011.

S. De Nicolò, E. Fabrizi, and A. Gardini. Extended Beta models for poverty mapping. An application integrating survey and remote sensing data in Bangladesh. In *Quaderni di Dipartimento. Serie Ricerche*. 2022.

D. B. Dunson and B. Neelon. Bayesian inference on order-constrained parameters in generalized linear models. *Biometrics*, 59(2):286–295, 2003.

C. Elbers, J. O. Lanjouw, and P. Lanjouw. Micro-level estimation of poverty and inequality. *Econometrica*, 71(1):355–364, 2003.

A. L. Erciulescu, N. B. Cruze, and B. Nandram. Model-based county level crop estimates incorporating auxiliary sources of information. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, (1):283–303, 2019.

M. S. Fabic, Y. Choi, and S. Bird. A systematic review of demographic and health surveys: data availability and utilization for research. *Bulletin of the World Health Organization*, 90:604–612, 2012.

E. Fabrizi, M. Ferrante, and C. Trivisano. Hierarchical Beta regression models for the estimation of poverty and inequality parameters in small areas. *Analysis of Poverty Data by Small Area Methods. John Wiley and Sons*, pages 299–314, 2016a.

E. Fabrizi, M. R. Ferrante, and C. Trivisano. Bayesian Beta regression models for the estimation of poverty and inequality parameters in small areas. *Analysis of poverty data by small area estimation*, pages 299–314, 2016b.

S.-g. Fan and E. E. Cho. Paths out of poverty: International experience. *Journal of Integrative Agriculture*, 20(4):857–867, 2021.

R. E. Fay and R. A. Herriot. Estimates of income for small places: an application of james-stein procedures to census data. *Journal of the American Statistical Association*, 74(366a):269–277, 1979.

S. Ferrari and F. Cribari-Neto. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7):799–815, 2004.

J. Gabry, D. Simpson, A. Vehtari, M. Betancourt, A. Gelman, et al. Visualization in Bayesian workflow. *Journal of the Royal Statistical Society Series A*, 182(2): 389–402, 2019.

E. Galasso and M. Ravallion. Decentralized targeting of an antipoverty program. *Journal of Public economics*, 89(4):705–727, 2005.

A. Gauci. Spatial maps. Targeting & mapping poverty. *London: United Nations. Economic Commission for Africa*, 2005.

M. Ghosh, T. Kubokawa, and Y. Kawakubo. Benchmarked empirical bayes methods in multiplicative area-level models with risk evaluation. *Biometrika*, 102(3):647–659, 2015.

J. Hájek. Discussion of 'An essay on the logical foundations of survey sampling, Part I', by D. Basu. *Foundations of Statistical Inference*, page 326, 1971.

O. Hall, F. Dompae, I. Wahab, and F. M. Dzanku. A review of machine learning and satellite imagery for poverty prediction: Implications for development research and applications. *Journal of International Development*, 2023.

A. Haque and S. Jahan. Impact of flood disasters in Bangladesh: A multi-sector regional analysis. *International Journal of Disaster Risk Reduction*, 13:266–275, 2015.

M. S. Hossain, G. M. Alam, S. Fahad, T. Sarker, M. Moniruzzaman, and M. G. Rabbany. Smallholder farmers' willingness to pay for flood insurance as climate change adaptation strategy in northern Bangladesh. *Journal of Cleaner Production*, 338: 130584, 2022.

M. F. Imam, M. A. Islam, M. A. Alam, M. J. Hossain, and S. Das. Small Area Estimation of Poverty in Rural Bangladesh. *The Bangladesh Journal of Agricultural Economics*, 40(1&2):1–16, 2019.

R. Janicki. Properties of the Beta regression model for small area estimation of proportions and application to estimation of poverty rates. *Communications in Statistics-Theory and Methods*, 49(9):2264–2284, 2020.

R. Janicki and A. Vesper. Benchmarking techniques for reconciling Bayesian small area models at distinct geographic levels. *Statistical Methods & Applications*, 26 (4):557–581, 2017.

N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301): 790–794, 2016.

S.-P. Kam, M. Hossain, M. L. Bose, and L. S. Villano. Spatial patterns of rural poverty and their relationship with welfare-influencing factors in Bangladesh. *Food Policy*, 30(5-6):551–567, 2005.

E. D. Kolaczyk and H. Huang. Multiscale statistical models for hierarchical spatial aggregation. *Geographical Analysis*, 33(2):95–118, 2001.

T. Krenzke, L. Mohadjer, J. Li, A. Erciulescu, R. Fay, W. Ren, W. Van de Kerckhove, L. Li, and J. Rao. Program for the international assessment of adult competencies (PIAAC): State and county estimation methodology report. NCES 2020-225. Technical report, 2020.

K. Lee and J. Braithwaite. High-resolution poverty maps in sub-saharan Africa. *World Development*, 159:106028, 2022.

B. Liu, P. Lahiri, and G. Kalton. Hierarchical bayes modeling of survey-weighted small area proportions. *Survey Methodology*, 40(1):1–13, 2014.

M. M. Louie and E. D. Kolaczyk. A multiscale method for disease mapping in spatial epidemiology. *Statistics in medicine*, 25(8):1287–1306, 2006.

I. Molina and J. Rao. Small area estimation of poverty indicators. *Canadian Journal of statistics*, 38(3):369–385, 2010.

I. Molina, B. Nandram, and J. Rao. Small area estimation of general parameters with application to poverty indicators: A hierarchical bayes approach. *The Annals of Applied Statistics*, 8(2):852–885, 2014.

B. Nandram, N. B. Cruze, A. L. Erciulescu, and L. Chen. Bayesian small area models under inequality constraints with benchmarking and double shrinkage. Technical report, 2022.

T. Okonek and J. Wakefield. A computationally efficient approach to fully bayesian benchmarking. *arXiv preprint arXiv:2203.12195*, 2022.

S. K. Pal, A. J. Adeloye, M. S. Babel, and A. Das Gupta. Evaluation of the effectiveness of water management policies in Bangladesh. *Water Resources Development*, 27(02):401–417, 2011.

S. Patra. *Constrained Bayesian inference through posterior projection with applications*. PhD thesis, 2019.

J. Piironen and A. Vehtari. Sparsity information and regularization in the Horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, 11(2):5018 – 5051, 2017. doi: 10.1214/17-EJS1337SI.

E. Pirani. Wealth index. *Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht. https://doi. org/10.1007/978-94-007-0753-5_3202*, 2014.

M. J. Poirier, K. A. Grépin, and M. Grignon. Approaches and alternatives to the wealth index to measure socioeconomic status using survey data: A critical interpretive synthesis. *Social Indicators Research*, 148(1):1–46, 2020.

M. Pratesi. Analysis of poverty data by small area estimation. 2015.

M. Pratesi and N. Salvati. Introduction on measuring poverty at local level using small area estimation methods. In *Analysis of poverty data by small area estimation*, chapter 1, pages 1–18. Wiley Online Library, 2016.

R. Puurbalanta. A clipped gaussian geo-classification model for poverty mapping. *Journal of Applied Statistics*, 48(10):1882–1895, 2021.

J. N. Rao and I. Molina. *Small Area Estimation*. John Wiley & Sons, 2015.

S. O. Rutstein and K. Johnson. The DHS Wealth Index. Technical report, DHS comparative reports no. 6, Calverton, Maryland: ORC Macro, 2004.

T. Schmid, F. Bruckschen, N. Salvati, and T. Zbiranski. Constructing sociodemographic indicators for national statistical institutes by using mobile phone data: Estimating literacy rates in Senegal. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(4):1163–1190, 2017.

D. Sen, S. Patra, and D. Dunson. Constrained inference through posterior projections. *arXiv preprint arXiv:1812.05741*, 2018.

T. P. Sohnesen, P. Fisker, and D. Malmgren-Hansen. Using satellite data to guide urban poverty reduction. *Review of Income and Wealth*, 68:S282–S294, 2022.

D. F. Souza and F. A. Moura. Multivariate Beta regression with application in small area estimation. *Journal of Official Statistics*, 32(3):747–768, 2016.

J. E. Steele, P. R. Sundsøy, C. Pezzulo, V. A. Alegana, T. J. Bird, J. Blumenstock, J. Bjelland, K. Engø-Monsen, Y.-A. De Montjoye, A. M. Iqbal, et al. Mapping poverty using mobile phone and satellite data. *Journal of The Royal Society Interface*, 14(127):20160690, 2017.

X. Tang, M. Ghosh, N. S. Ha, and J. Sedransk. Modeling random effects using global–local shrinkage priors in small area estimation. *Journal of the American Statistical Association*, 113(524):1476–1489, 2018.

M. Torabi and J. Rao. On small area estimation under a sub-area level model. *Journal of Multivariate Analysis*, 127:36–55, 2014.

N. Tzavidis, L.-C. Zhang, A. Luna, T. Schmid, and N. Rojas-Perilla. From start to finish: a framework for the production of small area official statistics. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 181(4):927–979, 2018.

A. Vehtari, A. Gelman, and J. Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432, 2017.

L. A. Waller and C. A. Gotway. *Applied spatial statistics for public health data*. John Wiley & Sons, 2004.

C. Yeh, A. Perez, A. Driscoll, G. Azzari, Z. Tang, D. Lobell, S. Ermon, and M. Burke. Using publicly available satellite imagery and deep learning to understand economic well-being in africa. *Nature communications*, 11(1):1–11, 2020.

Y. You and J. Rao. Small area estimation using unmatched sampling and linking models. *Canadian Journal of Statistics*, 30(1):3–15, 2002.

J. L. Zhang and J. Bryant. Fully Bayesian benchmarking of small area estimation models. *Journal of official statistics*, 36(1):197–223, 2020.

X. Zhao, B. Yu, Y. Liu, Z. Chen, Q. Li, C. Wang, and J. Wu. Estimation of poverty using random forest regression with multi-source data: A case study in Bangladesh. *Remote Sensing*, 11(4):375, 2019.

Y. Zhou and Y. Liu. The geography of poverty: Review and research prospects. *Journal of Rural Studies*, 93:408–416, 2022.