TEST
https://doi.org/10.1007/s11749-022-00822-1

ORIGINAL PAPER

Estimation of poverty and inequality in small areas: review and discussion

Isabel Molina[1] · **Paul Corral**[2] · **Minh Nguyen**[2]

Received: 11 November 2021 / Accepted: 24 May 2022

Abstract

Never better said, a correct diagnosis is crucial for patient recovery. In the eradication of poverty, which is the first of the sustainable development goals (SDGs) established by the United Nations, efforts in the form of social aid and programs will be useless if they are not directed where they are most needed. Nowadays, monitoring the progress on the SDGs is even more urgent after the sanitary crisis, which is reversing the global poverty reduction observed since 1990 and, given that social development funds are always limited, managing them correctly requires disaggregated statistical information on poverty of acceptable quality. But reliable estimates on living conditions are scarce due to sample size limitations of most official surveys. Common small area estimation procedures supplement the survey data with auxiliary data sources to produce more reliable disaggregated estimates than those based solely on the survey data. We describe the traditional as well as recent model-based procedures for obtaining reliable disaggregated estimates of poverty and inequality indicators, discussing their properties from a practical point of view, placing emphasis on real applications and describing software implementations. We discuss results from recent simulation experiments that compare some of the unit-level methods in terms of bias and efficiency, under model- and design-based setups. Finally, we provide some concluding remarks.

Keywords Small area estimation · ELL · Poverty map · Empirical best · Bootstrap

Mathematics Subject Classification 62D05 · 62E99 · 62G09

✉ Isabel Molina
isabelmolina8@gmail.com

[1] Department of Statistics, Universidad Carlos III de Madrid, C/Madrid 126, 28903 Getafe, Madrid, Spain

[2] Poverty and Equity Global Practice, The World Bank Group, Washington, DC, USA

Published online: 21 June 2022

1 Introduction

Design of effective economic development policies that are targeted to the correct population groups requires reliable disaggregated statistical information on living conditions. In fact, the global framework for the monitoring of the sustainable development goals (SDGs) across countries establishes the need to disaggregate statistical information on living conditions not only by geography, but also by income group, gender, age group, ethnicity, migratory status and disability condition.

For large regions or at the national level, *direct* survey estimators obtained using the survey data from the target region, usually have adequate quality. They are (at least approximately) unbiased with respect to the sampling replication mechanism and do not make any model assumptions (i.e., they are nonparametric). However, as long as one disaggregates by local areas or by certain population subgroups, the quality of these direct estimates decreases, owing to the small sample sizes of the target areas or domains of estimation. As a consequence, official statistical institutes cannot publish reliable wellbeing statistical figures at a local level. As an example, the Spanish Survey on Income and Living Conditions (SILC) is planned to produce estimates of poverty rates at the level of Spanish Autonomous Communities (large regions), but does not provide information of acceptable quality for Spanish provinces due to the small SILC sample sizes for some of the country's provinces. The Module of Socioeconomic Conditions of the Mexican National Survey of Household Income and Expenditure (ENIGH in Spanish) is designed to produce estimates of poverty and inequality at the national level and for the 32 federative entities (31 states and México DF) with disaggregation by rural and urban zones, every two years, but there is a mandate to produce estimates by municipality every 5 years, and the ENIGH alone cannot provide estimates for the municipalities with adequate quality.

This problem makes it necessary to use *indirect* techniques that combine auxiliary information from censuses, registers or other larger surveys. These techniques put together the data from all the areas of interest by specifying some homogeneity relationships across them; hence, they are based on the idea that "union is strength". Rao and Molina (2015) contains a comprehensive account of small area estimation (SAE) procedures. Other reviews are given by Ghosh and Rao (1994), Pfeffermann (2002, 2013), Jiang and Lahiri (2006) and more recently Ghosh (2020). Among SAE procedures, model-based techniques are popular because they yield estimates with much better quality even for areas with very small sample sizes. Of course, this can be achieved at the expense of making model assumptions. These assumptions can (and should) be validated using the available data.

Among small area estimation models, area-level models use only aggregated auxiliary information at the area level. Aggregated data are more easily available because it is typically not subject to confidentiality. These models have other advantages provided by the aggregation like less sensitivity to unit level outliers, but use less detailed information than unit-level models, which are specified at the microdata level.

The first unit-level model used for small area estimation was proposed by Battese et al. (1988), who estimated county crop areas under corn and soybeans. This model, called hereafter BHF model, includes, apart from the usual unit level errors, random effects for the areas representing the between-area heterogeneity or idiosyncracy that cannot be explained by the available auxiliary variables. These models can incorporate much more information in the estimation process than area level models, since area or subarea level (contextual) variables might also be included. Models with area effects belong to the general class of mixed models, widely used in many other fields such as Biostatistics, Engineering, Econometrics and other Social Sciences.

Poverty measurement is a multidimensional concept that is not easy to summarize, and some common poverty indicators actually depend on several dimensions of completely different nature, which often include some monetary measure of welfare, as well as other factors indicating availability or access to certain basic goods or services. Here we review the specific model-based SAE procedures that have been applied to estimate monetary poverty and inequality indicators. The complexity of the mathematical functions defining these indicators (see, e.g., Neri et al. 2005 for a review) make some of the traditional model-based SAE techniques, developed for linear indicators such as means or totals, not applicable. Often, several indicators depending on the same welfare measure need to be estimated, and, in this case, it seems reasonable to model the underlying welfare measure instead of each indicator separately, which would require finding useful covariates and a proper model for each one. Modeling the welfare measure and then estimating several indicators of different shapes based on its values for all the area units entails the application of more sophisticated SAE methods for nonlinear indicators. Most of the methods reviewed here may be applied for general indicators defined as functions of the welfare measure in the area units, including the area's empirical cumulative distribution function (cdf) of the welfare at given points. Nevertheless, very popular poverty indicators, including several of the indicators estimated by the World Bank (WB), are members of the FGT family introduced by Foster et al. (1984). If w_{di} denotes the welfare for individual i in area d, where $i = 1, \ldots, N_d$, and z is the poverty line, then the FGT poverty indicator of order $\alpha \geq 0$ is

$$F_{\alpha d} = \frac{1}{N_d} \sum_{i=1}^{N_d} \left(\frac{z - w_{di}}{z} \right)^{\alpha} I(w_{di} < z).$$

The poverty rate, which is also the empirical cdf of the welfare at z, is obtained for $\alpha = 0$. The poverty gap and severity are obtained, respectively, for $\alpha = 1, 2$.

Reviews on SAE for poverty mapping are given by Pratesi and Salvati (2016), Guadarrama et al. (2014), Rao and Molina (2016) and, more recently, Molina (2019, 2020) and Molina et al. (2019). Here we will discuss the three most popular groups of SAE approaches for estimation of poverty and inequality indicators. We compare them in terms of convenient properties, especially looking at their bias and efficiency, both under the model, where the sample units are kept fixed but population measurements follow the model, and under the design, where the population measurements are fixed but the sample units vary. Moreover, we focus on practical issues, trying to give advice to potential applicants of the techniques, providing examples of real applications undertaken around the world using each method and describing available software for each procedure. We also provide some insight on the latest extensions of the basic methods and other types of approaches, which are either not covered, or not covered in much detail, by the previous review papers. Actually, the basic approaches are designed to be applied under certain assumptions, such as when estimation is desired at a single disaggregation level, when there is no sample selection bias or under normality. We cover here extensions of the basic methods to non-standard but realistic cases, including the new approaches adopted by the WB. In addition, we discuss the results of some simulation experiments that compare several unit-level methods that have been implemented at the WB, both under the model-based and under the design-based setups. Finally, some conclusions are drawn.

2 Fay–Herriot model

Perhaps the most popular area-level model for SAE is the Fay–Herriot (FH) model, which was introduced by Fay and Herriot (1979) to estimate mean per capita income in small places in the USA. This model has been regularly used in the US Census Bureau, within the Small Area Income and Poverty Estimates (SAIPE) project (see http://www.census.gov/hhes/www/saipe), to produce estimates for states and counties of total persons in poverty by age groups, household median income and mean per capita income; for more details, see Bell (1997) or http://www.census.gov/hhes/www/saipe. Actually, these estimates are used by the U.S. Department of Education in determining annual allocations of federal funds to school districts. The allocations amounted to more than 14 billion in fiscal year 2013 (Bell et al. 2016). Estimates of median income for four-person families are also used in a formula to determine the eligibility for a program of energy assistance to low-income families administered by the U.S. Department of Health and Human Services (Rao and Molina 2015). In Europe, the FH model was applied under the EURAREA project (http://www.statistics.gov.uk/eurarea) to estimate linear parameters such as mean income. Other examples of application of the FH model for the estimation of poverty indicators are the works by Casas-Cordero Valencia et al. (2016), who applied this model to estimate poverty rates in Chilean comunas, by Molina and Morales (2009), who used it to predict poverty rates and gaps in Spanish provinces by gender, and by Jedrzejczak and Kubacki (2013), who estimated income inequality and poverty rates by regions and family type in Poland.

The FH model is defined in two stages. In the first stage, the true values of the indicators τ_d for all the areas $d = 1, \ldots, D$, are linked by establishing a linear regression model with a common vector of regression coefficients $\boldsymbol{\beta}$, called the *linking model*, that is,

$$\tau_d = \mathbf{x}_d' \boldsymbol{\beta} + u_d, \tag{1}$$

where $\mathbf{x}_d$ is a vector of area-level covariates that is linearly related with the target indicator τ_d. The random errors in this regression model, u_d, are called area effects because they represent the unexplained between-area heterogeneity, and are typically assumed to have zero mean and constant variance σ_u^2.

Model (1) cannot be fitted to the data, since the true values of the indicators, τ_d, are not observed. Direct estimators $\hat{\tau}_d^{DIR}$ are available from the survey microdata but are, obviously, subject to survey error, since they are calculated only with the area-specific survey data, which is of small size. Hence, in the second stage, this survey error is modeled by assuming that the direct estimators are centered around the true values of the indicators, as follows

$$\hat{\tau}_d^{DIR} = \tau_d + e_d, \tag{2}$$

with error variances $var(e_d|\tau_d) = var(\hat{\tau}_d^{DIR}|\tau_d) = \psi_d$, $d = 1, \ldots, D$, which are typically heteroscedastic, since the area sample sizes are usually unequal. Model (2) is called the *sampling model*. The D error variances ψ_d, $d = 1, \ldots, D$, need to be assumed as known even if they are not; otherwise, there would be a larger number of unknown parameters than the number of observations used to fit the FH model, $\hat{\tau}_d^{DIR}$, $d = 1, \ldots, D$. These variances ψ_d are typically estimated based on the survey micro-data from the target area and perhaps smoothed afterwards to reduce the instability resulting from the small area sample sizes.

The usual small area estimator obtained from this model is the *best linear unbiased predictor* (BLUP), which is the linear function $\hat{\tau}_d = \boldsymbol{\alpha}'\mathbf{y}$ of the data $\mathbf{y} = (\hat{\tau}_1^{DIR}, \ldots, \hat{\tau}_D^{DIR})'$ that is unbiased under the model, that is, $E(\hat{\tau}_d - \tau_d) = 0$, and minimizes the mean squared error (MSE) under the model. The BLUP is simply obtained by predicting τ_d through the model, that is, as $\hat{\tau}_d = \mathbf{x}_d'\hat{\boldsymbol{\beta}} + \hat{u}_d$, where $\hat{\boldsymbol{\beta}}$ is the weighted least squares (WLS) estimator of $\boldsymbol{\beta}$ under the FH model and $\hat{u}_d = \gamma_d(\hat{\tau}_d^{DIR} - \mathbf{x}_d'\hat{\boldsymbol{\beta}})$ is also the BLUP of the area effect u_d, for $\gamma_d = \sigma_u^2/(\sigma_u^2 + \psi_d)$.

Good theoretical properties of the BLUP are that it does not require normality assumptions on the model errors or area effects and it is unbiased under the model. Since the BLUP $\hat{\tau}_d$ depends on the unknown area effects' variance σ_u^2, the final small area estimator of τ_d, called *empirical* BLUP (EBLUP), is obtained by replacing σ_u^2 in the BLUP by an estimator $\hat{\sigma}_u^2$ with good properties (including consistency). The EBLUP preserves unbiasedness under symmetric distributions and under certain regularity assumptions on the considered estimator $\hat{\sigma}_u^2$ of σ_u^2 (even and translation invariant function of the sample data $\mathbf{y}$), which are satisfied by the usual estimators.

Another nice property of the BLUP under the FH model (1)–(2) is that, for known $\boldsymbol{\beta}$, its mean squared error (MSE) under the model is $\gamma_d\psi_d$, where $\gamma_d \in (0, 1)$, and hence it cannot be less efficient than the corresponding direct estimator (whose variance is ψ_d). For sufficiently large number of areas D, the WLS estimator of $\hat{\boldsymbol{\beta}}$ will be efficient enough and the EBLUP under the FH model that uses $\hat{\boldsymbol{\beta}}$ is likely to be more efficient than the corresponding direct estimator.

Another appealing property of the BLUP $\hat{\tau}_d$ for practitioners is that it can be expressed as a weighted average of the considered direct estimator $\hat{\tau}_d^{DIR}$ (customary estimator for areas with large sample size) and the regression-synthetic estimator, $\mathbf{x}_d'\hat{\boldsymbol{\beta}}$. The weight attached to the direct estimator is $\gamma_d = \sigma_u^2/(\sigma_u^2 + \psi_d)$, which grows as the error variance ψ_d decreases (or the area sample size increases), or when the unexplained variation σ_u^2 is large. This means that, for areas with large sample size or when the auxiliary information is not very powerful, the BLUP becomes close to the direct estimator; otherwise, it becomes close to the regression-synthetic estimator $\mathbf{x}_d'\hat{\boldsymbol{\beta}}$. This leads to the BLUP automatically borrowing strength only for the areas in which it is actually needed, and tending to the conventional survey direct estimators for the areas where the area sample size is large or in the cases in which the model is not useful.

Another good consequence of this weighted average form of the BLUP is that, since the weight attached to the direct estimator γ_d grows as the area sample size n_d increases, if the considered direct estimator is design unbiased (unbiased under the sampling replication mechanism), which is often the case, then the BLUP becomes also design unbiased as the area sample size increases. Hence, as long as the area sample size is positive, the BLUP accounts to some extent for the sampling design, provided that the direct estimator uses the survey weights. This fact protects the BLUP against informative survey designs, in which the selection mechanism of the units for the sample depends on the actual values of the variable of interest for those units, being potentially less affected by sample selection bias than the basic procedures described in the next sections.

The exact mean squared error (MSE) of the EBLUP does not have a closed form and, therefore, analytical approximations with good properties for large number of areas D have been obtained under normality. Similarly, an analytical estimator of the true MSE that is nearly unbiased for D large has been obtained under normality. This MSE estimator depends on the actual estimator of σ_u^2 that is applied; for the exact formulas, see Rao and Molina (2015), Section 6.2.1. For areas with moderate or large sample sizes and for indicators (and corresponding direct estimators) defined in terms of sums over the area units, the normality assumption is minimally ensured by the Central Limit Theorem for finite populations (see, e.g., Hájek 1964). However, for areas with small sample sizes or when target indicators are not defined in terms of sums over the area units, the normality assumption should be validated if these customary MSE estimates are considered; otherwise, they should be interpreted with care. Another good consequence of the use of aggregated data only is that unit-level outliers will have little effect on the EBLUP based on the FH model.

A clear drawback of the FH model is that the sampling variances of direct estimators, $\psi_d, d = 1, \ldots, D$, are assumed to be known even if they are actually estimated, and the customary MSE estimators do not account for the uncertainty due to estimation of these variances.

Note also that, if one wishes to estimate several indicators, a different model needs to be constructed for each indicator, even if they are all based on the same welfare variable. This means that one needs to find area level covariates that are linearly related to each indicator. It seems more reasonable to fit a model for the considered welfare variable at the unit level and then estimate all of the target monetary indicators based on that same model, as is done in the procedures based on unit-level models described in Sects. 3 and 4, provided that appropriate unit-level data sources are available.

Moreover, the area-level auxiliary information has most often the shape of area means or proportions. Hence, for indicators with a more complex shape than the simple mean welfare, it might be hard to find area-level covariates that are linearly related with that complex function of welfare for the area; hence, the linearity assumption in the FH model might fail, and finding an adequate transformation for linearity to hold might not be easy in some applications.

An additional drawback of the FH model is that it is not applicable for areas with zero sample size, where the direct estimators do not exist. Note also that, even when the area sample size is not zero, due to the small sample sizes, direct estimators of the poverty rates or gaps might be equal to zero or one, when either no one or everyone is below the poverty line. For those areas, the estimated sampling variances of the direct estimators are also equal to zero, which prevents from the application of the FH model, unless another approach is applied to obtain strictly positive variances. One solution for those areas is to apply the regression-synthetic counterparts $\mathbf{x}_d'\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is obtained fitting the FH model to all other areas, but these estimators do not account for the area effect.

In applications with several dependent target indicators, multivariate versions of the FH model may help providing even more efficient estimators. A multivariate version of the FH model was originally proposed by Fay (1987) and has been later used to improve the estimates of median income of four-person families by using direct estimators of median income for three and five-person families by Datta et al. (1991) and Datta et al. (1996). A bivariate FH model with t-distribution was considered by Bell and Huang (2006) to account for outliers in the poverty ratios for school-aged (5–17) children for US states in 2002. A bivariate FH model was also applied by Benavent and Morales (2016) to estimate poverty proportions and gaps at province level in 2005 and 2006 using the Spanish SILC.

Substituting (1) in (2), we obtain a single linear mixed model. However, if a link function $g(\tau_d)$ different from the identity is required on the left-hand side of (1), the two models are *unmatched*, in the sense that we cannot obtain a single linear mixed model. This happens for example when estimating poverty rates and applying the logit link on the left-hand side of (1). Note that no transformation is taken in (2) because, otherwise, design unbiasedness would fail for the transformed direct estimator. The arcsine transformation is a common alternative to the logit employed for proportions. This transformation of poverty rates was considered by Casas-Cordero Valencia et al. (2016). Unmatched sampling and linking models were studied originally by You and Rao (2002a, b). Franco and Bell (2013, 2015) considered alternatively a bivariate Binomial Logit Normal model for estimation of poverty rates of school-aged (5–17) children for counties.

Extension of the basic FH model to account for temporal and/or spatial correlation have also been considered in the poverty mapping context. For example, Esteban et al. (2012a, b) used the FH model with temporal correlation proposed originally by Rao and Yu (1994) to estimate small area poverty indicators. A spatial FH model was used by Giusti et al. (2017) to estimate mean income and poverty rates for the 57 Labor Local Systems (LLS) of the Tuscany region in Italy for the year 2011, similarly as was done by Salvati et al. (2014) in the LLS of Lombardy, Tuscany and Campania for 2008. Marhuenda et al. (2013) considered a spatio-temporal model for the estimation of poverty indicators for Spanish provinces in 2008, making use of survey data from years 2004–2008. On a different note, big data covariates were employed by Marchetti et al. (2015) under the measurement error model proposed by Ybarra and Lohr (2008), to estimate poverty rates and mean income in the ten provinces of the Tuscany region in Italy using data from the 2008 EU-SILC and the 2001 Italian Population Census.

The EBLUP based on the basic FH model is implemented in several R packages, namely the `sae` package (Molina and Marhuenda 2015), which includes also spatial and spatio-temporal extensions of the FH model, together with either analytical or resampling-based functions for MSE estimation, the `Josae` package (Breidenbach 2018), which allows to include heteroscedasticity in the model, `hbsae` (Boonstra 2012), which includes also Hierarchical Bayesian methods and `BayesSAE` (Shi 2018), which includes functions for unmatched models with log or logit transformations, and spatial models. The temporal Rao and Yu (1994) model is implemented in the R package `saery` (Esteban Lefler et al. 2014), multivariate FH models can be fit with the R package `msae` (Permatasari and Ubaidillah 2020) and, finally, measurement error models can be applied with the package `saeME` (Mubarak and Ubaidillah 2020). The EBLUP based on the basic FH model is also available as a command in Stata `fhsae` (Corral et al. 2018). The Stata command `fhsae` also allows for aggregation of estimators and obtains the mean cross product error detailed in Rao and Molina (2015), Section 6.2.6. Additionally, Halbmeier et al. (2019) introduced the `fayherriot` Stata command which allows for adjusting non-positive random effect variance estimates and also for transformation to deal with violations of the model's assumptions.

3 ELL method

The *ELL method*, due to Elbers et al. (2002, 2003), was until 2015 the de facto methodology used by the WB to obtain small area estimates of poverty and inequality indicators and perhaps has been so far the most applied SAE method across the globe. Good reviews of this approach with discussion and comparison to other methods can be found in Haslett (2016) and Das and Haslett (2019). For an early application this method, see, e.g., Bedi et al. (2007), and for a more recent one, see, e.g., Farris et al. (2017) on the estimation of FGT poverty indicators in Uganda.

The WB created a free software package with a simple point and click interface that could be easily used by any practitioner. The `PovMap` software (Zhao 2006) is also incredibly computationally efficient and fast, allowing users, even with limited computing power, to work with census data without facing memory limitations. In 2018, a Stata version of the `PovMap` software was released (Nguyen et al. 2018). The Stata command `sae` replicates most of the procedures and methods of the original `PovMap` software, and has become popular, because it facilitates replicability and inclusion of new methods.

The ELL method was designed to deal with general monetary poverty or inequality indicators defined as functions of the values of a welfare variable for the population units. It assumes a single unit-level model for the logarithm of household welfare; concretely, the nested error model of Battese et al. (1988), but where the random effects are customarily specified for the sampling clusters (or primary sampling units, often census tracts), which are typically nested within the areas. However, estimates are finally produced at the area level by aggregation. Specifically, let y_{dci} denote the log-welfare for household i within cluster c in the population from domain d and $\mathbf{x}_{dci}$ be a $1 \times p$ vector of auxiliary variables for that household, which might include unit, cluster and area level covariates (ELL recommend the inclusion of contextual, that is, aggregated, covariates). In the most basic setup with normality and homoscedastic error variances, the ELL method assumes that

$$y_{dci} = \mathbf{x}'_{dci}\boldsymbol{\beta} + u_c + e_{dci}, \quad i = 1, \dots, N_{dc},\ c = 1, \dots, M_d,\ d = 1, \dots, D, \quad (3)$$

where u_c and e_{dci} are, respectively, cluster and household-specific idiosyncratic errors, assumed to be independent from each other, following

$$u_c \overset{iid}{\sim} N(0, \sigma_u^2), \quad e_{dci} \overset{iid}{\sim} N(0, \sigma_e^2).$$

Here, M_d is the number of clusters in which area d is partitioned and N_{dc} is the number of population households in cluster c from area d, for $d = 1, \dots, D$. Finally, $\boldsymbol{\beta}$ is the $p \times 1$ vector of regression coefficients.

Traditionally, the ELL method considers heteroscedastic errors $e_{dci} \overset{iid}{\sim} N(0, \sigma_{dci}^2)$, and fits a preliminary model to predict these variances σ_{dci}^2, called the *alpha model*. The predicted error variances are then replaced as known values of σ_{dci}^2 in the model.

The model (3) is fitted to the survey data at the household level, and the resulting model parameter estimates are then used to generate multiple synthetic censuses by bootstrap from the fitted model, using the census auxiliary information. With each generated census, the target indicators τ_d are calculated for each area, and the averages of these indicators across the bootstrapped censuses are taken as ELL estimators, $\hat{\tau}_d^{ELL}$. Similarly, variances of the indicators obtained from the different censuses generated by bootstrap are taken as noise measures of ELL estimators. These bootstrap variances incorporate the uncertainty in the estimation of the model parameters because, in each bootstrap replicate, new model parameters are generated from an approximation to the distribution of the model parameter estimators obtained from the initial fit of the model to the original survey data, for details see (Corral et al. (2021b)), Section 2.

The traditional ELL method has met criticism almost since its publication. Banerjee et al. (2006), in a review of research conducted at the WB, raised concerns regarding accuracy of the method, as well as whether the ELL noise measures represent the real inaccuracy. They argued that knowledge of the values of the available covariates is not enough to produce an accurate estimate of poverty for a given area, because there are unobservable aspects toward an area's poverty. In a sense, the panel was suggesting to account for the unexplained between-area heterogeneity, which was eventually addressed by the Empirical Best prediction method of Molina and Rao (2010), described in Sect. 4.

Concerns regarding the estimation of the noise were not unfounded. In ELL (2003), the random effects are for the clusters (instead of the target areas), but the estimates are then reported for the areas. Banerjee et al. (2006) noted that ELL standard errors will not account for the correlation between the observations in different clusters within the same area. Producing estimates for the areas (indexed here by d), which represent a higher aggregation level than the clusters c, for which random effects are specified, may not be appropriate in cases of considerable between-area heterogeneity, and may underestimate the estimator's standard errors (Das and Chambers 2017). The inclusion of area-level covariates in the model, as suggested in ELL (2002), may reduce the unexplained between-area heterogeneity in the model and in turn the need for accounting for area effects but, based on empirical results, Marhuenda et al. (2017) recommend to put the random effects at least at the same aggregation level where estimation is desired.

In order to illustrate the main drawback of ELL method, consider the slightly simpler version that keeps fixed the model parameters used to generate each synthetic census of welfares in ELL's bootstrap procedure, to their true values (see details of the bootstrap procedure in Corral et al. 2021b, Section 2). Then, the above ELL estimator $\hat{\tau}_d^{ELL}$ reduces to a basic Monte Carlo (MC) approximation of the marginal expectation of the indicator, $E(\tau_d)$, with respect to the distribution of welfare (given the covariates) induced by the model (3). In this case, $\hat{\tau}_d^{ELL} = E(\tau_d)$ is exactly unbiased under the model.

Consider now the simpler case where clusters are equal to the areas (i.e., taking $M_d = 1$ and hence $N_{dc} = N_d$, $d = 1, \ldots, D$ in model (3)). Then, if we estimate simple area means $\tau_d = N_d^{-1} \sum_{i=1}^{N_d} y_{di}$ with a model for the welfare variable without transformation (i.e., where the welfare is the actual response variable in the model y_{di}), then the ELL estimator is approximating the expected value $E(\tau_d) = E(\bar{\mathbf{x}}_d' \boldsymbol{\beta} + u_d + \bar{e}_d) = \bar{\mathbf{x}}_d' \boldsymbol{\beta}$, where $\bar{\mathbf{x}}_d = N_d^{-1} \sum_{i=1}^{N_d} \mathbf{x}_{di}$ and $\bar{e}_d = N_d^{-1} \sum_{i=1}^{N_d} e_{di}$, which has zero expectation. Thus, the final ELL estimator reduces to the regression-synthetic estimator $\bar{\mathbf{x}}_d' \boldsymbol{\beta}$, which omits the area effect u_d and hence does not account for the idiosyncracy of the area. Note that, if $\boldsymbol{\beta}$ was actually known (best case), then $\bar{\mathbf{x}}_d' \boldsymbol{\beta}$ would not be using the actual survey data on the welfare variable y_{di}, only the census auxiliary information $\mathbf{x}_{di}$ and the model assumptions presumed by the statistician. Synthetic estimators rely very strongly on the model assumptions and might be misleading if the regression model is incorrectly specified. In fact, the uncertainty (noise) measures of the estimators are estimated assuming that the model actually holds and hence these measures do not account for model misspecification. Hence, the estimated noise may be small even if the model fails, giving in that case a lot of credibility to the (misleading) small area estimates. As we shall see later, even if the model is correctly specified, the MSE of ELL estimators can be substantial, especially if the explanatory power of the model is weak, and ELL's estimates of noise measures can severely understate the true MSEs under the model.

4 EB method

The empirical best/Bayes (EB) method for the estimation of general indicators in small areas was introduced by Molina and Rao (2010) under the support of the European project entitled "Small Area Methods for Poverty and Living Condition Estimates" (SAMPLE), (http://www.sample-project.eu/). This method is similar in spirit to ELL, in the sense that it combines survey data with census (or administrative records) auxiliary data, uses a unit-level model for the welfare variable and hence it can estimate general (and several) indicators depending on the welfare, based on that single model. Nevertheless, in contrast with ELL, which is not defined based on optimality ideas, EB gives approximately the *best* predictor, in the sense of minimum MSE under the model. Consequently, it provides estimators with better efficiency than ELL estimators when the model assumptions hold and, in certain cases, the gains in efficiency with respect to ELL may be remarkable, as illustrated by Molina and Rao (2010). This occurs because, as will be seen below, it uses more extensively the precious welfare information provided by the survey sample.

The EB method is implemented within the `sae` R package (Molina and Marhuenda 2015), as well as in an update to the `sae` package in Stata (Nguyen et al. 2018 found in https://github.com/pcorralrodas/SAE-Stata-Package). The EB method has been applied, e.g., to estimate poverty indicators in Spanish provinces by gender in 2006 using SILC data together with labor force survey (EPA in Spanish) data from the same year (Molina and Rao 2010), mean income in Mexican municipalities using the Module of Socio-economic Conditions from the 2010 ENIGH survey and using census data from the same year (Molina and Martín 2018), mean income and (non-extreme) poverty rates for census tracks by gender in Montevideo, Uruguay, using data from the 2011 Continuous Household Survey (ECH in Spanish) and the 2011 Population Census (Molina 2019) and poverty rates and gaps in Palestinian localities by gender using data from the Palestinian Expenditure Consumption Survey (PECS) corresponding to 2016/17 and the Population Census of 2017 (Molina Peralta and García Portugués 2020).

The EB method, as originally proposed in Molina and Rao (2010), considers that a one-to-one transformation (not necessarily the log) of the welfare measure in the population units (e.g., individuals or households), denoted here y_{di}, follows the BHF model with random effects for the areas of interest, that is,

$$y_{di} = \mathbf{x}'_{di}\boldsymbol{\beta} + u_d + e_{di}, \quad i = 1, \ldots, N_d,\ d = 1, \ldots, D, \tag{4}$$

where D is the number of small areas where estimation is desired and N_d is the population size of the dth area. Again, area effects u_d and errors e_{di} are independent, following usual assumptions, that is,

$$u_d \overset{iid}{\sim} N(0, \sigma_u^2), \quad e_{di} \overset{iid}{\sim} N(0, \sigma_e^2).$$

Heteroscedasticity is included in the description of the EB approach given in Rao and Molina (2015), but it is not implemented in the `sae` R package, so we describe here the basic homoscedastic case.

The target indicators τ_d are defined as functions of the welfare for all the units in the areas, that is, $\tau_d = h(\mathbf{y}_d)$, where $\mathbf{y}_d = (y_{d1}, \ldots, y_{dN_d})'$. Let us split the census welfare vector for any area d, $\mathbf{y}_d$, as $\mathbf{y}_d = (\mathbf{y}'_{ds}, \mathbf{y}'_{dr})'$, where $\mathbf{y}_{ds}$ and $\mathbf{y}_{dr}$ contain, respectively, the welfare for sampled and non-sampled units in area d. The best predictor of τ_d is then defined as the predictor $\hat{\tau}_d$ (i.e., the function of the sample data $\mathbf{y}_{ds}$) which minimizes the MSE given by $\mathrm{MSE}(\hat{\tau}_d) = E(\hat{\tau}_d - \tau_d)^2$. For a sampled area d (i.e., with sample size $n_d > 0$), the best predictor of τ_d is the conditional expectation $E(\tau_d|\mathbf{y}_{ds})$, where this expectation is taken with respect to the distribution of the unknowns (the non-sample vector $\mathbf{y}_{dr}$) given the observed sample data, $\mathbf{y}_{ds}$, obtained from model (4). For an out-of-sample area (with $n_d = 0$), the best predictor of τ_d reduces to $E(\tau_d)$, which is actually the expectation that is approximated by ELL estimator.

Good theoretical properties of the best predictor are that it is unbiased under the model, since by the law of iterated expectation, we have $E\,(E(\tau_d|\mathbf{y}_{ds})) = E(\tau_d)$ and it minimizes the MSE under the model. Unlike the ELL estimator, it accounts adequately for the area effect. To illustrate this fact, consider that the target parameter is the area mean $\tau_d = N_d^{-1}\sum_{i=1}^{N_d} y_{di}$, and there is no transformation of welfare. In this case, the best predictor of τ_d reduces to

$$E(\tau_d|\mathbf{y}_{ds}) = \bar{\mathbf{x}}_d'\boldsymbol{\beta} + \gamma_d(\bar{\mathbf{y}}_{ds} - \bar{\mathbf{x}}_{ds}\boldsymbol{\beta}), \tag{5}$$

which corrects the regression-synthetic estimator $\bar{\mathbf{x}}_d'\boldsymbol{\beta}$ by including the predicted area effect $\hat{u}_d = \gamma_d(\bar{\mathbf{y}}_{ds} - \bar{\mathbf{x}}_{ds}\boldsymbol{\beta})$, obtained based on the survey data from the area. Moreover, the size of this correction depends on the share of between-area heterogeneity σ_u^2 from the total variance, $\gamma_d = \sigma_u^2/(\sigma_u^2 + \sigma_e^2/n_d)$, which becomes null only when all the existing area heterogeneity is completely explained by the available auxiliary variables ($\sigma_u^2 = 0$) or when the area is not sampled ($n_d = 0$ and then γ_d is taken as zero). According to (5), this procedure makes more efficient use of the survey data for areas that are sampled and, for those areas, it does not rely so strongly on the regression-synthetic estimator $\bar{\mathbf{x}}_d'\boldsymbol{\beta}$, obtained from the linear regression that ignores the area effect.

Under the nested error model (4), the distribution of $\mathbf{y}_{dr}|\mathbf{y}_{ds}$, which is also normal due to the normality assumption on the area effects and errors, depends on the true values of the model parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma_u^2, \sigma_e^2)'$. In practice, these parameters are estimated using the survey data for all the areas, and the resulting estimators $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\sigma}_u^2, \hat{\sigma}_e^2)'$ are replaced in the best predictor $\hat{\tau}_d^B = E(\tau_d|\mathbf{y}_{ds}; \boldsymbol{\theta})$ of τ_d. The result of this is the EB predictor $\hat{\tau}_d^{EB} = E(\tau_d|\mathbf{y}_{ds}; \hat{\boldsymbol{\theta}})$ of τ_d. If the model parameter estimators in $\hat{\boldsymbol{\theta}}$ are consistent as the total survey sample size $n = \sum_{d=1}^{D} n_d$ grows, then the EB predictor approaches the best predictor. Note that the total sample size of official household surveys is typically very large so, in these type of applications, EB is expected to preserve approximately the unbiasedness and optimality of the best predictor.

When the expectation defining the EB predictor, $E(\tau_d|y_{ds})$, has no explicit form, MC simulation may be used to approximate it. MSEs of EB predictors may be approximated by the parametric bootstrap method for finite populations proposed by González-Manteiga et al. (2008). However, when EB predictors require MC simulation, the full procedure for estimation of the indicators and of their MSEs by bootstrap may be computationally much more intensive than the traditional ELL bootstrap approach. Note that, in that case, for every bootstrap replicate (which requires generation of a full census), the MC method needs to be reproduced (which entails generation of many censuses), so the number of generated censuses is multiplied. For details on the MC simulation method for approximating the EB predictor and on the bootstrap procedure for MSE estimation, see Molina and Rao (2010), Molina (2019) or Corral et al. (2021b).

Considering the clusters of the ELL procedure as the small areas of interest for more fair comparisons, Molina and Rao (2010) obtained substantial efficiency gains of EB estimators with respect to ELL ones in a (limited) simulation experiment. Actually, for known model parameters, EB estimators cannot be worse than ELL estimators under the same model assumptions. The main reason for the large gains in efficiency is that the best predictor conditions on the vector of survey welfares $\mathbf{y}_{ds}$, whereas ELL estimator approximates the unconditional expectation $E(\tau_d)$, which does not make use of the precious information on the actual welfare variable $\mathbf{y}_{ds}$, coming from the survey.

The EB predictor entails calculation of the conditional expectation $E(\tau_d|\mathbf{y}_{ds})$ with respect to the distribution of $\mathbf{y}_{dr}|\mathbf{y}_{ds}$ under model (4). This conditional distribution depends on the values of the auxiliary variables $\mathbf{x}_{di}$ for all the non-sampled units in the target area d. Hence, one of the earliest criticisms of the EB procedure was that it requires to identify which are the non-sampled units in the census file. The traditional ELL approach does not require to link the census and survey units, and many times linking the two data sets is not possible. In fact, often the survey sample is not really a subset of the population defined by the available census, possibly because they correspond to different time periods. This problem was solved by Correa et al. (2012), who defined the Census EB (CEB) predictor. This procedure is designed for the case when the survey sample is not a subset of the population defined by the available census. It considers that the vectors composed by attaching the survey and census welfare data $\mathbf{y}_d^* = (\mathbf{y}_{ds}', \mathbf{y}_d')'$ for each area $d = 1, \ldots, D$, follow the nested error model (4), but the target area indicator $\tau_d = h(\mathbf{y}_d)$ is defined in terms of the vector of census welfares $\mathbf{y}_d$. The Census Best predictor of $\tau_d = h(\mathbf{y}_d)$ is defined as the best predictor of $\tau_d = h(\mathbf{y}_d)$, that is, $\hat{\tau}_d^{CB} = E(\tau_d|\mathbf{y}_{ds}; \boldsymbol{\theta}) = \hat{\tau}_d^{B}$, under the nested error model for the enlarged vector $\mathbf{y}_d^*$. When the area sampling fraction n_d/N_d is negligible, the CEB predictor is practically equal to the EB predictor, as illustrated in Corral et al. 2021b. Molina (2019) presents a slight variation of the parametric bootstrap procedure of González-Manteiga et al. (2008) for estimation of the MSE of the CEB predictor, which avoids linking the survey and census data sets. This procedure gives the bootstrap approximation to the MSE of the CEB predictor under the nested error model for $\mathbf{y}_d^*$.

Since the bootstrap approach for MSE estimation of the EB predictor can be computationally very intensive for large populations and very complex indicators, Molina et al. (2014) proposed a hierarchical Bayes (HB) alternative that avoids performing the bootstrap, since posterior variances are obtained directly from the predictive distribution of the indicators of interest. For very complex indicators, such as those based on pairwise comparisons or that require sorting area units, or when the population is too large, Ferretti and Molina (2012) introduced a fast EB approach.

A problem that may affect the basic EB estimators is when the sampling design is informative, that is, when the selection mechanism depends on the values of the variable of interest, given the available covariates. These estimators do not use the survey weights and hence are not design unbiased under complex sampling designs. As a consequence, they may have substantial sample selection bias, as shown in Guadarrama et al. (2014). Similarly to the pseudo-EBLUP approach by You and Rao (2002a, b), which introduces the survey weights in the EBLUP of a linear indicator, Guadarrama et al. (2018) defined the pseudo-EB (PEB) predictor, which extends the EB predictor by incorporating the survey weights. The PEB is expected to reduce the design bias of EB under complex, possibly informative, survey designs.

For the case of two-stage sampling design, or when estimation is intended at two different (nested) aggregation levels, Marhuenda et al. (2017) extended the EB procedure based on the nested error model to the twofold nested error model that includes subarea (e.g., cluster) effects nested within the area effects (Stukel and Rao 1999). This procedure has now been incorporated to the `sae` Stata package. Marhuenda et al. (2017) obtained clear losses in efficiency of EB estimators of poverty indicators when the random effects are specified only for the subareas (e.g., clusters), but estimation is desired for areas. Recently, Guadarrama et al. (2020) have further extended the EB procedure by considering a twofold model with correlated time effects nested within the area effects, for the estimation of general nonlinear indicators.

A crucial assumption in the EB procedure, which actually affects also ELL procedure, is the normality assumption of area effects and unit level errors. There are different approaches to make this assumption hold (at least approximately). First of all, note that for whatever random variable X following a probability distribution F, it holds that $F(X) \sim U(0, 1)$, and then $\Phi^{-1}(F(X)) \sim N(0, 1)$, where $\Phi(u)$ is the c.d.f. of a standard normal distribution. This means that there is a transformation of the original variable X that follows a normal distribution. The problem is that the true F is unknown, and hence the challenge is to find a transformation that yields (at least approximate) normality. Hence, finding a good transformation of the welfare variable to be included as response in the model is a very important issue, both in EB and ELL methods, although under the PovMap implementation of ELL bootstrap procedure, bootstrap errors are drawn from the empirical distribution in cases where the normality assumption does not hold. The most widely used transformation for monetary variables is probably the log, since this kind of welfare variables are most often severely right-skewed and may show heteroscedasticity. When estimating area means of the original variables, Molina and Martín (2018) studied the analytical EB predictors under the model with log transformation and obtained second-order correct MSE estimators with closed form.

Other popular transformations are those from the power or Box-Cox families. Beyond the log, which is a special case in both families, an appropriate member of these families may be selected in the implemented function for the EB method `ebBHF()`, included in the R package `sae` (Molina and Marhuenda 2015). However, note that, in the presence of welfare values close to zero, the log transformation pushes these small values towards minus infinity, which may in turn produce a thin yet long left tail in the distribution. Hence, a serious right-skeweness problem is transformed into a mild left-skewness. A simple way of avoiding such an effect is just adding a positive shift to the welfare variable, before taking log. A drawback of this approach is that the selection of this shift, as well as choice of the specific Box-Cox or power transformation, needs to be based on the actual survey data, see Rojas-Perilla et al. (2020), who consider also other families of transformations. The most recent version of the `sae` Stata package also incorporates the Box-Cox transformation, as well as the log-shift transformation (available at https://github.com/pcorralrodas/SAE-Stata-Package). The R package `emdi` (Kreutzmann et al. 2019) also provides functions for the estimation, assessment and mapping of regional disaggregated indicators under these families of transformations.

An alternative approach to deal with non-normality is to consider a skewed distribution for the original welfare variable. Diallo and Rao (2018) extended the EB procedure to the skew normal distribution and Graf et al. (2019) considered the EB procedure under a generalized beta of the second kind (GB2). This distribution contains four parameters, one for each tail, offering a more flexible framework for modeling skewed data of different shapes.

When the deviation from normality appears only in isolated observations, which may occur even after transformation of the target variable, another possible approach is to consider models for outliers, such as models based on mixtures of Normal distributions. A mixture may be specified for the model errors, as proposed by Gershunskaya (2010) for the estimation of small area means, or for both the random effects and errors, as considered by Elbers and Van der Weide (2014) for the estimation of poverty and inequality indicators.

Finally, another approach to deal with representative outliers (i.e., not incorrect observations such as typing errors) is to reduce their impact in the small area estimation procedures by considering robust estimation procedures. The robust M-quantile SAE method introduced by Chambers and Tzavidis (2006) (see also Tzavidis et al. 2010) to deal with unit-level outliers has also been applied to the estimation of poverty indicators by Marchetti et al. (2012) and by Giusti et al. (2012). A robust version of the EBLUP based on BHF model was proposed by Sinha and Rao (2009), but it is suitable only for the estimation of linear indicators. Robust SAE methods for the basic area and unit level models are implemented in the R packages `rsae` (Schoch 2014) and `saeRobust` (Warnhol 2018).

In general, robust procedures deliver less efficient estimators when normality holds. In simulation experiments, Sinha and Rao (2009) obtained that outliers in the area effects have a minor effect on the EBLUP. On the other hand, since the total sample size of the usual household surveys employed for the estimation of poverty is typically very large and the effect of unit-level outliers in the fitted values tends to zero as the total sample size increases, the effect of a couple of isolated unit-level outliers is expected to be negligible, becoming significant only when the percentage of outliers is substantial (e.g., over 1%). The effect of very large or small isolated welfares is limited specially when estimating FGT measures, where welfares are simply classified to be above or below the (fixed) poverty line.

5 Other types of models

Generalized linear mixed models (GLMMs) have also been used for poverty estimation. Beyond the FH model, GLMMs at the area level have been applied for the estimation of particular monetary indicators, see, e.g., the hierarchical Beta mixed regression model applied by Fabrizi and Trivisano (2016) for the estimation of the Gini coefficient using Italian SILC data, or the Poisson mixed model with time effects applied by Boubeta et al. (2017) for the estimation of poverty rates in counties of the Spanish region of Galicia by gender.

GLMMs at the unit level have been applied as well, see, e.g., the Binomial logit model used by Hobza and Morales (2016) to estimate poverty proportions in counties from the Spanish region of Valencia, and the temporal extension of the same model considered by Hobza et al. (2018) to estimate poverty increments apart from poverty proportions with the same data. GLMs are also used by Isidro et al. (2016) in the so called extended structure preserving estimation (extended SPREE) method for updating small area estimates of poverty in intercensal years. A couple of methods for SAE estimation under GLMMs for counts are implemented in the R package `saeeb` (Fauziah and Wulansari 2020).

A drawback of the procedures based on GLMMs is that they are suitable for the estimation of totals or proportions such as the poverty rate, but are not designed for the estimation of other poverty or inequality indicators, such as the poverty gap. Moreover, reducing the continuous welfare to a binary variable may entail a loss of information, which might yield modest efficiency gains of the final small area estimators with respect to direct estimators. Finally, the optimal EB predictors under a unit level GLMM (with link function different from the identity) do not have closed-form and hence require the application of computationally intensive methods such as MC simulation. Joined to the fact that the likelihood does not either have a closed form and hence complex numerical methods are also needed for model fitting, the procedure for MSE estimation becomes even more computationally intensive, making the full approach suitable only for small populations. A way to alleviate this computational problem is to consider suboptimal estimators different from the EB predictor, but then good (bias and MSE) properties are not guaranteed.

6 Revised Word Bank methodology

The WB revised its methodology in 2014 according to the fitting method and the bootstrap procedure by Van der Weide (2014), intended to solve the main drawback of ELL by accounting for the area effects, similarly as EB method does. Virtues of this proposal are that it accounts for heteroscedasticity and incorporates the survey weights in the estimates of the regression coefficients and also in the variance components. For estimation of the variance components using the survey weights, he followed the approach of Huang and Hidiroglou (2003), based on Henderson method III. The survey weights were also included in the predicted area effects, in a similar fashion as in the pseudo-EBLUP proposed by You and Rao (2002a, b).

The bootstrap method proposed by Van der Weide (2014) generates synthetic censuses using the predicted area effects (and conditioning on them across the bootstrap procedure), unlike in the original ELL method, where random effects were generated in each bootstrap replicate from their assumed distribution. The considered predicted area effects are an extension of the ones in Guadarrama et al. (2018) to the case of heteroscedasticity. However, trying to make the procedure less dependent on distributional assumptions, the bootstrap procedure first draws samples of clusters with replacement from the original clusters in the survey data and then fits the model to the data from the bootstrapped sample of clusters, obtaining the set of model parameter estimates, along with predicted effects for the clusters that were sampled, and their corresponding estimated variance. Then, the model parameter estimates are used to simulate synthetic censuses of welfare. The indicator of interest is computed for each simulated census and, finally, the average of the indicators across the simulated censuses is taken as the final estimator, called here clustered-bootstrap EB (CB-EB) estimator and denoted as $\hat{\tau}_d^{CB-EB}$. Similarly as in the ELL method, the variance of the indicators across the simulated censuses is again taken as uncertainly measure of the CB-EB estimator.

As in the computational procedures of the previously described methods, the final estimator of the poverty indicator is obtained as an average across bootstrap replicates. By the MC principle, this average is approximating an expectation. However, here the expectation would be with respect to the joint distribution of all the measures that are generated in each bootstrap replicate: (a) the sample of clusters; (b) the model parameters, which are newly generated in each simulation, (c) the location effects, which are generated from their conditional distribution given the sample residuals (determined by the nested error model under normality), and (d) household-specific errors, which are generated from their distribution under the assumed nested error model with normality. Theoretical properties of the final estimator are not clear under this procedure. Moreover, the normality assumption is still needed.

7 Comparison of unit level methods

This section compares the unit-level model-based procedures for poverty estimation that have been implemented at the WB, based on bias and efficiency (MSE) of the predictors under the model and also under the design, as well as looking at the properties of the corresponding MSE estimators.

If model assumptions hold, EB estimators are nearly unbiased and nearly optimal in terms of MSE. Hence, under the same assumptions, they are likely to perform better than the traditional ELL and the new CB-EB estimators in terms of MSE, and part of their good performance is due to the fact that they use the survey data to estimate the area effects. Nevertheless, they become similar to ELL estimators for non-sampled areas, so the benefits of EB method are most salient in the sampled areas.

Molina and Rao (2010) showed an outstanding superiority of the EB approach over ELL in terms of a smaller MSE in a model-based simulation experiment. However, this simulation experiment has met some criticism, mostly because it considered scenarios under which ELL method is seldom applied. First, the total population size and the area population sizes were too small. Second, all the areas were sampled and the sample within each area represented 20% of the population, something hardly seen in real-world applications. Another point of departure from reality is that the considered model was really poor, with only two covariates yielding an adjusted R^2 of less than 0.01. Under the usual applications of the ELL approach, the number of covariates is much greater, reaching most of the times an adjusted R^2 of over 0.5. Moreover, contextual area-level covariates advocated by ELL (2002, 2003) to try to explain the variation across areas, were not used in the simulation experiment conducted in Molina and Rao (2010).

Recently, Corral et al. (2021b) have conducted new MC simulation experiments extending those of Molina and Rao (2010) in the following aspects: (i) including the area means of the covariates as additional variables in the model, playing the role of contextual variables; (ii) considering a model with much better explanatory power by including additional covariates; (iii) considering much larger area population sizes and much smaller sampling fractions; (iv) Generating errors from a Student's t_5 instead of a normal distribution; (v) decreasing the overall sample size and area sample sizes.

The MC simulations are model-based, where censuses of welfare are generated from the nested error model and the sample units are kept fixed. In a single simulation (one census), results in Corral et al. (2021b) show a rather flat nature of ELL and CB-EB estimators, which appear to be aligned to the national poverty rate, not respecting well the differences in the poverty levels of the different areas, in contrast with the Census EB estimates. This holds true for all the considered models, with weaker or stronger explanatory power, and with or without contextual variables in the form of area means of the covariates. When averaging across MC simulations, CB-EB estimators appear to be very seriously biased, regardless of whether the method is performed with clusters equal to areas or clusters nested within areas, and the best overall performance is for the Census EB estimators, which also show much smaller MSEs than all the other estimators. Even if Census EB predictors are based on normality, they still perform better than the traditional ELL (with bootstrap errors drawn from their empirical distribution) and the CB-EB estimators, when errors are generated from a symmetric distribution with longer tails than the normal and when the area sampling fractions are considerably reduced.

Since the Census EB method takes advantage of the area information on welfare from the survey, unlike ELL, and the two methods are in agreement when the sample size for a given area is zero, obviously their differences decrease as the sample size decreases. They also agree when the unexplained between-area heterogeneity, as measured by σ_u^2, is equal to zero.

The simulation experiments performed under the more realistic larger area population sizes and smaller area sampling fractions also show that MSEs are not correctly estimated under ELL or CB-EB approaches. On the other hand, the parametric bootstrap MSE estimates, based on the original procedure from González-Manteiga et al. (2008), but modified according to Molina (2019) for the case when the census and the survey units cannot be linked, track approximately the true MSEs for the Census EB. For the detailed results of the simulation experiments, see Corral et al. (2021b).

Corral et al. (2021, b) performed a design-based validation experiment using the Mexican Intracensal Survey. This large survey was regarded as a fixed census, and then 500 samples were drawn from it, with a sampling method similar to those applied in the surveys conducted by the Living Standards Measurement Study unit within the WB. This validation study also shows the superiority, in terms of design bias and MSE, of the EB and Census EB methods over the traditional ELL.

The better performance of the Census EB predictors in all these simulation experiments has led to a very recent revision of the methodology employed by the WB and the corresponding software (https://github.com/pcorralrodas/SAE-Stata-Package), which have incorporated the Census EB estimators of Correa et al. (2012), based on the original EB method by Molina and Rao (2010), and also the parametric bootstrap procedure of González-Manteiga et al. (2008), adapted for the case when the survey and census units in the data cannot be linked (Molina 2019).

The newly implemented estimators are extensions of the Census EB estimators that account for heteroscedasticity and include also survey weights in the model parameter estimators and in the predicted area effects, similarly as in Van der Weide (2014). Actually, they are the Census versions of the pseudo-EB estimators proposed by Guadarrama et al. (2018) to reduce the bias due to complex sampling designs, but also accounting for heteroscedasticity and using estimates of the variance components that include the survey weights. The adopted parametric bootstrap procedure is similar to that described in Molina (2019) for the Census EB predictor, where the synthetic censuses are generated in each bootstrap replicate separately from the survey data, although both census and survey data share the same model parameters and area effects.

8 Concluding remarks

Countries often set out to produce small area estimates of poverty, because precise estimates at a more granular level allow for improved allocation of resources, see Elbers et al. (2007). This research comes just in time for the 2020 round of population census and should provide relevant information for the operationalization of the SDGs at a sub-national level.

This paper reviews the basic methods for estimation of monetary poverty and/or inequality indicators in domains of small sample size, together with their variants, designed for scenarios that were not reflected by the assumptions of the basic methods. Concretely, we review the widely used FH area level model, discuss its benefits and drawbacks and mention extensions of this model in the context of poverty applications. We also go over the most common methods based on unit level models; concretely, the traditional ELL methods by Elbers et al. (2002, 2003), the EB method of Molina and Rao (2010), the updates by Van der Weide (2014) and the extended Census EB and parametric bootstrap methods incorporated recently by the WB.

As discussed in Sect. 7, simulation results obtained by Corral et al. (2021b) show substantial gains of the proposed Census EB method with respect to the previous ones in all the considered scenarios (varying area population sizes and sampling fractions, model prediction power and distributional assumptions) and also of the considered parametric bootstrap MSE estimators. In fact, for a single population, the traditional ELL estimates for the different areas tend to align with the national estimate and does not capture the area heterogeneity, although its performance improves when averaging across populations. In turn, the CB-EB update by Van der Weide (2014), as implemented in Povmap's update 2.5, gives estimates showing a considerable bias. In contrast, the bias of the revised Census EB estimators (very close to the original EB) is several orders of magnitude smaller. Furthermore, the true MSE of the Census EB method is substantially smaller. Based on these results, we can conclude that the extended Census EB and bootstrap MSE estimators and their accompanying software implementation represent a massive improvement to the previous WB's methods and overall poverty mapping agenda.

We have seen that EB estimators reduce to ELL ones for non-sample areas and that ELL might not reproduce the area differences in the poverty levels, although adequate contextual-level covariates could ameliorate the problem (Haslett, 2016). Then, caution should be taken when doing prediction for areas that are not sampled even when using the EB method. This is particularly relevant in cases where the explanatory power of the selected covariates is low. In fact, even if the best predictors are unbiased under the model, the bias under the sampling replication mechanism (design bias) may be substantial for those areas with zero sample size. This might occur for areas that deviate considerably from the assumed model, and model diagnostics cannot be done for areas without observations in the welfare. This fact also connects with the importance of model checking in EB method. The normality assumption for some transformation of the welfare variable and the remaining model assumptions need to be carefully checked with the survey data in question, and clear deviations from model assumptions call to making changes in the model so as to achieve the fulfilment of these assumptions, or perhaps to consider extensions of the basic EB method or even develop new ones depending on the particular deviation from the assumptions encountered.

One of the situations that make model assumptions fail is when the sampling design is informative. When the selection of the units depends on the particular values of the study variable even after conditioning on available covariates, the sample will likely represent to a larger extent the units with specific values of the study variable that have greater probabilities than the other units. In design-based inference, sampling weights protect against the sample selection bias by over-weighting the under-represented units and under-weighting the over-represented ones. Hence, design-based estimators that incorporate the sampling weights are (at least approximately) design unbiased. In the stata `sae` package, sampling weights are included in the implemented Census EB approach and in the model fitting method as proposed by Van der Weide (2014), in order to decrease the design bias in the sampled areas due to complex (possibly informative) sample selection mechanisms. This approach of accounting for the sampling design is somehow heuristic, in the sense that it is not based on formal theory or optimality ideas, but it does simple corrections on the estimators, which do not require additional modeling assumptions. Approaches that do correct for the sample selection mechanism in a formal way in the field of small area estimation but which require additional modelling assumptions, have been developed by Pfeffermann and Sverchkov (2007). Sverchkov and Pfeffermann (2018) extend it to correct also for non-ignorable nonresponse. These procedures have been applied to the estimation of small area means of the response variable in the model, but their extension for the estimation of more complex indicators is currently under study.

A problem that often appears with unit-level models is that the census used to obtain the auxiliary information might be outdated. This often occurs when estimation is desired for intercensal years and may yield to biased Census EB estimators. One way to avoid this problem is using area-level methods such as the FH model. Another attempt has been to consider the EB method based on the nested error model for the unit-level values of the welfare variable in terms of aggregates of the covariates, either at a subarea level such as clusters, or at the area level. Corral et al. (2021, b) showed a potential bias problem of this approach, but further research is probably needed on this important issue.

Finally, the parametric bootstrap procedure employed for MSE estimation of the Census EB estimators yields estimators of the MSE under the model, but the assumed models are seldom exactly correct in practice. Hence, the resulting MSE estimators do not account for the model uncertainty and hence might underestimate the error when the model does not hold exactly. Research in small area estimation is currently focusing on estimating instead the MSE under the sampling replication mechanism (design MSE), which accounts for model uncertainty, see Rao et al. (2018), Pfeffermann and Ben-Hur (2019) and Stefan and Hidiroglou (2021), who focus on linear area indicators. Molina and Strzalkowska-Kominiak (2020) proposed several bootstrap procedures for the estimation of the design MSE, based on the same idea of "borrowing strength" employed for the estimation of the small area indicators. These bootstrap procedures may be extended to the estimation of the design MSE of more general nonlinear indicators.

References

Banerjee AV, Deaton A, Lustig N, Rogoff K, Hsu E (2006) An evaluation of World Bank research, 1998–2005. Available at SSRN 2950327

Battese GE, Harter RM, Fuller WA (1988) An error-components model for prediction of county crop areas using survey and satellite data. J Am Stat Assoc 83:28–36

Bedi T, Coudouel A, Simler K (eds) (2007) More than a pretty picture: using poverty maps to design better policies and interventions. World Bank Publications

Bell W (1997) Models for county and state poverty estimates. Preprint, Statistical Research Division, U. S. Census Bureau

Bell WR, Huang ET (2006) Using the t-distribution to deal with outliers in small area estimation. In: Proceedings of statistics, Canada symposium on methodological issues in measuring population health, Statistics Canada, Ottawa, Canada

Bell WR, Basel WW, Maples JJ (2016) An overview of the U.S. Census Bureau's Small Area Income and Poverty Estimates Program. In: Pratesi M (ed) Analysis of poverty data by small area estimation. Wiley, New York

Benavent R, Morales D (2016) Multivariate Fay–Herriot models for small area estimation. Comput Stat Data Anal 94:372–390

Boonstra HJ (2012) hbsae: hierarchical Bayesian small area estimation. R package version 1.0

Boubeta M, Lombardía MJ, Morales D (2017) Poisson mixed models for studying the poverty in small areas. Comput Stat Data Anal 107:32–47

Breidenbach J (2018) Josae: unit-level and area-level small area estimation. R package version 0.3.0

Casas-Cordero Valencia C, Encina J, Lahiri P (2016) Poverty mapping for the Chilean comunas. In: Pratesi M (ed) Analysis of poverty data by small area estimation. Wiley, New York

Chambers R, Tzavidis N (2006) M-quantile models for small area estimation. Biometrika 93:255–268

Corral P, Seitz W, Azevedo JP, Nguyen MC (2018) FHSAE: stata module to fit an area level Fay–Herriot model

Corral P, Himelein K, McGee K, Molina I (2021a) A map of the poor or a poor map? World Bank Policy Research Working Paper 9620

Corral P, Molina I, Nguyen M (2021b) Pull your small area estimates up by the bootstraps. J Stat Comput Simul 91:3304–3357

Correa L, Molina I, Rao JNK (2012) Comparison of methods for estimation of poverty indicators in small areas. Unpublished report

Das S, Chambers R (2017) Robust mean-squared error estimation for poverty estimates based on the method of Elbers, Lanjouw and Lanjouw. J R Stat Soc Ser A 180(4):1137–1161

Das S, Haslett S (2019) A comparison of methods for poverty estimation in developing countries. Int Stat Rev 87(2):368–392

Datta GS, Fay RE, Ghosh M (1991) Hierarchical and empirical multivariate Bayes analysis in small area estimation. In: Proceedings of the seventh annual research conference, U.S. Census Bureau, pp 63–79
Datta GS, Ghosh M, Nangia N, Natarajan K (1996) Estimation of median income of four-person families: a Bayesian approach. In: Berry DA, Chaloner KM, Geweke JK (eds) Bayesian analysis in statistics and econometrics. Wiley, New York, pp 129–140
Diallo M, Rao J (2018) Small area estimation of complex parameters under unit-level models with skew-normal errors. Scand J Stat 45:1092–1116
Elbers C, Fujii T, Lanjouw P, Özler B, Yin W (2007) Poverty alleviation through geographic targeting: How much does disaggregation help? J Dev Econ 83(1):198–213
Elbers C, Van der Weide R (2014) Estimation of Normal mixtures in a nested error model with an application to small area estimation of poverty and inequality. World Bank Policy Research Working Paper 6962
Elbers C, Lanjouw JO, Lanjouw P (2002) Micro-level estimation of welfare. World Bank Policy Research Working Paper 2911
Elbers C, Lanjouw JO, Lanjouw P (2003) Micro-level estimation of poverty and inequality. Econometrica 71:355–364
Esteban MD, Morales D, Pérez A, Santamaría L (2012a) Two area-level time models for estimating small area poverty indicators. J Indian Soc Agric Stat 66(1):75–89
Esteban MD, Morales D, Pérez A, Santamaría L (2012b) Small area estimation of poverty proportions under area-level time models. Comput Stat Data Anal 56:2840–2855
Esteban Lefler MD, Morales González D, Pérez Martin A (2014) saery: small area estimation for Rao and Yu model. R package version 1.0
Farris J, Larochelle C, Alwang J, Norton GW, King C (2017) Poverty analysis using small area estimation: an application to conservation agriculture in Uganda. Agric Econ 48(6):671–681
Fay RE (1987) Application of multivariate regression to small domain estimation. In: Platek R, Rao JNK, Sarndal CE, Singh MP (eds) Small area statistics. Wiley, New York, pp 91–102
Fabrizi E, Trivisano E (2016) Small area estimation of the Gini concentration coefficient. Comput Stat Data Anal 99:223–234
Fauziah RA, Wulansari IY (2020) saeeb: small area estimation for count data. R package 0.1.0
Fay RE, Herriot RA (1979) Estimation of income for small places: an application of James-Stein procedures to census data. J Am Stat Assoc 74:269–77
Ferretti C, Molina I (2012) Fast EB method for estimating complex poverty indicators in large populations. J Indian Soc Agric Stat 66:105–120
Franco C, Bell WR (2013) Applying bivariate binomial/logit normal models to small area estimation. In: Proceedings of the American Statistical Association, section on survey research methods, pp 690–702. http://ww2.amstat.org/sections/srms/Proceedings/
Franco C, Bell WR (2015) Borrowing information over time in binomial/logit normal models for small area estimation. In: Statistics in transition (new series) and survey methodology, joint issue on small area estimation, vol 16, pp 563–584. http://stat.gov.pl/en/sit-en/issues-and-articles-sit/previous-issues/volume-16-number-4-december-2015/
Foster J, Greer J, Thorbecke E (1984) Accounting for dependent informative sampling in model-based finite population inference. Econometrica 52:761–766
Gershunskaya J (2010) Robust small area estimation using a mixture model. In: JSM proceedings of the survey research methods section of the Americal Statistical Association
Giusti C, Marchetti S, Pratesi M, Salvati N (2012) Robust small area estimation and oversampling in the estimation of poverty indicators. Surv Res Methods 6:155–163
Giusti C, Masserini L, Pratesi M (2017) Local comparisons of small area estimates of poverty: an application within the Tuscany region in Italy. Soc Indic Res 131(1):235–254
Ghosh M (2020) Small area estimation: its evolution in five decades. Stat Transit 21(4):40–44 (**with discussion**)
Ghosh M, Rao JNK (1994) Small area estimation: an appraisal. Stat Sci 9(1):55–76
González-Manteiga W, Lombardía MJ, Molina I, Morales D, Santamaría L (2008) Bootstrap mean squared error of a small-area eblup. J Stat Comput Simul 78(5):443–462
Graf M, Marín JM, Molina I (2019) A generalized mixed model for skewed distributions applied to small area estimation. TEST 28(2):565–597
Guadarrama M, Molina I, Rao JNK (2014) A comparison of small area estimation methods for poverty mapping. Stat Transit 17(1):41–66

Guadarrama M, Molina I, Rao JNK (2018) Small area estimation of general parameters under complex sampling designs. Comput Stat Data Anal 121:20–40
Guadarrama M, Morales D, Molina I (2020) Time stable small area estimates of general parameters under a unit-level model. Unpublished manuscript
Hájek J (1964) Asymptotic theory of rejective sampling with varying probabilities from a finite population. Ann Math Stat 35(4):1491–1523
Halbmeier C, Kreutzmann AK, Schmid T, Schroder C (2019) The Fay–Herriot command for estimating small-area indicators. Stand Genomic Sci 19(3):626–644
Haslett SJ (2016) Small area estimation using both survey and census unit record data. In: Pratesi M (ed) Analysis of poverty data by small area estimation. Wiley, New York, pp 325–348
Hobza T, Morales D (2016) Empirical best prediction under unit-level logit mixed models. J Off Stat 32(3):661–692
Hobza T, Morales D, Santamaría L (2018) Small area estimation of poverty proportions under unit-level temporal binomial-logit mixed models. TEST 27(2):270–294
Huang R, Hidiroglou M (2003) Design consistent estimators for a mixed linear model on survey data. Proc Surv Res Methods Sect Am Stat Assoc 2003:1897–1904
Isidro M, Haslett S, Jones G (2016) Extended structure preserving estimation (ESPREE) for updating small area estimates of poverty. Ann Appl Stat 10(1):451–476
Jedrzejczak A, Kubacki J (2013) Estimation of income inequality and the poverty rate in Poland, by region and family type. Stat Transit New Ser 14(3):359–378
Jiang J, Lahiri P (2006) Mixed model prediction and small area estimation. TEST 15:1–96
Kreutzmann A, Pannier S, Rojas-Perilla N, Schmid T, Templ M, Tzavidis N (2019) The R package emdi for estimating and mapping regionally disaggregated indicators. J Stat Softw 91(7):1–33
Marchetti S, Tzavidis N, Pratesi M (2012) Non-parametric bootstrap mean squared error estimation for M-quantile estimators of small area averages, quantiles and poverty indicators. Comput Stat Data Anal 56:2889–2902
Marchetti S, Giusti C, Pratesi M, Salvati N, Giannotti F, Pedreschi D, Rinzivillo S, Pappalardo L, Gabrielli L (2015) Small area model-based estimators using big data sources. J Off Stat 31(2):263–281. https://doi.org/10.1515/JOS-2015-0017
Marhuenda Y, Molina I, Morales D (2013) Small area estimation with spatio-temporal Fay-Herriot models. Comput Stat Data Anal 58:308–325
Marhuenda Y, Molina I, Morales D, Rao JNK (2017) Poverty mapping in small areas under a twofold nested error regression model. J R Stat Soc Ser A 180(4):1111–1136
Molina I (2019) Desagregación de datos en encuestas de hogares: metodologías de estimación en áreas pequeñas. Series of the Economic Commission for Latin America and the Caribbean (ECLAC) from United Nations, Estudios Estadísticos LC/TS.2018/ 82/Rev.1, CEPAL
Molina I (2020) Discussion on “Small area estimation: its evolution in five decades”, by M. Ghosh. Stat Transit 21(4):40–44
Molina I, Marhuenda Y (2015) sae: an R package for small area estimation. R Journal 7(1):81–98
Molina I, Martín N (2018) Empirical best prediction under a nested error model with log transformation. Ann Stat 46(5):1961–1993
Molina I, Morales D (2009) Small area estimation of poverty indicators. Boletín de Estadística e Investigación Operativa 25:218–225
Molina I, Rao JNK (2010) Small area estimation of poverty indicators. Can J Stat 38:369–385
Molina I, Strzalkowska-Kominiak E (2020) Estimation of proportions in small areas: application to the labour force using the Swiss Census Structural Survey. J R Stat Soc Ser A 183(1):281–310
Molina I, Nandram B, Rao JNK (2014) Small area estimation of general parameters with application to poverty indicators: a hierarchical bayes approach. Ann Appl Stat 8(2):852–885
Molina I, Rao JNK, Guadarrama M (2019) Small area estimation methods for poverty mapping: a selective review. Stat Appl 17(1):11–22
Molina Peralta I, García Portugués E (2020) Short guide for small-area estimation using household survey data: illustration to poverty mapping in Palestine with expenditure survey and census data. UN Economic and Social Commission for Western Asia (ESCWA), E/ESCWA/SD/2019/TP.4
Mubarak MR, Ubaidillah A (2020) saeME: small area estimation with measurement error. R package version 1.2.4
Neri L, Ballini F, Betti G (2005) Poverty and inequality in transition countries. Stat Transit 7:135–157

Nguyen MC, Corral P, Azevedo JP, Zhao Q (2018) Sae: a stata package for unit level small area estimation. World Bank Policy Research Working Paper 8630

Permatasari N, Ubaidillah A (2020) msae: multivariate Fay Herriot models for small area estimation. R package version 0.1.2

Pfeffermann D (2002) Small area estimation-new developments and directions. Int Stat Rev 70(1):125–143

Pfeffermann D (2013) New important developments in small area estimation. Stat Sci 28(1):40–68

Pfeffermann D, Ben-Hur D (2019) Estimation of randomisation mean square error in small area estimation. Int Stat Rev 87(S1):S31–S49

Pfeffermann D, Sverchkov M (2007) Small-area estimation under informative probability sampling of areas and within the selected areas. J Am Stat Assoc 102:1427–1439

Pratesi M, Salvati N (2016) Introduction on measuring poverty at local level using small area estimation methods. In: Pratesi M (ed) Analysis of poverty data by small area estimation. Wiley, New York, pp 1–18

Rao JNK, Molina I (2015) Small area estimation, 2nd edn. Wiley, Hoboken

Rao JNK, Molina I (2016) Empirical Bayes and hierarchical Bayes estimation of poverty measures for small areas. In: Pratesi M (ed) Analysis of poverty data by small area estimation. Wiley, New York, pp 315–324

Rao JNK, Yu M (1994) Small-area estimation by combining time-series and cross-sectional data. Can J Stat 22(4):511–528

Rao JNK, Rubin-Bleuer S, Estevao VM (2018) Measuring uncertainty associated with model-based small area estimators. Surv Methodol 44:151–166

Rojas-Perilla N, Pannier S, Schmid T (2020) Data-driven transformations in small area estimation. J R Stat Soc Ser A 183(1):121–148

Salvati N, Giusti C, Pratesi M (2014) The use of spatial information for the estimation of poverty indicators at the small area level. In: Betti G, Lemmi A (eds) Poverty and social exclusion. New methods of analysis. Routledge, London

Sinha SK, Rao JNK (2009) Robust small area estimation. Can J Stat 37(3):381–399

Shi C (2018) BayesSAE: Bayesian analysis of small area estimation. R package version 1.0-2

Schoch T (2014) rsae: robust small area estimation. R package version 0.1-5

Stefan M, Hidiroglou MA (2021) Estimation of design-based mean squared error of a small area mean model-based estimator under a nested error linear regression model. Can J Stat 49:1338–1363

Stukel D, Rao JNK (1999) On small-area estimation under two-fold nested error regression models. J Stat Plan Inference 78:131–147

Sverchkov M, Pfeffermann D (2018) Small area estimation under informative sampling and not missing at random non-response. J R Stat Soc Ser A 181(4):981–1008. https://doi.org/10.1111/rssa.12362

Tzavidis N, Marchetti S, Chambers R (2010) Robust prediction of small area means and distributions. Aust N Z J Stat 52:167–186

Van der Weide R (2014) Gls estimation and empirical bayes prediction for linear mixed models with heteroskedasticity and sampling weights: a background study for the povmap project. World Bank Policy Research Working Paper 7028

Warnhol S (2018) saeRobust: robust small area estimation. R package 0.2.0

Ybarra LMR, Lohr SL (2008) Small area estimation when auxiliary information is measured with error. Biometrika 95:919–931

You Y, Rao JNK (2002a) Small area estimation using unmatched sampling and linking models. Can J Stat 30:3–15

You Y, Rao JNK (2002b) A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. Can J Stat 30:431–439

Zhao Q (2006) User manual for povmap. World Bank. http://siteresources.worldbank.org/INTPGI/Resources/342674-1092157888460/Zhao_ManualPovMap.pdf