Pull your small area estimates up by the bootstraps

Paul Corral, Isabel Molina and Minh Nguyen

April 14, 2020

Abstract

After almost two decades of poverty maps produced by the World Bank and multiple advances in the literature, this paper presents a methodological update to the existing World Bank toolkit for small area estimation. The paper reviews the computational procedures of the current methods used by the World Bank: the traditional approach by Elbers, Lanjouw and Lanjouw (2003) and the Empirical Best/Bayes (EB) addition introduced by Van der Weide (2014), which extends the EB procedure of Molina and Rao (2010) by considering heteroscedasticity and includes the survey weights, and uses a different bootstrap approach, here referred to as clustered bootstrap, intended to be less dependent on distributional assumptions. Simulation experiments comparing these methods to the original EB approach of Molina and Rao (2010) provide empirical evidence of the shortcomings of the clustered bootstrap approach proposed by Van der Weide (2014), which, as currently implemented, yields biased point estimates. The main contributions of this paper are then two: 1) to adapt the original Monte Carlo simulation procedure of Molina and Rao (2010) for the approximation of the extended EB estimators that include heteroscedasticity and survey weights as in Van der Weide (2014); and 2) to adapt the parametric bootstrap approach for mean squared error (MSE) estimation considered by Molina and Rao (2010), and proposed originally by González-Manteiga et al. (2008), to these extended EB estimators. Simulation experiments illustrate that the revised Monte Carlo simulation method yields estimators that are considerably less biased and more efficient in terms of MSE than those obtained from the clustered bootstrap approach, and that the parametric bootstrap MSE estimators are in line with the true MSEs under realistic scenarios.

1 Introduction

It has been almost two decades since the publication of Elbers, Lanjouw and Lanjouw's (2003 - ELL henceforth) seminal paper on small area estimation (SAE). The methodology proposed by these authors has been the de facto methodology used by the World Bank to obtain small area estimates of poverty and inequality indicators and perhaps constitutes the most applied SAE method across the globe. The World Bank, in an effort to make the implementation of the method as simple as possible, 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 of great value to World Bank staff and to many statistical agencies, and its ease of use has allowed the ELL approach to be adopted worldwide. It has also permitted the World Bank to provide training and spread this knowledge to many statistical agencies.

After the work from Van der Weide (2014), several important modifications and additions were made to the PovMap software, and thus to the toolkit used by the World Bank. The first change was to add a new method for estimating the variances of the model errors, called Henderson's Method III (H3 - Henderson, 1953), which does not require normality because it is based on the method of moments. The second improvement involved adjusting the original Generalized Least Squares estimators of the regression parameters considered in Elbers, Lanjouw and Lanjouw's (2003) by including the survey weights following Huang and Hidiroglou (2003). Another important modification involved the addition of Empirical Best/Bayes (EB) prediction assuming normality. The modifications came about due to advances in the SAE literature (Molina and Rao, 2010 - MR) and criticism of the ELL methodology (Haslett et al., 2010 and MR, 2010). The final important change was to consider a different bootstrap approach for obtaining the EB estimators and their corresponding error measures.[1]

To make the application of the World Bank's method friendlier to more advanced practitioners and to those who seek more flexibility, a World Bank team produced a Stata version of PovMap (Nguyen et al., 2018) which is available under the name "sae". The development of the tool has made running multiple simulations and tests of the software much more straightforward. Since the creation of the "sae" Stata suite of commands, the World Bank has shifted towards training statistical agencies using this approach, which can facilitate replicability. Additionally, the code in the Stata package is open, so that curious practitioners can see in detail how methods have been implemented.

Small area estimation is a broad branch of statistics that focuses on improving estimates' precision when household surveys are not large enough to achieve a desired level of precision. Within small area estimation methods, model-based techniques "borrow strength" from a larger data set or auxiliary data across areas through models linking the areas (regression-type techniques), which yield indirect estimators (MR, 2010). Most of the model-based techniques fall into two groups, methods based on unit-level models and those based on area-level ones. The former are commonly used when data on units (e.g. households) are available; when only area-level data are available (e.g. area means), the latter are used, see Fay and Herriot (1979).

In the context of poverty, methods based on unit-level models typically rely on estimating the distribution of the household's welfare given a set of auxiliary variables or correlates. The model parameters are then used to simulate multiple welfare vectors from the fitted distribution for every single household in the census (Elbers et al., 2007). Using the simulated census vectors, it is possible to obtain poverty rates or any other welfare indicator, for every area (including the non-sampled ones). Perhaps the two most popular approaches for unit level small area estimation of poverty indicators are the traditional ELL

---

[1] The bootstrap is not discussed in Van der Weide's paper.

(2003) method used by the World Bank, and the EB approach by MR (2010).[2]

This paper focuses on unit-level models for small area estimation; in particular, on the traditional ELL (2003) approach, the latest additions to the World Bank toolkit by Van der Weide (2014), and the original EB approach introduced by MR (2010). It sheds light on the nuances of the traditional approach by ELL (2003) and the EB addition by Van der Weide (2014), and it updates these methods in line with the original EB approach by MR (2010). The paper goes in depth into the differences between the estimation of the noise in the ELL approach (as implemented in PovMap and the sae Stata command), and the estimation of the MSE according to MR (2010). Then, the paper proposes: 1) an adaptation of the Monte Carlo simulation procedure by MR (2010) for calculation of the extended EB estimators that incorporate heteroscedasticity and survey weights as proposed by Van der Weide (2014); 2) an adaptation of the parametric bootstrap method considered by MR (2010), which comes from González-Manteiga et al. (2008), for estimation of the corresponding MSE of the extended EB estimators. According to our simulation results, the adapted procedures for calculation of extended EB point estimates and corresponding MSEs represent a substantial improvement over the current approaches.

The paper reviews the current methods in chronological order. It begins describing the traditional ELL method in Section 2, then moves on to the EB method proposed by MR (2010) in Section 3 and finishes with the additions by Van der Weide (2014) in Section 4. It discusses these methods in Section 5. Then, in Section 6 it describes the proposed computational approaches for calculation of the extended EB estimators and for estimation of their MSEs. In Section 7, several simulation exercises, imitating those performed by MR (2010) but extending them to more realistic scenarios, are conducted to compare the different methods. Finally, conclusions are presented in Section 8.

2   Traditional ELL approach

The Elbers, Lanjouw and Lanjouw (2002, 2003) methodology is the de facto small area estimation approach utilized by the World Bank. The methodology has been widely applied across the globe to produce poverty maps conducted by the institution.[3] The popularity of the methodology can be attributed, to some degree, to the availability of the PovMap software (Zhao, 2006),[4] which was programmed in C and offers users a simple point and click interface. The PovMap software 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 allows for expansion of the methods available to users.[5]

The ELL method assumes that the welfare y_{ch} for each household h within each location c in the population is linearly related to a $1 \times K$ vector of characteristics (or correlates) x_{ch} for that household, according to the nested error model:

$$\ln(y_{ch}) = x_{ch}\beta + \eta_c + e_{ch},\ h = 1, \ldots, N_c,\ c = 1, \ldots, C, \tag{1}$$

where η_c and e_{ch} are respectively location and household-specific idiosyncratic errors, assumed to be

---

[2]A factor which likely contributes to these being well known is the availability of software, which can be easily used to apply these approaches; PovMap (Zhao, 2006) for ELL and R's sae package (Molina and Marhuenda, 2015)

[3]Poverty mapping is the common name within the World Bank for SAE methodology, where the obtained estimates are mapped for illustrative purposes

[4]downloadable from: http://iresearch.worldbank.org/PovMap/PovMap2/setup.zip

[5]Users should be aware that results from Stata and PovMap differ slightly due to the use of different random number generators

independent from each other, following

$$\eta_c \overset{iid}{\sim} N\left(0, \sigma_\eta^2\right), \ e_{ch} \overset{iid}{\sim} N\left(0, \sigma_e^2\right).$$

Here, C is the number of locations in which the population is divided and N_c is the number of households in location c, for $c = 1, \ldots, C$. Finally, β is the $K \times 1$ vector of regression coefficients.

Under the original ELL methodology, the locations indexed with c are supposed to be the clusters, or primary sampling units (PSUs), of the sampling design and do not necessarily correspond to the level at which the estimates need to be produced. In fact, clusters are typically nested within the areas of interest (e.g. census tracks). Presenting estimates at a higher aggregation level than the clusters (for which random effects are included in the model) may not be appropriate in cases of considerable between-area variability, and may underestimate the estimator's standard errors (Das and Chambers, 2017). A way to alleviate this is to include covariates that explain sufficiently well the between-area heterogeneity in the model (*ibid*). In this regard, ELL (2002) suggests the inclusion of area-level covariates as a way to improve precision, because they can explain the between-area variation in welfare and hence might reduce the contribution of area-specific residuals. Marhuenda et al. (2017) recommend to put the location effects at the same aggregation level where estimation is desired. According to this, in this manuscript we will consider that the clusters c are equal to the areas where estimation is desired.

A key feature of the implemented ELL approach is that it considers heteroscedasticity and fits a preliminary model for that, called alpha model. Very little research has gone into this aspect of the method. In most applications, the alpha model has usually a small adjusted R^2 (which may not reach 0.05) yet tends to play a considerable role in the obtained point estimates, especially for parameters beyond poverty rate such as Gini index, Theil index, Poverty Gap, etc. Heteroscedasticity is included in the description of the EB approach given in Rao and Molina (2015), but this is not implemented in the sae R package (Molina and Marhuenda, 2015). For simplicity, below we spell the steps of the traditional ELL approach only in the case of homoscedasticity.

The nested error model in (1) was originally fitted by Generalized Least Squares (GLS) accounting for heteroscedasticity, but Van der Weide (2014) updated this original GLS method to properly account for survey weights.[6] Currently, the implemented method fits the model by Feasible Generalized Least Squares (FGLS). In this procedure, a simple linear regression obtained by considering as model errors $u_{ch} = \eta_h + e_{ch}$ is fit using ordinary least squares (OLS), and afterwards the appropriate covariance matrix of regression parameter estimators is estimated.[7] The first stage of the process being an OLS is an important aspect, because most of the model testing and validation done for the ELL approach is in practice undertaken with the OLS fit. Nevertheless, the parameter estimates used for generating welfare in the census (see the steps below) come from the GLS fit accounting for heteroscedasticity and including the survey weights.

The variances of the model parameter estimators are required in the considered ELL bootstrap approach. To estimate these variances, ELL (2002) proposed to use the delta method. The actual implementation of the delta method is through a computationally intensive simulated numerical gradient. An estimate of the gradient vector is obtained by making perturbations of the parameters, and this is used to estimate the variances of the estimators via the delta method. ELL (2002) mentions also the possibility of drawing from the sampling distribution (parametric) as a way to incorporate the estimation error into the total prediction error. This later method has become the de facto approach used under the ELL methodology.[8]

---

[6]Haslett et al. (2010) presents the problems in the original GLS implementation of ELL.

[7]For a detailed look into the ELL approach, interested readers should refer to the original ELL papers (ELL, 2003; 2002) and Section 3 of Nguyen et al. (2018), which presents the current GLS estimator from Van der Weide (2014)

[8]In a comparison of the simulation methods proposed by ELL (2002), Demombynes et al. (2008) shows that the delta

The first implementation of the World Bank poverty mapping software was actually done in SAS (Demombynes, 2002). Poverty indicators and their corresponding standard errors were obtained by a bootstrap procedure relying on simulation. This simulation approach is very reminiscent of multiple imputation methods, where every single relevant parameter necessary for simulating vectors of welfare is drawn from its corresponding estimated asymptotic distribution (or an approximation to it). This approach is the one currently used in the World Bank's tools for implementing the ELL (2002) approach: Stata's sae package and PovMap. Specifically, the steps of the implemented bootstrap procedure designed to obtain the traditional ELL point estimates and their estimated noise are:

1. Fit the nested error model (1) to the survey data. This yields the vector of initial parameter estimates
$$\hat{\theta}_0 = (\hat{\beta}_0, \hat{\sigma}^2_{\eta 0}, \hat{\sigma}^2_{e0}),$$
where the 0 subscript is used hereafter to indicate that the estimates come from the original household survey. In the implemented version, the model is fit via FGLS as specified in Nguyen et al. (2018).

2. Draw new model parameters as follows. First, regression coefficients are drawn from
$$\beta^* \sim MVN\left(\hat{\beta}_0, \widehat{\text{vcov}}(\hat{\beta}_0)\right),$$
The variance of the household-level errors is drawn, according to Gelman et al. (2004, pp. 364-365), from
$$\sigma^{2*}_e \sim \hat{\sigma}^2_{e0} \frac{(n-K)}{\chi^{2*}_{n-K}},$$
where χ^{2*}_{n-K} denotes a random number from a chi-squared distribution with $n-K$ degrees of freedom. Here, n is the number of observations in the survey data used to fit the model and K is the number of correlates used in the model. The variance of the cluster effects σ^{2*}_η is drawn, according to Demombynes (2008) and Demombynes et al. (2002), from
$$\sigma^{2*}_\eta \sim Gamma\left(\hat{\sigma}^2_{\eta 0}, \widehat{\text{var}}(\hat{\sigma}^2_{\eta 0})\right).$$

3. Using the simulated model parameters in step 2, calculate the welfare for every household in the census y^*_{ch} from the model as
$$\ln(y^*_{ch}) = x_{ch}\beta^* + \eta^*_c + e^*_{ch},$$
where the household-specific errors are generated as
$$e^*_{ch} \overset{iid}{\sim} N\left(0, \sigma^{2*}_e\right),$$
and the location effects are generated as
$$\eta^*_c \overset{iid}{\sim} N\left(0, \sigma^{2*}_\eta\right).$$
Then construct the vector containing the N_c simulated welfares for the households in location c, denoted $y^*_c = (y^*_{c1}, \dots, y^*_{cN_c})^T$, where N_c is the number of census households in location c.

4. With the vectors y^*_c, $c = 1, \dots, C$, of simulated census welfares, indicators can be produced for all the locations. Let $\tau^*_c = f(y^*_c)$ be the indicator of interest calculated based on the simulated vector

method from ELL and the parametric drawing of the parameters provide similar results. In tests with pseudo surveys, the delta method seems to provide wider standard errors than the parametric approach (Demombynes et al. 2008), suggesting that perhaps the parametric estimates are too optimistic when compared to the delta method

for location c; for example, for the Foster, Greer and Thorbecke (1984 - FGT) class of decomposable poverty measures, the function $f(y_c^*)$ is defined as

$$f_\alpha(y_c^*) = \sum_{h=1}^{N_c} \frac{p_{ch}}{\sum_\ell p_{c\ell}} I\left(y_{ch}^* < z\right) \left(1 - \frac{y_{ch}^*}{z}\right)^\alpha, \quad \alpha \geq 0,$$

where z is the poverty line, p_{ch} is the size of household h from location c in the census and $I(y_{ch}^* < z) = 1$ if $y_{ch}^* < z$ and is equal to 0 otherwise.

5. Repeat steps 2 to 4 M times. Although traditionally $M = 100$, a larger number of replicates M is recommended. Let $\tau_c^{*(m)}$ be the indicator of interest obtained in m_{th} replicate of the bootstrap. The ELL estimator is then given by the average across the M bootstrap replicates,

$$\hat{\tau}_c^{ELL} = \frac{1}{M} \sum_{m=1}^M \tau_c^{*(m)}$$

and the ELL estimated variance of is given by

$$\mathrm{var}_{ELL}(\hat{\tau}_c^{ELL}) = \frac{1}{M-1} \sum_{m=1}^M \left(\tau_c^{*(m)} - \hat{\tau}_c^{ELL} - \right)^2.$$

As opposed to the EB method of MR (2010), which uses a Monte Carlo simulation procedure and a separate bootstrap procedure for MSE (see Section 3), in the above ELL procedure, a single computational algorithm tries to capture the noise of the initial model parameter estimates in the ELL standard error by varying the model parameters across simulations, which is aligned to Rubin's rules (Rubin, 2004). This means that, in each replicate of the bootstrap, different values of the model parameters β^*, σ_η^{2*} and σ_e^{2*} are used to generate the welfare data. However, this step might entail an increase of noise in the final ELL estimators.

Note that, if the model parameters were kept fixed to the initial survey estimates from step 1 (skipping step 2), the above ELL estimate $\hat{\tau}_c^{ELL}$ would be a basic Monte Carlo approximation to the marginal mean of the indicator, $E(\tau_c; \hat{\theta}_0)$, with respect to the distribution of the welfare (given the correlates) induced by the nested error model (1), with θ replaced by the initial estimate $\hat{\theta}_0$. Note that $E(\tau_c; \theta)$, evaluated at the true $\theta = (\beta, \sigma_\eta^2, \sigma_e^2)$, is unbiased by definition since it equals the expectation of the indicator under the true model.[9] According to this, the prediction error of ELL estimator can be decomposed as described in ELL (2003),

$$\hat{\tau}_c^{ELL} - \tau_c = \underbrace{\hat{\tau}_c^{ELL} - E(\tau_c; \hat{\theta}_0)}_{\text{computation error}} + \underbrace{E(\tau_c; \hat{\theta}_0) - E(\tau_c; \theta)}_{\text{model error}} + \underbrace{E(\tau_c; \theta) - \tau_c}_{\text{idiosyncratic error}},$$

where these sources of error are described as follows:

1. The **idiosyncratic error** is related to how the actual value of the expenditure for a given location, c, deviates from its expected value due to unobserved aspects in the expenditure for the location (ELL, 2002). For locations with smaller population sizes, the underlying distribution is not likely approximated when errors are drawn. This, however, is related to the explanatory power of the independent variables in the model, the smaller the unexplained portion of the model, the smaller the idiosyncratic error. With poorly fitting models, estimates for locations with smaller populations are likely to suffer from more variability across simulations. Therefore, in order to minimize this error, one of the goals of the modeling stage is to obtain the highest possible R^2.

---

[9]Note that under a model-based approach, the true value of the indicator τ_c is random. Under this setup, an estimator/predictor $\hat{\tau}_c$ of τ_c is said to be unbiased when $E(\hat{\tau}_c - \tau_c) = 0$ or, accordingly, when $E(\hat{\tau}_c) = E(\tau_c)$.

2. The **model error** is related to the properties of the model parameters and is unrelated to the size of the target population (ELL, 2002). The magnitude of the error is dependent on the precision of the β coefficients of the welfare model and the sensitivity of the indicator to deviations in welfare (ibid). Consequently, in order to minimize this source of error in the final estimates, it is recommended to remove all non-significant independent variables in the modeling stage.

3. The final source of error is the **computation error**, which is not related to the other two sources of error. This is related to the simulation and can be made as small as possible, as computational resources allow, by running a larger number of simulations.

Consider the case that model parameter estimates were kept fixed across simulations and equal to the initial survey estimates from step 1, $\hat{\theta}_0 = (\hat{\beta}_0, \hat{\sigma}_{\eta 0}^2, \hat{\sigma}_{e0}^2)$. In that case, consistency of these estimators as the total sample size n tends to infinity,[10] and of Monte Carlo averages to the actual expectations under the model as the number of bootstrap replicates M tends to infinity, would ensure that $\hat{\tau}_c^{ELL}$ approximates correctly the "theoretical" ELL estimator, $E(\tau_c; \theta)$, which is unbiased by definition. However, there is an additional source of error due to generation of welfare in each simulation from a different set of parameter values in step 3, instead of using the initial estimates from step 1.

Moreover, if our target indicator is the mean welfare in location c, $\tau_c = N_c^{-1} \sum_{h=1}^{N_d} y_{ch}$, and the welfare is not transformed with log, then $E(\tau_c; \theta) = E(\bar{x}_c \beta + \eta_c + \bar{e}_c) = \bar{x}_c \beta$, where $\bar{x}_c = N_c^{-1} \sum_{h=1}^{N_d} x_{ch}$ is the area mean of the correlates and $\bar{e}_c = N_c^{-1} \sum_{h=1}^{N_d} e_{ch}$ is the area mean of the household-level errors. This means that the area effect η_c vanishes in the final ELL estimator, which becomes basically the regression estimator $\bar{x}_c \beta$. This estimator is called "synthetic" in the SAE literature because it does not account for between-area heterogeneity (or idiosyncrasy of the areas) beyond that explained by the correlates. Note that, if the regression parameter β was actually known (best case), $\bar{x}_c \beta$ would not be using at all the actual survey data, only the census auxiliary information. Synthetic estimators rely very strongly on the assumed regression model and might be misleading if the model is incorrectly specified. In fact, standard errors are obtained assuming that the model actually holds and do not account for model misspecification. Hence, under model failure, they might still be small, giving a lot of credibility to the (misleading) estimates. As will be shown, even if the model is correctly specified, the MSE of the ELL estimators can be substantial, especially if the prediction power of the model is weak.

3 Molina and Rao's EB and Census EB estimators

Molina and Rao (2010) considers the nested error model (1) as in ELL procedure, but the location effects are originally defined for the areas of interest. A more recent paper by Guadarrama, Molina and Rao (2018) further extends the EB method of MR (2010) to complex sampling designs, but so far it has not been implemented in any software package as a library or command, unlike the traditional ELL (2002) with the subsequent updates by Van der Weide (2014). For simplicity, this section describes the original EB approach proposed by MR (2010).

The main difference with ELL approach is that the EB method of MR (2010) conditions on the survey sample data and thus makes a more efficient use of the survey data, which contains the only available (and hence precious) information on the actual welfare. Nevertheless, conditioning on the survey data requires areas and households across survey and census to be matched. Thus, if P_c denotes the area's population of size N_c, then the survey sample, s_c, is a sample of size $n_c \leq N_c$ drawn from P_c. The complement to the sampled population is referred to as r_c, which is just $P_c - s_c$. Thus, the census

---

[10]Note that the total survey sample size n is typically large

welfare vector for any area c is defined as $y_c = (y_{c,s}^T, y_{c,r}^T)^T$, which contains the welfare for sampled and non-sampled observations in area c, denoted $y_{c,s}$ and $y_{c,r}$ respectively. Typically, the non-sampled population (of size $N_c - n_c$) is much larger than the sampled population (of size n_c). It may also happen that $n_c = 0$ for some areas, meaning that the area is not sampled in the survey. The empirical best (EB) predictor of τ_c for an area c that is sampled ($n_c > 0$) is defined as the conditional expectation $E(\tau_c | y_{c,s}; \hat{\theta})$. For an out-of-sample area ($n_c = 0$), the empirical best (EB) predictor of τ_c is $E(\tau_c; \hat{\theta})$, which is similar to ELL estimator. Here, $\hat{\theta}$ is a consistent estimator of θ as the overall sample size n tends to infinity. In contrast with this procedure, the traditional ELL approach does not require to link the census and survey observations.

When the expectation $E(\tau_c | y_{c,s}; \hat{\theta})$ has no explicit form, the empirical best (EB) predictor of τ_c can be approximated using a Monte Carlo (MC) simulation method. The MC approximation to the EB predictor $\hat{\tau}_c^{EB} = E(\tau_c | y_{c,s}; \hat{\theta})$, as described in Rao and Molina (2015), is obtained as follows:

1. Using the survey data, fit model (1) via any method providing consistent estimators. This yields the vector of parameter estimates:
$$\hat{\theta}_0 = \left(\hat{\beta}_0, \hat{\sigma}_{\eta 0}^2, \hat{\sigma}_{e0}^2 \right).$$

   Usual fitting methods under this approach are maximum likelihood (ML) or restricted maximum likelihood (REML), both based on the normal likelihood, and H3 method, which does not require to specify a distribution. In the implemented sae R package, REML is used.

2. Use the parameter estimates obtained in step 1 as true values to simulate a vector of welfare in the census. First, the welfare from each out-of-sample household is generated as follows,

$$\ln(y_{ch,r}^*) = x_{ch,r} \hat{\beta}_0 + \eta_c^* + e_{ch,r}^*,$$

   where here we add the subscript r to indicate that the household is not sampled. Here, for an area c that is included in the sample, the area effect η_c^* is generated as

$$\eta_c^* \sim N \left(\hat{\eta}_{c0}, \hat{\sigma}_{\eta 0}^2 (1 - \hat{\gamma}_c) \right),$$

   where $\hat{\eta}_{c0}$ and $\hat{\gamma}_c$ are respectively given by

$$\hat{\eta}_{c0} = \hat{\gamma}_c \left(\bar{y}_{c,s} - \bar{x}_{c,s} \hat{\beta}_0 \right), \quad \hat{\gamma}_c = \frac{\hat{\sigma}_{\eta 0}^2}{\hat{\sigma}_{\eta 0}^2 + \hat{\sigma}_{e0}^2 / n_c}.$$

   Here, $\bar{y}_{c,s}$ and $\bar{x}_{c,s}$ are respectively the sample means of the welfare variable and the correlates in area c. If area c is not sampled, the area effect is generated as $\eta_c^* \sim N \left(0, \hat{\sigma}_{\eta 0}^2 \right)$. Finally, the household error $e_{ch,r}^*$ is generated as:

$$e_{ch,r}^* \sim N \left(0, \hat{\sigma}_{e0}^2 \right).$$

   For an area c that is sampled in the survey, the generated non-sample vector $y_{c,r}^*$ is then augmented by the survey data $y_{c,s}$. Therefore, the final vector for the whole census in area c is $y_c^* = (y_{c,s}^T, y_{c,r}^{*T})^T$, which is made up of the generated out-of-sample welfares and the survey ones.

3. The previous step yields a census vector of welfare y_c^*, simulated using the fitted model parameters. This census vector is then used to calculate the indicator for each area $c = 1, \ldots, C$, as

$$\tau_c^* = f \left(y_c^* \right),$$

   where $f(\cdot)$ can be any indicator function such as the FGT poverty indicators.

4. Repeat steps 1 to 3 a large number of times M. If $y_c^{*(m)}$ denotes the m_{th} replicate of the census vector $y_c^* = (y_{c,s}^T, y_{c,r}^{*T})^T$, then $\tau_c^{*(m)} = f(y_c^{*(m)})$ is the corresponding indicator, $m = 1, \ldots, M$. The final MC approximation to the EB estimator of τ_c is just the average across de M simulations of these indicators

$$\hat{\tau}_c^{EB} = \frac{1}{M} \sum_{m=1}^{M} \tau_c^{*(m)}.$$

Molina (2019) comments that the effect of adding the survey data is negligible when the sample is small relative to the census population of the area. Thus, a slight variation of the EB estimator, called Census EB (see e.g. Molina 2019), avoids appending the survey data in step 2 by generating all the census welfares (including the sample ones), similarly as the non-sampled ones are generated. Thus, the welfare for each household h in the census is generated from the model

$$\ln(y_{ch}^*) = x_{ch}\hat{\beta}_0 + \eta_c^* + e_{ch}^*,$$

with η_c^* and e_{ch}^* obtained exactly the same as in step 2. The result is the full census vector for area c given by $y_c^* = (y_{c1}^*, \ldots, y_{c,N_c}^*)^T$, which is used to compute the target indicator $\tau_c^* = f(y_c^*)$. Unfortunately, the Census EB estimator is not implemented in the sae R package. One possible approach for applying this software when survey units cannot be identified in the census would be to append the survey data to the available census data, which would lead to a different approximation of the EB estimator. A drawback of this approach is that, in this case, the total size of the area is then $N_c + n_c \geq N_c$.

The EB estimator $\hat{\tau}_c^{EB}$ approximates the so called "best" predictor of τ_c, which is the optimal predictor in the sense of minimizing the MSE under the model. This MSE of an estimator $\hat{\tau}_c$ is defined as the expectation of the squared prediction error under the considered model, $\text{MSE}(\hat{\tau}_c) = E(\hat{\tau}_c - \tau_c)^2$. The best predictor is given by the conditional expectation $\hat{\tau}_c^B = E(\tau_c|y_{c,s}; \theta)$, where the expectation is with respect to the conditional distribution of the non-sampled data $y_{c,r}$ given the sample data $y_{c,s}$, determined by the nested error model. Under normality of the random terms in the model, the conditional distribution is also normal. In step 2 of the above procedure, welfare values for non-sampled households are generated from that conditional distribution, but with estimated model parameters. Hence, the above MC simulation procedure actually approximates the EB estimator $\hat{\tau}_c^{EB} = E(\tau_c|y_{c,s}; \hat{\theta})$, which is the best predictor with model parameters replaced by the corresponding estimates based on the original sample survey data $\hat{\theta}_0 = (\hat{\beta}_0, \hat{\sigma}_{\eta0}^2, \hat{\sigma}_{e0}^2)$. Note that, unlike the ELL bootstrap procedure, here these estimates are kept constant across simulations because there is a unique DGP. Hence, this simulation procedure delivers a simple MC approximation for the EB predictor $\hat{\tau}_c^{EB} = E(\tau_c|y_{c,s}; \hat{\theta})$ of τ_c, where the expectation is replaced by the corresponding MC average.

If the target parameter is just the area mean $\tau_c = N_c^{-1} \sum_{h=1}^{N_d} y_{ch}$, and there is no log-transformation of welfare, then the best predictor reduces to $E(\tau_c|y_{c,s}; \theta) = \bar{x}_c\beta + \gamma_c(\bar{y}_{c,s} - \bar{x}_{c,s}\beta)$, which corrects the regression synthetic estimator $\bar{x}_c\beta$ by accounting for the area effect using the survey data. Moreover, the size of the correction depends on γ_c, which measures the share of between-area heterogeneity σ_η^2 from the total variance in the area, $\sigma_\eta^2 + \sigma_e^2/n_c$. Thus, the correction is stronger in the case of highly heterogeneous areas and it is weak otherwise, becoming null only when all the existing area heterogeneity is completely explained by the available auxiliary variables. In that case, the EB estimator reduces to the "theoretical" ELL estimator $E(\tau_c; \theta) = \bar{x}_c\beta$. Hence, the EB 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 assumed linear regression (which could be misspecified).

Concerning error measures of the estimators, ELL and EB approaches consider different error measures. In the EB approach of MR (2010), the noise is measured by the actual MSE of the EB estimator

under the nested error model, $\text{MSE}(\hat{\tau}_c^{EB}) = E(\hat{\tau}_c^{EB} - \tau_c)^2$. Under the model-based framework, the target indicators τ_c are regarded as random quantities (they are generated from the model) and hence incorporate uncertainty as well. Even if an estimator (or predictor) $\hat{\tau}_c$ is unbiased with respect to the model in the sense that the expectation of its prediction error is zero $E(\hat{\tau}_c - \tau_c) = 0$, it would still leave the variance of the prediction error $\text{var}(\hat{\tau}_c - \tau_c)$, which is not exactly equal to the variance of the predictor $\text{var}(\hat{\tau}_c)$ because of the randomness of the target indicator τ_c.

The MSE of the EB estimator is estimated using a parametric bootstrap procedure for finite populations introduced by González-Manteiga et al. (2008), which is computationally more intensive than the ELL bootstrap approach. This is because, in every bootstrap replicate, all the steps for obtaining the EB estimator are reproduced, including model fitting and the MC procedure for approximation of the conditional expectation defining the EB estimator. This means that the number of simulated census vectors is the product of the MC simulations and the number of bootstrap replicates, except when the target indicator has an explicit EB estimator (such as the poverty rate). In this latter case, the MC simulation procedure can be replaced by an explicit formula calculated without simulation, and then the number of simulated censuses will be only the number of bootstrap replicates.

The steps for the parametric bootstrap estimation of the MSE are as follows (MR, 2015):

1. Using the survey data, fit model (1) using a method providing consistent estimators. This yields the set of parameter estimates from the observed sample:

$$\hat{\theta}_0 = \left(\hat{\beta}_0, \hat{\sigma}_{\eta 0}^2, \hat{\sigma}_{e0}^2 \right).$$

   In the implemented sae R package, REML method is used.

2. Use the estimates in $\hat{\theta}_0$ to simulate census welfares[11] from the fitted model as follows:

$$\ln(y_{ch}^*) = x_{ch}\hat{\beta}_0 + \eta_c^* + e_{ch}^*,$$

   where the area effects η_c^* are generated as

$$\eta_c^* \sim N \left(0, \hat{\sigma}_{\eta 0}^2 \right)$$

   and the household-specific errors are generated as

$$e_{ch}^* \sim N \left(0, \hat{\sigma}_{e0}^2 \right).$$

3. From the simulated census vector y_c^*, we calculate the true value of the indicator of interest for each area c:

$$\tau_c^* = f \left(y_c^* \right),$$

   where $f(y_c^*)$ can be any indicator function such as the FGT indicators.

4. Since the survey is regarded as a subset of the census, the survey sample s_c is now extracted and, using the corresponding sample welfares $y_{c,s}^*$ (newly generated in step 2), one fits the model (1). This yields bootstrap estimates of the model parameters,

$$\hat{\theta}^* = \left(\hat{\beta}^*, \hat{\sigma}_\eta^{2*}, \hat{\sigma}_e^{2*} \right).$$

5. With the bootstrap vectors of sample welfare $y_{c,s}^*$, $c = 1, \ldots, C$, obtain the EB estimators $\hat{\tau}_c^{EB*}$, using MC simulation if needed. The estimation procedure is exactly the same EB procedure based on the original sample, but using the bootstrap sample data $y_{c,s}^*$, $c = 1, \ldots, C$ and the corresponding bootstrap model parameter estimates $\hat{\theta}^*$ from step 4.

---

[11]Note that here welfares are generated for all households in the census, sampled and non-sampled

6. Repeat steps 2 to 5 a sufficiently large number of times B. In each bootstrap replicate b, $\tau_c^{*(b)}$ is the true value of the indicator obtained from the b_{th} simulated census and $\hat{\tau}_c^{EB*(b)}$ is the corresponding EB estimator obtained from the extracted sample. The parametric bootstrap approximation of the MSE is then given by:

$$mse_B\left(\hat{\tau}_c^{EB}\right) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\tau}_c^{EB*(b)} - \tau_c^{*(b)}\right)^2.$$

4 Van der Weide's update to PovMap software

The addition of EB prediction developed by Van der Weide (2014) represented a landmark update to the PovMap project of the World Bank and its poverty mapping agenda. This section will highlight that this extension is not exactly the same as the EB estimator by MR (2010). In fact, it resembles the traditional ELL approach in the manner which point estimates and standard errors are computed according to the implementation in PovMap as well as in Stata's sae package.

EB prediction is implemented by generating censuses using the predicted location effects instead of generating them from their theoretical distribution. Let $e_{c,s}$ be the vector with the marginal residuals $e_{ch} = \ln(y_{ch}) - x_{ch}\beta$ for the households in the survey. The predictor of the location effect η_c proposed by Van der Weide (2014) is an extension of the best predictor $E\left[\eta_c | e_{c,s}\right]$ for the heteroscedastic nested error model with normality, incorporating also the survey weights to account for complex sampling designs. This predictor is given by

$$\hat{\eta}_c = \hat{\gamma}_c \left(\sum_h \frac{w_{ch}}{\hat{\sigma}_{ech}^2}\right)^{-1} \sum_h \frac{w_{ch}}{\hat{\sigma}_{ech}^2}\hat{e}_{ch} \tag{2}$$

where w_{ch} is the survey weight for household h in location c, $\hat{\sigma}_{ech}^2$ is the estimated household-specific error variance using the alpha model as in ELL (2002),

$$\hat{\gamma}_c = \frac{\hat{\sigma}_\eta^2}{\hat{\sigma}_\eta^2 + \sum_h w_{ch}^2 \left(\sum_h w_{ch}\sum_h \frac{w_{ch}}{\hat{\sigma}_{ech}^2}\right)^{-1}}$$

and $\hat{\sigma}_\eta^2$ is an estimator of σ_η^2. For this, Van der Weide (2014) considers the extended version of Henderson's method III (Henderson, 1953) obtained accounting also for heteroscedasticity and survey weights as proposed in Huang and Hidroglou (2003). Finally, $\hat{e}_{ch} = \ln(y_{ch}) - x_{ch}\hat{\beta}$ is the estimated marginal residual from the GLS estimator $\hat{\beta}$ that accounts for heteroscedasticity and survey weights, given by

$$\hat{\beta} = \left\{\sum_c \left(\sum_h \frac{w_{ch}}{\hat{\sigma}_{ech}^2}x_{ch}x_{ch}^T - \gamma_c \sum_h \frac{w_{ch}}{\hat{\sigma}_{ech}^2}\bar{x}_{c,w}\bar{x}_{c,w}^T\right)\right\}^{-1}\sum_c\left(\sum_h \frac{w_{ch}}{\hat{\sigma}_{ech}^2}x_{ch}y_{ch} - \gamma_c \sum_h \frac{w_{ch}}{\hat{\sigma}_{ech}^2}\bar{x}_{c,w}\bar{y}_{c,w}\right),$$

where all the sums are in the survey data, and $\bar{y}_{c,w}$ and $\bar{x}_{c,w}$ are the weighted sample means,

$$\bar{y}_{c,w} = \left(\sum_h \frac{w_{ch}}{\hat{\sigma}_{ech}^2}\right)^{-1}\sum_h \frac{w_{ch}}{\hat{\sigma}_{ech}^2}y_{ch}, \quad \bar{x}_{c,w} = \left(\sum_h \frac{w_{ch}}{\hat{\sigma}_{ech}^2}\right)^{-1}\sum_h \frac{w_{ch}}{\hat{\sigma}_{ech}^2}x_{ch}.$$

Van der Weide (2014) also proposed the following estimator of the variance of $\hat{\eta}_c$:

$$\widehat{\text{var}}\left[\hat{\eta}_c\right] = \hat{\sigma}_\eta^2 - \hat{\gamma}_c^2\left\{\hat{\sigma}_\eta^2 + \sum_h \left(\frac{w_{ch}}{\hat{\sigma}_{ech}^2}\right)^2\hat{\sigma}_{ech}^2\right\}. \tag{3}$$

Under homoscedasticity, the predicted location effects $\hat{\eta}_c$ are equal to those of the Pseudo EB in Guadarrama et al. (2018). They reduce to the estimated area effects $\hat{\eta}_c$ obtained from MR (2010) if additionally no survey weights are considered ($w_{ch} = 1$ for all h and c) as noted in Van der Weide (2014).

As mentioned before, for sampled clusters, EB prediction makes use of the estimated cluster effects from the survey data to improve the point estimates of the poverty indicators and their errors. A crucial assumption under EB prediction is that model errors e_{ch} and η_c are normally distributed. This yields normality for η_c given $e_{c,s}$, which leads to the proposed predicted cluster effect $\hat{\eta}_c$.

This EB addition implemented in the PovMap software uses an extended version of Henderson's method III (Henderson, 1953) to obtain estimators $\hat{\sigma}_\eta^2$ and $\hat{\sigma}_e^2$ of the variance components σ_η^2 and σ_e^2 that incorporate the sampling weights.[12] In step 2 of section 2 for the ELL, σ_η^{2*} is drawn from a *Gamma* distribution, however to avoid this assumption and because var$[\hat{\eta}_c]$ is unkonwn,[13] it is necessary to rely on bootstrap re-sampling to obtain the noise of the point estimates in similar fashion to ELL.[14]

Even if the estimator is based on the EB method of MR (2010), the bootstrap procedure used to generate the census data is completely different. This new bootstrap algorithm tries to avoid the distributional assumptions by generating in the first step samples of clusters from the survey data set. The full procedure for obtaining the estimator of τ_c, called here clustered bootstrap EB (CB-EB), is as follows:

1. Take a bootstrap sample of clusters (PSUs) with replacement from the survey data.[15]

2. Fit model (1) to the bootstrapped survey data from step 1. This yields the set of parameter estimates from the bootstrap sampled data $(\hat{\beta}^*, \hat{\sigma}_\eta^{2*}, \hat{\sigma}_e^{2*})$, along with predicted effects for the clusters that were sampled in step 1, $\hat{\eta}_c^*$, and their corresponding estimated variance $\widehat{\text{var}}_*[\hat{\eta}_c^*]$. In the implemented version, the model is fit by FGLS as specified in Nguyen et al. (2018).

3. Use the model parameter estimates from step 2 to simulate welfares y_{ch}^* in the census as[16]

$$\ln(y_{ch}^*) = x_{ch}\hat{\beta}^* + \eta_c^* + e_{ch}^*,$$

where, for a cluster c that is sampled in step 1, the cluster effect is generated as

$$\eta_c^* \sim N\left(\hat{\eta}_c^*, \widehat{\text{var}}_*\left[\hat{\eta}_c^*\right]\right),$$

where $\hat{\eta}_c^*$ and $\widehat{\text{var}}_*\left[\hat{\eta}_c^*\right]$ are the bootstrap versions of $\hat{\eta}_c$ and its corresponding variance, given in (2) and (3) respectively. For a cluster that is not sampled in step 1, generate

$$\eta_c^* \sim N\left(0, \hat{\sigma}_\eta^{2*}\right).$$

Under homoscedasticity, the household specific errors are generated as

$$e_{ch}^* \sim N\left(0, \hat{\sigma}_e^{2*}\right),$$

and, in the case of heteroscedasticity, as

$$e_{ch}^* \sim N\left(0, \hat{\sigma}_{ech}^{2*}\right).$$

Construct the vector $y_c^* = (y_{c1}, \ldots, y_{cN_c})^T$ of simulated welfare for every household within location c of size N_c, for all the locations $c = 1, \ldots, C$ in the census.

---

[12] Interested readers should refer to Van der Weide (2014) and/or Nguyen et al. (2018) for an in-depth look at how these are obtained under ELL and Henderson's method III.

[13] In the traditional ELL method of Section 2, $\hat{\sigma}_\eta^2$ is assumed to follow a Gamma distribution, which does not hold in this case.

[14] Actually, Van der Weide (2014) does not offer a method for the estimation of the standard errors, but the approach described below was the one implemented on PovMap and consequently also in the Stata sae package from Nguyen et al. (2018).

[15] Note that not all clusters are expected to appear

[16] Note that, in each bootstrap replicate, the census is generated from a different model.

4. With the simulated welfare vector y_c^*, calculate the indicator of interest as $\tau_c^* = f(y_c^*)$ such as the FGT indicator of order $\alpha \geq 0$, which is calculated as follows:

$$f_\alpha(y_c^*) = \sum_{h=1}^{N_c} \frac{p_{ch}}{\sum_\ell p_{c\ell}} I\left(y_{ch}^* < z\right) \left(1 - \frac{y_{ch}^*}{z}\right)^\alpha.$$

5. Repeat steps 1 to 3 M times. Even if traditionally $M = 100$, a larger number of replicates is recommended. Let $\tau_c^{*(m)}$ be the indicator obtained in m_{th} replicate, $m = 1, \ldots, M$. The CB-EB estimator $\hat{\tau}_c^{CB-EB}$ is then obtained as the average across the M bootstrap replicates,

$$\hat{\tau}_c^{CB-EB} = \frac{1}{M} \sum_{m=1}^{M} \tau_c^{*(m)}$$

and the variance of the estimator is approximated as

$$\text{var}_{CB-EB}(\hat{\tau}_c^{CB-EB}) = \frac{1}{M-1} \sum_{m=1}^{M} \left(\tau_c^{*(m)} - \hat{\tau}_c^{CB-EB}\right)^2.$$

Unlike the traditional ELL approach, this procedure requires to match the location codes in the survey and the census. However, in the application of the traditional ELL approach, the authors recommended the inclusion of contextual variables at the location level and thus the linking between the two data sources was still necessary.

As in the previous computational procedures, the final estimator of the poverty indicator for a location is obtained as an average across bootstrap replicates. By the Monte Carlo principle, this average is approximating an expectation. However, here it is not clear with respect to which DGP is the expectation taken, since the bootstrap procedure involves generation of several measures: In step 1, the sample of clusters varies, and this would define an expectation with respect to the sampling design, but considering only the first stage of the design. In step 2, model parameters are newly generated in each simulation. Finally, in step 3, location effects are generated from their conditional distribution given the sample residuals (determined by the nested error model under normality), and household-specific errors are generated from their distribution under the assumed nested error model with normality.

Even if the bootstrap procedure is not the same, standard errors are obtained similarly as in the traditional ELL approach, and are thus different from MSEs. Note that the traditional ELL approach was initially based on the multiple imputation literature. According to Rubin (1996, and noted in StataCorp, 2019), the objective of multiple imputation is not to re-create the missing data as closely as possible, but to handle it in a manner that allows for statistical inference.

5   Comparison of CB-EB, traditional ELL and EB methods

Because EB prediction uses the survey data to estimate the location effect, its benefits are evident only in locations present both in the census and the survey. In the traditional ELL approach, the location effect was originally at the cluster level and, in most surveys from developing countries where the World Bank focuses its efforts, the sampled clusters represent a small percentage of all the clusters in the country. This may also be the case for locations above clusters. For example, in a recently completed exercise in Moldova, the survey only contained 129 comunas out of a total of 901. Consequently, only a small share of these comunas benefit from EB prediction (Corral and Cojocaru, 2019). Nevertheless, differences also arise due to the computational procedures used to obtain the point estimates and their corresponding measures of noise.

MR (2010) presented evidence that their EB approach is superior to ELL in terms of a smaller MSE. This finding has met some criticism, mostly because the simulation experiment in MR (2010) was conducted under scenarios where ELL is seldom applied. First, the population size in MR (2010) was 20,000 uniformly spread across 80 areas. Second, all the areas were sampled and the sample within each area represented 20% of the population, something hardly seen in real world applications. ELL (2002) illustrates how the noise of the estimator falls as the size of the population increases, and advises against estimating below populations of 100 households. In practice, the level of disaggregation of ELL estimates depends on the model quality, and even with an R^2 not smaller than 0.6, it is seldom advised to go below 1,000 households (Elbers et al., 2007).

Another point of departure is that the model considered in the simulation experiments of MR (2010) was really poor, yielding an adjusted R^2 of less than 0.01. Under the usual applications of the ELL approach, the number of auxiliary variables is much greater, reaching most of the times an explanatory power, as measured by adjusted R^2, of over 0.5. Moreover, to improve precision, ELL (2002; 2003) also advocated for the use of contextual area-level variables[17] on the right-hand side, since these explain (and thus minimize) the variation across locations. These were not used in the simulation experiment conducted in MR (2010). In such a setup, the much noisier ELL estimates obtained in their simulation experiments are not surprising.

A final important difference is the fact that the error measures of estimators and the bootstrap procedures for obtaining them are considerably different in the three procedures. Currently, the error measures of the traditional ELL and of the more recent CB-EB procedure of Van der Weide (2014) are obtained under a very similar framework to that of multiple imputation. As an error measure, these consider the standard deviation of all the simulated point estimates of the indicators in the M bootstrap replicates, and so every simulated indicator is compared to the average of the simulated indicators. Moreover, in the bootstrap approaches used for that average of the simulated indicators, the initial model parameter estimates $\hat{\beta}_0$, $\hat{\sigma}_{\eta 0}^2$ and $\hat{\sigma}_{e 0}^2$ obtained using the original survey sample are not those used to generate the welfare vectors in the population. In fact, in each bootstrap replicate, model parameters $(\beta, \sigma_\eta^2, \sigma_e^2)$ are simulated from the estimated distribution of the initial estimators (or an approximation to it). In contrast, in MR's (2010) parametric bootstrap procedure, a simple MC simulation procedure is used to approximate the actual MSE under the nested error model with the average across bootstrap simulations of the squared prediction errors. In this bootstrap procedure, the parameter estimates used to generate the bootstrap censuses of welfare are kept fixed to the original survey estimates. Consistency of these initial survey estimators to their corresponding true values as the total sample size n tends to infinity ensures that the bootstrap MSE approximates the true MSE (under the model with the true values of the parameters).

6   Extended Census EB estimators

The EB estimator in MR (2010) requires to link the survey and census households, which may not be possible in real applications. In fact, the survey sample is likely not a subset of the available census. Hence, here we consider the Census EB (CEB) estimator, which avoids this step. In fact, in most cases, the number of sample households for a given area is much smaller than the number of census households and, in such cases, the CEB estimator is expected to perform very much like the original EB.

CEB small area estimators are obtained with a similar MC simulation approach as the one used by MR (2010) and described in Section 3, but simulating the vectors of welfare for all the census households

---

[17]These may come from the census or administrative records (ELL, 2003 p356).

(instead of the non-sampled ones only) and hence not appending the welfare for the survey units. The CEB is extended by accounting for the sampling design and heteroscedasticity, similarly as proposed by Van der Weide 2014. This procedure is implemented by also incorporating household expansion factors (household sizes) taken from the census. The proposed MC simulation procedure for the approximation of the extended CEB estimator is:

1. Fit model (1) to the survey data. This yields the set of parameter estimates:

$$\hat{\theta}_0 = \left(\hat{\beta}_0, \hat{\sigma}^2_{\eta 0}, \hat{\sigma}^2_{e0} \right).$$

   The currently implemented procedure fits the model either via FGLS and decomposing residuals as in the traditional ELL procedure, or using H3 method.

2. Use the model parameter estimates obtained in step 1 to simulate a vector of welfare for the N_c census households in area c, $y^*_c = (y^*_{c1}, \ldots, y^*_{cN_c})^T$, where each y^*_{c1} is obtained as

$$\ln(y^*_{ch}) = x_{ch} \hat{\beta}_0 + \eta^*_c + e^*_{ch},$$

   where, if area c is in the sample, its location effect is generated as

$$\eta^*_c \sim N \left(\hat{\eta}_{c0}, \widehat{\mathrm{var}} \left[\hat{\eta}_{c0} \right] \right),$$

   with $\hat{\eta}_{c0}$ given by

$$\hat{\eta}_{c0} = \hat{\gamma}_c \left(\sum_h \frac{w_{ch}}{\hat{\sigma}^2_{ech0}} \right)^{-1} \sum_h \frac{w_{ch}}{\hat{\sigma}^2_{ech0}} \hat{e}_{ch0},$$

   for $\hat{e}_{ch0} = \ln(y^*_{ch}) - x_{ch} \hat{\beta}_0$,

$$\hat{\gamma}_{c0} = \frac{\hat{\sigma}^2_{\eta 0}}{\hat{\sigma}^2_{\eta 0} + \sum_h w^2_{ch} \left(\sum_h w_{ch} \sum_h \frac{w_{ch}}{\hat{\sigma}^2_{ech0}} \right)^{-1}}.$$

   and

$$\widehat{\mathrm{var}} \left[\hat{\eta}_{c0} \right] = \hat{\sigma}^2_{\eta 0} - \hat{\gamma}^2_{c0} \left\{ \hat{\sigma}^2_{\eta 0} + \sum_h \left(\frac{w_{ch}}{\hat{\sigma}^2_{ech0}} \right)^2 \hat{\sigma}^2_{ech0} \right\}.$$

   If area c is not included in the sample, then its effect is generated as $\eta^*_c \sim N \left(0, \hat{\sigma}^2_{\eta 0} \right)$. In absence of heteroscedasticity, the household-specific residuals come from

$$e^*_{ch} \sim N \left(0, \hat{\sigma}^2_{e0} \right)$$

   and, in the case of heteroscedasticity, from

$$e^*_{ch} \sim N \left(0, \hat{\sigma}^2_{ech0} \right).$$

3. The previous step yields a simulated vector of census welfare for each area, $y^*_c = (y^*_{c1}, \ldots, y^*_{cN_c})^T$, which makes use of the fitted model parameters. With the census y^*_c, the indicator in area c is calculated as

$$\tau^*_c = f(y^*_c),$$

   where, for the FGT indicator of order $\alpha \geq 0$, we have

$$f_\alpha(y^*_c) = \sum_{h=1}^{N_c} \frac{p_{ch}}{\sum_\ell p_{c\ell}} I \left(y^*_{ch} < z \right) \left(1 - \frac{y^*_{ch}}{z} \right)^\alpha.$$

4. Repeat steps 1 to 3 a large number of times M. Let $\tau_c^{*(m)}$ be indicator obtained in m_{th} replicate, for $m = 1, \ldots, M$. The extended Census EB estimate is just the average of the M indicators,

$$\hat{\tau}_c^{CEB} = \frac{1}{M} \sum_{m=1}^{M} \tau_c^{*(m)}.$$

In contrast with the traditional ELL and CB-EB approaches, in the above MC procedure, the model parameter estimates $\hat{\beta}_0$, $\hat{\sigma}_{\eta 0}^2$ and $\hat{\sigma}_{e0}^2$, as well as $\hat{\eta}_{c0}$ and $\widehat{\text{var}}\,[\hat{\eta}_{c0}]$, are kept fixed across all the procedure; that is, there is a single DGP.

As a measure of noise of the extended CEB estimators, here the actual MSE under the nested error model is considered. To estimate this MSE, we propose to apply a parametric bootstrap procedure similar to the one in MR (2010), but adapted to the case where the survey is not necessarily a subset of the census. This procedure is also described in Molina (2019) for the CEB predictor, but here we apply it to the extended CEB predictor that includes heteroscedasticity and sampling weights:

1. Fit model (1) to the survey data. This yields the set of parameter estimates:

$$\hat{\theta}_0 = \left(\hat{\beta}_0, \hat{\sigma}_{\eta 0}^2, \hat{\sigma}_{e0}^2 \right).$$

   In the implemented version, the model is fit either via FGLS and decomposing residuals as in the traditional ELL fitting approach, or by H3.

2. Using the estimates obtained in step 1 as true values of the model parameters, simulate a vector of census welfare $y_c^* = (y_{c1}, \ldots, y_{cN_c})^T$ as follows:

$$\ln(y_{ch}^*) = x_{ch}\hat{\beta}_0 + \eta_c^* + e_{ch}^*, \ h = 1, \ldots, N_c,$$

   where the location effect is generated as

$$\eta_c^* \sim N\left(0, \hat{\sigma}_{\eta 0}^2\right).$$

   In the homoscedastic case, household-specific errors are generated as

$$e_{ch}^* \sim N\left(0, \hat{\sigma}_{e0}^2\right)$$

   and, in the case of heteroscedasticity, as

$$e_{ch}^* \sim N\left(0, \hat{\sigma}_{ech0}^2\right).$$

3. With the simulated census of welfare $y_c^* = (y_{c1}, \ldots, y_{cN_c})^T$ for each area $c = 1, \ldots, C$, produce indicators of interest:

$$\tau_c^* = f\left(y_c^*\right).$$

   where $f\left(y_c^*\right)$ can be any indicator function, such as one of the FGT poverty indicators.

4. Use the model parameter estimates obtained in step 1 to obtain new sample welfares, $y_{ch,s}^*$, for every location c as follows:
$$\ln(y_{ch,s}^*) = x_{ch,s}\hat{\beta}_0 + \eta_c^* + e_{ch,s}^*.$$

   (a) The estimates $\hat{\beta}_0$ come from step 1.

   (b) η_c^* is the same one generated in step 2. Specifically, all locations present in the survey are matched to the census locations and the same value of η_c^* that was simulated for that location in step 2 is applied to the survey households within the same location.

(c) The household-specific errors are simulated as

$$e^*_{ch,s} \sim N\left(0, \hat{\sigma}^2_{e0}\right)$$

in the homoscedastic case. When there is heteroscedasticity, they are simulated as

$$e^*_{ch,s} \sim N\left(0, \hat{\sigma}^2_{ech0}\right).$$

Unlike the original parametric bootstrap procedure in MR (2010) described in Section 3, here the sample errors $e^*_{ch,s}$ are not a subset of the census errors simulated in step 2, although they are generated from exactly the same model (with the same variance).

5. Fit the model (1) to the newly simulated survey data from step 4. This yields a bootstrap vector of estimated model parameters:

$$\hat{\theta}^* = \left(\hat{\beta}^*, \hat{\sigma}^{2*}_\eta, \hat{\sigma}^{2*}_e\right).$$

6. Obtain the bootstrap extended CEB estimator $\hat{\tau}^{CEB*}_{c,b}$ through MC simulation using the bootstrap parameter estimates $\hat{\theta}^*$ from step 5, by the approach described in the previous section.

7. Repeat steps 2 to 6 a sufficiently large number of times, B. Let $\tau^{*(b)}_c$ be the true value and $\hat{\tau}^{CEB*(b)}_{c,b}$ be the CEB estimator obtained in the b$_{th}$ replicate of the bootstrap procedure. A parametric bootstrap estimator of the MSE is given by:

$$mse_B\left(\hat{\tau}^{CEB}_c\right) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\tau}^{CEB*(b)}_c - \tau^{*(b)}_c\right)^2.$$

7 Simulation experiments for comparison of methods

7.1 Simulation experiment with poor model

Here we compare the procedures described in the previous sections for the estimation of poverty rates and gaps (FGT indicators with $\alpha = 0, 1$ respectively) and their corresponding error measures. To this aim, we perform simulation experiments under the same setup as in MR (2010), but extended to consider also more realistic scenarios.

Thus, like in MR (2010), a census data set of $N = 20,000$ observations is created, where all observations are uniformly spread among $C = 80$ areas, labeled from 1 to 80. This means that every area consists of $N_c = 250$ observations. Location effects are simulated as $\eta_c \overset{iid}{\sim} N\left(0, 0.15^2\right)$; note that every observation within a given area will have the same simulated effect. Then, values of two right-hand side binary variables are simulated. The first one, x_1, takes the value 1 if a generated random uniform value between 0 and 1 is less than or equal to $0.3 + 0.5\frac{c}{80}$. This means that observations in areas with a higher label are more likely to get value 1. The next one, x_2, is not tied to the area's label. This variable takes the value 1 if a simulated random uniform value between 0 and 1 is less than or equal to 0.2. The census welfare vectors $y_c = (y_{c,1}, \ldots, y_{c,N_c})^T$ for each area $c = 1, \ldots, C$, are then created as follows:

$$\ln(y_{ch}) = 3 + 0.03x_{1,ch} - 0.04x_{2,ch} + \eta_c + e_{ch},$$

where household-level errors are generated under the homoscedastic setup, as $e_{ch} \overset{iid}{\sim} N\left(0, 0.5^2\right)$. The poverty line is fixed at $z = 12$, which is roughly 60 percent of the median welfare of a generated population.

From the created "census" in each of the areas, 20% of the observations are sampled using simple random sampling without replacement;[18] this yields our "survey" data. Thus, in this experiment, survey weights are all equal. This generation process is repeated $L = 10,000$ times. In each simulation, the following quantities are computed for the poverty rates and gaps in each area:

1. True poverty indicators τ_c, using the "census".

2. Direct estimators $\hat{\tau}_c^{DIR}$ using the "survey", defined as the sample versions of τ_c.

3. Original EB estimators $\hat{\tau}_c^{EB}$ by MR (2010) with REML fitting as implemented in the sae R package, with $M = 50$ MC replicates.

4. Traditional ELL estimators $\hat{\tau}_c^{ELL}$ and their estimated standard errors, without location means and also including the location means of the considered auxiliary variables, where $M = 50$.

5. Census EB (CEB) estimators $\hat{\tau}_c^{CEB}$, obtained as in the MC procedure of MR (2010), but generating full census vectors from the conditional distribution and without appending the survey welfares, using REML fitting and with $M = 50$.

6. Extended CEB estimators using the MC and parametric bootstrap methods described in Section 6, using either H3 or ELL fitting methods with $M = 50$. Since in this simulation experiment model errors are homoscedastic and sampling weights are constant, here the extended CEB estimators only differ in the above CEB ones in the model fitting method.

7. Estimators based on the CB-EB approach from Section 4, $\hat{\tau}_c^{CB-EB}$ using either H3 or ELL fitting methods.[19] Since clusters are originally supposed to be nested within the areas and the clustered bootstrap method is supposed to sample clusters (PSUs) instead of areas in step 1, we computed additional CB-EB estimators, where every area is divided into 25 PSUs (with 10 households each) and selecting 5 PSUs within every area (instead of areas). Under this scenario $M = 50$ as well.

Model bias and MSE are approximated empirically as in MR (2010), as the averages across the $L = 10,000$ simulations of the prediction errors $\hat{\tau}_c^j - \tau_c$ and of the squared prediction errors respectively, where j stands for one of the methods: DIR, ELL, CEB or CB-EB.[20]

Before getting into the results of the full simulation experiment, the methods are initially compared using just a single generated population along with its corresponding sample. For this one population, the different methods are run by setting $M = 1000$ to get point estimates and $B = 1000$ for the MSE.[21] For expediency, results that do not conform to the main message of the paper are relegated to the appendix. However, from the figures in the appendix, we can draw the following conclusions: 1) The differences in the EB and Census EB point estimates are negligible even if the area sampling fractions $f_c = n_c/N_c$ are not so small in this experiment (20%), see Figure 17. Although differences appear in their estimated MSEs (see Figure 18), these differences become closer to zero as the sampling fractions f_c decrease. 2) Census EB estimates using H3 method as implemented in the sae Stata package are practically identical to the Census EB ones with REML fitting (Figure 19). Similarly, when comparing the Census EB estimators obtained with the proposed MC procedure using either H3 or ELL estimation methods, we can see that they are also aligned (Figure 20). This indicates that the actual method used

---

[18]The same units are sampled in every simulation and the values of x_1 and x_2 for all the census units are also kept fixed; this means that the x_1 and x_2 values for the sample units are always the same across simulations.

[19]For a detailed description of the difference between the methods readers should refer to Nguyen et al. (2018)

[20]$E\left(\hat{\tau}_c^j - \tau_c\right)$ for the bias, and $E\left(\hat{\tau}_c^j - \tau_c\right)^2$ for the MSE.

[21]For ELL from section 2 and H3 CBEB from section 4 since the process relies on a single computational algorithm $M = 1000$

to fit the model is not an important factor. In fact, in real applications, the overall survey sample is typically very large so any fitting method, as long as it provides consistent estimators, is expected to deliver estimates that will be very close to the corresponding true values.

The above conclusions support the use of the Census EB, using either REML, H3 or ELL estimation methods, instead of the original EB procedure in real world applications, where it is seldom possible to link the survey and census data. Moreover, since the large sampling fractions of this experiment favor the original EB method, in the following we show only results for the Census EB instead of the original EB, where welfares are generated in the MC procedure for all the census units and the survey welfares are not appended.

Figure 1 compares, for the poverty rates, Census EB estimators obtained from the proposed MC procedure of Section 6 using H3 estimation method (labeled "H3-CensusEB") with the analogues obtained using the CB-EB procedure of Section 4 (labeled "H3-CBEB (old bootstrap)"). We include also the results from the CB-EB method with selection of PSUs within the areas instead of selection of areas in step 1 (labeled "H3-CBEB-old with PSU"). Note that the CB-EB estimates presented in Section 4 differ considerably from the true poverty rates whereas the Census EB ones track approximately the true values. In fact, the sum of absolute differences of the CB-EB estimates to the true values is almost three times that of the Census EB estimates.

The main reason for this difference of behavior, even if both estimates are supposed to implement EB prediction is that, under the CB-EB approach in Section 4 with areas equal to clusters, a sample of areas is drawn in each simulation. This leads to a given area benefiting from EB prediction of the area effect only in a portion of the simulations, since it is likely that an area is not selected in every single bootstrap sample. This leads to CB-EB point estimates that are far off from the Census EB estimates obtained with the proposed MC procedure of Section 6. Another reason might be that the parameters used to simulate the census vectors in CB-EB approach are not kept fixed across simulations, making difficult the convergence to the conditional expectation defining the estimator. When using the CB-EB approach with PSUs, there is still a clear difference with the Census EB estimates obtained with the new bootstrap procedure of Section 6.

Turning now to the traditional ELL approach, Figure 2 shows ELL estimates for the poverty rates, based on the model with only x_1 and x_2 (labeled "ELL") and the same, but including the area means of x_1 and x_2 in the model (labeled "ELL with context"), together with true poverty rates and Census EB estimates obtained using the original MC procedure of MR (2010), without appending the survey data. In this plot, one finds that the traditional ELL estimates face the same troubles as the CB-EB ones. This is in line with what MR (2010) find in their simulations. Given the present simulation set-up with a poor model fit, ELL performs much worse than MR's Census EB method in terms of point estimates. Also note the rather flat nature of ELL, which is mostly due to the model providing very little information and, given the limited explanatory power of the correlates, the ELL estimates fall close to the average poverty rate of 16 for all the areas.[22] What is somewhat concerning is that the addition of contextual variables to the model for ELL, in the form of the area means, does little to improve the point estimates. Thus, even if the survey data are not preserved in the Census EB estimation method, the benefit of EB is quite considerable and caution should be taken when doing out of sample predictions in case of little explanatory power being brought in by the auxiliary variables.

The complete simulation experiment with $L = 10,000$ simulated populations are now discussed. Results in this case focus on the poverty gap, although the observed tendencies are the same for the poverty rates. Figures 3 and 4 show respectively the empirical bias and empirical MSE across simulated populations

---

[22]The ELL estimates are not really flat if sorted from smallest to largest

Figure 1: True poverty rates, proposed Census EB estimates of Section 6 using H3 method (labeled "H3-CensusEB"), CB-EB analogues from Section 4 (labeled "H3-CBEB (old bootstrap)") and CB-EB ones with selection of PSUs within the areas (labeled "H3-CBEB-old with PSU") in one simulated population and sample.
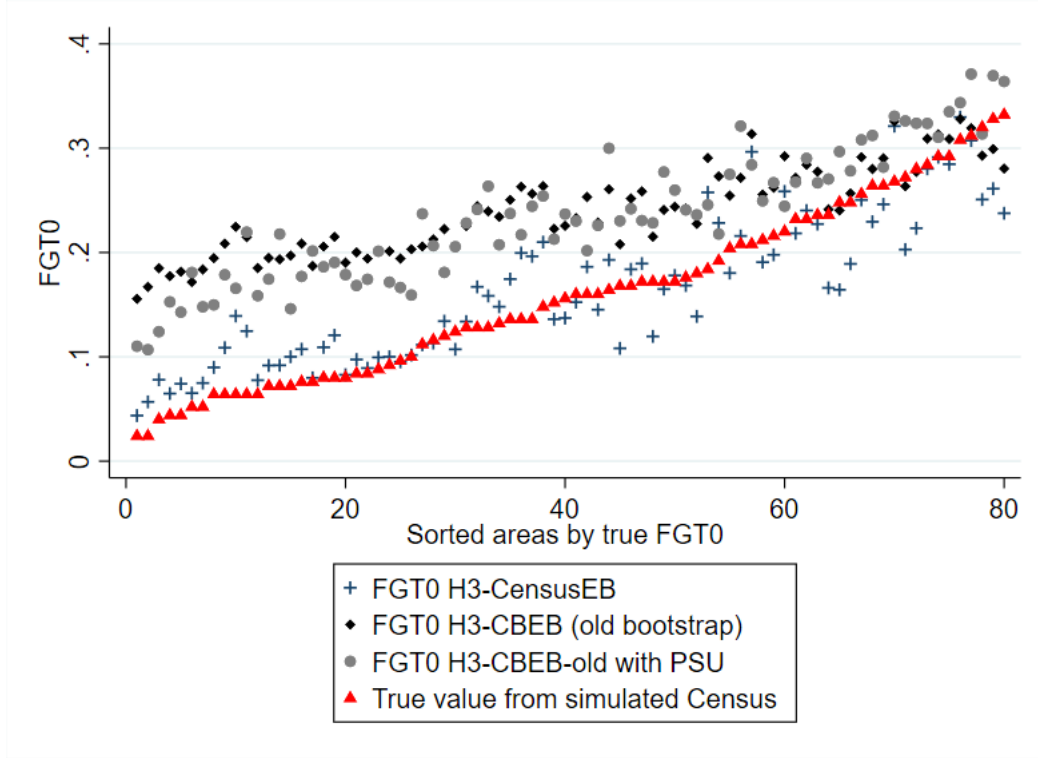

Figure 2: True poverty rates, ELL estimates (labeled "ELL"), ELL estimates including the area mean of x_1 in the model (labeled "ELL with context") and Census EB estimates obtained from MR (2010) without appending the survey data, in one simulated population and sample.
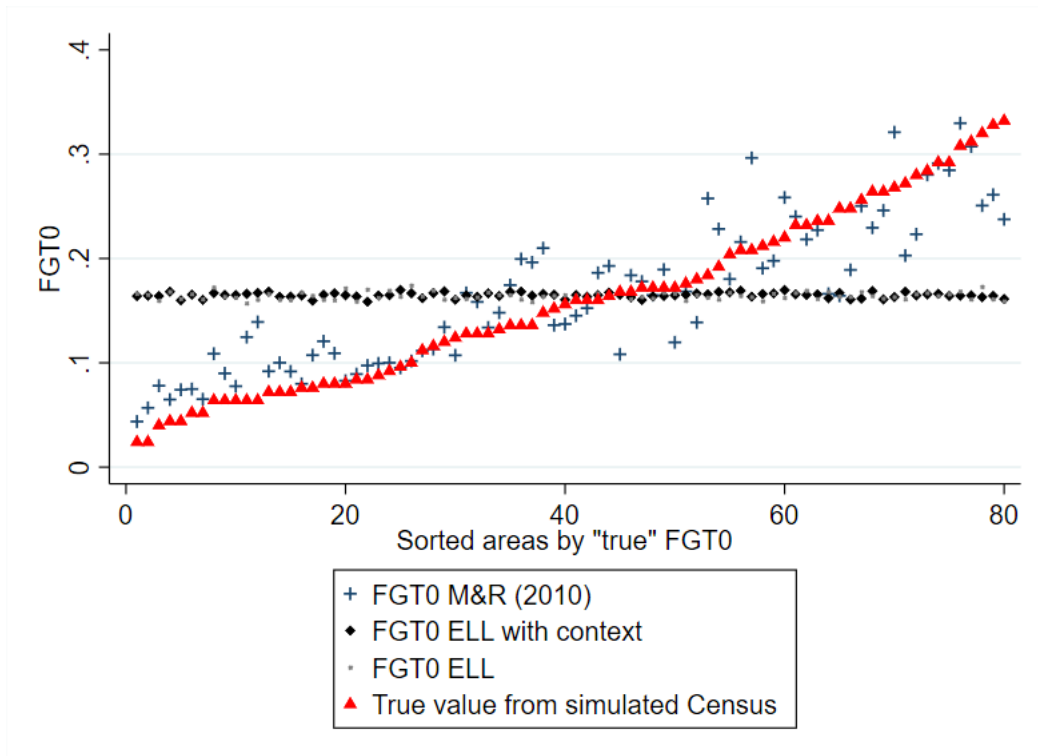
Figure 3: Empirical bias across simulated populations of Census EB from Section 6 using H3 method (labeled "H3-CensusEB"), traditional ELL (labeled "ELL"), CB-EB with H3 method (labeled "H3-CBEB (old bootstrap)"), CB-EB with PSUs (labeled "H3-CBEB (old bootstrap with PSU)") and Direct estimators (labeled "Direct") of the poverty gaps.
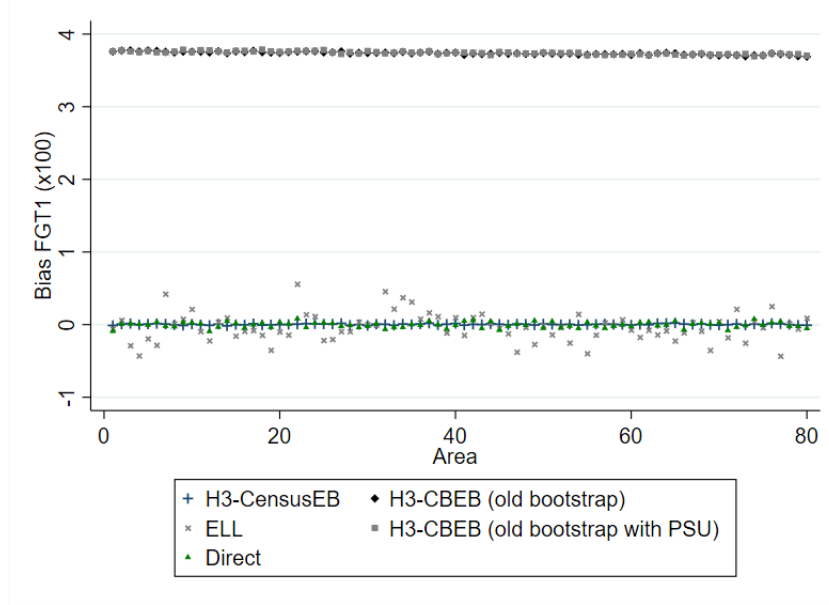
Figure 4: Empirical MSE across simulated populations of Census EB from Section 6 using H3 method (labeled "H3-CensusEB"), traditional ELL (labeled "ELL"), CB-EB with H3 method (labeled "H3-CBEB (old bootstrap)"), CB-EB with PSUs (labeled "H3-CBEB (old bootstrap with PSU)") and Direct estimators (labeled "Direct") of the poverty gaps.
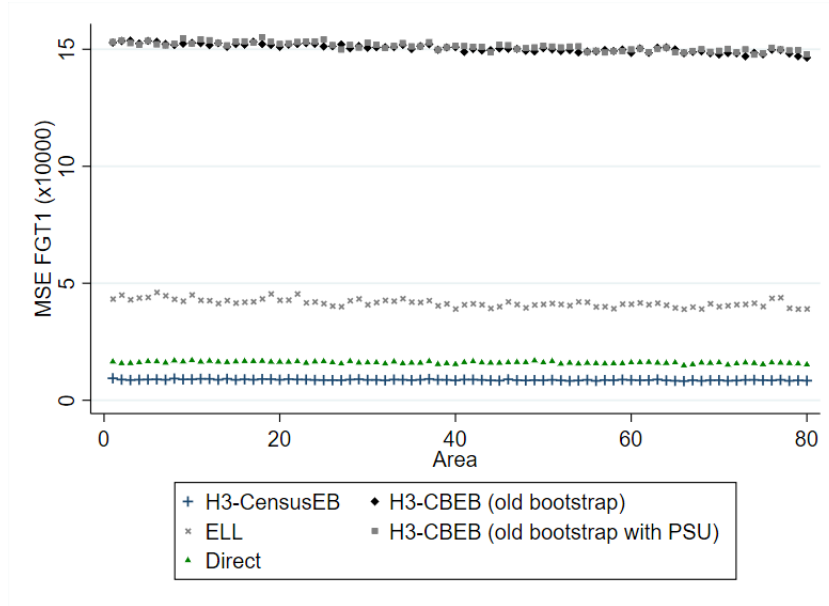
of Census EB from Section 6 using H3 estimation method (labeled "H3-CensusEB"), traditional ELL estimators (labeled "ELL"), estimators based on the CB-EB approach with H3 method (labeled "H3-CBEB (old bootstrap)"), the same with PSUs within areas selected in step 1 (labeled "H3-CBEB (old bootstrap with PSU)") and direct estimators (labeled "Direct"). These full results are quite sobering. While the bias for the traditional ELL estimators is not remarkable, the bias of the estimators obtained by the CB-EB method from Section 4 is substantial, showing upward bias for all areas. In contrast, the bias of the Census-EB estimators obtained from Section 6 seems negligible, ranging between -0.025 and 0.027, similar to the bias range of the original EB estimators from MR (2010).[23] This figure shows a considerable bias reduction of the proposed MC procedure for obtaining the Census EB estimators with respect to the CB-EB method.

In terms of MSE, results are aligned to those in MR (2010), see Figure 4. The traditional ELL yields an empirical MSE above the one of the direct estimator, which uses no model. Even if ELL estimators do not show a large bias when averaging across all the simulated populations (the ELL bootstrap procedure is supposed to approximate the marginal mean $E(\tau_c; \theta)$, which is exactly unbiased), as shown in Figure 2, it does not capture the area effect of the true poverty rates for a single population. Note that, under the model-based setup, welfares are generated from the model and in each simulation, new location effects are generated. This means that the true poverty rates vary for one population to another. In fact, a given area might have a positive location effect in one generated population and a negative effect in another. Since the location effects have an expected value of zero, the ELL method is tracking the average zero effect, which is the same for all the areas, but it does not capture the area effects in a given population. What is even more concerning is that the CB-EB method applied in Nguyen et al. (2018) and PovMap described in Section 4 yields an MSE that is more than 3 times greater that of the traditional ELL. This happens because a large amount of the MSE is the squared bias, which has been shown to be substantial for the CB-EB estimator. The proposed MC method for Census EB estimation presented here represents a massive improvement over the traditional ELL and the CB-EB approaches under the setup of this simulation experiment based on a poor model. The next section describes results when increasing the prediction power of the model.

7.2 Simulation experiment with improved model

This experiment addresses one of the concerns of the previous simulation experiment: the poor explanatory power of the model. Hence, in this section a more informed model is considered, yielding an adjusted R^2 of about 0.42, which is much closer to that of real world applications of ELL. The model includes six covariates. The first two, x_1 and x_2, are generated exactly as in the previous section. Four additional covariates are included, with values generated as follows:

1. x_3 is a binary variable, taking value 1 when a random uniform number between 0 and 1 is less than or equal to $0.1 + 0.2\frac{c}{80}$.

2. x_4 is a binary variable, taking value 1 when a random uniform value between 0 and 1 is less than or equal to $0.5 + 0.3\frac{c}{80}$.

3. x_5 is a discrete variable, simulated as the rounded integer value of the maximum value between 1 and a random Poisson variable with mean $\lambda = 3\left(1 - 0.1\frac{c}{80}\right)$.

4. x_6 is a binary variable, taking value 1 when a random uniform value between 0 and 1 is less than or equal to 0.4. Note that the values of x_6 are not related to the area's label.

---

[23]Results for the Census EB estimators obtained from the original procedure of MR's (2010) with REML estimation are not shown since they are aligned to the Census EB estimators from Section 6

The welfare vector for each area is created from the model with these covariates as follows

$$\ln y_{ch} = 3 + 0.09x_{1,ch} - 0.04x_{2,ch} - 0.09x_{3,ch} + 0.4x_{4,ch} - 0.25x_{5,ch} + 0.1x_{6,ch} + \eta_c + e_{ch},$$

where $e_{ch} \overset{iid}{\sim} N\left(0, 0.5^2\right)$ and $\eta_c \overset{iid}{\sim} N\left(0, 0.15^2\right)$. The poverty line in this scenario is fixed at $z = 10.2$. All other steps are the same as in the previous simulation.[24]

Even if the performance of CB-EB estimators improves when considering a better model, these are still clearly worse than the Census EB estimators obtained with the proposed MC approach of Section 6, see Figure 5. This conclusion remains true even after decreasing the location effect, by simulating the populations with $\eta_c \overset{iid}{\sim} N\left(0, 0.07^2\right)$.[25]

Similarly, despite the improved explanatory power of the model, Figure 6 still portrays a rather flat ELL hovering around the national average poverty rate. Under a single simulated population from the improved model, the traditional ELL method still performs poorly when compared with the Census EB according to MR (2010).

Results from the full simulation with $L = 10,000$ populations for the poverty gap also resemble those from the previous section, see Figures 7 and 8 showing empirical bias and empirical MSE respectively. The CB-EB method implemented as described in Nguyen et al. (2018), which replicates the methods of PovMap, still presents considerable bias in this new simulation exercise. In comparison, the proposed Census EB procedure seems to be nearly unbiased, although the obtained empirical biases actually range between -0.04 and 0.05. Moreover, even under a better regression model, the MSE of the proposed Census EB estimator is still much smaller. What is concerning is that the MSEs of the direct estimators are smaller than those of the traditional ELL estimators and of the CB-EB estimators.

Another criticism of the EB approach of MR (2010), which extends to the Census EB, is its dependence on normality. In order to study the sensitivity of the procedures to the normality assumption, an additional simulation experiment is run under the same setup of this section, but generating the household-specific errors from a Student's t distribution with 5 degrees of freedom and scaled by 0.5. This distribution is symmetric but has heavier tails than the normal distribution, which may represent the existence of outlying welfares in the data, often encountered in real applications.

Empirical bias and MSE across the $L = 10,000$ simulations are shown in Figures 9 and 10 respectively, for the proposed Census EB, traditional ELL and direct estimators. Note that the ELL method takes household-specific residuals from their empirical distribution, so its performance is not supposed to rely so strongly on normality. However, results suggest that, even under lack of normality, ELL's performance is still worse than that of the Census EB update. In terms of bias, some areas under the traditional ELL show quite considerable bias, yet on average across areas the bias can be near zero. This suggests that, even if the bias across all areas of ELL can be small, it is not an accurate estimator in each area.[26] In fact, the empirical MSEs of the traditional ELL estimators for many areas are still much greater than those of the proposed Census EB ones in this experiment.

Even after the better performance of the Census EB procedure compared with the traditional ELL method under normality departure in the form of higher tails, model checking and diagnostics are extremely important, as in any model-based procedure. In fact, good performance of EB and Census EB estimators in terms of bias and efficiency also relies on normality. Note that the sample size of household surveys is typically very large and therefore normality tests are all expected to reject the simple

---

[24] Another slight modification made in separate simulations is that the location effect is simulated as $\eta_c \overset{iid}{\sim} N\left(0, 0.07^2\right)$.

[25] Results available upon request. Although it must be noted that under this scenario under many populations $\hat{\sigma}_\eta < 0$, and thus many of the populations need to be re-simulated

[26] Note that the purpose in SAE is not obtaining estimators performing well on average across areas since, in that case, the overall sample mean at the population level would be sufficient.

Figure 5: True poverty rates, proposed Census EB estimates of Section 6 using H3 method (labeled "H3-CensusEB"), CB-EB analogues from Section 4 (labeled "H3-CBEB (old bootstrap)") and CB-EB ones with selection of PSUs within the areas (labeled "H3-CBEB-old with PSU") in one simulated population and sample, with improved model fit.
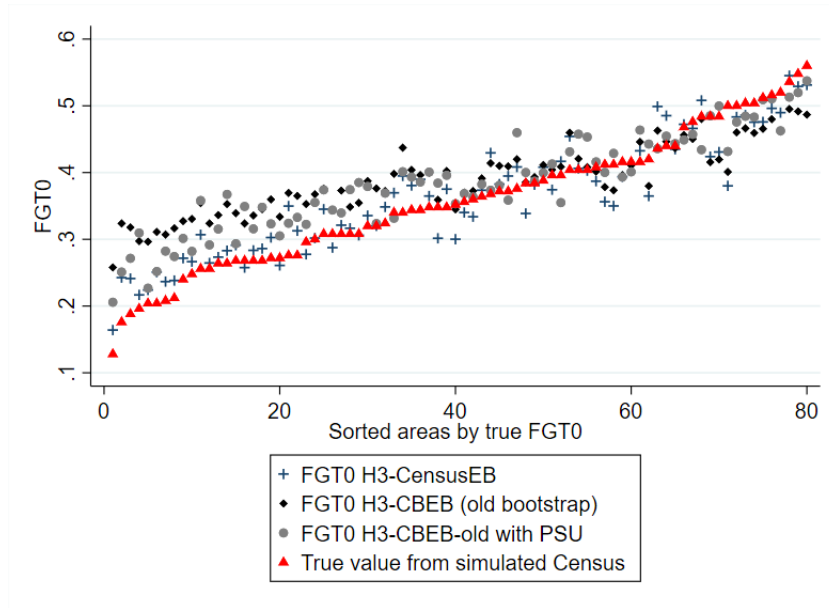

Figure 6: True poverty rates, ELL estimates (labeled "ELL"), ELL estimates including the area means of x_1 in the model (labeled "ELL with context") and Census EB estimates obtained from MR (2010) without appending the survey data, in one simulated population and sample, with improved model fit.
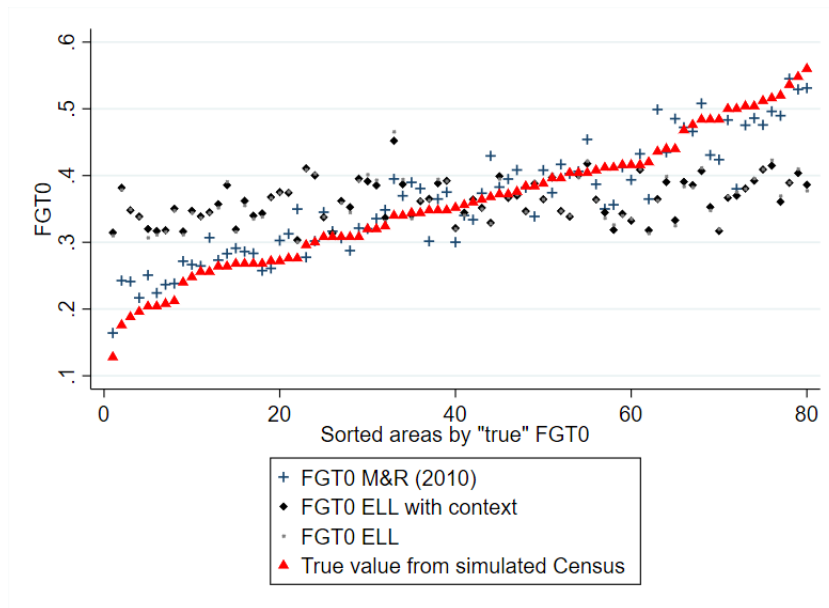
Figure 7: Empirical bias across simulated populations of Census EB from Section 6 using H3 method (labeled "H3-CensusEB"), traditional ELL (labeled "ELL"), CB-EB with H3 method (labeled "H3-CBEB (old bootstrap)"), CB-EB with PSUs (labeled "H3-CBEB (old bootstrap with PSU)") and Direct estimators (labeled "Direct") of the poverty gaps, with improved model.
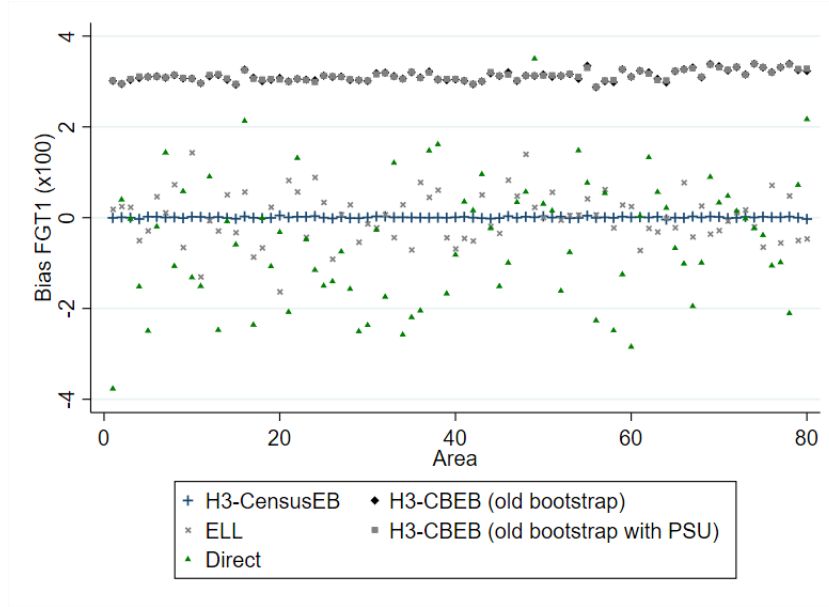
Figure 8: Empirical MSE across simulated populations of Census EB from Section 6 using H3 method (labeled "H3-CensusEB"), traditional ELL (labeled "ELL"), CB-EB with H3 method (labeled "H3-CBEB (old bootstrap)"), CB-EB with PSUs (labeled "H3-CBEB (old bootstrap with PSU)") and Direct estimators (labeled "Direct") of the poverty gaps, under improved model.
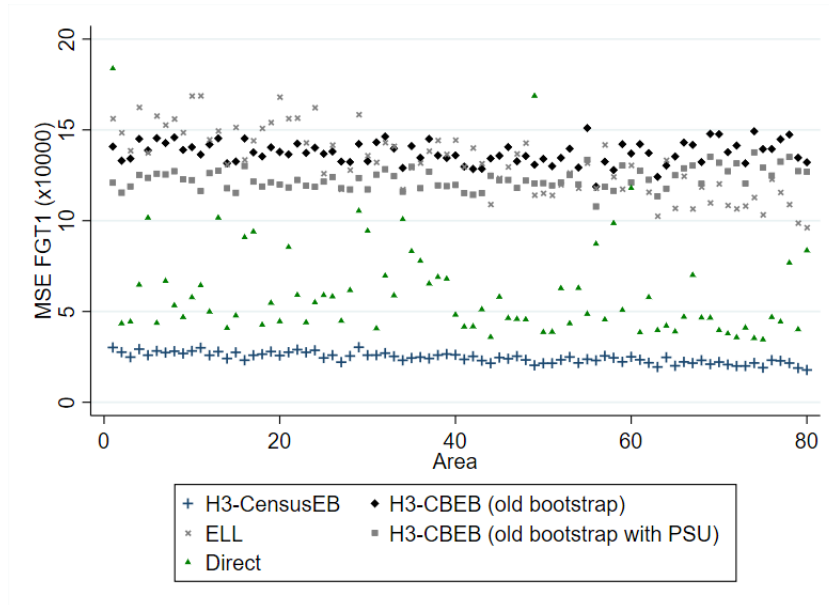
null hypothesis of normality due to the strong evidence against it (which is never exactly true in real applications). Hence, usual residual plots for checking the normality assumption using both predicted area effects and conditional unit-level residuals should be an important part of any application with real data.

In fact, log transformation of welfares is indeed conventionally taken because practically all welfare variables show strong right-skewness and heteroscedasticity. First of all, it is important to note that, even after log transformation, residuals in some applications still show a skewed distribution, where the skewness is transferred to the left tail. This occurs often in the presence of very small welfare values, which the log function actually shift to minus infinity. A simple way to solve this left skewness problem caused by the log function is to add a shift $c > 0$ to the original survey welfares, shifting them to values further apart from zero, where the slope of the log function is less pronounced. As Molina, Nandram and Rao (2014) recommends, this shift c can be obtained by fitting the nested error model to a grid of values of c in the range of the welfare values, and selecting the value for which the Fisher skewness of unit level residuals is closest to zero. Another possibility is to select a different transformation from the Box-Cox or power families of transformations (which contain the log), as implemented in the sae R package of Molina and Marhuenda (2015). In fact, Box and Cox (1964) included also the shift in their original family of transformations.

If one wishes to avoid finding the correct transformation to achieve normality[27] or if residuals still show skewness, even after selecting the best possible transformation of welfare, approaches for EB estimation have been recently developed for skewed distributions, see Graf, Marín and Molina (2019) for EB under general skewed distributions and Diallo and Rao (2018) considering skew-normal errors.

A final simulation experiment was done by increasing the size of the population to $N = 100,000$, but the sample size is kept as $n_c = 50$ households per area. In this case, each area is made up of $N_c = 1,250$ households, which means that sampling fractions are now 4% (much smaller than the previous 20%). Everything else in the simulation experiment is kept the same as in the simulation experiment with the improved model and assuming normality. Under this scenario, ELL still performs worse than the proposed Census EB, see Figures 11 and 12. However, it now shows improved performance in certain areas, where it achieves a smaller MSE than the direct estimators (Figure 12). This bodes well for real world scenarios, where the sampling fractions hardly reach 0.5% and thus ELL will likely be preferable to direct estimators.

Finally, under this same (more realistic) scenario, we analyze the performance of the MSE estimators obtained from each method. Figure 13 shows respectively the true MSEs of the traditional ELL estimators (approximated empirically with the $L = 10,000$ simulations), and the corresponding ELL estimated variances (section 2). We can see that estimated ELL variances are severely understating the true MSEs. This severe underestimation also affects the CB-EB procedure (see appendix Figure 21). On the other hand, according to Figure 14, showing true MSEs of Census EB estimators and their corresponding parametric bootstrap MSEs, where Census EB and bootstrap MSE estimators are obtained from the algorithms of Section 6, the parametric bootstrap MSE estimators are tracking adequately the true MSEs.

Similar conclusions were obtained for other indicators, such as the poverty severity (FGT indicator with $\alpha = 2$), although results are not shown for brevity.

---

[27]Note that for a random variable v, the transformation that leads to normality is given by $\Phi^{-1}(F(v))$, where $F(\cdot)$ is the true cumulative distribution function (cdf) of v and $\Phi(\cdot)$ is the standard normal cdf. The problem is that the true cdf $F(\cdot)$ is typically unknown

Figure 9: Empirical bias across simulated populations of proposed Census EB (labeled "H3-CensusEB"), ELL with location means (labeled "ELL with context") and direct estimators (labeled "Direct") of the poverty gaps, under non-normal errors.
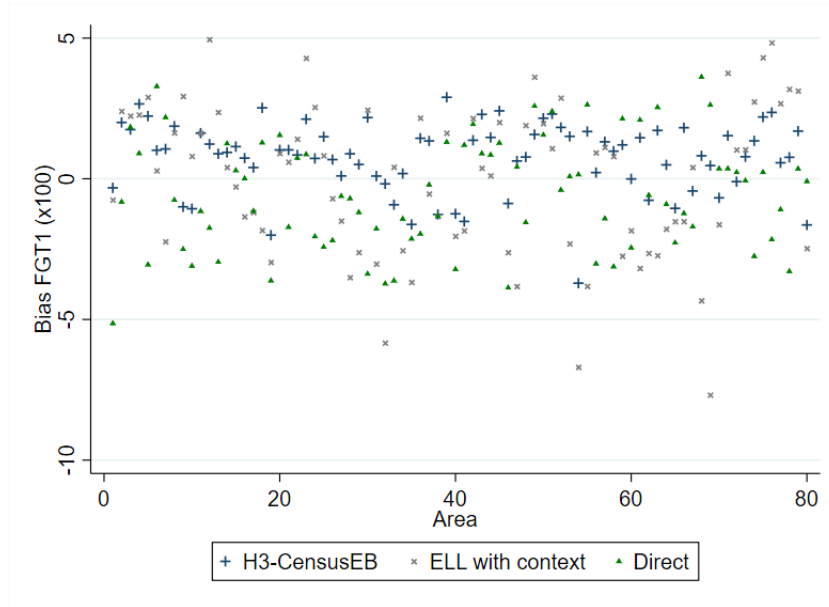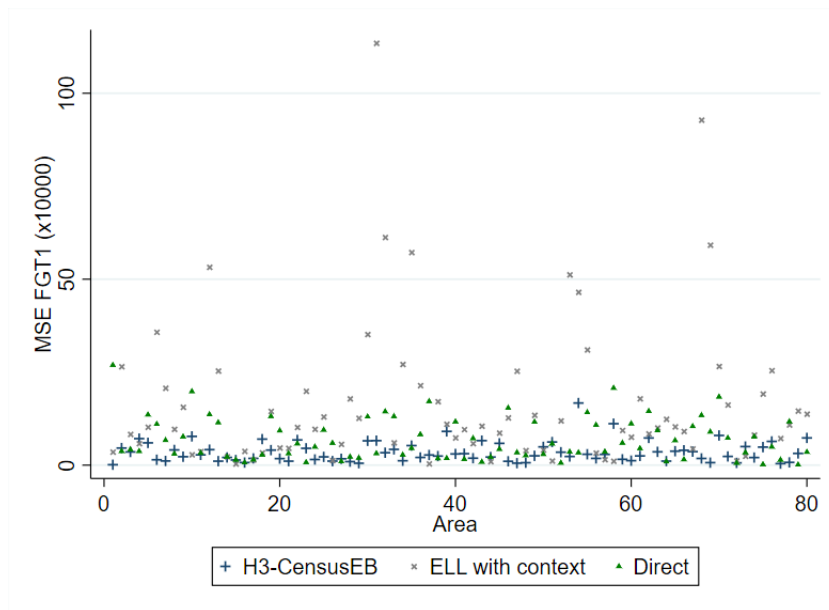

Figure 10: Empirical MSE across simulated populations of proposed Census EB (labeled "H3-CensusEB"), ELL with location means (labeled "ELL with context") and direct estimators (labeled "Direct") or the poverty gaps, under non-normal errors.

Figure 11: Empirical bias across simulated populations of Census EB estimators using H3 method (labeled "H3-CensusEB"), analogue CB-EB estimators sampling PSUs (labeled "H3-CBEB (old bootstrap with PSU)"), ELL with location means (labeled "ELL with context") and direct estimators (labeled "Direct"), with population size of 100K.
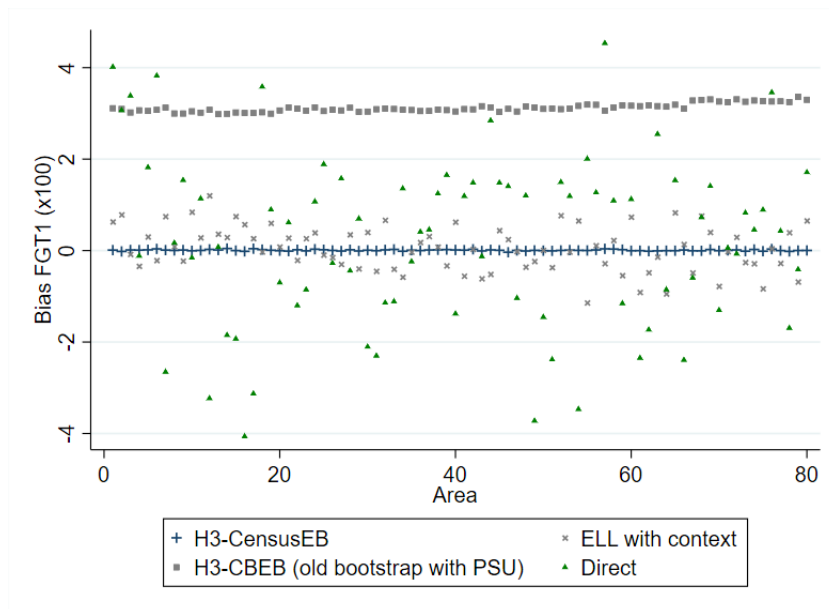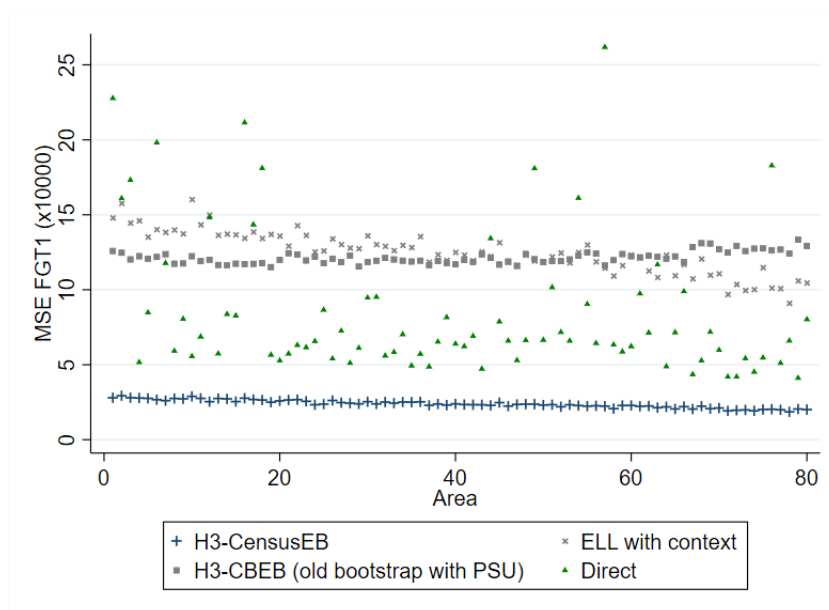

Figure 12: Empirical MSE across simulated populations of Census EB estimators using H3 method (labeled "H3-CensusEB"), analogue CB-EB estimators sampling PSUs (labeled "H3-CBEB (old bootstrap with PSU)"), ELL with location means (labeled "ELL with context") and direct estimators (labeled "Direct"), with population size of 100K.

Figure 13: Empirical MSEs of ELL estimators of poverty gaps (labeled "True ELL MSE") and corresponding ELL variance estimates (labeled "Estimated ELL Variance"), with population size of 100K.

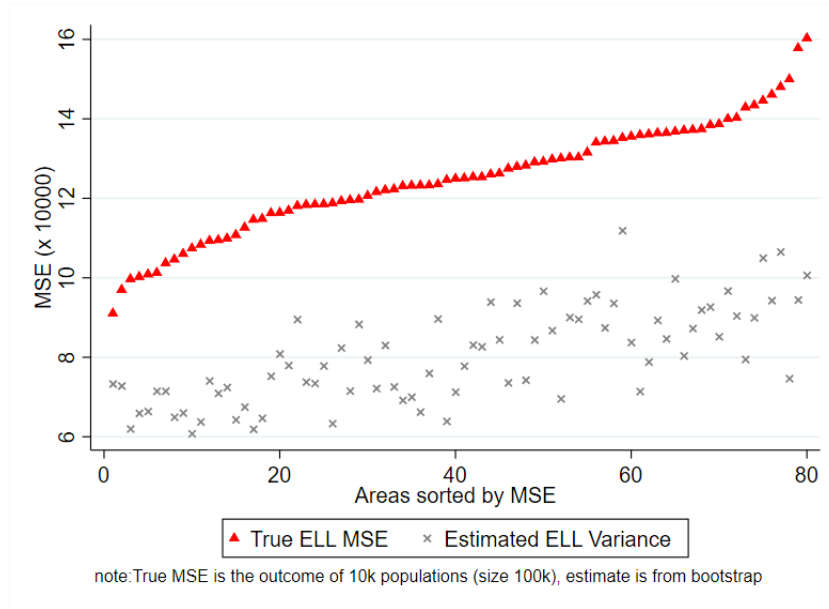note:True MSE is the outcome of 10k populations (size 100k), estimate is from bootstrap

Figure 14: Empirical MSEs of Census EB estimators of poverty gaps that use H3 estimation method (labeled "True MSE") and corresponding parametric bootstrap MSE estimates (labeled "Estimated H3-CensusEB MSE"), with population size of 100K.

note:True MSE is the outcome of 10k populations (size 100k), estimate is from bootstrap
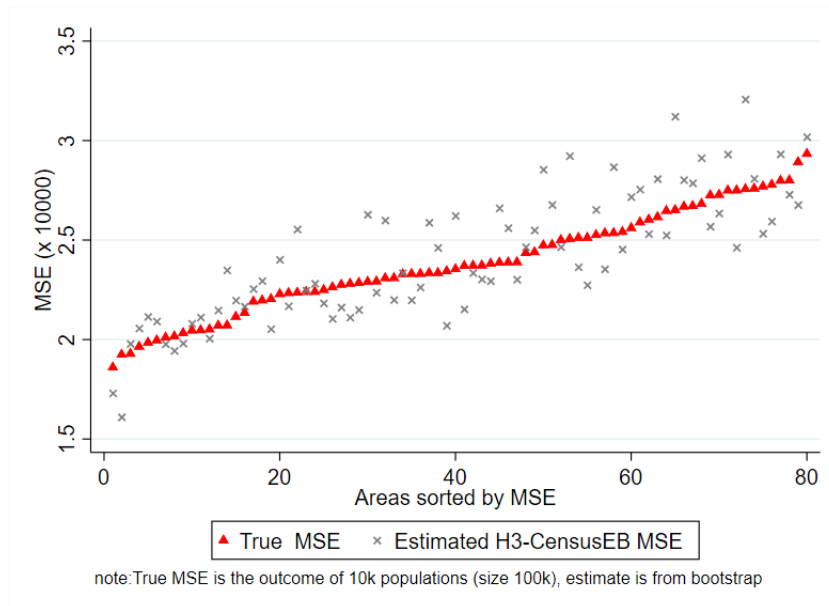
7.3 Examining the bias of CB-EB PovMap update

The bias of the CB-EB estimator observed in our simulation experiments is a reason of concern. In the CB-EB method with clusters equal to the areas, one could argue that the bias comes from the fact that not all areas are equally likely to be included in the sample that is selected in step 1 of this procedure. However, in our simulation experiments, bias also appears for the CB-EB method with PSUs selected within the areas in step 1. Thus, the question remains: what is the source of this bias?

Investigating whether part of the bias comes from the regression coefficients used to draw the welfares $\hat{\beta}^*$, which vary across the bootstrap replicates, we have plotted a histogram of the simulated regression intercepts $\hat{\beta}^*$ obtained in the M replicates of the CB-EB bootstrap procedure, with $M = 50, 100, 300$, for a single population. Figure 15 shows the results for $M = 50$ and $M = 300$ bootstrap replicates. This figure shows that the distribution of the simulated regression intercepts β_1^* for small M is not actually centered around the true value $\beta_1 = 3$; in fact, it is right-skewed with a mean clearly exceeding the true value. While increasing the bootstrap replicates M yields better results (right plot), the regression intercepts are still not centered around the true intercept.

Figure 15: Histogram of the regression intercepts $\beta^{*(m)}$, $m = 1, \ldots, M$ obtained in the M replicates of the CB-EB bootstrap method, for $M = 50$ (left) and $M = 300$ (right).
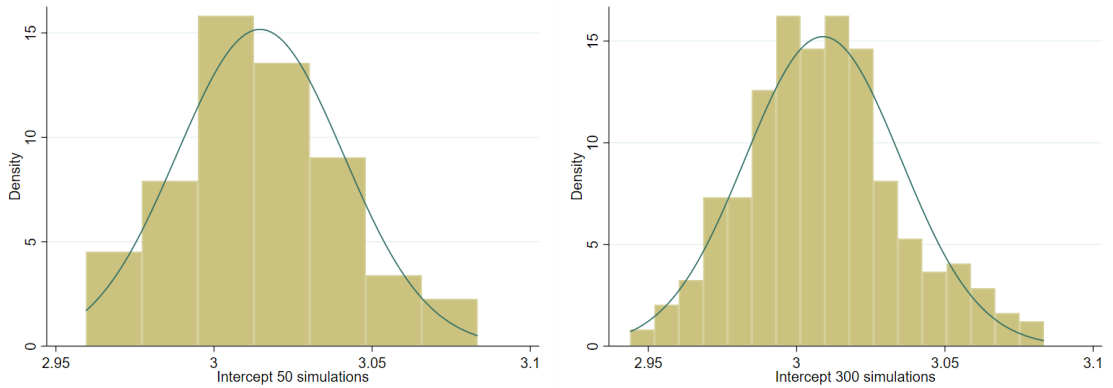

A second plausible source of bias is through the simulated location effects, η_c^*. Since the method selects PSUs with replacement, some areas might be selected more than once and others might not be selected. As a consequence, $\widehat{\text{var}}[\hat{\eta}_c]$ might be underestimated in an area selected more than once because there will artificially be more observations in that area. On the other hand, if an area is not selected, then $\eta_c^* \sim N\left(0, \hat{\sigma}_\eta^2\right)$,[28] and because $x_{ch}\beta^*$ is likely biased according to Figure 15, then the resulting estimate for this area will deviate considerably from the truth.

Figure 16 shows how the bias actually builds up in the predicted welfares.[29] The left plot in Figure 16 shows the empirical bias of the average across the census households of $\exp(x_{ch}\beta^*)$ for each area as an estimator of the corresponding average of the true values $\exp(x_{ch}\beta)$. The right plot includes η_c^* and thus shows the bias of the average of $\exp(x_{ch}\beta^* + \eta_c^*)$ as estimator of the true average based on $\exp(x_{ch}\beta + \eta_c)$, for each area c. We can see that the fixed part of the regression already shows a considerable upward bias, which agrees with the bias observed for the regression intercept in Figure 15. Once the location effect is added, things get much worse, with bias becoming substantial for most areas.

---

[28]Note that $\hat{\sigma}_\eta^2$ is not the one estimated from the sample, it comes from a bootstrap sampling of the data.

[29]This simulation is executed under the same scenario as that of Subsection 7.2, except that $L = 5,000$ instead of $L = 10,000$

Figure 16: Empirical bias of the average across the census households of $\exp(x_{ch}\beta^*)$ for each area $c = 1, \ldots, C$, as an estimator of the corresponding average of the true values $\exp(x_{ch}\beta)$ (left) and of the average of $\exp(x_{ch}\beta^* + \eta_c^*)$ as estimator of the true average based on $\exp(x_{ch}\beta + \eta_c)$, for each area c.
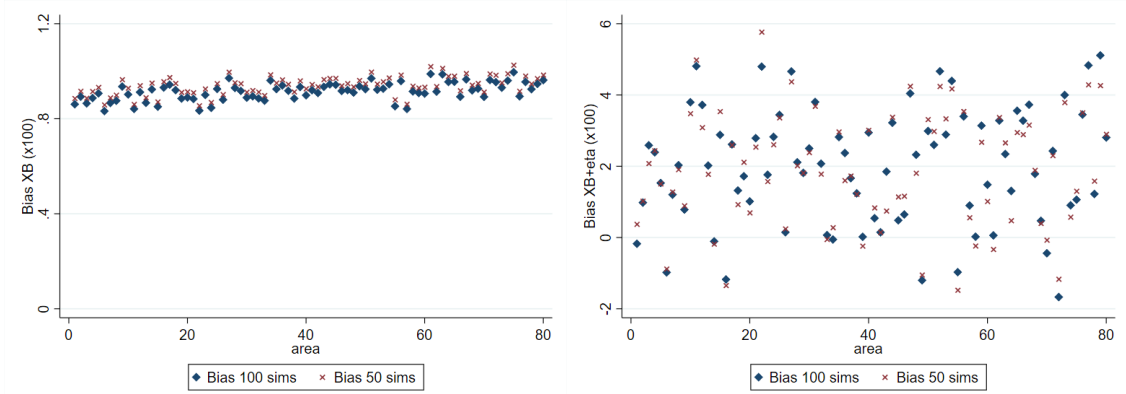
8 Conclusions

Since the turn of the 21st century, the World Bank has been obtaining small area estimates of poverty and inequality indicators using the ELL approach. One of the reasons for the method's popularity was the existence of software that made applying the method to real data relatively easy. This began with an implementation in SAS by Demombynes (2002), which was followed by PovMap implemented by Zhao (2006), and finally a Stata version ("sae" package by Nguyen et al., 2018). Along those lines, the research effected in this paper also represents a considerable update to the "sae" package in Stata.[30]

Countries often set out to obtain small area estimates because more precise poverty 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 an improved tool for the operationalization of the SDGs at a sub-national level.

An important aspect of the small area estimation procedures is how parameter estimates from an assumed population model are applied to census data to obtain the small area estimates. As noted in this document, the traditional ELL procedure is aligned to the approach used in multiple imputation, where the main aim is not prediction. Under multiple imputation, the method that yields the lowest MSE yields invalid statistical inference (Van Buuren, 2018). On the other hand, the goal of small area estimation is to improve precision.

This paper presents a substantial revision to the traditional methods by ELL (2002, 2003) and the updates by Van der Weide (2014) for small area estimation.[31] Our results 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. Thus, the updated methodology provides a considerable improvement in the quality of the SAE estimates obtained under the World Bank's agenda. We show that, for a single population, the original ELL method tends to align with the national poverty estimates and does not capture the area heterogeneity, although its performance improves when averaging across populations. Nevertheless, caution should be taken when doing out of sample prediction even under EB methods, since in that case estimates are very similar to the traditional

---

[30] An update to Nguyen et al. (2018) is in progress, but all Stata codes and commands used in this document are available at https://github.com/pcorralrodas/SAE-Stata-Package

[31] The revision is also accompanied by an update to the 2018 Stata "sae" package by Nguyen et al. (2018)

ELL ones and can thus deviate considerably from the truth. This is particularly relevant in cases where the explanatory power of the chosen correlates is low.

An additional finding of this research is the considerable bias of the CB-EB update by Van der Weide (2014) and as implemented by Nguyen et al. (2018) to the World Bank's toolkit. The revised Census EB method is aligned with MR's (2010) method, which shows a level of bias that is several orders of magnitude smaller than the previous CB-EB method. Furthermore, the MSE for the Census EB method is also considerably smaller. This research and its accompanying software implementation represent a massive improvement to the current World Bank's toolkit and overall poverty mapping agenda.

References

Box, G. and Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26:211–252.

Corral, P. and Cojocaru, A. (2019). Moldova poverty map: an application of small area estimation. mimeo.

Das, S. and Chambers, R. (2017). Robust mean-squared error estimation for poverty estimates based on the method of elbers, lanjouw and lanjouw. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(4):1137–1161.

Demombynes, G. (2002). A manual for the poverty and inequality mapper module. *University of California, Berkeley and Development Research Group, the World Bank.*

Demombynes, G., Elbers, C., Lanjouw, J., Lanjouw, P., Mistiaen, J., and Ozler, B. (2002). Producing an Improved Geographic Profile of Poverty: Methodology and Evidence from Three Developing Countries. WIDER Working Paper Series 039, World Institute for Development Economic Research (UNU-WIDER).

Demombynes, G., Elbers, C., Lanjouw, J. O., and Lanjouw, P. (2008). How good is a map? putting small area estimation to the test. *Rivista Internazionale di Scienze Sociali*, pages 465–494.

Diallo, M. and Rao, J. (2018). Small area estimation of complex parameters under unit-level models with skew-normal errors. *Journal of the Royal Statistical Society, Series A*, 45:1092–1116.

Elbers, C., Fujii, T., Lanjouw, P., Ozler, B., Yin, W., et al. (2007). Poverty alleviation through geographic targeting: How much does disaggregation help? *Journal of Development Economics*, 83(1):198–213.

Elbers, C., Lanjouw, J. O., and Lanjouw, P. (2002). Micro-level estimation of welfare. *World Bank Policy Research Working Paper*, (2911).

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

Fay III, R. E. and Herriot, R. A. (1979). 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.

Foster, J., Greer, J., and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica: Journal of the Econometric Society*, pages 761–766.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2004). *Bayesian Data Analysis.*, volume 2nd ed of *Texts in Statistical Science.* CRC Press [CAM].

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., and Santamaría, L. (2008). Bootstrap mean squared error of a small-area eblup. *Journal of Statistical Computation and Simulation*, 78(5):443–462.

Graf, M., MarÃn, J. M., and Molina, I. (2019). A generalized mixed model for skewed distributions applied to small area estimation. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, 28(2):565–597.

Guadarrama, M., Molina, I., and Rao, J. (2018). Small area estimation of general parameters under complex sampling designs. *Computational Statistics & Data Analysis*, 121:20–40.

Haslett, S., Isidro, M., and Jones, G. (2010). Comparison of survey regression techniques in the context of small area estimation of poverty. *Survey Methodology*, 36(2):157–170.

Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics*, 9(2):226–252.

Huang, R. and Hidiroglou, M. (2003). Design consistent estimators for a mixed linear model on survey data. *Proceedings of the Survey Research Methods Section, American Statistical Association (2003)*, pages 1897–1904.

Marhuenda, Y., Molina, I., Morales, D., and Rao, J. N. K. (2017). Poverty mapping in small areas under a twofold nested error regression model. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(4):1111–1136.

Molina, I. (2019). Desagregación de datos en encuestas de hogares: metodologías de estimación en áreas pequeñas.

Molina, I. and Marhuenda, Y. (2015). sae: An r package for small area estimation. *The R Journal*, 7(1):81–98.

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

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

Nguyen, M. C., Corral, P., Azevedo, J. P., and Zhao, Q. (2018). sae: A stata package for unit level small area estimation. *World Bank Policy Research Working Paper*, (8630).

Rao, J. N. and Molina, I. (2015). *Small area estimation*. John Wiley & Sons, 2nd edition.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American statistical Association*, 91(434):473–489.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons.

StataCorp, L. (2019). Stata mi reference manual.

Van Buuren, S. (2018). *Flexible imputation of missing data*. Chapman and Hall/CRC.

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).

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

Appendix: Additional simulation results

Figure 17: True poverty rates, EB and Census EB estimates obtained according to MR ((2010)) with REML fitting, in one population and sample, with poor model fit.
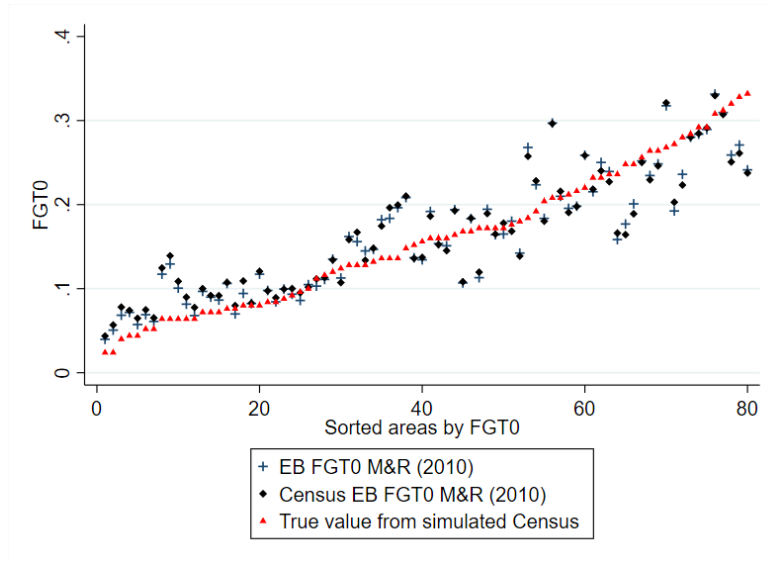
Figure 18: Difference in estimated MSE between EB and Census EB obtained according to MR (2010) with REML fitting, in one population and sample, with poor model fit.
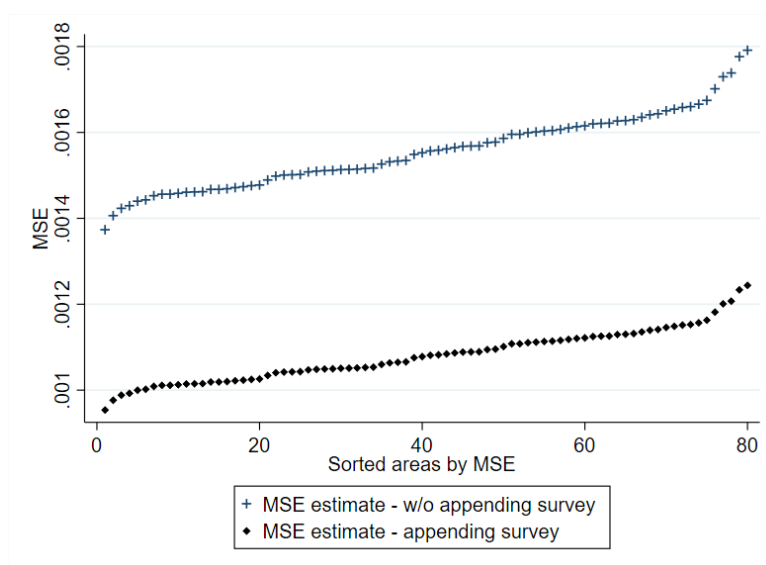

Figure 19: Census EB estimates with H3 estimation method and with REML fitting according to MR (2010), in one population and sample, with poor model fit.
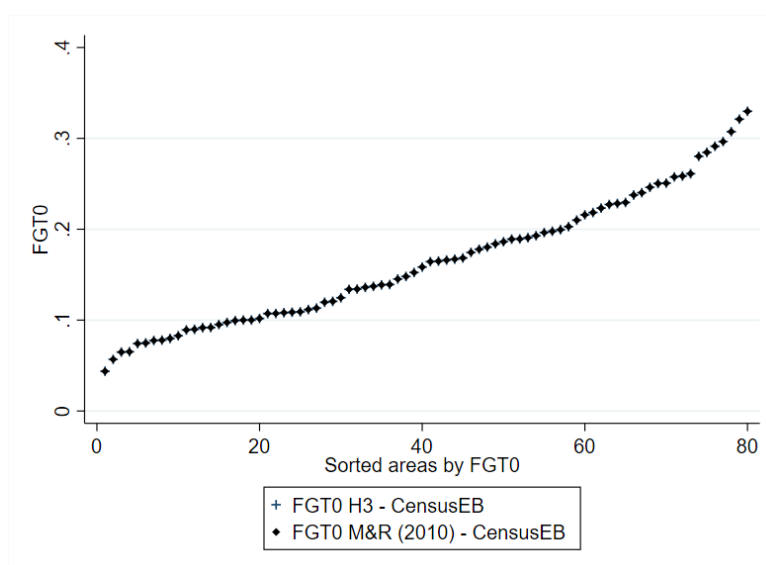
Figure 20: Empirical bias across simulated populations of Census EB with H3 or ELL estimation methods, compared with direct estimators.
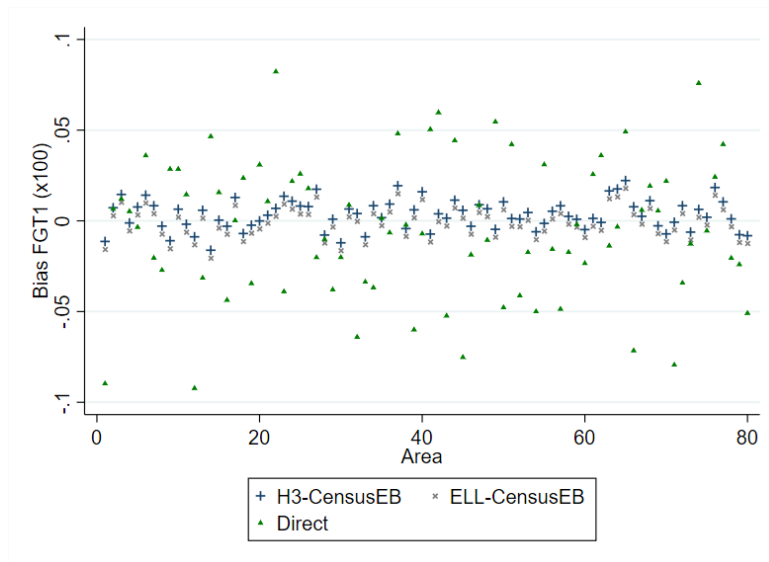

Figure 21: Empirical MSEs of H3-CBEB estimators of poverty gaps (labeled "True ELL MSE") and corresponding H3-CBEB variance estimates (labeled "Estimated H3-CBEB variance"), with population size of 100K.