Pull your small area estimates up by the bootstraps

Paul Corral, Isabel Molina, and Minh Nguyen

March 13, 2020

Abstract

After almost 2 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 presents, in detail, the simulation approach of the current methods used by the World Bank: Elbers, Lanjouw, and Lanjouw (2003 - ELL), and the Empirical Bayesian (EB) update done by Van der Weide (2014). These methods are then compared to those from Molina and Rao (2010). The paper presents evidence of the shortcomings of the simulation approach undertaken for Van der Weide's (2014) EB update, which, as is currently implemented, yields biased point estimates. The paper adapts the bootstrap approach used by Molina and Rao (2010), which comes from González-Manteiga et al. (2008), to the methods of Van der Weide (2014) and ELL (2002). The revised simulation methods yield lower mean squared errors, and in the case of the EB prediction presented in Nguyen et al. (2018) estimates which are considerably less biased.

Key words: Small area estimation, ELL, Poverty mapping, Poverty map, Empirical best, parametric bootstrap

JEL classification: C55, C87, C15

1 Introduction

It has been almost 2 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 SAE methodology used by the World Bank to obtain small area estimates of poverty, 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 which could be easily used by any practitioner. The software, PovMap (Zhao, 2006) is of great value to World Bank staff and many statistical agencies due to its easy point and click interface. The software has allowed the ELL approach to be adopted worldwide, and has permitted the World Bank to provide training and spread this knowledge to many statistical agencies due to its ease of use.

After the work from Van der Weide (2014), the PovMap software was revised with one very important modification and two substantial additions made to the software, and thus to the toolkit used by the World Bank. The first modification added a new approach for obtaining the unconditional variance of the residual parameters, Henderson's Method III (H3). The second improvement involved adjusting the Generalized Least Squares estimation of ELL with a modification of Huang and Hidroglou (2003). The final modification involved the inclusion of Empirical Bayesian (EB) prediction assuming normality. The modifications came about due to advances in the SAE literature (Molina and Rao, 2010 - M&R) and criticism of the ELL methodology (Haslett et al., 2010 and Molina and Rao, 2010).

To make the application of the World Bank's method friendlier to more advanced practitioners, and 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 is open and more curious practitioners can see in detail how methods have been implemented.

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

A very summarized description of the standard small area estimation methods based on unit-level models for poverty estimation is that they rely on estimating the joint distribution of the household's welfare given the 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 vectors in the census, it is possible to obtain poverty rates for every small area, including the non-sampled ones. Perhaps the two most well known approaches for unit level small area estimation of poverty are the traditional ELL (2003) method used by the World Bank, and Molina and Rao's EB approach (2010).[1]

This paper focuses on unit-level models for small area estimation; in particular, on the traditional ELL approach (2003), the latest additions to the World Bank toolkit by Van der Weide (2014), and the approach presented by Molina and Rao (2010). The purpose is to present a considerable update to the

[1] 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)

method of how point estimates and MSE is estimated under ELL and Van der Weide's (2014) update. The document sheds light on the nuances of the traditional ELL approach (2003) and the H3 and EB updates by Van der Weide (2014) , and it updates these methods in line with Molina and Rao's approach (2010). The paper goes in depth into the difference between the estimation of the noise and point estimates of the ELL approach (as implemented in PovMap and the sae Stata command), and the estimation of the MSE presented in Molina and Rao (2010). The paper adapts the bootstrap approach used by Molina and Rao (2010), which comes from González-Manteiga et al. (2008), to the World Bank methods. According to our simulation results, the adaptation of the bootstrap approach to ELL and the H3 methods presents a considerable improvement over the existing approach used by the institution.

The paper reviews the current methods in chronological order. It begins describing the traditional ELL method, and then moves on to the EB method proposed by Molina and Rao (2010), and finishes with the updates by Van der Weide (2014) to the ELL methodology. The paper then proposes an adaptation of the bootstrap method from González-Manteiga et al. (2008) to the World Bank methods (traditional ELL and its corresponding update, called here H3-EB). In the penultimate section, a simulation exercise, similar to the one performed by Molina and Rao (2010), is conducted and the different methods are compared. Finally, conclusions are presented.

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.[2] The popularity of the methodology can be attributed, to some degree, to the availability of the PovMap software (Zhao, 2006),[3] which was programmed in C and offers users a simple point and click interface. The PovMap software is also incredibly 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.[4]

The ELL method assumes the following nested error model for all the households in the population:

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

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

where η_c and e_{ch} are assumed to be independent from each other with different data generating processes. The indexes c and h in equation (1) correspond to cluster and household respectively, where N_c is the number of households in cluster c and C is the number of clusters. Additionally, x_{ch} is a $1 \times K$ vector of characteristics (also called correlates), and β is a $K \times 1$ vector of regression coefficients.

Under the original ELL methodology, the clusters indexed with c are supposed to be based on the sampling design and do not necessarily correspond to the level at which the small area estimates are

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

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

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

presented. Presenting estimates at a higher level than at which we have the nested error (η) may not be suitable in cases of considerable between-area variability, and may underestimate the estimator's standard errors (Das and Chambers, 2017). A way to alleviate this (*ibid*) is to include covariates which sufficiently explain the between-area heterogeneity in the model. In this regard, ELL (2002) suggests the inclusion of area level covariates which can explain the variation in welfare between areas as a way to improve precision, since it reduces location specific residuals. Marhuenda et al. (2017) recommend to put the location effects at the level where estimation is desired. According to this, here we consider that the clusters c are equal to the areas of interest.

Van der Weide (2014) updated the original GLS fitting method proposed in the traditional ELL approach, by making use of You and Rao (2002) for the inclusion of sampling weights, and Huang and Hidroglou (2003) for inclusion of heteroskedasticity. The unit level model in equation (1) is fit by Feasible Generalized Least Squares (FGLS). This implies that, initially, the residuals are assumed to not be nested as in equation (1) and thus the model is fit using ordinary least squares (OLS), after which the appropriate covariance matrix is estimated.[5] 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 an OLS model. Nevertheless, the parameter estimates used for generating welfare in the census come from the GLS fit. Given the way ELL implements its bootstrap procedure for estimation of the noise of ELL estimators, this is a crucial aspect because, if not done carefully, it may lead to considerably wide standard errors of the final point estimates.

A key feature of the ELL approach is that it also allows for the modeling of heteroskedasticity. Very little research has gone into this aspect of the method. The alpha model usually has a small adjusted R^2 (below 0.05) yet tends to play a considerable role in the point estimates obtained - particularly for parameters beyond poverty such as Gini, Theil, Poverty Gap, etc. Molina and Rao's (2010) approach also notes the possibility of allowing for heteroskedasticity, but this is not implemented in the sae R package (Molina and Marhuenda, 2015).

ELL proposed in their 2002 paper to use the delta method to calculate the variance of the estimators. The actual implementation of the method is through a computationally intensive simulated numerical gradient. The method relied on making perturbations to the parameters which is then used to derive an estimate of the gradient vector that is used for the variance estimation of the small area estimators via the delta method. The advantage the method provides is that it allows for a decomposition of the prediction error into components. ELL (2002) also presents the possibility of drawing from the sampling distribution (parametric) as a way to incorporate model error into the total prediction error. This later method has become the de facto approach used under the ELL methodology.[6]

The first implementation of the World Bank poverty mapping software was done in SAS (Demombynes, 2002), and for the estimation of poverty estimates and standard errors it relied on simulations, where parameters are drawn from their respective sampling distribution. This approach is the one currently used in the World Bank's tools for implementing the ELL (2002) small area approach: Stata's sae package and PovMap. The simulation approach of drawing from the sampling distribution is very reminiscent of multiple imputation methods. Every single relevant parameter necessary for simulating vectors of welfare is drawn from its (estimated) asymptotic distribution (or an approximation to it). To see this, we present the model used in ELL, which is a nested error regression model estimated via FGLS.

[5] 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).

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

The steps for the bootstrap designed to obtain the traditional ELL point estimates and ELL estimated noise is:

1. Fit the model (1) via FGLS as specified in Nguyen et al. (2018) using the survey data. This yields the set of parameter estimates from the sampled data: $\hat{\beta}, \hat{\sigma}^2_\eta, \hat{\sigma}^2_e$

2. Use the obtained parameter estimates as true parameter values to randomly draw from their respective asymptotic distributions as follows:

$$\beta^* \sim MVN\left(\hat{\beta}, \text{vcov}(\hat{\beta})\right)$$

where β^* is a $K \times 1$ vector of randomly drawn parameters. Generate household specific residuals as:

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

where (Gelman et al. 2004, 364-365)

$$\sigma^{2*}_e \sim \hat{\sigma}^2_e \frac{(n-K)}{\chi^2_{n-K}}.$$

Here, χ^2_{n-K} denotes a random number from a chi-squared distribution with $n-K$ degrees of freedom, where 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 location specific error is generated as

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

where (Demombynes, 2008; and Demombynes et al., 2002)

$$\sigma^{2*}_\eta \sim Gamma\left(\hat{\sigma}^2_\eta, \text{var}(\hat{\sigma}^2_\eta)\right).$$

Note that for a given simulation only one value of η^*_c is needed for every location, but all η^*_c within a simulation is drawn from the same σ^{2*}_η.

3. The simulated welfare y^* for a given household in the census is then obtained as

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

The vector of simulated welfares $y^*_c = (y^*_{c1}, \ldots, y^*_{cN_c})^T$ for every household within location c will be of size N_c, the number of census households in the location.

4. Repeat steps 2 and 3 M times. The standard value has been so far $M = 100$, although in practice a larger number of simulations are required to approximate well the distributions and so a larger number should be executed.

5. With all vectors y^*_c, $c = 1, \ldots, C$ of simulated welfare in hand, indicators can be produced. Define the indicator of interest for a given simulated vector in location c as $\tau^{ELL*}_c = f(y^*_c)$; for example, for the Foster, Greer, Thorbecke (1984) class of decomposable poverty measures, we use the vector of simulated welfare for location c, y^*_c, as:

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

where p_{ch} is the size of household h of location c in the census and $I(y^*_{ch} < z) = 1$ if $y^*_{ch} < z$ and equal to 0 otherwise. The ELL estimator $\hat{\tau}_c^{ELL}$ is given by:

$$\hat{\tau}_c^{ELL} = \frac{1}{M}\sum_{m=1}^{M} \tau_{c,m}^{ELL*}$$

and the estimated variance of the ELL estimator is given by:

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

This traditional ELL method uses the same computational procedure for both estimation of the small area indicators of interest and their respective mean squared errors (MSEs), unlike the EB procedure of Molina and Rao's (2010), which uses a Monte Carlo method for estimation of the small area indicators and a separate bootstrap procedure for MSE estimation. The standard error in ELL procedure is captured by the variation of estimates across simulations, which is aligned to Rubin's rules (Rubin, 2004). As will be detailed in a latter section, the ELL (2002) method tends to yield less precise estimates than Molina and Rao's (2010) approach.

The standard errors of the final indicator estimates are impacted by different sources of error. ELL (2002) describe three different sources of error in the final estimates, namely idiosyncratic error, model error and computation error, 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 poor 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.

2. The **model error** is related to the properties of the model parameters, β 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 our 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.

3 Molina and Rao's Empirical Best Predictor

The nested error model considered by Molina and Rao (2010) is similar to ELL's model (1), but it is typically fit by maximum likelihood (ML) or restricted ML. Moreover, the implemented version of ELL incorporates survey weights[7] for the estimation of the model parameters and the location effect. ELL (2002), and subsequently Van der Weide (2014), show how complex survey weights are incorporated

[7]The weighted GLS for ELL is presented in footnote 8 of ELL (2002)

into the procedure.[8] A more recent paper by Guadarrama, Molina and Rao (2018) further extends the method of Molina and Rao (2010) to complex sampling design, but so far it has not been implemented in any software package as a library or command. This section focuses on the original approach proposed by Molina and Rao (2010).

The main difference to ELL's approach is that Molina and Rao's (2010) method conditions on the sample data and thus, requires areas across sample and census to be matched. Even if the goal is to estimate indicators defined in terms of the welfare for all the population households, the survey is regarded as a subset of the census. The authors define the area's population as P_c and the survey sample, s_c, is a sample of size n_c drawn from P_c. The complement to the sampled population is referred to as r_c, which is just $P_c - s_c$. Consequently, the final welfare vector for any area c is defined as $y_c = (y_{cs}^T, y_{cr}^T)^T$, which has the welfare for sampled and non-sampled observations in area c.

Typically, the non-sampled population $(N_c - n_c)$ is much larger than the sampled population (n_c). It may also happen that $n_c = 0$ for some areas, meaning that the area is not sampled in the survey. In contrast, the ELL approach does not require to link the census and survey observations.

The method for obtaining empirical best (EB) predictors through Monte Carlo approximation given below is based on the description in Rao and Molina (2015). The Monte Carlo approximation for the EB small area estimator is obtained as follows:

1. Using the survey data, fit model (1) via restricted maximum likelihood or any method providing consistent estimators. This yields the set of parameter estimates from the observed sample:

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

   where the 0 subscript indicates that the parameters come from the original household survey.

2. Use the parameter estimates obtained in step 1 as true values and simulate M vectors of welfare in the census. First, the welfare from each out-of-sample household, $y_{ch,r}$, is generated as follows:

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

   where the index ch, r stands for household h of the non-sampled population in area c, and η^*_c is generated as

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

   where $\hat{\eta}_{c0}$ is

$$\hat{\eta}_{c0} = \hat{\gamma}_c \left(\bar{y}_{c,s} - \bar{x}_{c,s}\hat{\beta}_0\right)$$

   for the areas c included in the sample, and $\eta^*_c \sim N\left(0, \hat{\sigma}^2_{\eta 0}\right)$ for the out-of-sample ones. Finally, $e^*_{ch,r}$ is generated as:

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

   In the above,

$$\hat{\gamma}_c = \frac{\hat{\sigma}^2_{\eta 0}}{\hat{\sigma}^2_{\eta 0} + \hat{\sigma}^2_{e0}/n_c},$$

   where n_c is the number of sampled observations in the cluster.

   For an area c that is sampled in the survey, the generated vector $y^*_{c,r}$ is augmented by the survey data $y_{c,s}$. Therefore, the final vector for the whole census in area c is $y^*_c = (y^*_{c,r}, y_{c,s})^T$, which is

[8] Haslett et al. (2010) show that, in the case of the original ELL with GLS fitting, the variance covariance matrix of the GLS estimator may not be symmetric due to the presence of the survey weights. Consequently, the GLS fitting procedure was revised by Van der Weide (2014).

made up of the census data that is out of the sample and the sample observations coming from the survey. 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. If one wishes to use the implemented version of the EB method in the `sae` package but cannot identify the survey units in the census, one possible approach is to append the survey to the census, which would lead to an approximate EB estimator. However, in this case, the total size $N_c + n_c$ would be greater than the actual population size of the area.

3. The previous step yields M simulated vectors of welfare $y^*_{c,m}$, $m = 1, \ldots, M$, which make use of the fitted model parameters, where $y^*_{c,m}$ is the m_{th} replicate of $y_c = (y_{cs}^T, y_{cr}^{*T})^T$. The true census indicator for area c in Monte Carlo simulation m is given by:

$$\hat{\tau}^*_{c,m} = f\left(y^*_{c,m}\right),$$

where $f()$ can be any indicator function such as the FGT poverty indicators. The final Monte Carlo approximation to the EB estimator of τ_c is just the average across simulations:

$$\hat{\tau}_c^{EB} = \frac{1}{M}\sum_{m=1}^{M} \hat{\tau}^*_{c,m}$$

Note that the best estimator of τ_c is the conditional expectation $\hat{\tau}_c^{EB} = E(\tau_c | y_{c,s})$, where the expectation is with respect to the conditional distribution of $y_{c,r}$ given $y_{c,s}$. Under normality of the random terms in the nested error model, this is a normal distribution as well. Welfare values for non-sampled households are generated from that conditional distribution, which is completely determined by the assumed nested error model (with normality). The EB estimator is obtained by simply replacing the true values of the model parameters (regression coefficients β and variance components) in the above expectation by the estimators obtained based on the original sample data, but these are constant across simulations because there is a unique DGP. Note that this simulation procedure is a basic Monte Carlo approximation of the EB predictor of τ_c, where the expectation defining the empirical best predictor is replaced by the corresponding Monte Carlo average. In the traditional ELL approach, the final estimator and also the corresponding variance are also supposed to approximate an expectation, but it is not sufficiently clear with respect to which processes is the expectation taken since there are several sources of variability in the generation process: generation of model parameters and generation of welfare values. Moreover, under ELL each generated census (in a given replicate of the simulation) is obtained from a different DGP since the model parameters are changing, but this is done in order to incorporate the uncertainty due to the estimation of the model parameters in the ELL standard error. As we are going to see in the next section, in Molina and Rao's approach, the mean squared error is obtained through a parametric bootstrap procedure that imitates the entire estimation procedure (model fitting and calculation of the conditional expectation) within each bootstrap simulated census.

The first thing to mention is that Molina and Rao actually estimate the mean squared error defined as the expectation of the squared prediction error under the considered model. Note that, under the model-based framework, the target indicators τ_c are random quantities and hence incorporate uncertainty as well. Even if the considered estimator (or predictor) was unbiased with respect to the model in the sense that the expectation of the prediction error is zero, it would still leave the variance of the prediction error, which does not equal the variance of the predictor because of the randomness of the target indicator.

The mean squared error of the EB estimator is obtained via a bootstrap procedure, which is computationally more intensive than the ELL bootstrap approach. This is because, in every bootstrap replicate, all the estimation steps done to obtain the EB estimator are reproduced, including model fitting and the Monte Carlo simulation 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 Monte Carlo simulations and the number of bootstraps, except when the target indicator has an explicit EB estimator (such as the poverty rate). In this latter case, the Monte Carlo simulation procedure can be replaced by an explicit formula calculated without any simulation procedure and then the number of simulated censuses will be the number of bootstrap replicates. The bootstrap method presented by Molina and Rao (2010) comes from González-Manteiga et al. (2008). The steps for the bootstrap estimation of the MSE are as follows (Molina and Rao, 2015):

1. Using the survey data, fit model (1) via restricted maximum likelihood or any method providing consistent estimators. This yields the set of parameter estimates from the observed sample:
$$\hat{\theta}_0 = \left(\hat{\beta}_0, \hat{\sigma}^2_{\eta 0}, \hat{\sigma}^2_{e0}\right)$$
2. Use the estimates in $\hat{\theta}_0$ to simulate census welfare[9] in the following manner:
$$\ln(y^*_{ch}) = x_{ch}\hat{\beta}_0 + \eta^*_c + e^*_{ch},$$
where the area effects η^*_c are generated as:
$$\eta^*_c \sim N\left(0, \hat{\sigma}^2_{\eta 0}\right)$$
and the household specific errors are generated as
$$e^*_{ch} \sim N\left(0, \hat{\sigma}^2_{e0}\right).$$
3. Produce indicators of interest for this vector of welfare $y^*_{c,b}$ in the simulated census, for every area c:
$$\tau^*_{c,b} = f\left(y^*_{c,b}\right)$$
where $f(y^*_{c,b})$ can be any indicator function such as the FGT indicators and b is the index for the bootstrap iteration.
4. Since the survey is regarded as a subset of the census, the survey sample is now extracted and, using this sample (note that this sample now has a new y vector simulated in step 2), one fits the model (1). This yields a bootstrap estimate $\hat{\theta}_b$:
$$\hat{\theta}_b = \left(\hat{\beta}_b, \hat{\sigma}^2_{\eta,b}, \hat{\sigma}^2_{e,b}\right)$$
5. Obtain the EB estimator $\hat{\tau}^{EB}_{c,b}$ through Monte Carlo simulation (if needed to approximate the conditional expectation because it has no analytical form), using the bootstrap estimates of the model parameters $\hat{\theta}_b$ in step 4. The estimation procedure is exactly the same as the one used to obtain the EB estimator based on the original sample, but using the bootstrap estimates $\hat{\theta}_b$.
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 $\mathrm{b_{th}}$ simulated census and $\hat{\tau}^{EB}_{c,b}$ is the EB estimator obtained from the corresponding extracted sample.
7. Once a sufficiently large number of bootstrap iterations, B, have been made, the parametric bootstrap approximation of the MSE is given by:
$$mse_B\left(\hat{\tau}^{EB}_c\right) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\tau}^{EB}_{c,b} - \tau^*_{c,b}\right)^2$$

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

4 Henderson's Method III - PovMap update - EB

Henderson's method III (Henderson, 1953) decomposition of the variance components and the EB addition (H3-EB) 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 the extension is not necessarily the same as Molina and Rao's (2010) EB, and it is considerably different from the ELL (2002; 2003) approach. The similarity to the ELL approach comes through in the manner which point estimates and standard errors are computed in the implementation in PovMap as well as in Stata's `sae` package.

The method is different from ELL (2003) since Henderson's method III (Henderson, 1953) decomposition of the variance components, implemented by Van der Weide (2014), yields different variance components from the ones estimated using the traditional ELL approach. This in turn yields a different variance covariance matrix for the $\hat{\beta}$ parameters, which yields different $\hat{\beta}$ than those of the ELL approach. The EB addition is implemented by generating the censuses using the estimated values for the location effect (η) instead of generating them from their theoretical distribution. The estimated location effect proposed by Van der Weide (2014) incorporates survey weights and heteroskedasticity, and is given by:

$$\hat{\eta}_c = \hat{E}\left[\eta_c | e_c\right] = \hat{\gamma}_c \left(\frac{\sum_h \left(\frac{w_{ch}}{\hat{\sigma}^2_{ech}}\right) \hat{e}_{ch}}{\sum_h \left(\frac{w_{ch}}{\hat{\sigma}^2_{ech}}\right)} \right), \tag{2}$$

where w_{ch} represents the survey weight for household h in location c and

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

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

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

Note that heteroskedasticity in the household specific residuals is allowed, where $\hat{\sigma}^2_{ech}$ represents the household specific parameter. The heteroskedasticity is estimated in the same manner to the one presented in ELL (2002). This same approach can be applied to the ELL (2003) method of obtaining the variance parameters (σ^2_η and σ^2_e), where in the equations presented one can substitute the relevant parameters. Thus, the updates presented by Van der Weide (2014) also present a way to do ELL variance decomposition with EB methods. An additional thing to note in the EB method from Van der Weide (2014) is that when there are no survey weights and assuming homoskedasticity, the EB parameters collapse to those from Molina and Rao (2010).

Empirical best prediction makes use of the estimated location effects from survey data in order to improve the point estimates during the process of obtaining poverty estimates and their standard errors. A crucial assumption under EB is that the residuals (e and η) are normally distributed, and consequently the distribution of η conditional on e will also be normally distributed (Van der Weide, 2014). Note that the use of empirical best has no bearing on the GLS $\hat{\beta}$ of the chosen method for obtaining the unconditional variance of the residual parameters.

One of the main differences, as noted in Nguyen et al. (2018), between the ELL method of decomposing the variance components of the residuals and the Henderson's method III adaptation is that under the ELL method the distribution of the variance components is assumed to be known. Therefore in order to overcome this lack of knowledge it is necessary to use bootstrap re-sampling of data to circumvent the issue of unknown distributions.

Van der Weide (2014) updated the traditional ELL method by changing the model fitting method and also the way that census data are generated in the bootstrap procedure by imitating the Empirical Best method. Note that the fitting method and the use of EB are independent from one another. For estimation of the variance components σ_η^2 and σ_e^2, it uses a variation of Henderson's method III that incorporates the sampling weights.[10] Because in this case the distribution ofvar $[\hat{\eta}_c]$ is unkown,[11] it is necessary to rely on bootstrap samples of the data to obtain standard errors:[12]

1. Take a bootstrap sample of clusters (primary sampling units) with replacement from the survey data set.

2. Fit model (1) via FGLS as specified in Nguyen et al. (2018), using 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 estimated effects for the sampled clusters in step 1, $\hat{\eta}_c$, and their corresponding estimated variance $\widehat{\text{var}}[\hat{\eta}_c]$. Note that these are not from the survey sample, but come from a bootstrap sample of the survey.

3. Use the obtained parameter estimates from step 2 to simulate y^* for a given household h in the census as:[13]

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

   Construct the vector y_c^* of simulated welfare for every household within location c of size N_c, for all the locations $c = 1, \dots, C$ in the census. For every location c in the census which is also present in the survey, generate

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

   where $\widehat{\text{var}}\left[\hat{\eta}_c\right]$ is given in (3) and $\hat{\eta}_c$ in (2). For locations not in the survey, generate

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

   The household specific errors are generated as

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

   and in the case of heteroskedasticity, as

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

4. Repeat steps 1 to 3 M times. As already said, even if traditionally $M = 100$, a larger number of simulations is recommended.

5. With the simulated welfare vector y_c^*, calculate the indicator of interest as $\tau_c^{H3-EB*} = f(y_c^*)$ such as the FGT indicator, for each location $c = 1, \dots, C$, as follows simulated y^*:

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

[10]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.

[11]Note that in ELL the var $\left[\sigma_\eta^2\right]$ is assumed to follow a Gamma distribution, and thus the same approach as that of section 2 is not possible

[12]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)

[13]Note how each population is generated from a different model.

The H3-EB estimator $\hat{\tau}_c^{H3-EB}$ is obtained as the average across bootstrap simulations of the indicators:

$$\hat{\tau}_c^{H3-EB} = \frac{1}{M} \sum_{m=1}^{M} \tau_{c,m}^{H3-EB*}$$

and the variance of the estimator is approximated as:

$$\text{var}_{H3-EB}(\hat{\tau}_c^{H3-EB}) = \frac{1}{M-1} \sum_{m=1}^{M} \left(\tau_{c,m}^{H3-EB*} - \hat{\tau}_c^{H3-EB} \right)^2 .$$

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 and expectation with respect to the sampling design, but considering only one stage of the design. Then, in step 3, location effects are generated from their conditional distribution given the sample data (determined by the nested error model), and household specific errors are generated from their distribution under the assumed nested error model.

Standard errors are obtained similarly as in the traditional ELL approach even if the bootstrapped data are generated differently, and are thus different from mean squared errors. 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 Differences between H3-EB, ELL and Molina and Rao's EB

Because EB prediction uses the survey data to obtain a value for the location effect and its error, its benefits will be only evident in locations present in the census and the survey. In the original ELL approach, the location effect was at the cluster level, which in most surveys in developing countries where the World Bank focuses its efforts, the sampled clusters represent a small percentage of all the clusters in the country. This can 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 the locations benefit from EB (Corral and Cojocaru, 2019). Nevertheless, differences arise because of the way the point estimates and their noise are obtained as will be seen in the next section.

Molina and Rao (2010) present evidence that their EB approach is superior to ELL since it yielded less biased and more precise estimates than ELL. This finding has met some criticism, specifically because of how the simulation experiment was conducted under Molina and Rao (2010). The simulation was conducted under scenarios where seldomly ELL is applied. The population for the simulation was 20,000 uniformly spread across 80 areas. Moreover the sample was taken in every area and was 20 percent of the area, 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 going below populations of 100 households. In practice the level of disaggregation of the ELL method depends on the model quality, and even with an R^2 close to 0.6 it is seldomly advised to go below 1,000 households (Elbers et al., 2007).[14] Another point of departure is that ELL recommends the use of location means of the correlates[15] to improve precision since these explain, and thus minimize, the variation due to location - something not used in the simulation conducted in Molina and Rao (2010). A final difference, and the one addressed in the following section is the fact that the bootstrap method for MSE estimation of Molina and Rao (2010) is considerably different from that of the original ELL (2003) and the updates made by Van der Weide (2014).

6 Census EB and parametric bootstrap MSE estimator using H3 & ELL fitting

Currently, the MSE of the traditional ELL and of the more recent H3 (Van der Weide, 2014) procedure are estimated under a very similar framework to that of multiple imputation, where parameters $\left(\beta, \sigma_\eta^2, \sigma_e^2\right)$ are simulated from their estimated distribution. Consequently, the final estimate of the standard error of the indicator is just the standard deviation of all the simulated point estimates, and so every simulated indicator is compared to the the average of the indicators.[16] In contrast, in Molina and Rao's (2010) bootstrap procedure, the average across bootstrap simulations of the squared differences between the EB predictors and the true values of the indicators, approximates the MSE under the nested error model.

Applying a parametric bootstrap procedure to the EB method presented by Van der Weide (2014) will allow us to compare the methods used by the World Bank to Molina and Rao (2010) on a more even footing, since a similar approach is used to obtain the estimated MSEs.

However, the way that EB estimates in Molina and Rao (2010) are obtained requires to link the survey and census households. Note that the EB estimator is the conditional mean of the indicator given the sample's welfare. This process is avoided here because it is often not possible to link both data sets. In fact, in most cases, the number of sample households for a given area is much smaller than the number of census households. Moreover, Molina and Marhuenda's (2015) R package for SAE does not allow for the inclusion of household expansion factors in the census data (household sizes). This implies that, if the households are the units, then in the implemented R function for EB estimation, every household counts equally as if the household sizes are equal, which may yield biased estimates because the target indicators τ_c are usually defined for individuals, and households have generally different number of individuals. On the other hand, if the units are the individuals, we might underestimate the MSE because the strong dependency (in fact, equality) among the welfare of individuals within the same household is ignored.[17]

Thus, here we apply the EB procedure without linking the survey and census data files. Consequently, the small area estimators of the target indicators will be obtained with a computation procedure similar to the Monte Carlo simulation approach used in Molina and Rao (2010), but simulating the vectors of welfare for all census households and not including the welfare for the survey units. This is similar

[14] When simulating the errors the distirbution is assumed to be normal, and when doing this for small populations across simulations there will be considerable differences.

[15] These may come from the census or admin data (ELL, 2003 p356)

[16] Another interesting point is that the $\hat{\beta}$ and $\hat{\sigma}s$ estimated using the original survey sample are not used to generate a welfare vector in the population. This means that the coefficients used across the simulated vectors are likely not equal to those estimated using the survey, and in fact may differ considerably.

[17] Note that welfare is usually considered a household level indicator, thus if one individual in the household is poor then every member of the household is poor. Simulating at the individual level would violate this premise, and it is possible that one individual's simulated error pushes her above the poverty line while for another the simulated error can push her below the poverty line.

to the Census EB procedure described in Molina (2019), but we account for the sampling design and heteroskedasticiy as in the H3-EB procedure (Van der Weide, 2014). For estimation of the MSE of the resulting small area estimators, we propose to apply the same parametric bootstrap method described in Molina and Rao (2010), adapted to the new small area estimators.

7 Extended Census EB estimators

Here we propose an adaptation of the EB prediction and bootstrapped MSE that is very similar to the Monte Carlo approximation and the parametric bootstrap presented by Molina and Rao (2010), but it excludes the linking of observations between survey and census. The proposed EB procedure is an extension of the Census EB estimator (see e.g. Molina, 2019) that accounts for the sampling design and heteroskedasticity according to Van der Weide (2014) The implemented version of this procedure also allows for the use of household expansion factors (household sizes) taken from the census. The proposed Monte Carlo simulation procedure for the approximation of the extended Census EB estimator is:

1. Fit model (1) via FGLS using the survey data. Then, decompose residuals and their variance parameter estimates similarly as in the original ELL procedure or using H3 method. This yields the set of parameter estimates from the observed sample:

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

   where the 0 subscript indicates that the parameters come from the original household survey.

2. Use the parameter estimates obtained in step 1 to simulate M vectors of welfare for all the census households, where each of the welfare vectors $y_c^* = (y_{c1}^*, \ldots, y_{cN_c}^*)^T$ is constructed as follows (note the absence of the sub-index r, since now it is simulated for all census households):

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

   where the location effects are generated for areas c in the sample as

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

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

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

   and

$$\widehat{\text{var}}\left[\hat{\eta}_{c0}\right] = \hat{\sigma}^2_{\eta 0} - \hat{\gamma}^2_{c0} \left(\hat{\sigma}^2_{\eta 0} + \sum_h \left(\frac{w_{ch}}{\hat{\sigma}^2_{e_{ch0}}}\right)^2 \hat{\sigma}^2_{e_{ch0}} \right).$$

   Here,

$$\hat{\gamma}_{c0} = \frac{\hat{\sigma}^2_{\eta 0}}{\hat{\sigma}^2_{\eta 0} + \sum_h w^2_{ch} \left(\sum_h w_{ch} \sum \frac{w_{ch}}{\hat{\sigma}^2_{e_{ch0}}}\right)^{-1}}.$$

   For an area c not included in the sample, the location effect is generated as $\eta_c^* \sim N\left(0, \hat{\sigma}^2_{\eta 0}\right)$. In absence of heteroskedasticity, the household specific residuals come from

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

   and, in the case of heteroskedasticity,

$$e_{ch}^* \sim N\left(0, \hat{\sigma}^2_{e_{ch0}}\right).$$

3. The previous step yields M simulated vectors of welfare for each area, $y^*_{c,m} = (y^*_{c1,m}, \ldots, y^*_{cN_c,m})^T$, which make use of the fitted model's parameters. The Monte Carlo approximation of the extended Census EB predictor for an indicator for a given area c, for each Monte Carlo simulation is given by

$$\hat{\tau}^*_{c,m} = f\left(y^*_{c,m}\right),$$

where, for the FGT indicators, we have

$$f(y^*_{c,m}) = \sum_{h=1}^{N_c} \frac{p_{ch}}{\sum_h p_{ch}} I\left(y^*_{ch,m} < z\right)\left(1 - \frac{y^*_{c,m}}{z}\right)^{\alpha}$$

and the extended Census EB estimate is just the average of the Monte Carlo indicators:

$$\hat{\tau}^{EB}_c = \frac{1}{M}\sum_{m=1}^{M} \hat{\tau}^*_{c,m}.$$

Note that $f\left(y^*_{c,m}\right)$ can be any target parameter beyond FGT poverty indicators. Note also how across the M simulated welfare vectors, there is no variation in $\hat{\beta}_0$, $\hat{\eta}_{c0}$, $\hat{\sigma}^2_{e0}$, $\widehat{\text{var}}\left[\hat{\eta}_c\right]$ or $\hat{\sigma}^2_{\eta 0}$.

Unlike the original EB approach of Molina and Rao (2010), areas that are sampled in the survey are not augmented with the survey welfare data. Therefore, the census welfare vectors y^*_c generated in step 2 are made up of solely the census data. Molina (2019) comments that the effect of adding the survey data when the sample is small relative to the census population of the area is negligible, and thus this omission should have a small impact.

Here we propose the following adapted version of the parametric bootstrap for the estimation of MSE described in Molina and Rao (2015) and in Molina (2019):

1. Fit model (1) via FGLS using the survey data and decompose residuals and their variance parameters via the traditional ELL fitting approach or H3. This yields the set of parameter estimates from the observed sample:
$$\hat{\theta}_0 = \left(\hat{\beta}_0, \hat{\sigma}^2_{\eta 0}, \hat{\sigma}^2_{e0}\right).$$

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
$$\eta^*_c \sim N\left(0, \hat{\sigma}^2_{\eta 0}\right).$$
The household specific residuals come from
$$e^*_{ch} \sim N\left(0, \hat{\sigma}^2_{e0}\right)$$
and, in the case of heteroskedasticity, from
$$e^*_{ch} \sim N\left(0, \hat{\sigma}^2_{e_{ch0}}\right).$$

3. Produce indicators of interest using the generated census welfare vector for every area:
$$\tau^*_{cb} = f\left(y^*_{cb}\right),$$
where $f\left(y^*_{cb}\right)$ can be any indicator function such as the FGT poverty indicators, and b is the index for the bootstrap iteration.

4. Use the parameters estimated in step 1 to obtain new sample welfare, $y^*_{ch,s}$, for every location 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 residuals are simulated as
$$e^*_{ch,s} \sim N\left(0, \hat{\sigma}^2_{e0}\right)$$
   and, in the case of heteroskedasticity, as
$$e^*_{ch,s} \sim N\left(0, \hat{\sigma}^2_{e_{ch0}}\right).$$
   Note that, because these are household specific, there is no link to the simulated vector in step 2, apart from the variance parameters $\hat{\sigma}^2_{e0}$ or $\hat{\sigma}^2_{e_{ch0}}$.
   (d) The process ensures that the survey households could plausibly be an extract of the census simulated in step 2. In any given location, two households with the same characteristics will have a different value of welfare only due to the household specific idiosyncratic term $e^*_{ch,s}$.
5. Make use of the simulated survey data from step 4 (note that the survey now has a new y vector) and fit the model (1) via FGLS. This yields a bootstrap vector of estimated model parameters:
$$\hat{\theta}_b = \left(\hat{\beta}_b, \hat{\sigma}^2_{\eta,b}, \hat{\sigma}^2_{e,b}\right)$$
6. Obtain the EB estimator ($\hat{\tau}^{EB}_{c,b}$) through Monte Carlo simulation for the model fit in step 5 using the same approach as described in the previous section. Specifically, make use of the parameters obtained from the model $\hat{\theta}_b$ to obtain a new EB estimator.
7. Repeat steps 2 to 6 a sufficiently large number of times B. Let $\tau^*_{c,b}$ be the true value and $\hat{\tau}^{EB}_{c,b}$ be the EB estimator obtained from the generated census in the $\mathrm{b_{th}}$ iteration of the bootstrap procedure.
8. Once a sufficiently large number of bootstrap iterations have been made, a parametric bootstrap estimator of the MSE is given by:
$$mse_B\left(\hat{\tau}^{EB}_c\right) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\tau}^{EB}_{c,b} - \tau^*_{c,b}\right)^2.$$

8 Simulation experiments comparing the different methods

8.1 Model-based simulation under poor model fit

Here we compare the different procedures for estimation of the poverty rates (FGT indicators) and the resulting MSEs using the parametric bootstrap described above by replicating the simulation exercise done by Molina and Rao (2010).

A census data set of 20,000 observations is created, where all observations are uniformly split among 80 areas, labeled from 1 to 80. This means that every area consists of 250 observations. A location effect for each area is simulated as $\eta_c \sim N\left(0, 0.15^2\right)$; note that every observation within a given area will have the same simulated effect. Then, values of two right-hand side variables are simulated. Both are binary variables. 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 yields values for which observations in areas with a higher label are more likely to take value 1. The next one, x_2 is not tied to the area's label. This variable is taken to be equal to 1 if a simulated random uniform value between 0 and 1 is less than or equal to 0.2. The welfare vector for each area is created as follows:

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

where $e_{ch} \overset{iid}{\sim} N\left(0, 0.5^2\right)$. The poverty threshold is fixed at 12, which is roughly 60 percent of the median welfare.

From the created "census" in each of the areas, 20 percent of the observations are sampled using simple random sampling without replacement;[18] this yields our "survey" data. This set-up is repeated 10,000 times. For each replicate, the following steps are done:

1. Calculate true FGT measures using the "census".
2. Obtain the direct estimator for the FGT measures using the "survey".
3. Fit the model to the survey data, and obtain the estimator for the FGT measures by running 50 Monte Carlo simulations using the original empirical best approach proposed by Molina and Rao (2010).
4. Obtain ELL estimators for FGT measures and their MSEs using the traditional ELL method.

The simulation exercise presented by Molina and Rao (2010) is extended by including the following steps:

1. Fit the model to the survey data and obtain the estimator for the FGT measures by running 50 Monte Carlo simulations using the empirical best approach proposed by Molina and Rao (2010), but without appending the survey to the simulated census vectors (Census EB).
2. Apply the H3-EB (EB predicting with H3 fitting) and ELL-EB (EB prediction with ELL fitting method), considering equal survey weights and without heteroscedasticity:[19]
   (a) using the bootstrap method detailed in Section 4.
   (b) using the extended Census EB and parametric bootstrap methods described in Section 7.

Model bias and MSE are estimated in simlar fashion to Molina and Rao (2010). Model bias of the chosen indicator in a given area is obtained as $E\left(\hat{\tau}_c^j - \tau_c\right)$ where E is the expectation and j indicates the method being used (Molina and Rao, 2010). The MSE for a given area is equal to $MSE = E_\xi\left(\hat{\tau}_c^j - \tau_c\right)^2 (ibid)$.

Before analyzing the obtained results, a few issues have to be noted regarding the simulation performed. The only right-hand side variables used by Molina and Rao (2010) in their simulation are x_1 and x_2. Under the usual applications of the ELL approach, the number of auxiliary variables is greater, and ELL (2002; 2003) also advocated for the use of contextual area-level variables in the right-hand side. Because of the large amount of explanatory variables, it is not uncommon for the explanatory power, as measured by adjusted R^2, of a typical ELL application to be over 0.5 (although sometimes it might

[18]The same observations are sampled every iteration, ensuring x_1 and x_2 vectors are the same.
[19]For a detailed description of the difference between the methods readers should refer to Nguyen et al. (2018).

be smaller). Under the traditional ELL bootstrap method, a low explanatory power of the model will result in a noisier poverty estimate. The simulation conducted by Molina and Rao (2010) and replicated here yields an adjusted R^2 of less than 0.01. In such situation, the much noisier ELL estimates found by Molina and Rao (2010) are not surprising.

Another factor adding to the lower precision of the traditional ELL approach is the small area population size, which is also a considerably smaller number than that recommended by ELL (2002; 2003). Note also that a big factor adding to the higher precision of Molina and Rao's (2010) approach is the appending of the survey observations, which in this simulation experiment represent 20 percent of the area population sizes. This is quite a high number that is hardly observed in practice under SAE applications, where the fraction of surveyed households in a given location is typically much smaller.

8.2 Results

Before getting into the results of the full population simulation, a single population is used to do a baseline comparison of the methods. More explicitly, the methods are initially compared using just one of the populations generated along with its corresponding sample. Each method is run by setting $M = 1000$ and $B = 1000$, the results are plotted and compared.

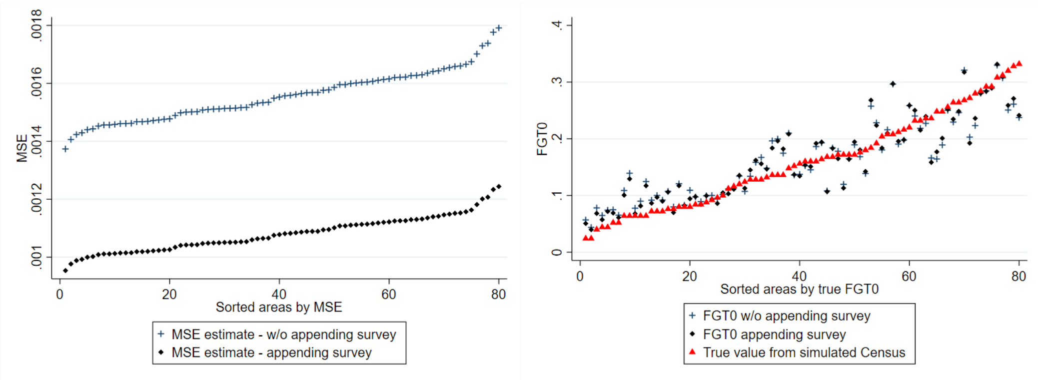

Figure 1: Comparison between appending survey and not appending

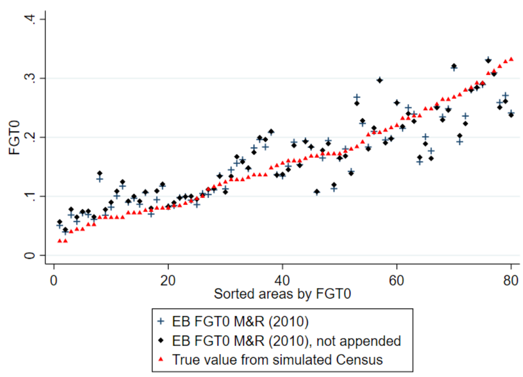

Figure 2: Difference with truth of M&R appended and not appended

One of the first things that pops out is that the impact on the estimated MSE of appending the survey is not as small as expected. In fact, our results indicate that the impact of not appending the survey is an increase on the resulting estimated MSE for poverty headcount of 44 percent for every single area compared to appending the survey (Figure 1 left), although the difference in the point estimates is quite

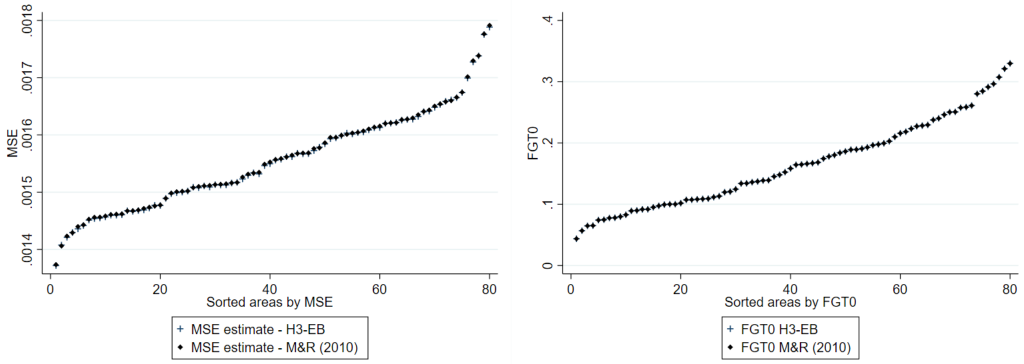

Figure 3: Comparison between H3-EB and M&R (2010)

small (Figure 1 right)[20] This value of 44 percent holds for each of the three main FGT indicators (0,1, and 2).[21] Appending the survey also seems to have a small impact on the overall difference between the estimated FGT0 (headcount rate) and the truth, with the absolute difference to the truth for this one simulation being 7 percent larger for the non-appended version (Figure 2). This is surprising given the little amount of information being brought in by the model. This result is shown to illustrate how much of an impact appending the survey has on the estimated MSE. Nevertheless, under most practical applications the size of the sample of any given area is less than 1 percent and thus the impact of appending the survey is considerably smaller. In the adapted methods presented henceforth the appending of the survey is not implemented, and instead what is shown is the outcome of CensusEB, which does not append or link survey households to census households.

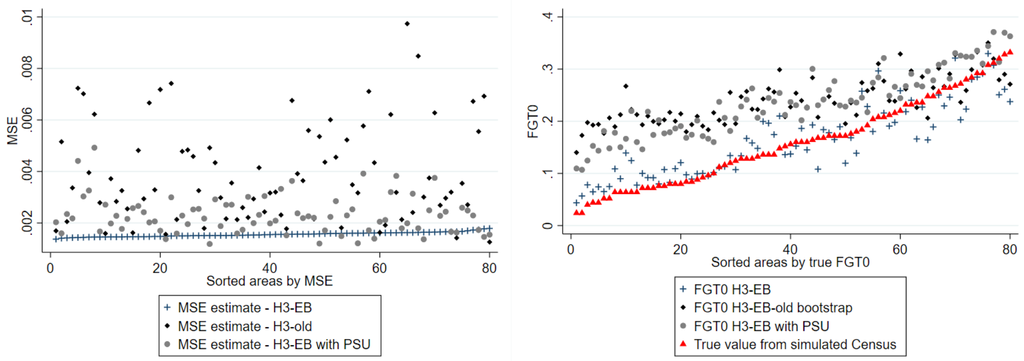

Figure 4: Comparison between H3 bootstrap methods

A second result of interest is that the difference between the H3-CensusEB using the procedures in Section 7 and Molina and Rao's (2010) CensusEB (without appending the survey) is almost non-existent. This can be observed for the estimated MSE as well as for the point estimates (Figure 3 left and right, respectively). Under the original H3-EB procedure presented in Section 4, the results differ considerably, with the original simulation approach yielding considerably different estimates (the absolute difference to the true value is almost three times that of the modified method in Section 7). The reason for this is that, under the original simulation approach in Section 4, a bootstrap sample of clusters is needed for each simulation and here the clusters are the areas. This leads to a given area benefiting from EB only in a portion of the simulations, since it is likely that an area is not selected in every single bootstrap

[20]This is because the simulation is fully controlled, thus the simulated vectors under both scenarios come from the same distribution.

[21]Figures for these results are available from the authors upon request.

sample. This leads to point estimates that are far off from those of the updated bootstrap of Section 7 (Figure 4 right). Another reason is that the parameters used across all simulated vectors used to obtain the point estimates differ across simulations, and if these are unable to converge to the mean, then it will lead to differences. Furthermore, the estimated MSE is considerably larger under the old approach of Section 4, as can be seen in Figure 4 left.[22]

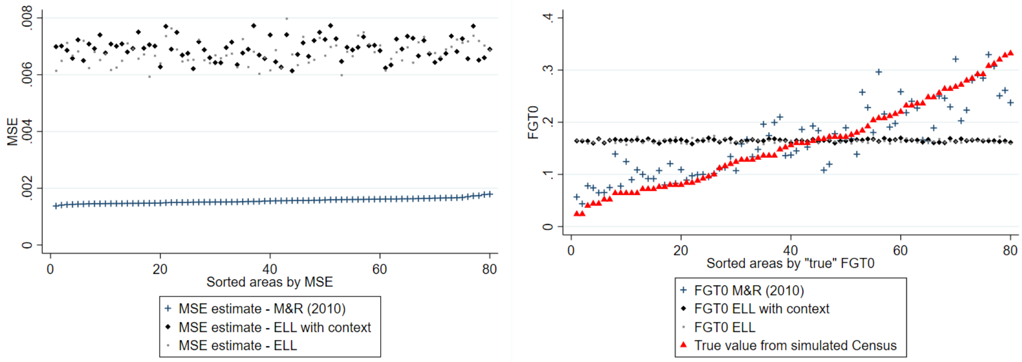

Figure 5: Comparison between ELL and M&R (2010)

In an alternative simulation experiment, we create 25 PSUs (clustered) within every area, each with 10 households. Five PSUs within every area are selected and the H3-EB with the old bootstrap approach of Section 4 is executed. The results are also shown in Figure 4. There appears to be a gain in estimated MSE; however, there is still a considerable difference to the poverty rate estimated under the new bootstrap procedure of Section 7. Nevertheless, some of the areas have MSE estimates which are below those of the new approach, this is because of how the bootstrap samples of data are obtained: across simulations some areas may be over-sampled and thus these areas have a larger number of observations. This would lead to a larger value of $\hat{\gamma}_c$, which then leads to a lower value of $\hat{\sigma}^2_{\eta 0}(1-\hat{\gamma}_c)$.

Turning to the original ELL approach, what one finds is that it faces the same troubles as the H3-EB with the original bootstrap procedure. This is in line with what Molina and Rao (2010) find in their simulations. Given the present simulation set-up, ELL performs much worse than Molina and Rao's EB method in terms of estimated MSE and point estimates (Figure 5, left and right, respectively). Also note the rather flat nature of ELL, this is mostly due to the model providing very little information, and given the lack of EB and limited explanatory power of the correlates, all estimates fall close to the average poverty rate of 16.[23] What is somewhat concerning is that the addition of a context level variable to the model for ELL, in the form of the average area value for x_1, does little to improve the point estimates (Figure 5, right). Thus, 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 contextual area variables.

The full results of the complete simulation after 10,000 populations are quite sobering. The bias of the method from section 4 used to obtain estimates under the H3-EB approach yields estimates that show considerable bias. The results also appear to be upward biased for all areas. Figure 6 presents the bias of the methods, and while the bias of the revision for H3-EB (see section 7) method may seem to be 0 these range between -0.025 and 0.027 (similar ranges to those from Molina and Rao's (2010) method).[24] The gains in terms of reduced bias of the revision can be easily appreciated in this figure. A closer look at

[22]What in reality is being compared here is the MSE presented in Section 7, to the variance presented in Section 4

[23]The estimates are not in fact flat, if sorted from smallest to largest the estimates range between XX and XX

[24]Results for Molina and Rao's (2010) are not shown since the Census EB results are aligned to the H3-Census EB results, as shown in figure 4.3

the bias of the revised H3-EB method and the revised ELL-EB method, show that these are also aligned (Figure 7).

Moreover, in terms of the estimated MSE, results found are aligned to results found by Molina and Rao (2010). The traditional ELL yields an MSE estimate which is above the one for the direct estimate. What is even more concerning is that the MSE yielded by the method applied in Nguyen et al. (2018) and PovMap for H3-EB (see section 4) to obtain point estimates is more than 3 times greater than the one of the traditional ELL. The revision to the method presented here represents a massive improvement.

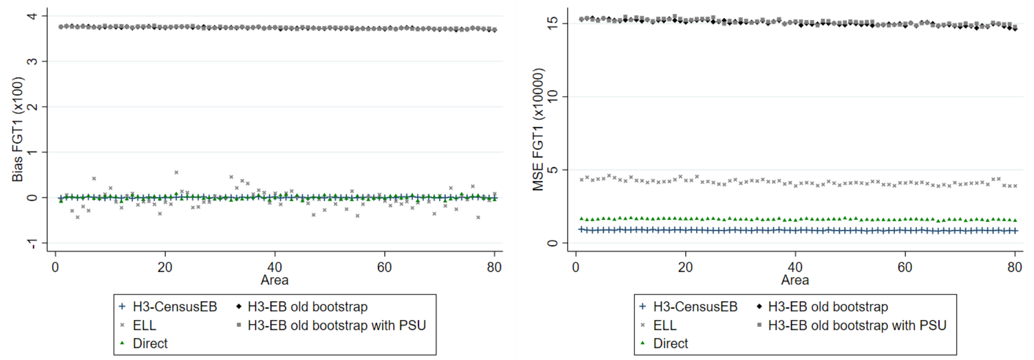

Figure 6: Bias and MSE over simulated populations

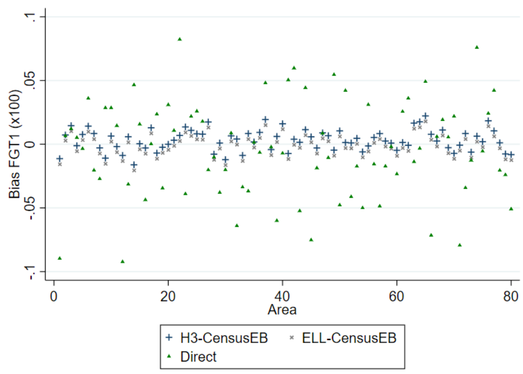

Figure 7: Bias of different methods

8.3 Model-based simulation with improved model fit

In order to address one of the concerns of the previous simulation experiment, the poor explanatory power of the model, this section creates a more informed model which yields an adjusted R^2 of 0.42, which is much closer to that of real world applications of ELL. The model includes six covariates which are generated as follows:

1. x_1 is a binary variable where a random uniform value between 0 and 1 is simulated, and if the value is less than or equal to $0.3 + 0.5\frac{c}{80}$ it is equal to 1.

2. x_2 is a binary variable but it is not tied to the area's label; a random uniform value between 0 and 1 is simulated and if it is less than or equal to 0.2 it will be equal to 1.

3. x_3 is a binary variable where a random uniform value between 0 and 1 is simulated, and if the value is less than or equal to $0.1 + 0.2\frac{c}{80}$ it is equal to 1.

4. x_4 is a binary variable where random uniform value between 0 and 1 is simulated, and if the value is less than or equal to $0.5 + 0.3\frac{c}{80}$ it is equal to 1.
5. x_5 is a discrete variable simulated as the rounded integer value of the maximum value between 1 and a random Poisson distribution with mean 3 times $1 - 0.1\frac{c}{80}$
6. x_6 is a binary variable but it is not tied to the area's label; a random uniform value between 0 and 1 is simulated and if it is less than or equal to 0.4 it will be equal to 1.

The welfare vector for each area is created 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 10.2. All other steps are similar to those of the previous simulation.[25]

8.4 Results

The argument towards executing this simulation is that the original simulation presented very poor explanatory power of the model, whereas under usual applications of ELL approach, the adjusted R^2 is considerably higher. The model under this simulation yields and adjusted R^2 of roughly 0.42. However, despite the improved explanatory power of the model, the traditional ELL method still performs poorly. The result in Figure 8 still portrays a rather flat ELL hovering around the national average poverty rate and a considerably higher estimated MSE than that of Molina and Rao's (2010) EBP method.[26]

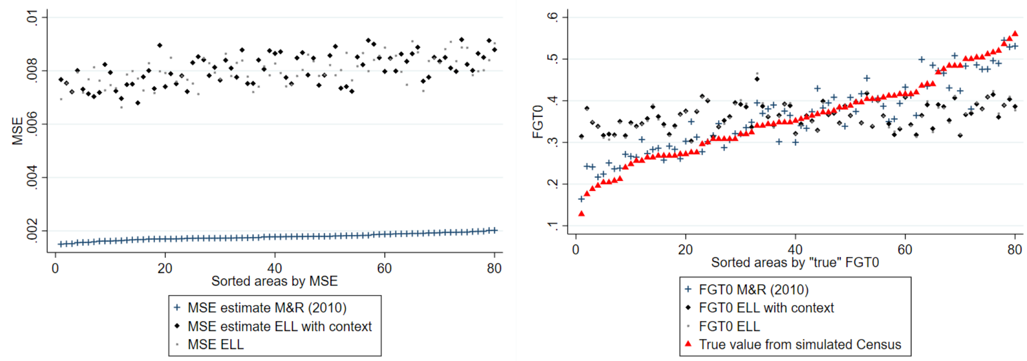

Figure 8: Comparison of ELL and M&R EBP estimates under improved model fit

Unsurprisingly, the results under the revised H3-EB method, when the model is better, still resemble those from the previous section (Figure 9). The revised bootstrap approach performs considerably better than the current methods in use and detailed in Nguyen et al. (2018). This conclusion remains true even when the simulated location effect is much smaller, simulated as $\eta_c \overset{iid}{\sim} N\left(0, 0.07^2\right)$.[27]

Moreover, the magnitude of the difference in estimated MSE due to appending the survey is the same as in the previous simulation, 44 percent larger for the Census EB (Figure 10), suggesting that the benefit is specifically tied to the size of the area's survey sample. Note that the magnitude of the estimated MSE under this simulation is not comparable to that of the previous simulation, because of different data generating processes.

[25] Another slight modification made in separate simulations is that the location effect is simulated as $\eta_c \overset{iid}{\sim} N\left(0, 0.07^2\right)$.
[26] The results shown are for Census EB for Molina and Rao's (2010) method.
[27] Results available upon request.

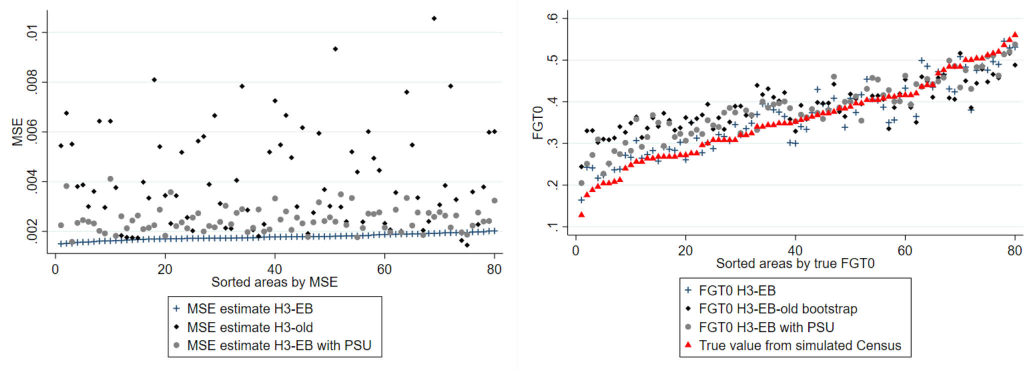

Figure 9: Comparison between previous bootstrap and new bootstrap for H3

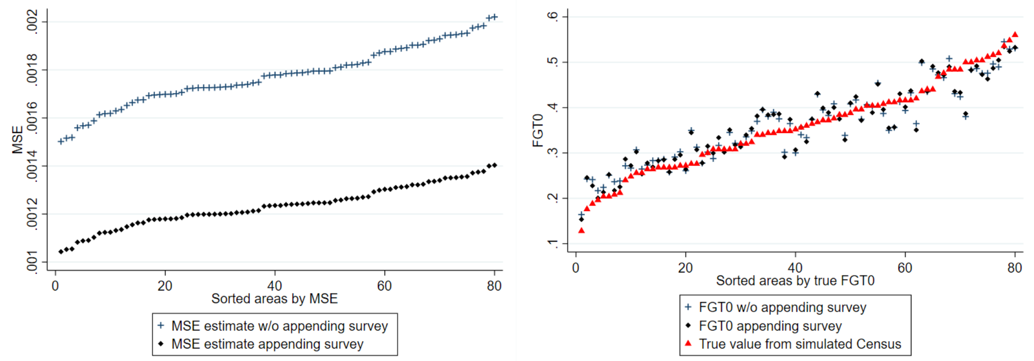

Figure 10: Comparison between appending survey and not appending

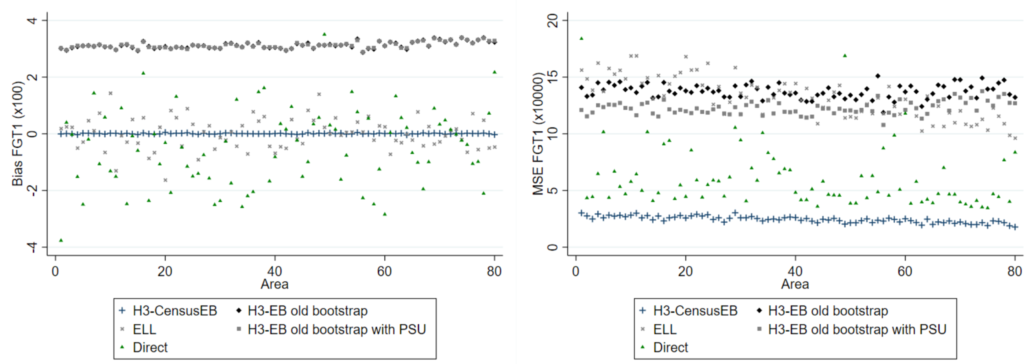

Figure 11: Bias and MSE over simulated populations

Results from the full simulation with 10,000 population vectors resemble those from the previous section (Figure 11). The H3-EB method implemented as described in Nguyen et al. (2018), which replicates the methods of PovMap, also presents considerable bias in this exercise. Due to the magnitude of the bias for the old H3-EB method, the revised H3-EB procedure seems to be 0, but in fact ranges between -0.04 and 0.05. Moreover, the MSE under this simulation is also much smaller for the method's revision. What is concerning is that the MSEs of the direct estimates are not only smaller than those of the traditional ELL approach, but also smaller than those of the H3-EB presented in Van der Weide (2014) and detailed in Nguyen et al. (2018).

An additional simulation is run where the only difference is that the household specific residual is simulated as non-normal. Specifically the residuals are simulated as a $t - distribution$ with 5 degrees of freedom and scaled by 0.5 (Figure 12). In this simulation the household specific residuals for ELL are taken from its empirical distribution. The results suggest that even in the face of non-normality of the household specific errors ELL's performance still lags that of the H3-CensusEB update. In terms of bias, some areas under the traditional ELL show quite considerable bias, yet on average across areas the bias is near zero. This suggests that even if the bias across all areas is small it is not an accurate measure since it hides heterogeneity. Moreover, the MSE of the ELL is still much greater than that of the H3-CensusEB update.

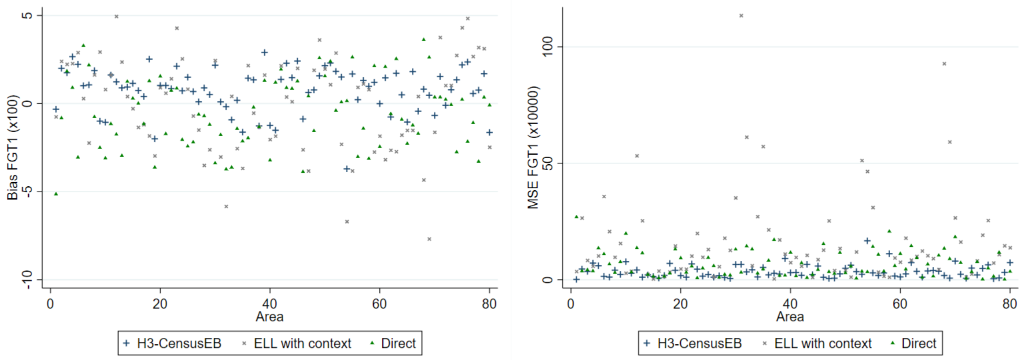

Figure 12: Bias and MSE over simulated populations with non-normal residuals

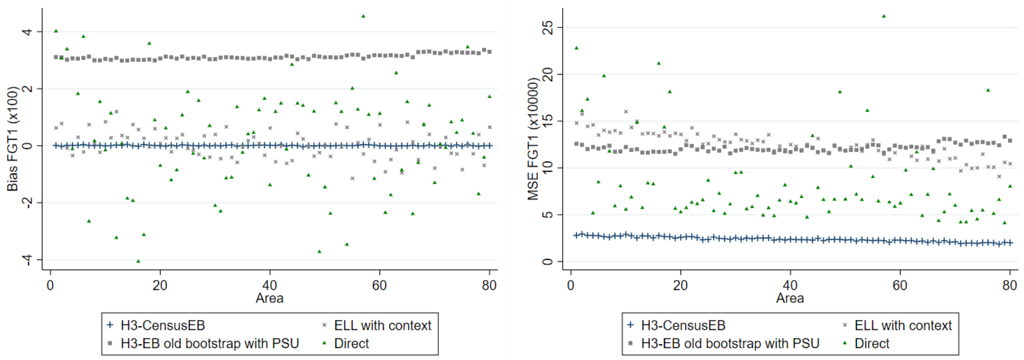

Figure 13: Bias and MSE over simulated populations (Pop of 100K)

A final simulation is done where the size of the population is expanded to 100,000 and the sample for each area is still 50 households per area.[28] Under this scenario ELL still lags the H3-CensusEB update (Figure 13). However, under a larger population and a much smaller sample (4%) the traditional ELL

[28]Note that now each area is made up of 1,250 households, and the sample is still 50 households from each area. Everything else is the same as the simulation proposed in section 8.3.

now shows imporved performance in some areas where it can be observed to have a smaller MSE than the direct estimates (Figure 13, right). This bodes well for real world scenarios, where the sample hardly ever is near half a percent of the population and ELL is applied and thus it likely is preferable to the direct estimates.

8.5 Examining the bias of H3-EB PovMap update

The H3-EB bias from the method of section 4 observed in the experiments in this section is considerable and a cause for concern. A possible explanation given is that the bias is introduced because not all areas are equally likely to be included. However, under this scenario all areas are equally likely to be included and thus there should be no contribution due to the selection. Thus, the question remains: what is leading to the bias?

A plausible explanation under this simulation scenario is that a good part of the bias is coming from the β^* over the bootstrap samples taken while obtaining EB estimates. Note that the $\hat{\beta}$ estimated from the original sample are never used to obtain the simulated vectors (see section 4). Figure 14 shows that under the method of section 4 a small number of simulations yield β^* which are not normally distributed and not centered around the truth, while running more simulations yields better results but still not ideal since it is not centered around the truth.

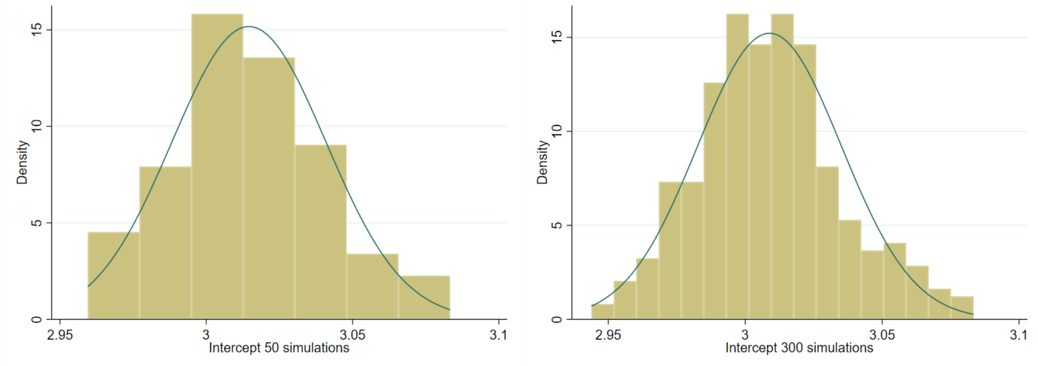

Figure 14: Bias in the coefficients due to previous bootstrap method

A second plausible source of bias is through the simulated η^*. Because the method presented in section 4 selects clusters randomly with replacement, it is likely that some areas are selected more than once and others are not selected. As a consequence of this, an area selected more than once may underestimate $\widehat{\text{var}}\left[\hat{\eta_c}\right]$ because there would artificially be more observations in that area. When an area is not selected then $\eta_c^* \sim N\left(0, \hat{\sigma}_\eta^2\right)$,[29] and because $\bar{X}\beta^*$ is likely biased due to the issue noted in figure 14, then the resulting estimate for this area will deviate considerably from the truth.

Figure 15 shows how the bias builds up.[30] Figure 12, on the left, shows the bias for the average of $e^{(X\beta^*)}$ for the areas; the figure on the right adds $\eta^*$and thus shows $e^{(X\beta^*+\eta_c^*)}$ for all areas. The linear fit already shows a consideable bias, this is also aligned to the coefficient observed in figure 14, where the value seems to be upward biased. Once the location effect is added, things get considerably worse for most areas.

[29]Note that $\hat{\sigma}_\eta^2$ is not the one estimated from the sample, it comes from a bootstrap sampling of the data.
[30]This simulation is executed under the same scenario as that of subsection 8.3, except that 5,000 populations are executed instead of 10,000

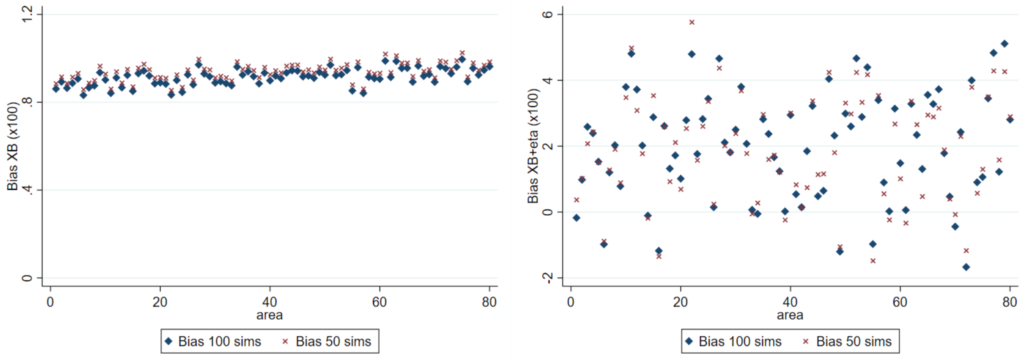

Figure 15: Bias and MSE over simulated populations due to previous bootstrap methods

9 Conclusions

Since the turn of the 21st century, the World Bank has been obtaining small area estimates of poverty using the ELL approach. A large reason for the method's popularity was the implementation of software that made executing the method much easier. 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.[31]

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.

A considerable aspect of small area estimation is how parameter estimates from an assumed population model are applied to census data to obtain small area estimates. As noted in this document, the traditional ELL procedure is aligned to the approach used in multiple imputation. However, multiple imputation's goal is not prediction. Multiple imputation's goal is to obtain unbiased parameters and their appropriate standard errors under regression analysis, and 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 considerable revision to the implementation of ELL (2002 and 2003) and the updates by Van der Weide (2014) for small area estimation.[32] The results show that the gains due to the change in methods are considerable; this is the case for mean squared error (MSE) and bias. The update should represent a considerable improvement in the quality of the SAE obtained under the World Bank's agenda. In the document, it is shown that the original ELL method tends to align with the national poverty estimates and shows little alignment to estimates at the area level, yet on the whole is nearly unbiased. This also serves as a note of caution for out of sample prediction under EB methods, since these will be similar to those of the original ELL and likely will 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 H3-EB update by Van der Weide (2014) and as implemented by Nguyen et al. (2018) to the World Bank's toolkit. The revised H3-CensusEB method is aligned with Molina and Rao's (2010) method, which shows a level of bias that is

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

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

several orders of magnitude smaller than before the revision. Furthermore, the MSE for the H3-CensusEB 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

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.

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.

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. 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).

You, Y. and Rao, J. (2002). A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. *Canadian Journal of Statistics*, 30(3):431–439.

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