Supplementary Material for : A note on Simultaneous Confidence Intervals for Direct, Indirect and Synthetic Estimators

Christophe Quentin Valvason and Stefan Sperlich
University of Geneva, Geneva School of Economics and Management
Bd du Pont d'Arve 40, 1211 Genève 4, Switzerland
{christophe.valvason , stefan.sperlich}@unige.ch

March 15, 2024

Abstract

This Supplement contains three Sections: First we discuss some miscellaneous issues like the notation in survey sampling, sample size requirements for approximation by normality, and bootstrap procedures for finite populations. This is followed by a Section that revisits the considered estimators with their properties and variance estimators employed in the main article. Finally, the third Section presents a simulation study for complex indicators in domains with moderate and small samples. We see there that again the max-type statistic approach yields a dramatic improvement compared to the classical ones. However, for the considered sample sizes it would still be recommendable to switch to mixed model-based methods, i.e. the so-called small area estimation techniques.

Keywords: domain estimation; simultaneous confidence intervals; uniform inference; comparative statistics

1 Miscellaneous: Notation, Normal approximation, and Bootstrap for finite populations

1.1 Notations in Survey Sampling

This Section provides a brief introduction to the used notation, and to some principles of design-based estimation; for more details you may consult (Cochran, 1977; Fuller, 2009; Tillé, 2020). Recall that we are interested in estimating θ_d, depending on observations y_k, for all domains d in a finite population $U = \{1, \dots, k, \dots, N\}$ with N subjects. Population U is partitioned into D domains, say U_d, from which we select a random sample S. Direct

estimators for θ_d only use the observations made in domain d, i.e. subsamples denoted by $s_d = s \bigcap U_d$ with sampling size n_d, where s is the realized sample of size $n = \sum_{d=1}^{D} n_d$, i.e. a specific realization of S.

A sampling design $p(\cdot)$ is a probability distribution on $\mathcal{S} = \{s : s \subset U\}$ such that $p(s) \geqslant 0$ for all $s \in \mathcal{S}$ and $\sum_{s \in \mathcal{S}} p(s) = 1$. The following notation and assumption is somewhat unusual in other fields of statistics: While sample S is considered to be a multivariate random variable, each y_k is constant. Instead, for each unit k in the population, one denotes by I_k a random variable indicating its inclusion in the sample, and by π_k its first order inclusion probability, i.e. the probability that unit k is selected:

$$\pi_k = \mathbb{P}(k \in S) = \mathbb{E}(I_k) = \sum_{s \in \mathcal{S}: s \ni k} p(s).$$

Similarly, one denotes by $\pi_{k\ell}$ the probability that units k and ℓ are selected together:

$$\pi_{k\ell} = \mathbb{P}(k \in S \text{ and } \ell \in S) = \mathbb{E}(I_k I_\ell) = \sum_{s \in \mathcal{S}: s \ni k,\ell} p(s),$$

where $\pi_{k\ell}$ is called the second order (inclusion) probability with $\pi_{kk} = \pi_k$. An important quantity for variance estimation, see Section 2, is the covariance between two indicator variables I_k and I_ℓ:

$$\Delta_{k\ell} := \operatorname{Cov}(I_k, I_\ell) = \mathbb{E}(I_k I_\ell) - \mathbb{E}(I_k)\,\mathbb{E}(I_\ell) = \pi_{k\ell} - \pi_k \pi_\ell \ , \quad \Delta_{kk} := \pi_k - \pi_k^2 = \pi_k(1 - \pi_k).$$

In the main article the considered parameter θ_d is the total of all y_k in domain d, denoted as Y_d with its mean $\bar{Y}_d$.

1.2 Sample Size for Normality Approximation

In the main article, a major reason for the failure of classical approaches for simultaneous inference methods like Bonferroni and Šidák correction is the poor distributional approximation by normality or Student's t. Clearly, there must be a relation between the distribution of the original variable Y in the population, the complexity of the parameter and its estimator $\hat{\theta}_d$, and finally the sample size n_d. The question we briefly discuss in this subsection is: What does the literature say about the sample size to rely on a normality approximation?

The problem is that practical rules will ignore that $\hat{\theta}_d$ prevents us from linking the distribution of Y directly to n_d. Yet, like in the main article we can limit our considerations here to parameters that are just linear combinations of the y_k, like for instance the total Y_d and the mean $\bar{Y}_d$. Consequently, it should be sufficient to look at the literature discussing the estimation of the mean. For such a problem, different authors have proposed different rules to determine an appropriate sample size for making the normal approximation work. We only sketch here a few of them. Not surprisingly, the all relate the sample size to higher orders of the underlying distribution.

More specifically, Cochran (1977) recommends to apply the rule $n_d > 25 G_1^2 \ \forall d$, where

$$G_1 := \frac{1}{N \sigma_y^3} \sum_{k \in U} (y_k - \bar{y})^3, \text{ (or where appropriate, for } N_d, U_d, \text{ etc.)}$$

is a measure of skewness, and σ_y the standard deviation of $\{y_1, \ldots, y_N\}$. It is said that this rule could be applied whenever the sample distribution is positively skewed.

Similarly, Dalén (1986) modifies this rule concentrating on the α-quantiles, suggesting s a consequence to take $n_d > K_\alpha G_2^2$, where

$$G_2 := \frac{1}{N\sigma_y^3} \sum_{k \in U} |y_k - \bar{y}|^3$$

is the standardized absolute third moment of the population, and K_α a constant depending on α, see his article for more details.

Sugden et al. (2000) propose other refinements of Cochran (1977) that are independent of the α quantiles. One of these refinements refinement is $n_d > 28 + 25G_1^2$, i.e. to add the constant 28 which results from some higher order approximations.

Recalling the problem considered in our main article, any of these rules if employed, would have to be applied to all domains. Unfortunately, we would need then to find for each domain estimators for G_1 or G_2, potentially together with an adequate K_α. While already this is typically unfeasible, in some domains sample sizes or even population sizes may be smaller than the required (or say, suggested) sample size(s). All together, this may render these methods little helpful in practice.

1.3 Bootstrap for Finite Population

As explained in the main article, we cannot use any of the bootstrap methods proposed in the papers Reluga et al. (2023a,b), since they were made for mixed effects regression models. In contrast, we need a bootstrap method that directly resamples from the y_k and accounts for the fact that we consider finite populations.

Admittedly, there exist various suggestions for bootstrapping in finite populations, although in different contexts, and developed for other purposes. For general reviews on bootstrap methods for finite populations see Gross (1980); Booth et al. (1994). Further, Rao et al. (1992) considered procedures for complex survey and multi-stage sampling designs that rescaled the sampling weights, while Antal and Tillé (2011) proposed a so-called direct bootstrap for complex designs.

In our main article we follow Chauvet (2007): he recommends that for a low sampling rate $f = n/N$, with n being the overall sample size and N the population size, one should use a method derived in Gross (1980); for larger sampling rates one should use the method of Booth et al. (1994). Consequently, in our main article we apply the former for sampling rate $f = 1/6$ and the latter for $f = 2/3$. Where necessary, these methods are adapted to each particular estimator considered; keep in mind that we are studying different estimators for the estimation of the domain parameters θ_d, $d = 1, ..., D$.

The bootstrap procedure works generally as follows: The algorithms for finite populations aim to construct a pseudo-population U^* from a sample S taken from population $U = \bigcup_{d=1}^{D} U_d$, or more precisely from its realisation s. From U^* one draws B samples S^b, $b = 1, \ldots, B$ with the same sampling design as one had for S. Then a bootstrap analogue of

$\hat{\theta}$, i.e. $\hat{\theta}_b^*$, is computed from each sample S^b. From the bootstrap estimates one obtains an empirical bootstrap distribution $F_{\hat{\theta}}^*$ of $\hat{\theta}$.

The specific resampling procedure is one of the following two, designed for small, respectively large sampling rates f: Our procedures construct pseudo-populations U^* by replicating the sampled data. Gross (1980) proposes to replicate each observation in S exactly $p = [N/n]$ times to obtain U^*, whereas Booth et al. (1994) proposes to replicate each observation in S only $p = [N/n - 0.5]$ times, and to add a random subsample of S of size $r = N - np$ to obtain U^*. Clearly, the former bootstrap population is of size np, whereas the latter is of size N. For the statistical properties of these, or to learn more about alternative methods, you may consult Chauvet (2007).

2 Revisiting Direct and Indirect Domain Estimators and Their Properties

2.1 Direct Estimators

Even though the definitions of the considered estimators are already given in the main article, for the ease of reading and presentation we list them here again, this time together with their statistical properties.

Recall that the Horvitz-Thompson estimator ($\hat{\theta}_d$) of the domain total $\theta_d = Y_d$ is

$$\hat{Y}_d^{ht} := \sum_{k \in s_d} \frac{y_k}{\pi_k} \tag{2.1}$$

and similarly $\hat{\bar{Y}}_d^{ht} = N_d^{-1} \hat{Y}_d^{ht}$ for the mean with N_d the domain size. Both are design unbiased, $\mathbb{E}(\hat{Y}_d^{ht}) = Y_d$. Their variance and covariance(s) can be estimated by

$$\widehat{\mathrm{Var}}\big[\hat{Y}_d^{ht}\big] = \sum_{k \in s_d} \sum_{\ell \in s_d} \frac{\Delta_{k\ell} y_k y_\ell}{\pi_{k\ell} \pi_k \pi_\ell}\,, \qquad \widehat{\mathrm{Cov}}\big[\hat{Y}_d^{ht}, \hat{Y}_{d'}^{ht}\big] = \sum_{k \in s_d} \sum_{\ell \in s_{d'}} \frac{\Delta_{k\ell} y_k y_\ell}{\pi_{k\ell} \pi_k \pi_\ell}. \tag{2.2}$$

When auxiliary variables $\boldsymbol{x}_k \in \mathbb{R}^p,\ p \geqslant 1$ are available, the direct-GREG estimator assisted by the regression model with errors ϵ_{kd} is defined as

$$y_{kd} = \boldsymbol{x}_{kd}^\top \boldsymbol{\beta}_d + \epsilon_{kd}, \qquad \mathrm{Var}\,(\epsilon_{kd}) = \sigma_{kd}^2, \tag{2.3}$$

where $\boldsymbol{\beta}_d$ is a parameter vector associated with domain U_d, typically estimated by

$$\hat{\boldsymbol{\beta}}_d = \left(\sum_{k \in s_d} \frac{\boldsymbol{x}_k \boldsymbol{x}_k^\top}{\pi_k} \right)^{-1} \left(\sum_{k \in s_d} \frac{\boldsymbol{x}_k y_k}{\pi_k} \right). \tag{2.4}$$

Then, for $\hat{y}_{kd} = \boldsymbol{x}_{kd}^\top \hat{\boldsymbol{\beta}}_d$ and residuals $e_{kd} = y_{kd} - \hat{y}_{kd}$, the direct-GREG is

$$\hat{Y}_d^{dgreg} := \sum_{k \in U_d} \hat{y}_k + \sum_{k \in s_d} \frac{e_k}{\pi_k} \;=\; \hat{Y}_d^{ht} + (\boldsymbol{X}_d - \hat{\boldsymbol{X}}_d^{ht})^\top \hat{\boldsymbol{\beta}}_d, \tag{2.5}$$

where $\boldsymbol{X}_d$ is the vector of true domain totals for each auxiliary variable, and $\hat{\boldsymbol{X}}_d^{ht}$ its Horvitz-Thompson estimator. Note that it is not necessary to know $\boldsymbol{x}_k$ for each unit in the population; the (true) totals and their values for the observed sample are sufficient. This is certainly due to the linear form of (2.3), it cannot be gerenalised that easily to more complex models.

The variance of the direct-GREG estimator can be approximated by Taylor's linearization which in turn can be estimated by

$$\widehat{\mathrm{Var}}\left[\hat{Y}_d^{dgreg}\right] = \sum_{k\in s_d}\sum_{\ell\in s_d}\frac{e_k e_\ell \Delta_{k\ell}}{\pi_k \pi_\ell \pi_{k\ell}}. \tag{2.6}$$

An alternative variance estimator was given in Särndal et al. (1989) who used the *g-weights*,

$$g_k := I_{dk} + I_{dk}(\boldsymbol{X}_d - \hat{\boldsymbol{X}}_d^{ht})^\top \left(\sum_{k\in s_d}\frac{\boldsymbol{x}_k \boldsymbol{x}_k^\top}{\pi_k}\right)^{-1} \boldsymbol{x}_k,$$

where I_{dk} denotes the domain indicator. The resulting estimator is

$$\widehat{\mathrm{Var}}_2\left[\hat{Y}_d^{dgreg}\right] = \sum_{k\in s_d}\sum_{\ell\in s_d}\frac{g_k e_k g_\ell e_\ell \Delta_{k\ell}}{\pi_k \pi_\ell \pi_{k\ell}}. \tag{2.7}$$

In practice the latter (alternative) is often preferable, see Särndal et al. (1989) for details.

2.2 Indirect Estimators

Indirect estimators borrow strength from other domains (Ghosh and Rao, 1994). In order to do so, we need to introduce the notion of groups that are different from domains. These groups are subsets $U_g, g = 1, \ldots, G$, not necessarily of interest, that partition the population, i.e. $\bigcup_{g=1}^G U_g = U$, with $U_g \bigcap U_{g'} = \varnothing$ for $g \neq g'$. Typically, G is small as in our simulations in the main article. Obviously, an estimator of a group's mean $\bar{Y}_g$ is given by

$$\hat{\bar{Y}}_g := \frac{1}{\hat{N}_g^{ht}}\sum_{k\in s_g}\frac{y_k}{\pi_k} = \frac{\hat{Y}_g^{ht}}{\hat{N}_g^{ht}} \quad \text{where} \quad \hat{N}_g^{ht} := \sum_{k\in s_g}\frac{1}{\pi_k}, \tag{2.8}$$

also known as the Hajek estimator of the mean (Hájek, 1971). Then, a synthetic estimator for the total is given by the combination

$$\hat{Y}_d^{synth} := \sum_{g=1}^G N_{dg}\hat{\bar{Y}}_g, \tag{2.9}$$

where N_{dg} are the crossed population sizes between domain d and group g. The expectation is $\mathbb{E}(\hat{Y}_d^{synth}) \approx \sum_{g=1}^G N_{dg}\bar{Y}_g$ with bias $\mathrm{B}(\hat{Y}_d^{synth}) \approx \sum_{g=1}^G N_{dg}(\bar{Y}_g - \bar{Y}_{dg})$ which is zero for $\bar{Y}_g = \bar{Y}_{dg}$. Again, its variance can be estimated by linearization,

$$\widehat{\mathrm{Var}}\big[\hat{Y}_d^{synth}\big] = \sum_{g=1}^G\sum_{g'=1}^G N_{dg}N_{dg'}\ \widehat{\mathrm{Cov}}\big[\hat{\bar{Y}}_g, \hat{\bar{Y}}_{g'}\big], \tag{2.10}$$

where the covariance between two Hajek estimators can be estimated by

$$\frac{\widehat{\text{Cov}}(\hat{Y}_g^{ht}, \hat{Y}_{g'}^{ht}) - \hat{\bar{Y}}_{g'}\widehat{\text{Cov}}(\hat{Y}_g^{ht}, \hat{N}_{g'}^{ht}) - \hat{\bar{Y}}_g\widehat{\text{Cov}}(\hat{Y}_{g'}^{ht}, \hat{N}_g^{ht}) + \hat{\bar{Y}}_g\hat{\bar{Y}}_{g'}\widehat{\text{Cov}}(\hat{N}_g^{ht}, \hat{N}_{g'}^{ht})}{\hat{N}_g^{ht}\hat{N}_{g'}^{ht}}. \tag{2.11}$$

A modification of the synthetic estimator is the so-called post-stratified version. Here, instead of using the group's mean, one uses the means in subsets $U_{dg} = U_d \bigcap U_g$,

$$\hat{Y}_d^{psts} := \sum_{g=1}^{G} N_{dg}\hat{\bar{Y}}_{dg}, \qquad \text{where} \quad \hat{\bar{Y}}_{dg} := \frac{1}{\hat{N}_{dg}^{ht}} \sum_{k \in s_{dg}} \frac{y_k}{\pi_k} \qquad \text{with} \quad \hat{N}_{dg}^{ht} := \sum_{k \in s_{dg}} \frac{1}{\pi_k}. \tag{2.12}$$

The post-stratified synthetic estimator is generally unbiased, see Morales et al. (2021) for details, and performs better than the basic synthetic estimator when y_k has a large variation within groups.

A variance estimate is

$$\widehat{\text{Var}}\big[\hat{Y}_d^{psts}\big] = \sum_{g=1}^{G}\sum_{g'=1}^{G} N_{dg}N_{dg'}\widehat{\text{Cov}}\big[\hat{\bar{Y}}_{dg}, \hat{\bar{Y}}_{dg'}\big], \tag{2.13}$$

where the covariance can be estimated by

$$\frac{\widehat{\text{Cov}}(\hat{Y}_{dg}^{ht}, \hat{Y}_{dg'}^{ht}) - \hat{\bar{Y}}_{dg'}\widehat{\text{Cov}}(\hat{Y}_{dg}^{ht}, \hat{N}_{dg'}^{ht}) - \hat{\bar{Y}}_{dg}\widehat{\text{Cov}}(\hat{Y}_{dg'}^{ht}, \hat{N}_{dg}^{ht}) + \hat{\bar{Y}}_{dg}\hat{\bar{Y}}_{dg'}\widehat{\text{Cov}}(\hat{N}_{dg}^{ht}, \hat{N}_{dg'}^{ht})}{\hat{N}_{dg}^{ht}\hat{N}_{dg'}^{ht}}. \tag{2.14}$$

Another alternative is the so-called indirect-GREG estimator (still for the total). It can also be considered as an indirect estimator under regression $y_k = \boldsymbol{x}_k^\top\boldsymbol{\beta} + \epsilon_k$, i.e. with a common parameter vector β for the population instead of one for each domain (Lehtonen and Veijanen, 2009). Similarly to the direct-GREG estimator, the regression parameter can be estimated by

$$\hat{\boldsymbol{\beta}} = \left(\sum_{k \in s} \frac{\boldsymbol{x}_k\boldsymbol{x}_k^\top}{\pi_k}\right)^{-1} \sum_{k \in s} \frac{\boldsymbol{x}_k y_k}{\pi_k}.$$

Then the indirect-GREG estimator of the total is for $\hat{y}_k = \boldsymbol{x}_k^\top\hat{\boldsymbol{\beta}}$ and $e_k = y_k - \hat{y}_k$ given as

$$\hat{Y}_d^{igreg} := \sum_{k \in U_d} \hat{y}_k + \sum_{k \in s_d} \frac{e_k}{\pi_k} = \hat{Y}_d^{ht} + \left(\boldsymbol{X}_d - \hat{\boldsymbol{X}}_d^{ht}\right)^\top \hat{\boldsymbol{\beta}} = \sum_{k \in s} \frac{g_{dk}y_k}{\pi_k}, \tag{2.15}$$

where $g_{dk} := I_{dk} + (\boldsymbol{X}_d^{ht} - \hat{\boldsymbol{X}}_d^{ht})^\top(\sum_{k \in s} \boldsymbol{x}_k\boldsymbol{x}_k^\top/\pi_k)^{-1}$.

Variance estimators for this estimator are similar to the ones for the direct-GREG estimator, i.e.

$$\widehat{\text{Var}}\left[\hat{Y}_d^{igreg}\right] = \sum_{k \in s}\sum_{\ell \in s} \frac{e_k e_\ell \Delta_{k\ell}}{\pi_k\pi_\ell\pi_{k\ell}}, \qquad \widehat{\text{Var}}_2\left[\hat{Y}_d^{igreg}\right] = \sum_{k \in s}\sum_{\ell \in s} \frac{g_k e_k g_\ell e_\ell \Delta_{k\ell}}{\pi_k\pi_\ell\pi_{k\ell}},$$

where the sums run over the entire sample (Särndal et al., 1992). If N_d is known, (2.15) can be simplified (Lehtonen and Veijanen, 2009) to

$$\hat{Y}_d^{igreg} := \sum_{k \in U_d} \hat{y}_k + \frac{N_d}{\hat{N}_d^{ht}} \sum_{k \in s_d} \frac{e_k}{\pi_k}. \tag{2.16}$$

For more details about these estimators consult Morales et al. (2021); Rao and Molina (2015); Särndal et al. (1992); Lehtonen and Veijanen (2009).

3 A Simulation Study for Poverty Indicators

3.1 Indicator and Simulation Design

In the main article we conducted simulations for estimating totals, which forms the basis of survey sampling theory. While our findings should apply to all kind of linear functions of them, often more complex estimators are considered in practice.

Here our interest is to see if our results carry over to those, and if so, to what extent. Quite frequently used complex indicators are those related to poverty studies (Pratesi, 2016). Although there exist many proposals in the literature, we restrict here to a generalized measure of poverty of Foster et al. (1984) (FGT):

$$F(\lambda, \gamma) := \frac{1}{N} \sum_{k \in U} \left(\frac{\gamma - y_k}{\gamma} \right)^{\lambda} \mathbb{1}_{y_k \leqslant \gamma}, \tag{3.1}$$

where γ a given poverty line, y_k a measure of income for individual k, and $\lambda \geqslant 0$ a parameter of sensitivity: $\lambda = 0$ gives the Head Count Ratio (HCR), which is the percentage of individuals under the poverty line, and $\lambda = 1$ gives the Poverty Gap (PG) measuring how far the poor are from this line. Both, the HCR and PG are popular measures, as well as the methods we study in the following, are applicable for any value $\lambda > 0$. If we replace the indicator functions by a dichotomous variable $\chi_k = \mathbb{1}_{y_k \leqslant \gamma}$ it is easier to see that actually even the HCR could be interpreted as a (special case of a) mean estimator. In this sense, we have already examined this case in the main article. Consequently, in the following we concentrate on the PG version of the FGT.

A direct domain estimator for the PG is obviously

$$\hat{F}_d^{ht}(1, \gamma) := \frac{1}{N_d} \sum_{k \in s_d} \frac{1}{\pi_k} \left(\frac{\gamma - y_{dk}}{\gamma} \right) \mathbb{1}_{y_{dk} \leqslant \gamma} = \hat{PG}_d^{ht}(\gamma). \tag{3.2}$$

As it is common for complex parameters, an analytic expression of the variance is hard to obtain. In practice one estimates the variance of PG estimators by linearization or bootstrap. In the former case, one hopes that, among other assumptions (Deville, 1999), the variance of a linearized variable $\tilde{z}_k$ approximates well the variance of the original estimator. In case of the above PG estimators, one could use for instance

$$\tilde{z}_{kd} = \frac{1}{N_d} \left[\left(\frac{\gamma - y_{kd}}{\gamma} \right) \mathbb{1}_{y_{dk} \leqslant \gamma} - \frac{\hat{PG}_d^{ht}(\gamma)}{N_d} \right].$$

Then the variance of interest is approximated by the variance of $\sum_{k\in s_d} \tilde{z}_{kd}/\pi_k$; for more details see Verma and Betti (2011). However, today it has become much more popular to use a bootstrap approximation of the variance. As we will use bootstrap for approximating the critical values of the max-type statistic anyway, it is more natural to use it also here.

We next describe the data generating process applied for our simulations. A common variable for determining any measure of poverty is the (logarithm of) wage of individuals or households since the alternative 'expenditures' is typically much harder to measure. To generate some income observations, consider the popular *Mincer earnings function* (Mincer, 1958, 1974) with a synthetic population of individuals with wages:

$$\log w_{kd} = \log w_0 + \beta_1 ed_{kd} + \beta_2 ex_{kd} + \beta_3 ex_{kd}^2 + u_d + \epsilon_{kd}, \quad k = 1, \dots, N_d, d = 1, \dots, D, \quad (3.3)$$

where w_0 indicates wage without any experience and education, ed the years of education and ex labour experience. Following the literature, variable ed is generated by a discrete uniform distribution with parameters $\min = 0, \max = 14$, and ex by a Weibull distribution with parameters $\text{shape} = 3, \text{scale} = 40$. As before, u and ϵ follow mean-zero normal distributions with σ_u^2, σ_ϵ^2. We do not vary the parameters since we have seen that those have only little impact on the coverage probability of our simultaneous confidence intervals (SCI). For all simulations we set $\gamma = \log 4000$, $w_0 = 2000$, $\sigma_u = 0.5$, $\sigma_\epsilon = 0.2$, $\beta_1 = 0.07$, $\beta_2 = 0.05$ and $\beta_3 = -0.001$. Those parameters loosely reproduce the distribution of wages in Switzerland (see Federal Statistical Office, 2022b,a). We generate data for $D = \{10, 100\}$ domains with population sizes $N = \{300, 3000\}$ respectively. As before, consider sampling rates $f = \{1/6, 2/3\}$ giving samples of sizes $n = \{(50, 200), (500, 2000)\}$, and employ the bootstrap algorithms described in Gross (1980) and Booth et al. (1994), cf. Section 1, the former for both f, the latter $f = 2/3$. The nominal error level α is always set to 0.05 as it was the case in the main article.

3.2 Simulation Results

Like we did for the totals, we compare the three methods (Bonferroni and Šidák correction as well as the max-type statistic) used to construct SCI. We hope to see from our simulations, in which situations they achieve or fail to deliver an appropriate coverage for the true parameters (i.e. our generated poverty rates). An important difference to the main article is that here we have only one PG estimator at hand. The results shown are in Table 3.1. Recall that bootstrap is use for the Bonferroni and Šidák correction when estimating the standard errors.

For this complex estimator and the given sample sizes n_d, none of the proposed methods is able to deliver SCI with the desired nominal coverage. For the low sampling rate ($f = 1/6$) all methods have an uniform coverage close or equal to 0. For a high sampling rate ($f = 2/3$) the situation is much better for the max-type approach but clearly not satisfactory, and of no use for large D. While Bonferroni and Šidák corrections fail completely, probably because the distribution of the estimator is too far from normality, cf. Figure 3.1, also the max-type method exhibits serious problems now, though its coverage is by far higher than the ones of the other methods.

		D = 10			D = 100		
f	bootstrap method	Bonferroni	Šidák	Max-type	Bonferroni	Šidák	Max-type
1/6	Gross	0.019	0.019	0.079	0	0	0
2/3	Gross	0.458	0.458	0.716	0.001	0.001	0.061
2/3	Booth	0.506	0.506	0.735	0.004	0.004	0.052

Table 3.1: Results of the simulations for the PG estimator. Each cell is the estimated joint coverage of our 95% SCI.

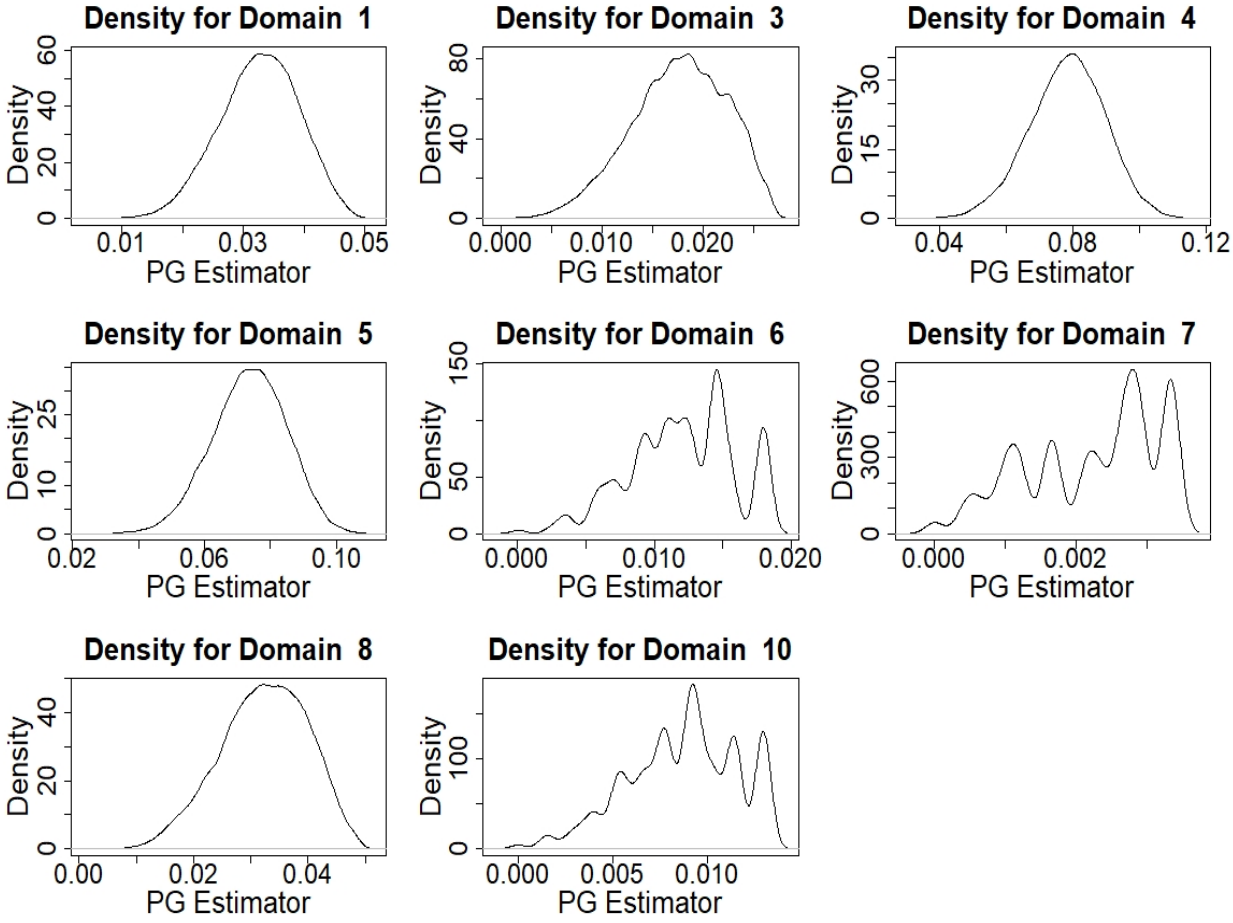

Figure 3.1: Kernel densities of the PG estimators for $N = 300, f = 2/3$ and $D = 10$. The two domains not shown gave even discrete probability distributions. The values are divided by 10^9 for presentation purpose.

In Figure 3.1 you also see that the PG's are quite small quantities. One may argue then correctly, that to obtain reliable quantiles for them, large domain sample sizes are required. We therefore changed the distribution producing much larger PG, and effectively, the results for those income distributions gave better results for the max-type statistic but not so for the others. As you always can simulate populations and situations in which your method works very well, let us continue the discussion along our first choice.

A larger domain sample size can be obtained through larger domain population sizes and high sampling rates. Therefore we conducted simulations increasing the populations size to $N = 5000$, and examining the cases for $D = 10$ and $D = 50$, with results in Table 3.2.

		$D = 10$			$D = 50$		
f	bootstrap method	Bonferroni	Šidák	Max-type	Bonferroni	Šidák	Max-type
1/6	Gross	0.68	0.68	0.92	0	0	0.01
2/3	Booth	0.957	0.957	0.984	0.27	0.27	0.65

Table 3.2: Additional results for the PG estimator; population size is $N = 5000$. Each cell is the estimated uniform coverage probability of our 95% SCI.

Now the results look much better, though mainly for the larger sampling rate. There is even some hope that for a sufficiently large population with high sampling rate, it also works for $D = 50$. But it is hard to say what the necessary N and f are. For $D = 10$ we see that the two simple, classical methods work only for large sampling rates. Then they are even less conservative than the max-type one, which is expected since the max-type method accounts for the worst case (Reluga et al., 2023b).

Supposing you have a much larger population size N, even with an unlimited budget that allows to take samples of size n close to N, the desired coverage of an SCI is not guaranteed if D is big. Certainly, only the number of domains simultaneously studied matters; you may have 100 domains but study only $D = 10$ of them jointly, then there is some hope for valid SCI. That is, practitioners have to be cautious when computing SCI for a complex problem with direct estimators. Obviously, for uniform or simultaneous inference you either have to drop domains with small samples or to switch to model-based methods (Reluga et al., 2023b,a). It is true that this issue is where SAE is all about, but comparative statistics seems to already call for model-based or -assisted estimation when for individual domain analysis samples are still sufficiently large.

References

Antal, E. and Y. Tillé (2011). A direct bootstrap method for complex sampling designs from a finite population. *Journal of the American Statistical Association 494*, 534–543.

Booth, J. G., R. W. Butler, and P. Hall (1994). Bootstrap methods for finite populations. *Journal of the American Statistical Association 89*(428), 1282–1289.

Chauvet, G. (2007, dec). *Méthodes de Bootstrap en Population Finie*. Ph. D. thesis, Université de Rennes 2.

Cochran, W. G. (1977). *Sampling Techniques* (third ed.). John Wiley & Sons.

Dalén, J. (1986). Sampling from finite populations : Actual coverage probabilities for confidence intervals on the population mean. *Journal of Official Statistics 2*(1), 13–24.

Deville, J.-C. (1999, dec). Variance estimation for complex statistics and estimators: Linearization and residual techniques. *Survey Methodology 25*(2), 193–203.

Federal Statistical Office (2022a, nov). Poverty. https://www.bfs.admin.ch/bfs/en/home/statistics/economic-social-situation-population/economic-and-social-situation-of-the-population/poverty-and-material-deprivation/poverty.html.

Federal Statistical Office (2022b, nov). Work and income. https://www.bfs.admin.ch/bfs/en/home/statistics/work-income.html.

Foster, J., J. Greer, and E. Thorbecke (1984, may). A class of decomposable poverty measures. *Econometrica 52*(3), 761–766.

Fuller, W. A. (2009). *Sampling Statistics*. John Wiley & Sons.

Ghosh, M. and J. Rao (1994). Small area estimation : An appraisal. *Statistical Science 9*(1), 55–93.

Gross, S. (1980, jan). Median estimation in sample surveys. *Proceedings of the Section on Survey Research Methods*, 181–184.

Hájek, J. (1971). Discussion of an essay on the logical foundations of survey sampling, part ones by d. basu. In V. P. Godambe and D. A. Sprott (Eds.), *Foundations of Statistical Inference*. Toronto, Holt, Rinehart and Winston of Canada.

Lehtonen, R. and A. Veijanen (2009). *Design-based Methods of Estimation for Domains and Small Areas*, Volume 29B of *Handbook of Statistics*, Chapter 31, pp. 219–249. Elsevier B.V.

Mincer, J. (1958, aug). Investment in human capital and personal income distribution. *The Journal of Political Economy 66*(4), 281–302.

Mincer, J. (1974). *Schooling, Experience and Earnings*. Human Behavior and Social Institutions. National Bureau of Economic Research.

Morales, D., M. D. E. Lefler, A. Pérez, and T. Hobza (2021). *A Course on Small Area Estimation and Mixed Models*. Statistics for Social and Behavioral Sciences. Springer.

Pratesi, M. (2016). *Analysis of poverty data by small area estimation*. Wiley Series in Survey Methodology. John Wiley.

Rao, J. and I. Molina (2015). *Small Area Estimation* (second ed.). Wiley series in survey methodology. John Wiley & Sons, Inc.

Rao, J., C. Wu, and K. Yue (1992, dec). Some recent work on resampling methods for complex surveys. *Survey Methodology 18*(2), 209–217.

Reluga, K., K. Lombardía, and S. Sperlich (2023a). Simultaneous inference for empirical best predictors with a poverty study in small areas. *Journal of the American Statistical Association 118*(541), 583–595.

Reluga, K., K. Lombardía, and S. Sperlich (2023b). Simultaneous inference for linear mixed model parameters with an application to small area estimation. *International Statistical Review 91*(2), 193–217.

Särndal, C.-E., B. Swensson, and J. Wretman (1992). *Model Assisted Survey Sampling* (first ed.). Springer Series in Statistics. Springer-Verlag New York, Inc.

Särndal, C.-E., B. Swensson, and J. H. Wretman (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. *Biometrika 76*(3), 527–537.

Sugden, R. A., T. M. F. Smith, and R. P. Jones (2000). Cochran's rule for simple random sampling. *Journal of the Royal Statistical Society : Series B 62*(4), 787–793.

Tillé, Y. (2020). *Sampling and Estimation from Finite Populations.* Wiley Series in Survey Methodology. John Wiley & Sons.

Verma, V. and G. Betti (2011, aug). Taylor linearization sampling errors and design effects for poverty measures and other complex statistics. *Journal of Applied Statistics 38*(8), 1549–1576.