

1 Introduction

The EU Statistics on Income and Living Conditions (EU-SILC), launched 2004, represents a European standardized survey to generate comparative measures of poverty and social exclusion among the EU Member States. The survey is conducted annually in each country with a rotating panel design of 4 waves (see Verma and Betti, 2006). In combination with a harmonized survey a set of common indicators, the so called Laeken indicators, was adopted for the countries of the EU (see Atkinson et al., 2002).

Due to sampling design and sample size the EU-SILC delivers qualitatively high well-being indicators at national or NUTS1 level, but usually lacks the capability to do the same for regional indicators, on for example NUTS2 or NUTS3 level. Due to the need of regional indicators for policy makers many methods have already been developed that aim to calculate statistical significant indicators on lower NUTS levels. (Betti et al., 2012; Qinghua and Peter, 2009) Many of these methods are based on models for small area estimation, pooling the data for consecutive years or applying a jack knife procedure for consecutive years and taking the mean over the jackknife replicates. for variance estimation. Others also use administrative data to impute the variable of interest, given a specialized, model onto a large data set, like the population census. Especially the last was has the downside of being very country specific and the risk of delivering estimates with high variance. Furthermore the use of administrative data also brings with it the problem of timeliness of the information in the administrative data.

In this work we present a method for estimating statistically significant estimates on lower NUTS levels using multiple years of EU-SILC in combination with bootstrapping techniques. In contrast to jack knife replicates, the bootstrap replicates yield consistent estimates for variance even for non-smooth estimators like the median. Due to the nature of our method it can easily be applied to EU-SILC data of any county given the constraint that EU-SILC data for at least 3 consecutive years is available. Furthermore the households in EU-SILC must be linked over the years through a household ID to ensure the applicability of our method.

2 Methodology

In the following we present the used methodology that is applied on multiple consecutive years of EU-SILC data for one country. The methodology contains the following steps, in this order

1. Draw B bootstrap replicates from EU-SILC data for each year y_t , $t = 1, \dots, n_y$ separately. Since EU-SILC has a rotating panel design the bootstrap replicate of a household is carried forward through the years. That is, the bootstrap replicate of a household in the follow-up years is set equal to the bootstrap replicate of the same household when it first enters EU-SILC.
2. Multiply each set of bootstrap replicates by the sampling weights to obtain uncalibrated bootstrap weights and calibrate each of the uncalibrated bootstrap weights

using iterative proportional fitting.

3. Estimate the point estimate of interest θ , for each year and each calibrated bootstrap weight to obtain $\tilde{\theta}^{(i,y_t)}$, $t = 1, \dots, n_y$, $i = 1, \dots, B$. For fixed y_t apply a filter with equal weights for each i on $\tilde{\theta}^{(i,y^*)}$, $y^* \in \{y_{t-1}, y_t, y_{t+1}\}$, to obtain $\tilde{\theta}^{(i,y_t)}$. Estimate the variance of θ using the distribution of $\tilde{\theta}^{(i,y_t)}$.

2.1 Bootstrapping

Bootstrapping has long been around and used widely to estimate confidence intervals and standard errors of point estimates. (Efron, 1979) Given a random sample (X_1, \dots, X_n) drawn from an unknown distribution F the distribution of a point estimate $\theta(X_1, \dots, X_n; F)$ can in many cases not be determined analytically. However when using bootstrapping one can simulate the distribution of θ .

Let $s_{(\cdot)}$ be a bootstrap sample, e.g. drawing n observations with replacement from the sample (X_1, \dots, X_n) , then one can estimate the standard deviation of θ using B bootstrap samples through

$$sd(\theta) = \sqrt{\frac{1}{B-1} \sum_{i=1}^B (\theta(s_i) - \bar{\theta})^2} \quad ,$$

with $\bar{\theta} := \frac{1}{B} \sum_{i=1}^B \theta(s_i)$ as the sample mean over all bootstrap samples.

In context of sample surveys with sampling weights one can use bootstrapping to calculate so called bootstrap weights. These are computed via the bootstrap samples s_i , $i = 1, \dots, B$, where for each s_i every unit of the original sample can appear 0- to m -times. With f_j^i as the frequency of occurrence of observation j in bootstrap sample s_i the uncalibrated bootstrap weights \tilde{b}_j^i are defined as:

$$\tilde{b}_j^i = f_j^i w_j \quad ,$$

with w_j as the calibrated sampling weight of the original sample. Using iterative proportional fitting procedures one can recalibrate the bootstrap weights \tilde{b}_j^i , $j = 1, \dots, B$ to get the adapted or calibrated bootstrap weights b_j^i , $j = 1, \dots, B$.

2.1.1 Rescaled Bootstrap

Since EU-SILC is a stratified sample without replacement drawn from a finite population the naive bootstrap procedure, as described above, does not take into account the heterogeneous inclusion probabilities of each sample unit. Thus it will not yield satisfactory results. Therefore we will use the so called rescaled bootstrap procedure introduced and investigated by (Rao and Wu, 1988). The bootstrap samples are selected without replacement and do incorporate the stratification as well as clustering on multiple stages (see (Chipperfield and Preston, 2007), (Preston, 2009)).

For simplistic reasons we will only describe the rescaled bootstrap procedure for a two stage stratified sampling design. For more details on a general formulation please see (Preston, 2009).

Sampling design Consider the finite population U which is divided into H non-overlapping strata $\bigcup_{h=1,\dots,H} U_h = U$, of which each strata h contains of N_h clusters. For each strata h , C_{hc} , $c = 1, \dots, n_h$ clusters are drawn, containing N_{hc} households. Furthermore in each cluster C_{hc} of each strata h simple random sampling is performed to select a set of households Y_{hcj} , $j = 1, \dots, n_{hc}$.

Bootstrap procedure In contrast to the naive bootstrap procedure where for a stage, containing n sampling units, the bootstrap replicate is obtained by drawing n sampling units with replacement, for the rescaled bootstrap procedure $n^* = \lfloor \frac{n}{2} \rfloor$ sampling units are drawn without replacement. (Chipperfield and Preston, 2007) have shown that the choice of either $\lfloor \frac{n}{2} \rfloor$ or $\lceil \frac{n}{2} \rceil$ is optimal for bootstrap samples without replacement, although $\lfloor \frac{n}{2} \rfloor$ has the desirable property that the resulting uncalibrated bootstrap weights will never be negative.

At the first stage the i -th bootstrap replicate, $f_{hc}^{i,1}$, for each cluster $C_{hc}, c = 1, \dots, n_h$, belonging to strata h , is defined by

$$f_{hc}^{i,1} = 1 - \lambda_h + \lambda_h \frac{n_h}{n_h^*} \delta_{hc} \quad \forall c \in \{1, \dots, n_h\}$$

with

$$n_h^* = \left\lfloor \frac{n_h}{2} \right\rfloor$$

$$\lambda_h = \sqrt{\frac{n_h^* (1 - \frac{n_h}{N_h})}{n_h - n_h^*}} \quad ,$$

where $\delta_{hc} = 1$ if cluster c is selected in the sub-sample of size n_h^* and 0 otherwise. The i -th bootstrap replicate at the second stage, $f_{hcj}^{i,2}$, for each household Y_{hcj} , $j = 1, \dots, n_{hc}$, belonging to cluster c in strata h is defined by

$$f_{hcj}^{i,2} = f_{hc}^{i,1} - \lambda_{hc} \sqrt{\frac{n_h}{n_h^*}} \delta_{hc} \left[\frac{n_{hc}}{n_{hc}^*} \delta_{hcj} - 1 \right] \quad \forall c \in \{1, \dots, n_h\}$$

with

$$n_{hc}^* = \lfloor \frac{n_{hc}}{2} \rfloor$$

$$\lambda_{hc} = \sqrt{\frac{n_{hc}^* N_h (1 - \frac{n_{hc}}{N_{hc}})}{n_{hc} - n_{hc}^*}} \quad ,$$

where $\delta_{hcj} = 1$ if household j is selected in the sub sample of size n_{hc}^* and 0 otherwise.

Single PSUs When dealing with multistage sampling designs the issue of single PSUs, e.g. a single response unit is present at a stage or in a strata, can occur. When applying bootstrapping procedures these single PSUs can lead to a variety of issues. For the methodology proposed in this work we combined single PSUs at each stage with the next smallest strata or cluster, before applying the bootstrap procedure.

Taking bootstrap replicates forward The bootstrap procedure above is applied on the EU-SILC data for each year y_t , $t = 1, \dots, n_y$ separately. Since EU-SILC is a yearly survey with rotating penal design the i -th bootstrap replicate at the second stage, $f_{hcj}^{i,2}$, for a household Y_{hcj} is taken forward until the household Y_{hcj} drops out of the sample. That is, for the household Y_{hcj} , which enters EU-SILC at year y_1 and drops out at year $y_{\bar{t}}$, the bootstrap replicates for the years $y_2, \dots, y_{\bar{t}}$ are set to the bootstrap replicate of the year y_1 .

Split households Due to the rotating penal design so called split households can occur. For a household participating in the EU-SILC survey it is possible that one or more residents move to a new so called split household, which is followed up on in the next wave. To take this dynamic into account we extended the procedure of taking forward the bootstrap replicate of a household for consecutive waves of EU-SILC by taking forward the bootstrap replicate to the split household. That means, that also any new individuals in the split household will inherit this bootstrap replicate.

The following Tables illustrate the mechanism for taking bootstrap replicates forward as well as dealing with split households. Consider a 4 member household that enters EU-SILC in the year 2013, as show in Table 1. For this household the bootstrap replicate was set to 2.

Table 2 shows the household members from Table 1 including a split household for the year 2014. We see that the household member with PID=4750003 has moved to a new household, thus creating the split household, HID=47501. Furthermore the bootstrap procedure, since it is applied on each year separately, did not select households HID=47500 or HID=47501 for the year 2014, e.g. BOOTSTRAP.REP=0. Following our rule for taking bootstrap replicates forward and dealing with split households the bootstrap replicates for the year 2014 will be set equal to the bootstrap replicates for 2013. This is also true for the split household since it was created through the household

Table 1: Household with 4 members in year 2013.

YEAR	PID	HID	BOOTSTRAP.REP
2013	4750001	47500	2
2013	4750002	47500	2
2013	4750003	47500	2
2013	4750004	47500	2

Table 2: Original household with 3 members and 1 split household in year 2014.

YEAR	PID	HID	BOOTSTRAP.REP
2014	4750001	47500	0
2014	4750002	47500	0
2014	4750004	47500	0
2014	4750003	47501	0
2014	4750101	47501	0

member PID=4750003, which moved out of household HID=47500 between 2013 and 2014. Every Household member in of HID=47501 will therefore inherit the bootstrap replicate which household member PID=4750003 had in the year 2013. Table 3 shows the bootstrap replicates after for 2014 after they have been taken forward from 2013.

Taking bootstrap replicates forward as well as considering split households ensures that bootstrap replicates are more comparable in structure with the actual design of EU-SILC.

Uncalibrated bootstrap weights Using the i -th bootstrap replicates at the second stage one can calculate the i -th uncalibrated bootstrap weights b_{hcj}^i for each household Y_{hcj} in cluster c contained in strata h by

Table 3: Bootstrap replicates taken forward to the next year as well as to the split household.

YEAR	PID	HID	BOOTSTRAP.REP
2014	4750001	47500	2
2014	4750002	47500	2
2014	4750004	47500	2
2014	4750003	47501	2
2014	4750101	47501	2

$$\tilde{b}_{hcj}^i = f_{hcj}^{i,2} w_{hcj} \quad ,$$

where w_{hcj} corresponds to the original household weight contained in the sample.

For ease of readability we will drop the subindices regarding strata h and cluster c for the following sections, meaning that the j -th household in cluster c contained in strata h , Y_{hcj} , will now be denoted as the j -th household, Y_j , where j is the position of the household in the data. In accordance to this the i -th uncalibrated bootstrap replicates for household j are thus denoted as \tilde{b}_j^i and the original household weight as w_j .

2.2 Iterative proportional fitting (IPF)

The uncalibrated bootstrap weights \tilde{b}_j^i computed through the rescaled bootstrap procedure yields population statistics that differ from the known population margins of specified sociodemographic variables for which the base weights w_j have been calibrated. To adjust for this the bootstrap weights \tilde{b}_j^i can be recalibrated using iterative proportional fitting as described in (Meraner et al., 2016).

Let the original weight w_j be calibrated for v_1, \dots, v_u sociodemographic variables which can take on V_1, \dots, V_u values each and N_{v_i} be the population margins of the i -th sociodemographic variables. The iterative proportional fitting procedure for the weights \tilde{b}_j^i is made up of 5 Steps. Starting with $k = 0$ the 5 Steps are repeated and k is raised by 1 if the constraints are not satisfied after step 5.

2.2.1 Steps 1-3

The uncalibrated bootstrap weights \tilde{b}_j^i for the j -th observation is iteratively multiplied by a factor so that the projected distribution of the population matches the respective calibration specification N_{v_i} , $i = 1, \dots, u$. For each $i \in \{1, \dots, u\}$ the calibrated weights against $N_{(v_i, y)}$ are computed as

$$\tilde{b}_j^{(u+2)k+i} = \tilde{b}_j^{(u+2)k+i-1} \frac{N_{v_i}}{\sum_l \tilde{b}_l^{(u+2)k}},$$

where the summation in the denominator expands over all observations which have the same value as observation j for the sociodemographic variable v_i . If any weights $\tilde{b}_j^{(u+2)k+i}$ fall outside the range $[\frac{w_j}{4}; 4w_j]$ they will be recoded to the nearest of the two boundaries. The choice of the boundaries results from expert-based opinions and restricts the variance of which has a positive effect on the sampling error. This procedure represents a common form of weight trimming where very large or small weights are trimmed in order to reduce variance in exchange for a possible increase in bias (Potter, 1990, 1993).

2.2.2 Step 4

Since the sociodemographic variables v_1, \dots, v_u can include personal as well as household specific variables, the weights \tilde{b}_j^{5k+u} resulting from the iterative multiplication can be unequal for members of the same household. This can lead to inconsistencies between results projected with household and person weights. To avoid such inconsistencies each household member is assigned the mean of the household weights. That is for each person j in household p with H_p household members, the weights are defined by

$$\tilde{b}_j^{(u+2)k+u+1} = \frac{\sum_{l=1}^{H_p} \tilde{b}_{p(l)}^{(u+2)k+u}}{H_p}$$

This can result in loosing the population structure performed in steps 1-3.

2.2.3 Step 5

To get weights that do conform with the population margins defined by v_1, \dots, v_u the weights \tilde{b}_j^{5k+u+1} are again updated according to the uncertainty parameter p_h . The parameters p_h represent the allowed deviation from the population margins using the weights $\tilde{b}_j^{(u+2)k+u+1}$ compared to N_{v_i} , $i = 1, \dots, u$ where v_i corresponds to a household variable.

The updated weights are computed as

$$\tilde{b}_j^{(u+2)k+u+2} = \begin{cases} \tilde{b}_j^{(u+2)k+u+1} \frac{N_{v_i}}{\sum_l \tilde{b}_l^{(u+2)k+u+1}} & \text{if } \sum_l \tilde{b}_l^{(u+2)k+u+1} \notin ((1-p_h)N_{v_i}, (1+p_h)N_{v_i}) \\ \tilde{b}_j^{(u+2)k+u+1} & \text{otherwise} \end{cases}$$

with the summation in the denominator ranging over all households l which take on the same values for v_i as observation j . As described in the previous subsection the new weight are recoded if they exceed the interval $[\frac{w_j}{4}; 4w_j]$ and set to the upper or lower bound, depending of $\tilde{b}_j^{(u+2)k+u+2}$ falls below or above the interval respectively.

2.2.4 Convergence

After these 5 steps we check if the population margins defined by v_1, \dots, v_u and calculated with $\tilde{b}_j^{(u+2)k+u+2}$ do not deviate too much from N_{v_i} , $i = 1, \dots, u$. That is

$$\frac{\sum_l \tilde{b}_l^{(u+2)k+u+1} - N_{v_i}}{N_{v_i}} < \begin{cases} 0.05 & \text{if } v_i \text{ is a household variable} \\ 0.01 & \text{if } v_i \text{ is a personal variable} \end{cases}$$

holds true for all $i, i = 1, \dots, u$, where the sum in the denominator expands over all observations which have the same value for the variable v_i . If this inequality holds true the algorithm reaches convergence, otherwise k is raised by 1 and Steps 1-5 are repeated.

The above described calibration procedure is applied on each year y_t of EU-SILC separately, $t = 1, \dots, n_y$, thus resulting in so called calibrated bootstrap sample weights $b_j^{(i, y_t)}$, $i = 1, \dots, B$ for each year y and each household j .

2.3 Variance estimation

Applying the previously described algorithms to EU-SILC data for multiple consecutive years y_t , $t = 1, \dots, n_y$, yields calibrated bootstrap sample weights $b_j^{(i, y_t)}$ for each year y_t . Using the calibrated bootstrap sample weights it is straight forward to compute the standard error of a point estimate $\theta(\mathbf{X}^{y_t}, \mathbf{w}^{y_t})$ for year y_t with $\mathbf{X}^{y_t} = (X_1^{y_t}, \dots, X_n^{y_t})$ as the vector of observations for the variable of interest in the survey and $\mathbf{w}^{y_t} = (w_1^{y_t}, \dots, w_n^{y_t})$ as the corresponding weight vector, with

$$sd(\theta) = \sqrt{\frac{1}{B-1} \sum_{i=1}^B (\theta^{(i, y_t)} - \overline{\theta^{(\cdot, y_t)}})^2}$$

with

$$\overline{\theta^{(\cdot, y_t)}} = \frac{1}{B} \sum_{i=1}^B \theta^{(i, y_t)} \quad ,$$

where $\theta^{(i, y_t)} := \theta(\mathbf{X}^{y_t}, \mathbf{b}^{(i, y_t)})$ is the estimate of θ in the year y_t using the i -th vector of calibrated bootstrap weights.

As already mentioned the standard error estimation for indicators in EU-SILC yields high quality results for NUTS1 or country level. When estimation indicators on regional or other sub-aggregate levels one is confronted with point estimates yielding high variance.

To overcome this issue we propose to estimate θ for 3, consecutive years using the calibrated bootstrap weights, thus calculating $\{\theta^{(i, y_{t-1})}, \theta^{(i, y_t)}, \theta^{(i, y_{t+1})}\}$, $i = 1, \dots, B$. For fixed i one can apply a filter with equal filter weights on the time series $\{\theta^{(i, y_{t-1})}, \theta^{(i, y_t)}, \theta^{(i, y_{t+1})}\}$ to create $\tilde{\theta}^{(i, y_t)}$

$$\tilde{\theta}^{(i, y_t)} = \frac{1}{3} \left[\theta^{(i, y_{t-1})} + \theta^{(i, y_t)} + \theta^{(i, y_{t+1})} \right] \quad .$$

Doing this for all i , $i = 1, \dots, B$, yields $\tilde{\theta}^{(i, y_t)}$, $i = 1, \dots, B$. The variance of θ can then be estimated with

$$sd(\theta) = \sqrt{\frac{1}{B-1} \sum_{i=1}^B (\tilde{\theta}^{(i, y_t)} - \overline{\tilde{\theta}^{(\cdot, y_t)}})^2}$$

with

$$\overline{\tilde{\theta}^{(\cdot, y_t)}} = \frac{1}{B} \sum_{i=1}^B \tilde{\theta}^{(i, y_t)} \quad .$$

Applying the filter over the time series of estimated $\theta^{(i,y_t)}$ leads to a reduction of variance for θ since the filter reduces the noise in $\{\theta^{(i,y_{t-1})}, \theta^{(i,y_t)}, \theta^{(i,y_{t+1})}\}$ and thus leading to a more narrow distribution for $\tilde{\theta}^{(i,y_t)}$.

It should also be noted that estimating indicators from a survey with rotating panel design is in general not straight forward because of the high correlation between consecutive years. However with our approach to use bootstrap weights, which are independent from each other, we can bypass the cumbersome calculation of various correlations, and apply them directly to estimate the standard error.

(Martin et al., 2013) showed that using the proposed method on EU-SILC data for Austria the reduction in resulting standard errors corresponds in a theoretical increase in sample size by about 25%. Furthermore this study compared this method to the use of small area estimation techniques and on average the use of bootstrap sample weights yielded more stable results.

2.3.1 Improvement of 3 year mean

In the following we will show the improvement of precision of using the mean over several years on the UDB data of EU-SILC for Spain and Austria. We created 250 bootstrap replicates and calibrated for Spain each of the replicates for the variable DB040, household size and gender by age group. For Austria we calibrated by DB040 and RB090. After that the weighted ratio for AROPE was estimated, including it's standard error for each DB040 by DB100 for each year as well as using 3 consecutive years.

Figure 1 shows for the results of this calculation for the year 2015 and the mean over years 2014, 2015 and 2016. The identity line corresponds to the estimates using a single year as basis for standard error estimation and the points show the results for using the years 2014 to 2016. The dotted line represents the 25% improvement that is to be expected for using the mean over 3 consecutive years, as shown by (Martin et al., 2013). It is clear to see that using the UDB the improvement of the the point estimate for standard error by about 25%, when using 3 consecutive years, holds true as well.

To conclude the above presented methodology represents a reasonable approach for lowering the variance of point estimates if the estimates stay relatively stable over time. However for point estimates which change trend over time or exhibit seasonal or cyclical behavior this methodology can not be recommended. At least not with applying a filter with equal filter weights.

2.4 R-Package vardpoor

For the work-package on NetSilc 2 a different methodology was proposed for estimating standard errors for EU-SILC by (Berger et al., 2017) and even implemented this methodology in the R-Package vardpoor, see (Breidaks et al., 2017). This approach uses the so called ultimate cluster approximation, where the variance between the sampling units at the first sampling stage is used as approximation for the total sampling variance (see (Särndal et al., 1992)). Using this approximation the standard error estimates

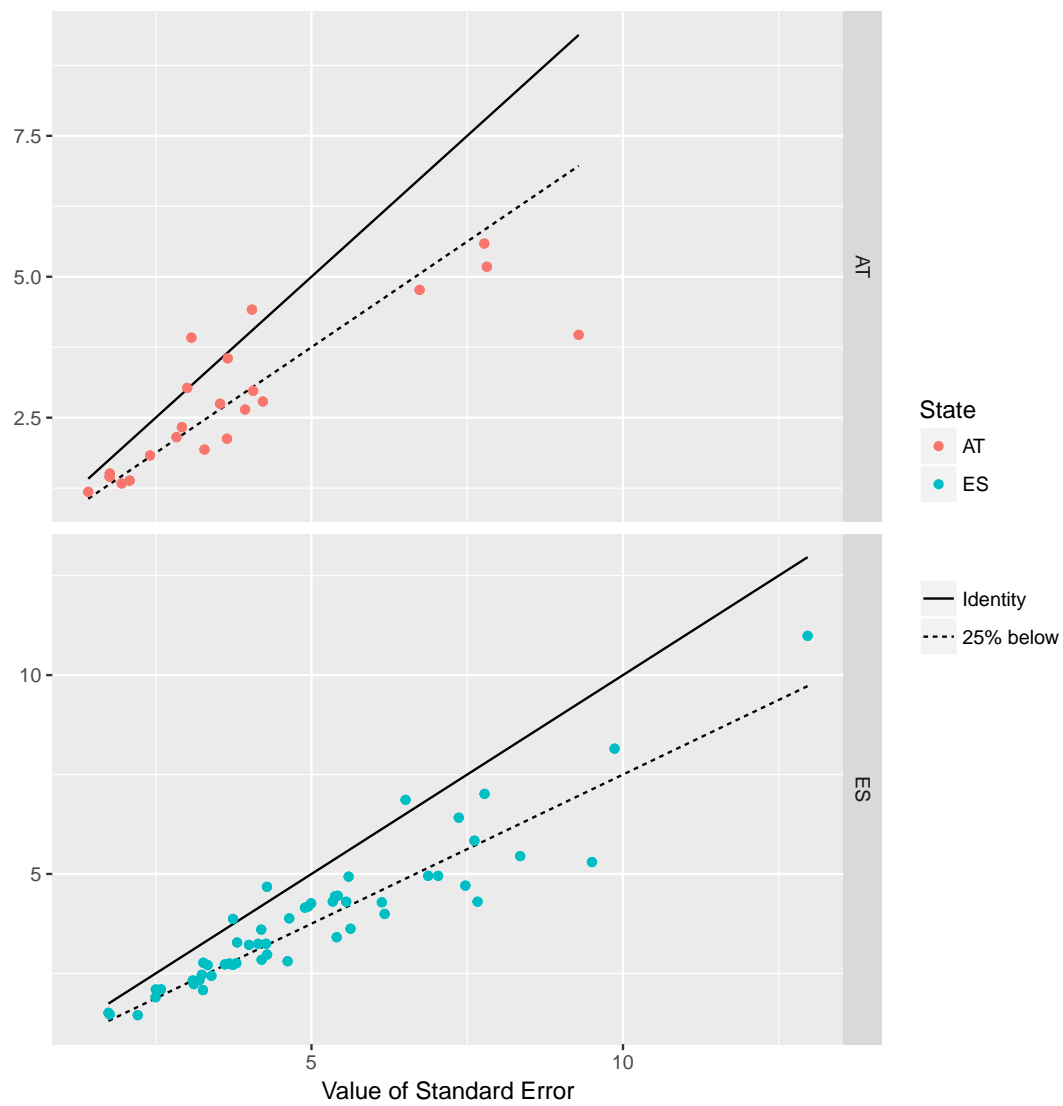


Figure 1: Results for setup 1 regarding the sensitivity analysis on standard error estimations.

Table 4: Standard error estimates with ultimate cluster approximation and bootstrapping.

	DB040	vardpoor	surveysd_250	surveysd_1000
ES11		1.723206	1.869290	1.829328
ES12		2.032826	2.333783	2.224175
ES13		3.219196	3.087267	3.036866
ES21		1.725623	1.765524	1.895363
ES22		2.497444	2.695355	2.717238
ES23		2.684678	2.851551	2.795135
ES24		2.397791	2.441098	2.604816
ES30		1.311531	1.582262	1.651133
ES41		1.762461	1.757140	1.795810
ES42		2.655581	2.865607	2.914314
ES43		2.565243	2.728938	2.793706
ES51		1.057729	1.241227	1.202477
ES52		2.153203	2.540085	2.375432
ES53		2.726446	3.107017	2.965827
ES61		1.517506	1.880508	1.971840
ES62		2.610391	2.923325	2.994599
ES63		6.556073	6.841691	6.753498
ES64		4.698478	6.735764	6.818549
ES70		3.718441	3.888209	3.822487

are analytically calculated and if needed approximated through linearization ((Deville, 1999)).

To show that our proposed method is in accord to the methodology proposed in NetSilc 2 we applied both methods on UDB data of EU-SILC for Spain to estimate standard errors for the weighted ratio of AROPE on NUTS2-Level for the year 2016.

Table 4 shows the estimated standard errors for both methodologies. Our proposed methodology was run with 250 and 1000 bootstrap sample weights, columns surveysd_250 and surveysd_1000, respectively. We see that for almost all standard errors the results for our methodology are slightly higher. This is to be expected since we do incorporate the complete sampling design in our calculations.

2.5 Number of bootstrap replicates

When using bootstrap replicates with survey data it is not always clear how many bootstrap replicates are needed to get stable results. Obviously it is more beneficial to use too many bootstrap replicates instead of too little. Nevertheless using a very large number of bootstrap replicates on large data sets can heavily increase memory usage as well as run-time for the computation. To get a bit of insight on how many bootstrap

replicates suffice for reasonable results we conducted a sensitivity analysis.

2.5.1 Setup

For the sensitivity analysis we used the UDB data of EU-SILC for Spain for the years 2008 till 2016. We created 1000 bootstrap replicates and calibrated each of the replicates for the variable DB040, household size and gender by age group. The estimate of interest was the weighted Ratio in % of AROPE for each year at a national level as well as the grouping DB040 by DB100. The sensitivity analysis was then conducted with two different approaches

1. Starting with 50 bootstrap sample weight the standard error, 0.025 and 0.975 quantile were calculated. After that the next bootstrap sample weight was added and standard error, 0.025 and 0.975 quantile were calculated again. This was done until all 1000 bootstrap sample weights were used for the calculation.
2. We sampled i bootstrap sample weight from the original set of 1000 bootstrap sample weight and calculated standard error, 0.025 and 0.975 quantile of the point estimate. i was chosen to be in $I := \{i : i \in 50, 100, 150, \dots, 700, 750, 800\}$ and for each k this procedure was done 100 times over.

Calculating the estimates for year, DB040 and DB100 can in some cases yield very little observations in the data. Since we expect the behavior of estimates for standard error, 0.025 and 0.975 quantile to be depending on the number of observations used we divided the results into groups. The groups were $G := \{g : g \in \{[0 - 50), [50, 100), [100, 250), [250, 500), [500, 1000), 1000+\}\}$, displaying the range for number of observations used from the data for each year and each year by DB040 and DB100.

2.5.2 Results Setup 1

Let $\hat{\sigma}_{(k,g)}^{(i)}$ be the k -th estimate for standard error using i bootstrap sample weights in group g which contains n_g estimates, $g \in G$. We first calculated for each g the absolute change, $\tilde{\sigma}_{(k,g)}^{(i)}$, for the point estimate

$$\tilde{\sigma}_{(k,g)}^{(i)} = \hat{\sigma}_{(k,g)}^{(i+1)} - \hat{\sigma}_{(k,g)}^{(i)} \quad i = 1, \dots, 999.$$

Afterwards for each group g and number of bootstrap sample weights i we estimated the minimum, maximum and median over $\tilde{\sigma}_{(k,g)}^{(i)}$. In Figure 2 this minimum, maximum and median is displayed for each group and number of bootstrap sample weights used. The y axis was, for better readability, cut off below and above -1 and 1, respectively. The same calculations were done for the 0.025 and 0.975 quantile and the results are displayed in Figure 3 and Figure 4.

Comparing Figures 2 to 4 we conclude that for estimating the standard error a value of 250 bootstrap sample weights does suffice for stable estimates through all group g . When it comes to the estimates for the 0.025 and 0.975 quantile the results change drastically.

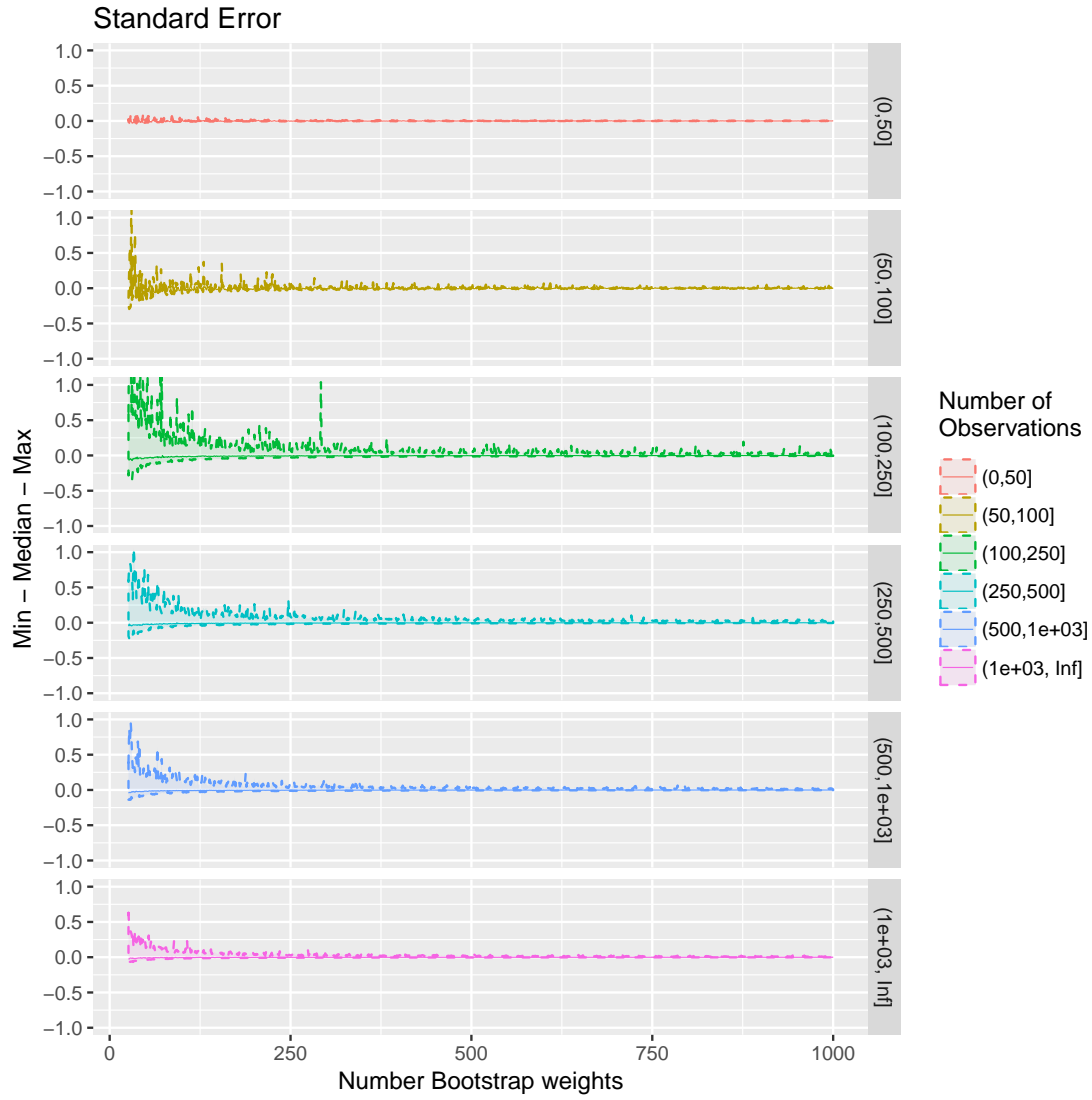


Figure 2: Results for setup 1 regarding the sensitivity analysis on standard error estimations.

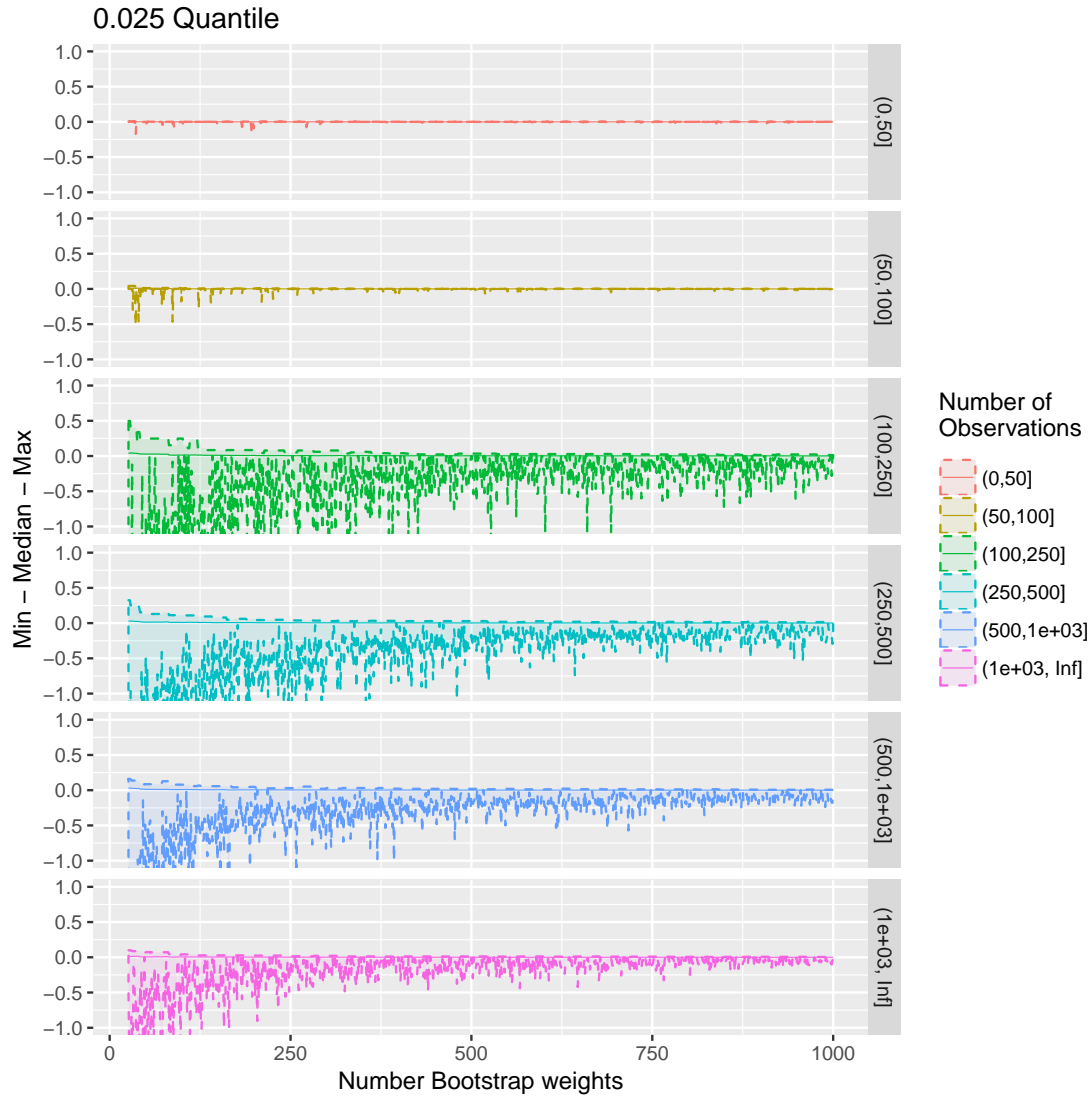


Figure 3: Results for setup 1 regarding the sensitivity analysis on estimations for the 0.025 quantile.

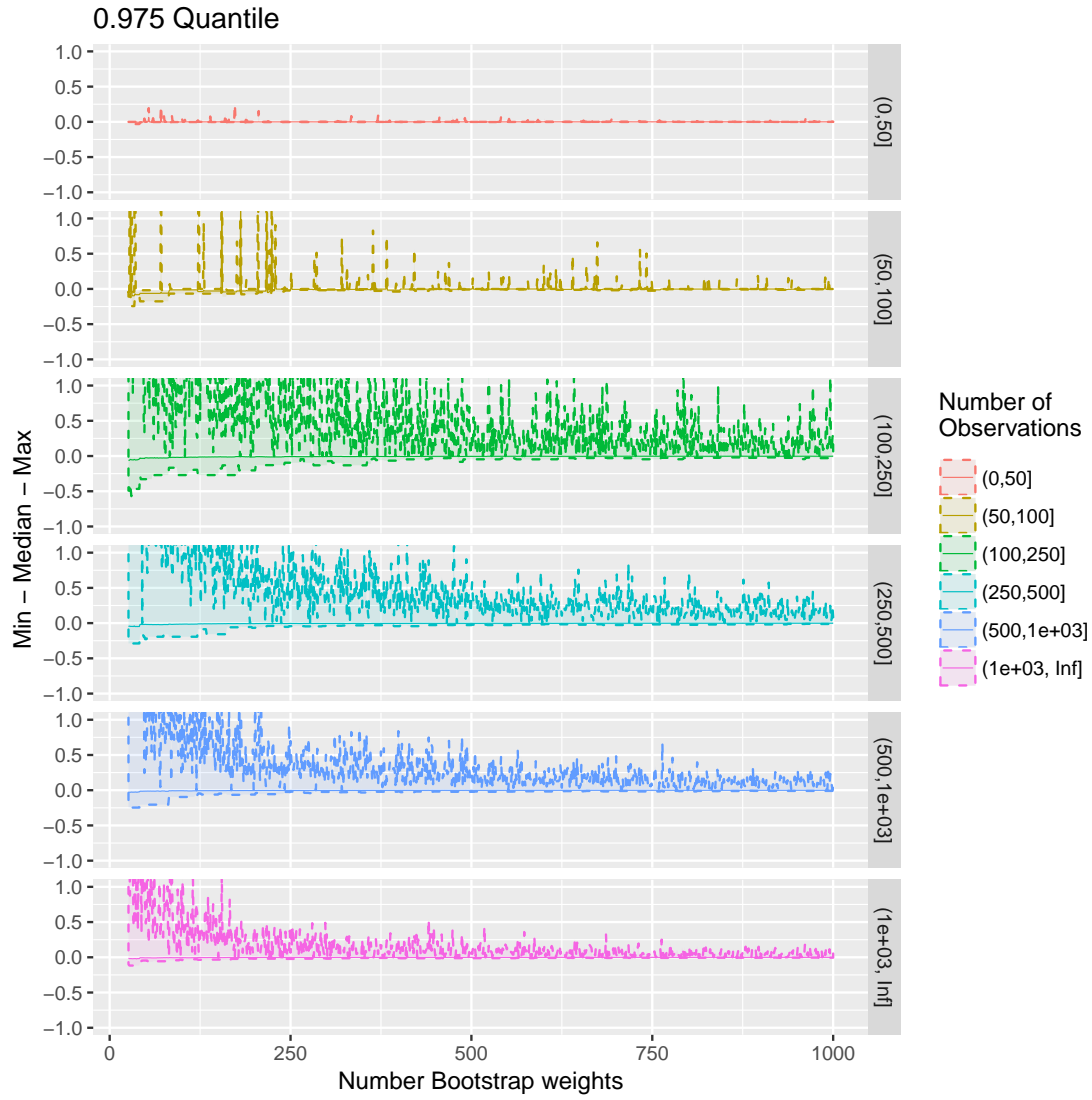


Figure 4: Results for setup 1 regarding the sensitivity analysis on estimations for the 0.975 quantile.

Regardless of the number of observations used for the estimation we suggest to never use less than 500 bootstrap sample weights. But even this can be insufficient when looking at the group (100, 250] in our analysis. Ideally one should at least take a 1000 bootstrap sample weights for those kind of estimates.

2.5.3 Results Setup 2

For the second setup consider $\hat{\sigma}_{(k,g)}^{(i,j)}$ as j -th result, $j = 1, \dots, 100$, for the k -th estimate, using i bootstrap sample weights, $i \in I$ in group g , $g \in G$. We then calculated for fixed i , k and g the range over all $\hat{\sigma}_{(k,g)}^{(i,j)}$, $R_{(k,g)}^{(i)}$

$$R_{(k,g)}^{(i)} = \max(\hat{\sigma}_{(k,g)}^{(i,j)}) - \min(\hat{\sigma}_{(k,g)}^{(i,j)}) \quad ,$$

From the sets of $R_{(k,g)}^{(i)}$ we then calculated for each g and each i the median, 0.01 and 0.99 quantile. Figure 5 displays the results for median, 0.01 and 0.99 quantile for each g over i . We followed the same procedure for the estimates regarding 0.025 and 0.975 quantile and the corresponding results are displayed in Figure 6 and 7.

Looking at Figure 5 we conclude that for larger groups, e.g above 500 observations, a number of 250 bootstrap sample weights does suffice for getting relatively stable estimates, that is, the variance of the estimator itself is not too large. For groups regarding smaller number of observations it would be beneficial to not use less than 500 observations.

Similar to the results from setup 1, the results for the 0.025 and 0.975 quantile, shown in Figure 6 and 7, show a very different picture. Even for larger groups one should at least use 500 bootstrap sample weights, although even that might not be enough. As stated for setup one for those kind of estimates it is necessary to use at least 1000 bootstrap sample weights.

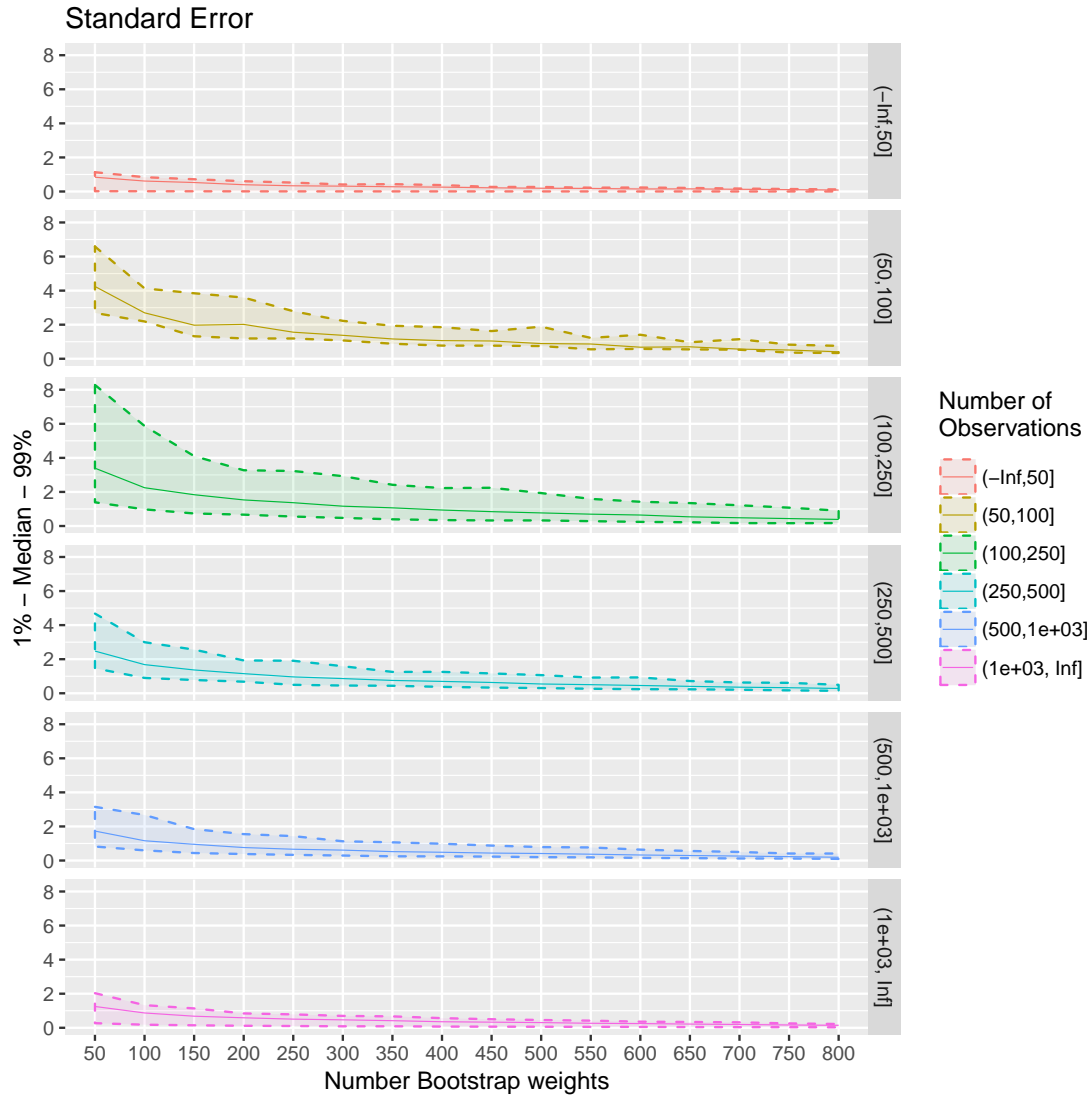


Figure 5: Results for setup 2 regarding the sensitivity analysis on standard error estimations.

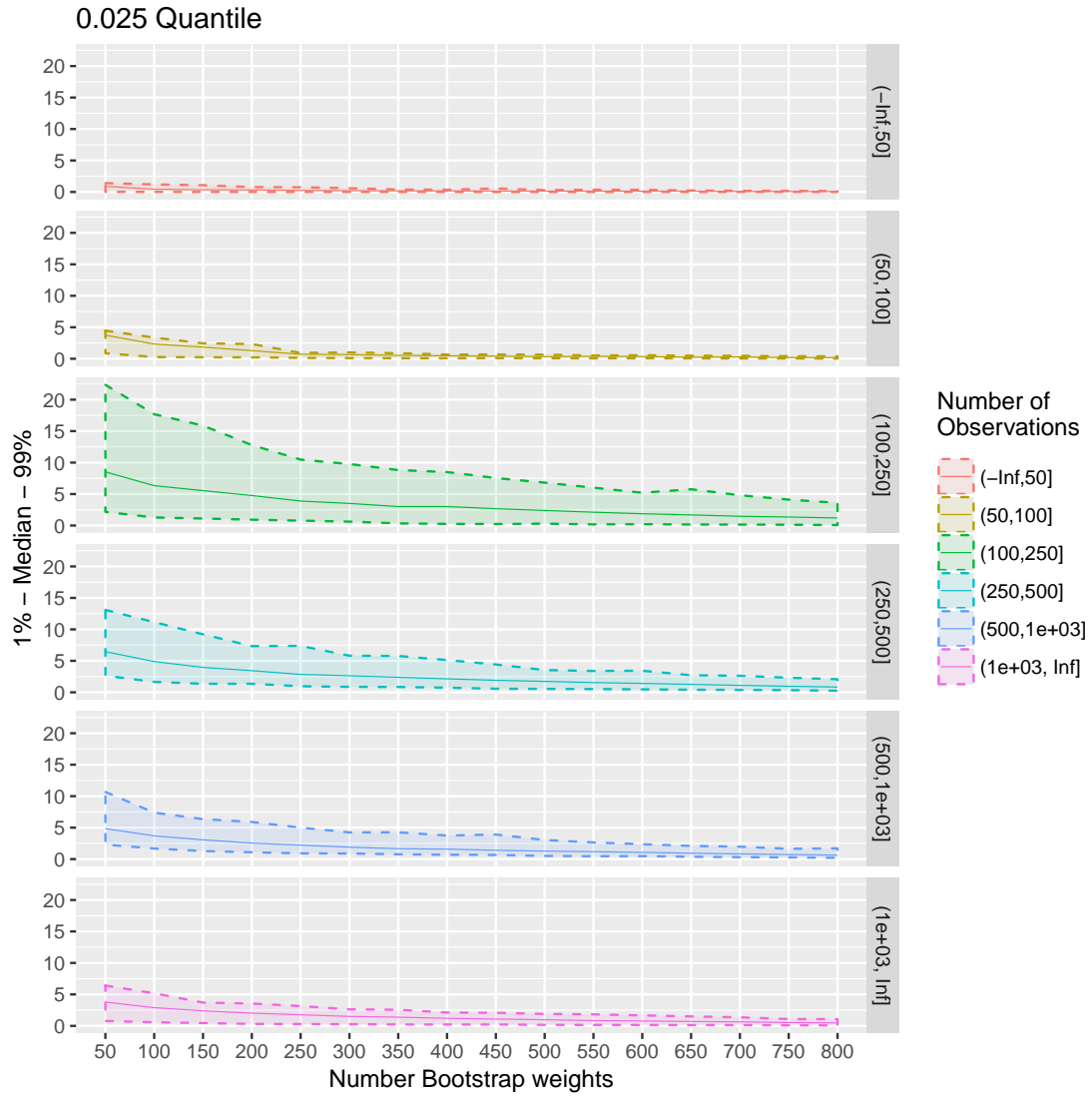


Figure 6: Results for setup 2 regarding the sensitivity analysis on estimations for the 0.025 quantile.

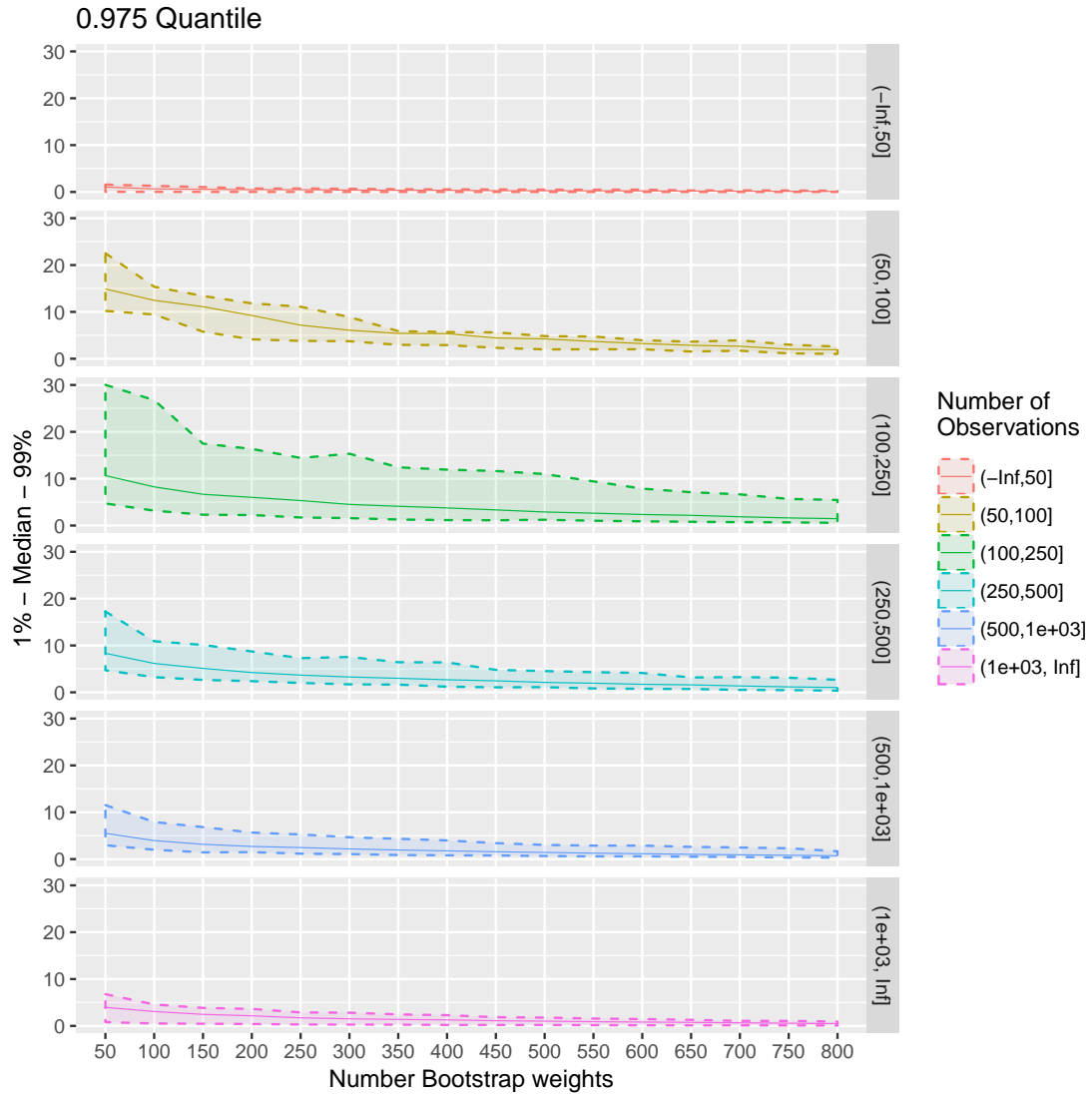


Figure 7: Results for setup 2 regarding the sensitivity analysis on estimations for the 0.975 quantile.

References

- Anthony Atkinson, Bea Cantillon, Eric Marlier, and Brian Nolan. *Social Indicators: The EU and Social Inclusion*. 01 2002. ISBN 978-0-19-925349-4.
- Yves Berger, Guillaume Osier, and Tim Goedeme. Standard error estimation and related sampling issues. In Anthony B. Atkinson, Anne-Catherine Guio, and Eric Marlier, editors, *Monitoring Social Inclusion in Europe*, pages 465–478. Publications Office of the European Union, 2017. URL <https://eprints.soton.ac.uk/416829/>.
- Gianni Betti, Francesca Gagliardi, Achille Lemmi, and Vijay Verma. Subnational indicators of poverty and deprivation in europe: Methodology and applications. 5:129–147, 03 2012.
- Juris Breidaks, Martins Liberts, and Santa Ivanova. *vardpoor: Estimation of indicators on social exclusion and poverty and its linearization, variance estimation*. Riga, Latvia, 2017. URL <https://csblatvia.github.io/wardpoor/>. R package version 0.9.11.
- James Chipperfield and John Preston. Efficient bootstrap for business surveys. 33: 167–172, 12 2007.
- Jean-Claude Deville. Variance estimation for complex statistics and estimators: Linearization and residual techniques. *Techniques d’EnquÃˆte*, 25(2):219–230, 1999.
- B. Efron. Bootstrap methods: Another look at the jackknife. *Ann. Statist.*, 7(1):1–26, 01 1979. doi: 10.1214/aos/1176344552. URL <https://doi.org/10.1214/aos/1176344552>.
- Bauer Martin, Till Matthias, Heuberger Richard, Bilgili Marcel, Glaser Thomas, Kafka Elisabeth, Klotz Johannes, Kowarik Alexander, Lamei Nadja, Meraner Angelika, Oismüller Anneliese, Plate Marc, Scheikl Stefanie, and Zucha Vlasta. Studie zu armut und sozialer eingliederung in den bundesländern. Technical report, Statistik Austria [in German], 05 2013.
- Angelika Meraner, Daniela Gumprecht, and Alexander Kowarik. Weighting procedure of the austrian microcensus using administrative data. 45:3, 06 2016.
- Frank J Potter. A study of procedures to identify and trim extreme sampling weights. In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, volume 225230, 1990.
- Frank J Potter. The effect of weight trimming on nonlinear survey estimates. In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, volume 758763, 1993.
- J Preston. Rescaled bootstrap for stratified multistage sampling. 35:227–234, 12 2009.
- Zhao Qinghua and Lanjouw Peter. *Using PovMap2*, 03 2009. URL <http://iresearch.worldbank.org/PovMap/PovMap2/PovMap2Manual.pdf>.

- J. N. K. Rao and C. F. J. Wu. Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83(401):231–241, 1988.
- C.E. Särndal, B. Swensson, and J.H. Wretman. *Model Assisted Survey Sampling*. Springer series in statistics. Springer-Verlag, 1992. ISBN 9780387975283. URL <https://books.google.at/books?id=MWCzngEACAAJ>.
- Vijay Verma and Gianni Betti. Eu statistics on income and living conditions (eu-silc): Choosing the survey structure and sample design. 7:935–970, 07 2006.