**The R-Package ‘surveysd’ - Estimating Standard Errors in Complex Surveys with Rotating Panel Design**

Johannes Gussenbauer, Statistics Austria, Johannes.Gussenbauer@statistik.gv.at

Alexander Kowarik, Statistics Austria,  [Alexander.Kowarik@statistik.gv.at](mailto:%20Alexander.Kowarik@statistik.gv.at)

Matthias Till, Statistics Austria, Matthias.Till@statistik.gv.at

**Abstract**

*There is an urgent need for regional indicators, especially on poverty and social exclusion, by policy makers in the EU. Surveys which were designed to estimate these indicators on national level, such as EU-SILC, usually do not provide the required precision for reliable estimates on regional levels like NUTS2 and below.*

*With the R-Package ‚surveysd‘, we present a package to improve precision and estimate standard errors for social indicators on regional levels in a straightforward way. Regional estimates from subsequent waves are simply cumulated over time, assuming that these structural patterns remain fairly robust. Variance estimation for pooled data is complicated due to a high correlation within the pooled data. The package resolves the problem by using bootstrap techniques that incorporate pooling of correlated samples, like annual waves of EUSILC. In addition to variance estimation for point estimates the variance estimation for differences is supported. Usability of the package and variance improvement, using this bootstrap methodology, is demonstrated on EU-SILC UDB-data of selected countries with various sampling designs.*

**Keywords:** Variance Estimation, Bootstrapping, EU-SILC, R-Programming

**1. Introduction**

At EU level, social inclusion policies implemented by Member States are monitored mainly with comparative indicators based on EU-SILC data (prior to EU-SILC, they were based on the European Community Household Panel [ECHP]). Over time the original set of so-called Laeken indicators endorsed at the Laeken European Summit in 2001 has been further developed (Atkinson et al. 2002, Marlier et al 2007 and European Commission 2015). Since 2010, EU-SILC indicators have come to guide the “social inclusion target”, i.e. a reduction by 20 million of the number of people in the EU who are at risk of poverty or social exclusion, as one of the five headline targets of the Europe 2020 strategy to promote smart, sustainable and inclusive growth (Atkinson et al. 2017).

EU-SILC is conducted annually in each country. In the great majority of countries EU-SILC is implemented as a sample survey with a rotating panel design of 4 waves (Trindade, Goedemé 2016).

Design and size of the sample ensure adequate precision of EU-SILC indicators at the level of the Member States. EU-SILC has however so far not delivered sufficient subnational precision as has been subject of an earlier assessment of the robustness of EU-SILC based indicators at regional level (Verma, Betti and Gagliardi 2010b).

When it comes to regional indicators based on sample data such as EU-SILC, broadly three approaches are known to improve their precision (Verma et al. 2017, p. 176):

1. Improved size, allocation or design of regional samples;
2. Improved estimation techniques which use auxiliary information;
3. Relaxing requirements of reporting by aggregating information over space, time or indicators.

The first approach requires changes in the data collection which imply considerable cost and needs time for implementation. In any case size and (regional) allocation of the EU-SILC sample have indeed become essential elements for the future revision of EU-SILC (Eurostat 2013).

Small Area Estimation (SAE) has been developed as an alternative (or complementary) strategy to changes in sample design (Tzavidis et al 2016). Various approaches are known and each may yield different results, precision gains and also potential bias. This can be challenging for users. A model based approach which is widely used, has for example been developed by the World Bank (Qinghua and Lanjouw 2009). It combines sample data and census information.

This paper addresses primarily cumulation over three years. The cumulation refers to estimates, instead of pooling data at micro level. Those estimates are weighted equally. The cumulated estimate can thus be considered as an estimate for the middle of the observation period, even if the sample size of adjacent years happens to be larger.[[1]](#footnote-1) To distinguish such estimates with enhanced precision from single year direct estimates, they are here referred as the Approximate Annual Average (AAA). The approach is particularly flexible and may in fact be applied to any estimate that is based on population subgroups with small sample size, such as certain groups of migrants, occupations etc…Also the approach can easily accommodate any type of indicator, including complex nonlinear indicators such as the Gini coefficient.

**2. Assessing the Precision of AAA Estimates**

In the first round of Net-SILC Verma, Betti and Gagliardi (2010a) had originally suggested Jackknife Repeated Replications (JRR) to be used for the estimation of standard errors in EU-SILC.[[2]](#footnote-2) As an extension of their approach, Verma, Betti and Gagliardi (2010b) had suggested to also assess the variance of cumulated EU-SILC estimates by common set of Jackknife Repeated Replications. Using EU-SILC longitudinal data for 2005 and 2006 from Czech Republic and Poland Verma, Betti and Gagliardi (2010b p45) could demonstrate that the standard error of the at-risk-of poverty (AROP) indicator is reduced by 17% when cumulating over two successive years.

During Net-SILC2 the ultimate cluster approach has been suggested as a pragmatic solution for estimating standard errors (Berger, Osier, and Goedemé 2017). The approach simplifies variance estimation for complex samples by approximating total sampling variance from the variance between clusters at the first stage (‘ultimate clusters’) of the sampling process (Särndal, Swensson, and Wretman 1992). The approach is considered easy to implement in standard software and has also been implemented in an R-package called vardpoor (Breidaks, Liberts and Ivanova 2017). In this method, standard errors for nonlinear indicators are calculated upon linearised variables which have asymptotically the same variance properties as the indicators in question (Deville 1999, Osier 2009).

In a comprehensive study of poverty in NUTS 2 regions (Bauer et al 2013) Statistics Austria has derived standard errors with a bootstrap replication method which partly drew on experience made in the AMELI project (Alfons, Templ 2012). Results were compared over several years, indicators and estimation methods. Standard errors of EU-SILC estimates which are cumulated over three successive years were found to be about 25% below that of single years. This implies an increase of effective sample size by approximately 78%. Martin et al (2013) also showed that results from more advanced Small Area Estimation techniques could generally not provide more stable results.

*2.2. Validating the precision of AAA estimates with the surveysd package*

Based on prior experience for estimating standard errors in EU-SILC (Bauer et al 2013) and the Labour Force Survey (LFS) (Meraner, Gumprecht, and Kowarik 2016) this paper presents a bootstrap replication approach to calculate standard errors for survey data which is especially useful to assess the precision of AAA estimates in EU-SILC.

The algorithm involves replication of the sample design. Because the longitudinal component is integrated into the cross sectional sample longitudinal identifiers must also be available to trace the origin of sample units. This information was not available in the UDB before the 2014 wave. Linkable longitudinal identifiers for all cross sectional data since 2008 were provided by INE to Statistics Austria. The surveysd package is implemented in the free R-software and its current version can be downloaded together with a detailed manual.[[3]](#footnote-3)

The approach presented below goes beyond the ultimate cluster approach as it can handle multistage sampling designs, provided that this information is specified in the micro data. it also allows to consider the effect of calibration. The methodology can be summarised in three stages:

1. Draw bootstrap replicates from EU-SILC data for years , .
2. Recalibrate weights by multiplying each set of bootstrap replicates by the sampling weights and calibrate each of the uncalibrated bootstrap weights using iterative proportional fitting.
3. Estimate the point estimate of interest , for each year and each calibrated bootstrap weight to obtain , , . For fixed apply a filter with equal weights for each on , , to obtain . Estimate the variance of using the distribution of .

*2.3. Draw bootstrap replicates*

Bootstrapping has been used widely to estimate confidence intervals and standard errors of point estimates (Efron 1979). Most EU-SILC samples are stratified and drawn without replacement from a finite population. Naive bootstrap procedure does not take into account the specific sample design whereas rescaled bootstrap procedure (Rao and Wu 1988) allows for a more appropriate representation of inclusion probabilities. In the surveysd package, bootstrap samples are selected without replacement and can incorporate stratification as well as clustering on multiple stages (Chipperfield and Preston 2007, Preston 2009).

Replication methods require certain assumptions to be met. These include the presence of more than one PSU within each stratum. In practice, this condition is met for most EU-SILC samples. However, there may be exceptions which can lead to the occurrence of single PSUs within strata (Verma and Betti 2010). Once such PSUs are detected, the algorithm implemented in the surveysd package automatically redefines the computational sample structure by combining this stratum with the next smallest stratum before bootstrapping.

The bootstrap procedure considers the rotational panel design by taking bootstrap replicates forward for all household members in consecutive waves. This is also done for household members which were interviewed in the EU-SILC survey and form so called split households in subsequent waves.

Tables 1 illustrate the mechanism. Consider a household that enters EU-SILC in the year 2013. The household, HID=47500, contained 4 household members of which one household member, PID=4750003, moved to a new household in year 2014. Thus creating the split household, HID=47501, in year 2014. The bootstrap procedure is applied on each year separately so that in this example the household was selected in the year 2013 (BR.ORIGINAL=2) but in this example neither HID=47500 or HID=47501 were included in the random selection for the year 2014 (BR.ORIGINAL=0).

**Table 1. Example of independent bootstrap replicates for a split household**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **YEAR** | **PID** | **HID** | **BR.ORIGINAL** | **BR.NEW** |
| 2013 | 4750001 | 47500 | 2 | 2 |
| 2013 | 4750002 | 47500 | 2 | 2 |
| 2013 | 4750003 | 47500 | 2 | 2 |
| 2013 | 4750004 | 47500 | 2 | 2 |
| 2014 | 4750001 | 47500 | 0 | 2 |
| 2014 | 4750002 | 47500 | 0 | 2 |
| 2014 | 4750004 | 47500 | 0 | 2 |
| 2014 | 4750003 | 47501 | 0 | 2 |
| 2014 | 4750101 | 47501 | 0 | 2 |

When bootstrap replicates are taken forward they 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 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. Column BR.NEW shows the bootstrap replicates 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 longitudinal design of EU-SILC.

*2.3. Recalibrate weights*

Using the -th bootstrap replicate and multiplying it with the -th household weights yields the -th uncalibrated bootstrap weight . The uncalibrated bootstrap weights 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 have been calibrated. To adjust for this, the bootstrap weights can be recalibrated using iterative proportional fitting (Meraner, Gumprecht, and Kowarik 2016).

*2.4. Estimate variance*

Applying the algorithms mentioned above to EU-SILC data for multiple consecutive years , , yields calibrated bootstrap sample weights for each year . Using the calibrated bootstrap sample weights it is straight forward to compute the standard error of a point estimate for year with as the vector of observations for the variable of interest in the survey and as the corresponding weight vector, with

where is the estimate of in the year using the -th vector of calibrated bootstrap weights.

When we have cumulated estimates for three consecutive years we calculate , . For fixed a filter with equal filter weights is applied on the time series to create

Doing this for all , , yields , . The standard error of can then be estimated with

with

Applying the filter over the time series of estimated leads to a reduction of variance for since the filter reduces the noise in and thus leads to a more narrow distribution for .

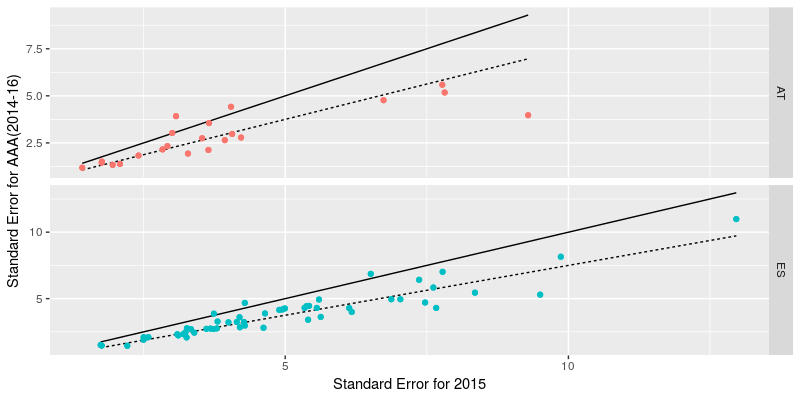
**3. Precision gains**

To evaluate AAA estimates empirically, we have applied the above methodology to UDB data of EU-SILC for Spain and Austria. In addition to the variables available in the UDB, INE provided sample design variables; regional classifications as well as longitudinal identifiers for all waves since 2008. Likewise, Statistics Austria augmented its UDB data with the same information.

In this exercise 250 bootstrap replicates were calibrated within each NUTS2 region (DB040) for household size and gender by age group. After that, the weighted ratio for AROPE was estimated, including its standard error for each region (DB040) by degree of urbanisation (DB100) for each year separately as well as cumulating over 3 adjacent years. Figure 2 compares 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 (Martin et al. 2013). Most estimates are indeed gathering around that line.

**Figure 1. AROPE standard error for 2015 compared with AAA (2014-16)**



*3.1. Benchmarking surveysd against R-package vardpoor*

The approach implemented in the surveysd algorithm appears to be consistent with the approximations suggested in Net-SILC2. Table 4 compares the estimated standard errors for the AROPE indicator 2016 for NUTS2 regions in Spain resulting from both methodologies. Even when additional sample design information is used to account for higher level clustering, standard errors are only marginally increased. Furthermore, there also appear to be differences of similar magnitude when the number of replicates is increased from 250 to 1000 (indicated by columns surveysd\_250 and surveysd\_1000).

**Table 2. Comparison of regional standard errors (AROPE, Spain, 2016) calculated with vardpoor and surveysd**

|  |  |  |  |
| --- | --- | --- | --- |
| **DB040** | **vardpoor** | **surveysd\_250** | **surveysd\_1000** |
| ES11 | 1.749898 | 1.869290 | 1.829328 |
| ES12 | 2.125473 | 2.333783 | 2.224175 |
| ES13 | 2.913933 | 3.087267 | 3.036866 |
| ES21 | 1.812530 | 1.765524 | 1.895363 |
| ES22 | 2.490208 | 2.695355 | 2.717238 |
| ES23 | 2.485835 | 2.851551 | 2.795135 |
| ES24 | 2.386610 | 2.441098 | 2.604816 |
| ES30 | 1.550924 | 1.582262 | 1.651133 |
| ES41 | 1.657396 | 1.757140 | 1.795810 |
| ES42 | 2.714761 | 2.865607 | 2.914314 |
| ES43 | 2.656992 | 2.728938 | 2.793706 |
| ES51 | 1.106510 | 1.241227 | 1.202477 |
| ES52 | 2.225909 | 2.540085 | 2.375432 |
| ES53 | 2.669413 | 3.107017 | 2.965827 |
| ES61 | 1.875178 | 1.880508 | 1.971840 |
| ES62 | 2.880729 | 2.923325 | 2.994599 |
| ES63 | 5.577841 | 6.841691 | 6.753498 |
| ES64 | 6.401432 | 6.735764 | 6.818549 |
| ES70 | 3.598554 | 3.888209 | 3.822487 |

**4. Conclusion and Outlook**

Above all, it is of utmost importance for any future work related to the assessment of sampling precision, that regional identifiers and sample design variables are accessible to researchers. Great progress has already been made by making longitudinally linkable identifiers available in the cross sectional files of the EU-SILC UDB from 2014 onwards. Given the importance of these identifiers for the replication method described in this paper, it would be suggested to revise sectional identifiers also for previous waves.

The experience presented in this paper can be summarised as follows.

1. Multi-Annual Averages Approximation (AAA) promises easy gains compared to alternatives for precision enhancement though sample size increase or SAE.
2. Standard errors obtained from the proposed replication method are slightly higher than those obtained by the ultimate cluster approach proposed in Net-SILC 2.

This paper has put an emphasis on methodological aspects of estimation of indicators and their variance. The focus was on the AROPE-indicator only. In the future it will be worthwhile to address also multidimensional aspects which, for example may also take into account differences in housing cost (Annoni and Weziak-BIalowolska 2014) and inequalities of income and purchasing powers between regions.

Given the computational flexibility of the replication approach, the surveysd package may be easily extended, including the following examples:

1. Assessing the precision gain through consolidated indicators, such as AROP with different income thresholds. (Verma, Betti and Gagliardi 2010b)
2. Similar to the cumulation over time, the number of unreliable regional AAA estimates may also be reduced by constructing averages over space.
3. Technical improvement of the replication method may be necessary to represent the actual longitudinal sampling by rotational group. In the present version of the algorithm, bootstrap replicates are generated independently for each wave and only afterwards taken forward for individuals from the same original sample household.

**5. References**

Annoni, P. and Weziak-Bialowolska, D. (2016), “A measure to target anti-poverty policies in the European Union regions”, *Applied Research in Quality of Life*, 11(1): 181-207

Alfons, A. and Templ, M. (2012), Estimation of social exclusion indicators from complex surveys: The R package Laeken.

Atkinson, A. B., Guio, A. C. and Marlier, E. (2017), Monitoring social inclusion in Europe. In *Luxembourg: Publications Office of the European Union*.

Atkinson, T., Cantillon, B., Marlier, E. and Nolan, B. (2002), Social indicators: The EU and social inclusion: OUP Oxford.

Bauer, Martin., Till M., Heuberger R. Bilgili M., Glaser T., Kafka E. and Klotz J. (2013), “Studie zu Armut und sozialer Eingliederung in den Bundesländern.” Statistics Austria.

Betti, G., Gagliardi, F., Lemmi A., and Verma V. (2012), Sub-national indicators of poverty and deprivation in Europe: methodology and applications, Cambridge Journal of Regions, Economy and Society, 5(1), pp. 149-162.

Berger, Y., Osier, G. and Goedemé, T. (2017), “Standard error estimation and related sampling issues” in Atkinson, A.B., Guio, A.-C. and Marlier, E. (eds.) *Monitoring social inclusion in Europe*, Luxembourg: Publications Office of the European Union, p. 465-478.

Chipperfield, James, and John Preston. 2007. “Efficient Bootstrap for Business Surveys” 33 (December): 167–72.

Deville, Jean-Claude. 1999. “Variance Estimation for Complex Statistics and Estimators: Linearization and Residual Techniques.” *Techniques d’Enquête* 25 (2): 219–30.

Efron, B. (1979), “Bootstrap Methods: Another Look at the Jackknife.” *Ann. Statist.* 7 (1). The Institute of Mathematical Statistics: 1–26. doi:[10.1214/aos/1176344552](https://doi.org/10.1214/aos/1176344552).

Eurostat (2013a), “Handbook on precision requirements and variance estimation for ESS household surveys” Eurostat methodologies and working papers, Eurostat, Luxembourg.

Eurostat (2013b), Outcomes of the TF on the Revision of the EU-SILC legal basis. Working paper for the Working Group Meeting Statistics on Living Conditions 11-13 June 2013 (Doc. LC/88/13/EN).

Marlier, E., Atkinson, A.B., Cantillon, B. and Nolan, B. (2007), The EU and social inclusion: Facing the challenges, The Policy Press, Bristol.

Meraner, A., Gumprecht, D. and Kowarik, A. (2016), Weighting Procedure of the Austrian Microcensus using Administrative Data. In *Austrian Journal of Statistics* 45 (3), pp. 3–14.

Osier, G. (2009), Variance estimation for complex indicators of poverty and inequality using linearization techniques. Survey Research Methods (3), 167-195.

Preston, J. (2009), “Rescaled Bootstrap for Stratified Multistage Sampling” 35 (December): 227–34.

Qinghua, Z., and Lanjouw P. (2009), *Using Povmap2*. <http://iresearch.worldbank.org/PovMap/PovMap2/PovMap2Manual.pdf>.

Rao, J. N. K., and Wu, C. F. J. (1988), “Resampling Inference with Complex Survey Data.” *Journal of the American Statistical Association* 83 (401). Taylor & Francis: 231–41.

Trindade, L. Z. and Goedemé, T, (2016), Notes on updating the EU-SILC UDB sample design variables 2012-2014.

Tzavidis, N., Zhang, L. C., Luna Hernandez, A., Schmid, T., and Rojas-Perilla, N. (2016), *From start to finish: A framework for the production of small area official statistics* (No. 2016/13). Discussion Paper, School of Business & Economics: Economics.

Verma, V. and Betti, G. (2010), Data accuracy in EU-SILC. in A.B. Atkinson and E. Marlier (eds.), Income and living conditions in Europe, Publications Office of the European Union, Luxembourg.

Verma, V., Betti, G. and Gagliardi, F. (2010a), ‘An assessment of survey errors in EU-SILC’, Eurostat methodologies and working papers, Eurostat, Luxembourg.

Verma, V., Betti, G. and Gagliardi, F. (2010b), Robustness of some EU-SILC based indicators at regional level. Eurostat methodologies and working papers, Eurostat, Luxembourg.

Verma, V., Gagliardi, F., and Ferretti, C. (2013), Cumulation of poverty measures to meet new policy needs. In Advances in Theoretical and Applied Statistics (pp. 511-522). Springer, Berlin, Heidelberg.

Verma, V., Lemmi, A, Betti, G., Gagliardi, F. and Piacentini, M, (2017), How precise are poverty measures estimated at the regional level? In *Regional Science and Urban Economics* 66, pp. 175–184.

1. Alternatively, weights may be determined with regard to the distance of the centre point or in proportion to their variances to improve efficiency. Assessment of sensitivity with regard to such refinements may be addressed in further research. [↑](#footnote-ref-1)
2. the SAS code is provided by Eurostat at:

   <https://circabc.europa.eu/webdav/CircaBC/ESTAT/eusilc/Library/tools/estimation_jackknife> [↑](#footnote-ref-2)
3. The current version of the code is available here: <https://github.com/statistikat/surveysd> . To install the package, it is necessary to download a zipfile, open the R-project contained in it and rebuild the code. It is recommended to use R Studio for that purpose. [↑](#footnote-ref-3)