STATISTICS AUSTRIA
The Information Manager

Johannes Gussenbauer
Alexander Kowarik
Gregor de Cillia
Matthias Till
Qualitymanagement and Methods

The R Package `surveysd`

Florence
May 2019

Motivation

- EU Statistics on Income and Living Conditions (EU-SILC)
- Annual survey with a rotating panel design of 4 waves (usually)
- Produce poverty rates
  - high quality indicators on national level
  - estimates on sub-national level have poor accuracy
- Difference in poverty rates over time
  - potentially high correlation between consecutive years
  - variance estimation for many estimators, variables and domains (tedious)
- Created R-Package `surveysd` during NetSILC-3

- R-package for variance estimation on surveys with rotating panel design
- Variance estimation through bootstrap techniques (Preston 2009)
- Improve accuracy by using multiple (consecutive) waves of the survey
  - Average bootstrap replicates over waves (Verma et al. 2017)
- Why not Small Area Estimates?
  - needs external data - might not be accessible
  - lack of comparability when applying different models to data from different country

- Package aims to have streamlined functionality
- Combines all necessary steps to use calibrated bootstrapping with custom estimation functions
  - draw bootstrap replicates
  - calibrate bootstrap replicates
  - calculate point estimates and standard errors

Draw bootstrap replicates

```
# UDB-Data from Spain
silcESboot <-
  draw.bootstrap(silcES, REP = 1000, hid = "DB030", weights = "RB050",
                 period = "RB010", strata = c("DB050", "I"),
                 cluster = c("DB060", "DB030"),
                 totals = c("NumClust", "NumHouse"),
                 split = TRUE, pid = "RB030")
```

- Sampling design can be arbitrary (1-Stage, 2-Stage,. . .)
- Replicates are taken forward to mimic rotational panel design
  - Split households are considered
- Single PSUs are automatically detected and dealt with

Calibrate Bootsrap Replicates

```
silcEScalib <- recalib(silcESboot,
                       conP.var = c("AgeSex"),
                       conH.var = c("hsize", "DB040"))
```

- Calibration with iterative proportional fitting (Meraner, Gumprecht, and Kowarik 2016)
  - function migrated from R-Package `simPop`
- Define households and/or personal variables to be used in calibration
- Calibration totals are computed directly from the sample

Estimate standard errors

```
AroseEst <- calc.stError(silcEScalib, var = "arose",
                         fun = weightedRatio, group = c("DB040"))
```

- Supply any number of grouping variables
- Predefined point estimates available or supply function from other package or custom function
- Many optional parameter:
  - average over k years `period.mean`
  - difference between years `period.diff`
  - quantiles from distribution of bootstrap replicates

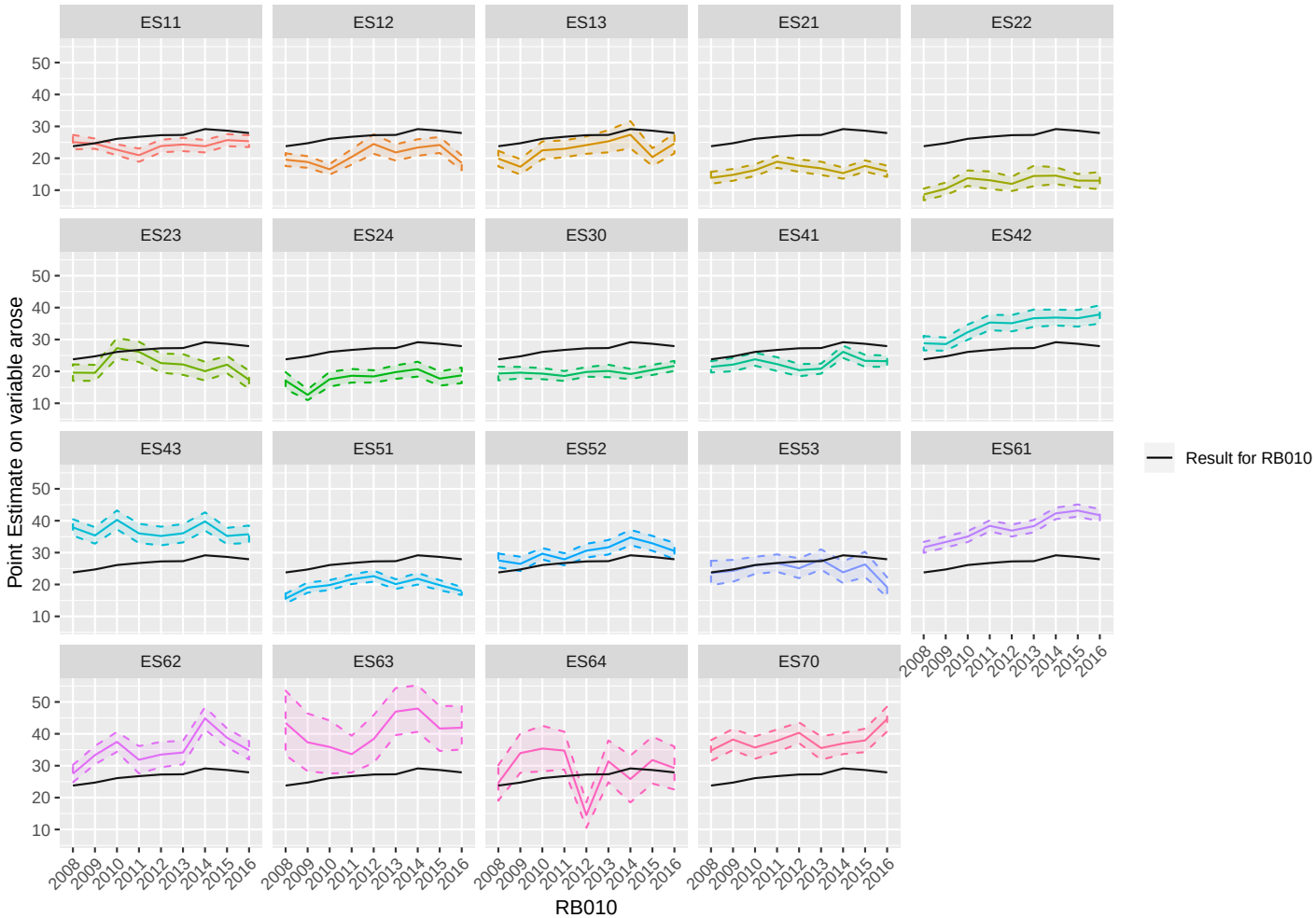
Estimate standard errors

```
AroseEst
```

```
## Calculated point estimates for variable(s)
##
##  arose
##
## Results hold 180 point estimates for 9 periods in 20 subgroups
##
## Estimted standard error exceeds 10 % of the the point estimate
```

Plot-Function

```
plot(AroseEst, type = "grouping", groups = "DB040", sd.type="ribbon")
```



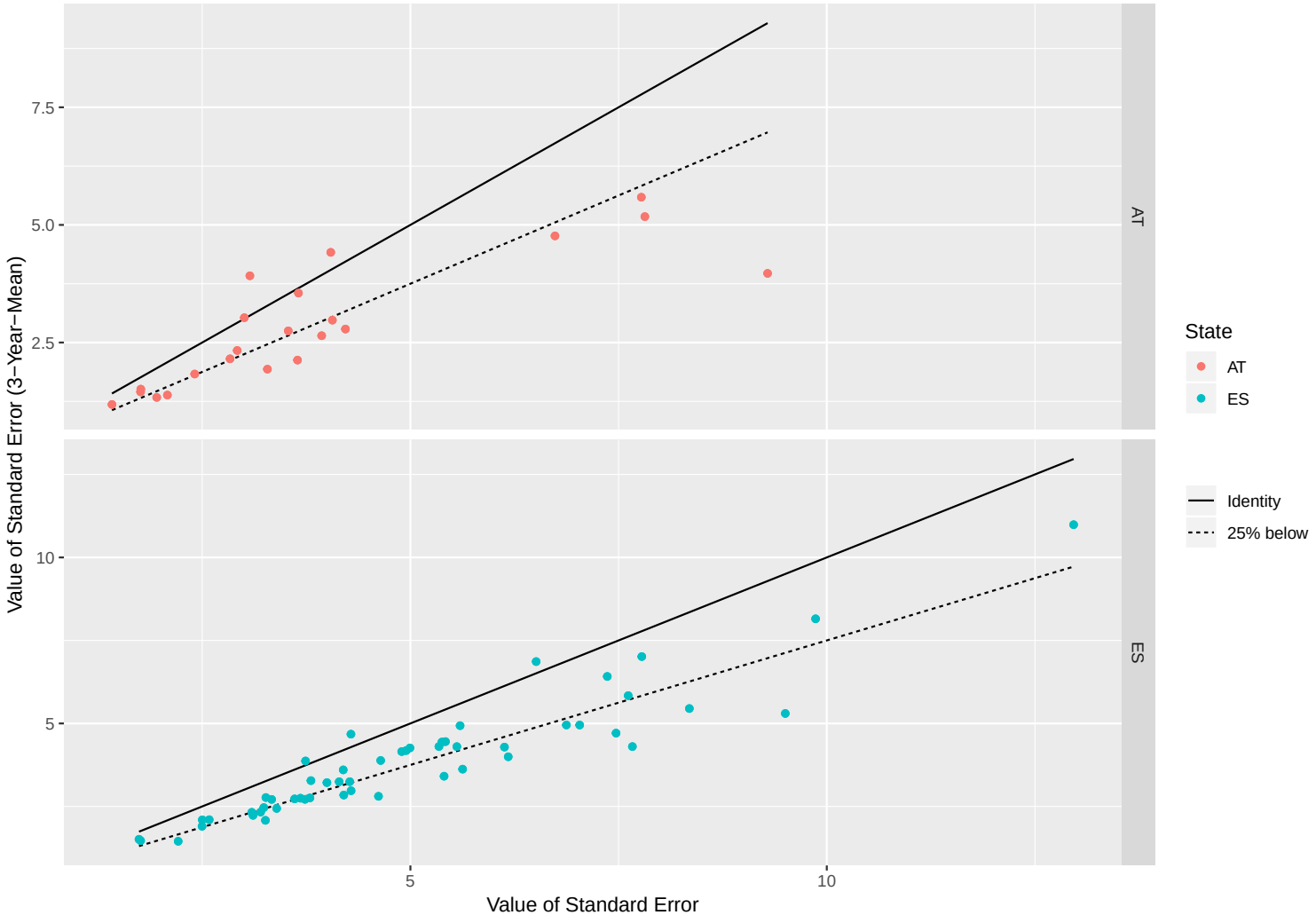
Impact of 3-Year Mean

- Use estimates for 3 (or more) consecutive years $\sim$ filter with equal filter weights

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

- Estimate standard error and mean from distribution of $\tilde{\theta}^{(i,y_t)}$
- Reduction in resulting standard errors by 25% on average (Bauer et al. 2013)

Impact of 3-Year Mean



Final Remarks

- Simple to use R-Package
- Supports harmonious approach for estimating standard errors on surveys with rotating panel design
  - Achieve more accuracy by averaging over multiple years
  - No need for administrative data or modelling assumptions
- On CRAN and Github
  - https://cran.r-project.org/web/packages/surveysd/index.html
  - https://github.com/statistikat/surveysd
  - https://statistikat.github.io/surveysd/articles/methodology.html

References

Bauer, Martin, Matthias Till, Richard Heuberger, Marcel Bilgili, Thomas Glaser, Elisabeth Kafka, Johannes Klotz, et al. 2013. "Studie Zu Armut Und Sozialer Eingliederung in Den Bundesländern." Statistik Austria [in German].

Meraner, Angelika, Daniela Gumprecht, and Alexander Kowarik. 2016. "Weighting Procedure of the Austrian Microcensus Using Administrative Data." Austrian Journal of Statistics 45 (June): 3. https://doi.org/10.17713/ajs.v45i3.120.

Preston, J. 2009. "Rescaled Bootstrap for Stratified Multistage Sampling." Survey Methodology 35 (December): 227–34. https://www.researchgate.net/publication/281735659.

Verma, Vijay, Achille Lemmi, Gianni Betti, Francesca Gagliardi, and Mario Piacentini. 2017. "How Precise Are Poverty Measures Estimated at the Regional Level?" Regional Science and Urban Economics 66 (July). https://doi.org/10.1016/j.regsciurbeco.2017.06.007.

STATISTICS AUSTRIA
The Information Manager

Please address queries to:
Johannes Gussenbauer

Contact information:
Guglgasse 13, 1110 Vienna
phone: +43 (1) 71128-7327
Johannes.Gussenbauer@statistik.gv.at

The R Package `surveysd`