

Variance estimation for annual point estimates and net changes for LFS using R package vardpoor

Juris Breidaks

June 20, 2019

Table of contents

- 1 Introduction
- 2 Methodology
- 3 Function vardannual

Introduction

The methodology used for the production of variance estimates for annual point estimates and net changes was developed, and it is possible to evaluate whether net changes and annual point estimates are statistically significant. The methodology and calculations were applied for the following LFS indicators:

- employment rate 20-64, total;
- employment rate 20-64, men;
- employment rate 20-64, women.

Theoretical basis of 'vardannual'

- Guillaume Osier and Virginie Raymond article "Development of methodology for the estimate of variance of annual net changes for LFS-based indicators" (Osier, Raymond, 2015)
- Guillaume Osier and Pauline Perray article "Variance estimators of annual levels and net changes for a defined set of LFS-based indicators" (Osier, Perray, 2016).

Both papers showed an estimator proposed by Berger and Priam (2013) and Berger and Oguz Alper (2013). This estimator has also been recommended by the Second Network for the analysis of the EU-SILC data ("Net-SILC2" project) to deal with measures of changes between the EU-SILC reference indicators of poverty, inequality and social exclusion (Eurostat, 2013).

Methodology description plan

- LFS survey
- The variance estimation of quartely estimates
- Correlation estimation of two quarter change estimates
- Variance estimation for annual point estimate
- Variance estimation for annual net change estimates

LFS survey

Methodology for variance estimation for annual point estimates and net changes was developed taking into account sampling design (including overlap between quarters) and calibration. The indicator based on the LFS is sample estimate. The sample was formed as a rotating panel sample, where sampling units took part in the survey several times. Thus, we cannot assume that quarterly or annual estimates of the LFS are time-independent and, therefore, temporal correlations between cross-sectional estimates also reflected in variance calculations.

The variance estimation of quartely estimates

The estimate of the ratio is calculated as:

$$\hat{\theta}_t = \frac{\hat{Y}_t}{\hat{Z}_t} = \frac{\sum_{k \in s_t} w_{t,i} \cdot y_{t,i}}{\sum_{k \in s_t} w_{t,i} \cdot z_{t,i}},$$

where s_t denotes the sample of individuals at quarter t , the w_t are the survey weights at quarter t , the $y_{t,i}$ is the variable y for person i at quarter t , the $z_{t,i}$ is the variable z value for person i at quarter t .

$\hat{\theta}_t$ is a non-linear statistic and assumes that can be expressed as a smooth function of estimated population total $\hat{\tau}_t$. First-order Taylor series approximation of $\hat{\theta}_t$ leads to:

$$\hat{\theta}_t = f(\hat{\tau}_t) \cong \sum_{i \in s_t} w_{t,i} \cdot u_{t,i} + R_{t,n},$$

The variance estimation of quartely estimates

Where the remainder $R_{t,n}$ tends towards zero as the sample size tends towards infinity. The variance of $\hat{\theta}_t$ is asymptotically equal to the linear part $\hat{\tau}_t = \sum_{i \in s_t} w_{t,i} \cdot u_{t,i}$.

$$\widehat{Var}(\hat{\theta}_t) \cong \widehat{Var}(\hat{\tau}_t),$$

Each quarter linearized variable for the ratio estimator (Särndal et al, 1992) is calculated as

$$\hat{u}_{t,i} = \frac{1}{\widehat{Z}_t} (y_{t,i} - \hat{\theta}_t z_{t,i}).$$

The variance estimation of quartely estimates

If the weights are calibrated, then calibration residual estimates \hat{e}_k are calculated (Lundstrom, Särndal, 2001) by formula

$$\hat{\varphi}_{t,k} = \sum_{i \in \text{household}_k} \hat{u}_{t,i} - x'_{t,k} \hat{B}_t,$$

where

$$\hat{B}_t = \left(\sum_{k \in s} d_{t,k} q_{t,k} x_{t,k} x'_{t,k} \right)^{-1} \left(\sum_{k \in s} d_{t,k} q_{t,k} x_{t,k} \sum_{i \in \text{household}_k} \hat{u}_{t,i} \right)$$

X_t - the matrix of auxiliary variables in quarter t , $q_{t,k}$ - variable, which is vector of positive values accounting for heteroscedasticity.

The variance estimation of quartely estimates

The estimation of variance was implemented by ultimate cluster approach and multivariate regression. With the ultimate cluster approach, the Primary Sampling units (PSUs) will play the role of sampling units. We compute the following estimates for each PSU with such an approach.

$$\check{y}_{t,j} = \sum_{k \in s_t \cap PSU_{t,j}} w_{t,k} \cdot \begin{cases} \hat{\varphi}_{t,k}, & \text{calibration is used,} \\ \sum_{i \in household_k} \hat{u}_{t,i}, & \text{calibration is not used.} \end{cases} \quad (1)$$

where $j = \overline{1, m}$ denotes total of PSUs at quarter t , $PSU_{t,j}$ denotes $j - th$ PSU at quarter t .

Variance estimation with the ultimate cluster method

$$\widehat{Var}(\hat{\tau}_t) = \sum_{h=1}^{H_t} \frac{n_{t,h}}{n_{t,h} - 1} \sum_{j=1}^{n_{t,h}} (\check{y}_{t,j} - \check{y}_{t,h*})^2$$

where

- $\check{y}_{t,h*} = \frac{\sum_{j=1}^{n_{t,h}} \check{y}_{t,j}}{n_{t,h}}$
- h is the stratum number, with a total of H strata in quarter t ,
- j is the label of the PSU within stratum h , with a total of $n_{t,h}$ PSUs in quarter t .

Variance estimation with the ultimate cluster method

Variance estimation is implemented with ultimate cluster approach and regression model, where difference

$$(\check{y}_{t,j} - \acute{y}_{t,h*})^2$$

is regarded as the estimated residuals of regression model.

Regression model:

$$\check{y}_{t,j} = \sum_{h=1}^{H_t} \beta_{t,h} \cdot z_{th,j} + \varepsilon_{t,j}$$

Variance estimation with the ultimate cluster method

where $\check{y}_{t,j}$ was defined by (1) and dummy component $z_{th,j}$ is given by:

$$z_{th,j} = \begin{cases} 1, & PSU_j \in s_{t,h}, \\ 0, & PSU_j \notin s_{t,h}. \end{cases}$$

where $s_{t,h}$ is sample for PSU within stratum h at quarter t .
Estimated residuals are used in variance estimator for variable.

$$\widehat{Var}(\hat{\tau}_t) = \sum_{h=1}^{H_t} \frac{n_{t,h}}{n_{t,h} - 1} \sum_{j=1}^{n_{t,h}} (\hat{\varepsilon}_{t,j})^2$$

Variance estimation of two changes

Variance estimation of two quarter changes for LFS deals with measuring change between complex non-linear indicators. The idea is to apply the linearization method (Särndal et al, 1992; Deville, 1999) to approximate a complex non-linear statistic with a linear function of the observations, justified by asymptotic properties of the estimator. Then, a variance estimator of the linear approximation gives an estimator of the variance of the non-linear statistic.

Variance estimation of two changes

The estimator of change Δ between non-linear parameter $\hat{\theta}_1$ of quarter 1 and non-linear parameter $\hat{\theta}_2$ of quarter 2 is

$$\hat{\Delta} = \hat{\theta}_2 - \hat{\theta}_1,$$

where $\hat{\theta}_t$ are the standard estimator of θ_t based on sample s_t of quarter t , where t is first or second. If θ_t is the ratio of two population totals at quarter t , where t is first or second, then we have

$$\hat{\theta}_t = \frac{\sum_{i \in s_t} y_{t,i} \cdot w_{t,i}}{\sum_{i \in s_t} z_{t,i} \cdot w_{t,i}}, t = 1, 2$$

Variance estimation of two changes

Let $\hat{\theta}_t$ be a non-linear statistic and let $\hat{\theta}_t$ us assume that can be expressed as a smooth function of estimated population total $\hat{\tau}_t$. First-order Taylor series approximation of $\hat{\theta}_t$ leads to:

$$\hat{\theta}_t = f(\hat{\tau}_t) \cong \sum_{i \in S_t} w_{t,i} \cdot u_{t,i} + R_{t,n},$$

Variance estimation of two changes

Where the remainder $R_{t,n}$ tends towards zero as the sample size tends towards infinity.

Let $\hat{\tau}_t = \sum_{k \in S_t} w_{t,k} \cdot \varphi_{t,k}$ the linear approximation of $\hat{\theta}_t$ or both t value. We obtain the following formula:

$$\hat{\varphi}_{t,k} = \begin{cases} \sum_{i \in \text{household}_k} \hat{u}_{t,i} - x'_{t,k} \hat{B}_t, & \text{when variable is linearized} \\ & \text{and calibration is used,} \\ \sum_{i \in \text{household}_k} \hat{u}_{t,i}, & \text{when variable is linearized.} \end{cases}$$

The variance estimation of quartely estimates

where

$$\hat{u}_{t,i} = \frac{1}{\hat{u}_t} (y_{t,i} - \hat{\theta}_t z_{t,i})$$

$$\hat{B}_t = \left(\sum_{k \in s} d_{t,k} q_{t,k} x_{t,k} x'_{t,k} \right)^{-1} \left(\sum_{k \in s} d_{t,k} q_{t,k} x_{t,k} \sum_{i \in \text{household}_k} \hat{u}_{t,i} \right)$$

The variance estimation of quartely estimates

Hence, the variance of $\hat{\Delta}$ is given by:

$$\begin{aligned} \text{Var}(\hat{\Delta}) &= \text{Var}(\hat{\theta}_2 - \hat{\theta}_1) \cong \text{Var}(\hat{\tau}_2 - \hat{\tau}_1) = \\ &= \text{Var}(\hat{\tau}_1) + \text{Var}(\hat{\tau}_2) - 2 \cdot \text{Cov}(\hat{\tau}_2, \hat{\tau}_1) = \\ &= \text{Var}(\hat{\tau}_1) + \text{Var}(\hat{\tau}_2) - 2 \cdot \sqrt{\text{Var}(\hat{\tau}_1)} \cdot \sqrt{\text{Var}(\hat{\tau}_2)} \cdot \rho(\hat{\tau}_2, \hat{\tau}_1) \end{aligned}$$

The variance estimates of and is calculated using the estimator proposed by Net-SILC2 (Berger et al, 2016).

The variance estimation of quartely estimates

As the survey quarters in LFS are time-correlated, the correlation $\rho(\hat{\tau}_2, \hat{\tau}_1)$ between the cross-sectional estimators cannot be ignored. The estimator proposed by Berger and Priam (2016) is based on the residual matrix of a multivariate regression model. The model includes covariates which specify the stratification. In addition, interaction terms specify the rotation of the sampling designs:

$$\begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} = \sum_{i=1}^H \begin{pmatrix} \beta_h^1 \cdot z_{1h,j} + \beta_h^2 \cdot z_{2h,j} + \beta_h^{12} \cdot z_{1h,j} \cdot z_{2h,j} \\ \gamma_h^1 \cdot z_{1h,j} + \gamma_h^2 \cdot z_{2h,j} + \gamma_h^{12} \cdot z_{1h,j} \cdot z_{2h,j} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,j} \\ \varepsilon_{2,j} \end{pmatrix}.$$

where $s \in s = s_1 \cup s_2$ and residuals $\varepsilon_j = (\varepsilon_{1j}, \varepsilon_{2j})^T$ follow a bivariate distribution with mean zero and an unknown variance-covariance matrix.

The variance estimation of quartely estimates

The $z_{1h,j}$ and $z_{2h,j}$ are (dummy) design variables which specify the stratification:

$$z_{1h,j} = \begin{cases} 1, & PSU_j \in s_{1h}, \\ 0, & PSU_j \notin s_{1h}. \end{cases}$$

$$z_{2h,j} = \begin{cases} 1, & PSU_j \in s_{2h}, \\ 0, & PSU_j \notin s_{2h}. \end{cases}$$

The quantities β_h^1 , β_h^2 , γ_h^1 , γ_h^2 , β_h^{12} and γ_h^{12} are regression parameters that need to be included into the model.

The variance estimation of quartely estimates

The estimated regression residuals ε_{1j} and ε_{2j} gives an estimate of the correlation $\rho(\hat{\tau}_2, \hat{\tau}_1)$ between τ_1 and τ_2 , which can be respectively defined by:

$$\rho(\hat{\tau}_2, \hat{\tau}_1) = \frac{\sum_{j \in s_1 \cup s_2} \hat{\varepsilon}_{1j} \cdot \hat{\varepsilon}_{2j}}{\sqrt{\sum_{j \in s_1 \cup s_2} (\hat{\varepsilon}_{1j})^2} \cdot \sqrt{\sum_{j \in s_1 \cup s_2} (\hat{\varepsilon}_{2j})^2}}$$

Variance estimation for annual point estimates

In order to calculate annual point estimates, let

$$\hat{R}_t^{(l)} = \frac{\hat{Z}_t^{(l)}}{\hat{Y}_t^{(l)}} = \frac{\sum_{i \in s_{t,l}} y_{t,i}^{(l)} \cdot w_{t,i}^{(l)}}{\sum_{i \in s_{t,l}} z_{t,i}^{(l)} \cdot w_{t,i}^{(l)}},$$

be the estimated ratio at the quarter $l \in 1, 2, 3, 4$ of year t . Assuming that the size of the active population remains (almost) unchanged from one quarter to another within the same year, we obtain:

$$\hat{\theta}_t = \frac{\hat{R}_t^{(1)} \hat{Z}_t^{(1)} + \hat{R}_t^{(2)} \hat{Z}_t^{(2)} + \hat{R}_t^{(3)} \hat{Z}_t^{(3)} + \hat{R}_t^{(4)} \hat{Z}_t^{(4)}}{\hat{Z}_t^{(1)} + \hat{Z}_t^{(2)} + \hat{Z}_t^{(3)} + \hat{Z}_t^{(4)}},$$

$$\hat{\theta}_t \cong \frac{\hat{R}_t^{(1)} + \hat{R}_t^{(2)} + \hat{R}_t^{(3)} + \hat{R}_t^{(4)}}{4}$$

Variance estimation for annual point estimates

The variance of the annual point estimates $\hat{\theta}_t$ in year t can be written as:

$$\text{Var}(\hat{\theta}_t) = \frac{1-f}{16} X_t^T A_t X_t$$

where f is the sampling rate and the matrices A and X contain the quarterly variances and correlations, respectively:

$$X_t = \begin{pmatrix} \sqrt{\text{Var}(\hat{R}_t^{(1)})} \\ \sqrt{\text{Var}(\hat{R}_t^{(2)})} \\ \sqrt{\text{Var}(\hat{R}_t^{(3)})} \\ \sqrt{\text{Var}(\hat{R}_t^{(4)})} \end{pmatrix}$$

Variance estimation for annual point estimates

$$A_t = \begin{pmatrix} 1 & \hat{\rho}(\hat{R}_t^1, \hat{R}_t^2) & \hat{\rho}(\hat{R}_t^1, \hat{R}_t^3) & \hat{\rho}(\hat{R}_t^1, \hat{R}_t^4) \\ \hat{\rho}(\hat{R}_t^2, \hat{R}_t^1) & 1 & \hat{\rho}(\hat{R}_t^2, \hat{R}_t^3) & \hat{\rho}(\hat{R}_t^2, \hat{R}_t^4) \\ \hat{\rho}(\hat{R}_t^3, \hat{R}_t^1) & \hat{\rho}(\hat{R}_t^3, \hat{R}_t^2) & 1 & \hat{\rho}(\hat{R}_t^3, \hat{R}_t^4) \\ \hat{\rho}(\hat{R}_t^4, \hat{R}_t^1) & \hat{\rho}(\hat{R}_t^4, \hat{R}_t^2) & \hat{\rho}(\hat{R}_t^4, \hat{R}_t^3) & 1 \end{pmatrix}$$

All the covariance terms in the matrix A can be estimated using the approach proposed by Berger and Priam (2016).

Variance estimation for annual net change estimates

In order to calculate annual net changes, let

$$\hat{R}_t^l = \frac{\hat{Z}_t^l}{\hat{Y}_t^l} = \frac{\sum_{i \in s_{t,l}} y_{t,i}^l \cdot w_{t,i}^l}{\sum_{i \in s_{t,l}} z_{t,i}^l \cdot w_{t,i}^l},$$

be the estimated ratio at the quarter $l \in 1, 2, 3, 4$ of year $t \in 1, 2$. Assuming that the size of the active population remains (almost) unchanged from one quarter to another within the same year, we obtain:

$$\hat{\theta}_1 = \frac{\hat{R}_1^1 \hat{Z}_1^{(1)} + \hat{R}_1^2 \hat{Z}_1^2 + \hat{R}_1^3 \hat{Z}_1^3 + \hat{R}_1^4 \hat{Z}_1^4}{\hat{Z}_1^{(1)} + \hat{Z}_1^2 + \hat{Z}_1^3 + \hat{Z}_1^4},$$

Variance estimation for annual net change estimates

$$\hat{\theta}_t = \frac{\hat{R}_2^1 \hat{Z}_2^1 + \hat{R}_2^2 \hat{Z}_2^2 + \hat{R}_2^3 \hat{Z}_2^3 + \hat{R}_2^4 \hat{Z}_2^4}{\hat{Z}_2^1 + \hat{Z}_2^2 + \hat{Z}_2^3 + \hat{Z}_2^4},$$

The estimate of the annual net changes is written as:

$$\hat{\Delta} = \hat{\theta}_2 - \hat{\theta}_1$$

The variance of the annual net changes estimates $\hat{\Delta}$ can be written as:

$$\text{Var}(\hat{\Delta}) = \frac{1-f}{16} X^T A X$$

Variance estimation for annual net change estimates

$$X = \begin{pmatrix} \sqrt{\text{Var}(\hat{R}_t^1)} \\ \sqrt{\text{Var}(\hat{R}_t^2)} \\ \sqrt{\text{Var}(\hat{R}_t^3)} \\ \sqrt{\text{Var}(\hat{R}_t^4)} \\ \sqrt{\text{Var}(\hat{R}_t^5)} \\ \sqrt{\text{Var}(\hat{R}_t^6)} \\ \sqrt{\text{Var}(\hat{R}_t^7)} \\ \sqrt{\text{Var}(\hat{R}_t^8)} \end{pmatrix}$$

Variance estimation for annual net change estimates

$$A = \begin{pmatrix} \widehat{\widehat{1}} & \widehat{\widehat{\rho(R_1^1, R_2^1)}} & \widehat{\widehat{\rho(R_1^1, R_3^1)}} & \widehat{\widehat{\rho(R_1^1, R_4^1)}} & -\widehat{\widehat{\rho(R_1^1, R_2^1)}} & -\widehat{\widehat{\rho(R_1^1, R_2^2)}} & -\widehat{\widehat{\rho(R_1^1, R_2^3)}} & -\widehat{\widehat{\rho(R_1^1, R_2^4)}} \\ \widehat{\widehat{\rho(R_2^1, R_1^1)}} & \widehat{\widehat{1}} & \widehat{\widehat{\rho(R_2^1, R_3^1)}} & \widehat{\widehat{\rho(R_2^1, R_4^1)}} & -\widehat{\widehat{\rho(R_2^1, R_2^1)}} & -\widehat{\widehat{\rho(R_2^1, R_2^2)}} & -\widehat{\widehat{\rho(R_2^1, R_2^3)}} & -\widehat{\widehat{\rho(R_2^1, R_2^4)}} \\ \widehat{\widehat{\rho(R_3^1, R_1^1)}} & \widehat{\widehat{\rho(R_3^1, R_2^1)}} & \widehat{\widehat{1}} & \widehat{\widehat{\rho(R_3^1, R_4^1)}} & -\widehat{\widehat{\rho(R_3^1, R_2^1)}} & -\widehat{\widehat{\rho(R_3^1, R_2^2)}} & -\widehat{\widehat{\rho(R_3^1, R_2^3)}} & -\widehat{\widehat{\rho(R_3^1, R_2^4)}} \\ \widehat{\widehat{\rho(R_4^1, R_1^1)}} & \widehat{\widehat{\rho(R_4^1, R_2^1)}} & \widehat{\widehat{\rho(R_4^1, R_3^1)}} & \widehat{\widehat{1}} & -\widehat{\widehat{\rho(R_4^1, R_2^1)}} & -\widehat{\widehat{\rho(R_4^1, R_2^2)}} & -\widehat{\widehat{\rho(R_4^1, R_2^3)}} & -\widehat{\widehat{\rho(R_4^1, R_2^4)}} \\ \widehat{\widehat{\rho(R_4^1, R_1^1)}} & \widehat{\widehat{\rho(R_4^1, R_2^1)}} & \widehat{\widehat{\rho(R_4^1, R_3^1)}} & \widehat{\widehat{\rho(R_4^1, R_4^1)}} & -\widehat{\widehat{\rho(R_4^1, R_2^1)}} & -\widehat{\widehat{\rho(R_4^1, R_2^2)}} & -\widehat{\widehat{\rho(R_4^1, R_2^3)}} & -\widehat{\widehat{\rho(R_4^1, R_2^4)}} \\ -\widehat{\widehat{\rho(R_2^1, R_1^1)}} & -\widehat{\widehat{\rho(R_2^1, R_2^1)}} & -\widehat{\widehat{\rho(R_2^1, R_3^1)}} & -\widehat{\widehat{\rho(R_2^1, R_4^1)}} & \widehat{\widehat{1}} & \widehat{\widehat{\rho(R_2^1, R_2^2)}} & \widehat{\widehat{\rho(R_2^1, R_2^3)}} & \widehat{\widehat{\rho(R_2^1, R_2^4)}} \\ -\widehat{\widehat{\rho(R_2^2, R_1^1)}} & -\widehat{\widehat{\rho(R_2^2, R_2^1)}} & -\widehat{\widehat{\rho(R_2^2, R_3^1)}} & -\widehat{\widehat{\rho(R_2^2, R_4^1)}} & \widehat{\widehat{\rho(R_2^2, R_2^1)}} & \widehat{\widehat{1}} & \widehat{\widehat{\rho(R_2^2, R_2^2)}} & \widehat{\widehat{\rho(R_2^2, R_2^3)}} \\ -\widehat{\widehat{\rho(R_2^3, R_1^1)}} & -\widehat{\widehat{\rho(R_2^3, R_2^1)}} & -\widehat{\widehat{\rho(R_2^3, R_3^1)}} & -\widehat{\widehat{\rho(R_2^3, R_4^1)}} & \widehat{\widehat{\rho(R_2^3, R_2^1)}} & \widehat{\widehat{\rho(R_2^3, R_2^2)}} & \widehat{\widehat{1}} & \widehat{\widehat{\rho(R_2^3, R_2^3)}} \\ -\widehat{\widehat{\rho(R_2^4, R_1^1)}} & -\widehat{\widehat{\rho(R_2^4, R_2^1)}} & -\widehat{\widehat{\rho(R_2^4, R_3^1)}} & -\widehat{\widehat{\rho(R_2^4, R_4^1)}} & \widehat{\widehat{\rho(R_2^4, R_2^1)}} & \widehat{\widehat{\rho(R_2^4, R_2^2)}} & \widehat{\widehat{\rho(R_2^4, R_2^3)}} & \widehat{\widehat{1}} \end{pmatrix}$$

Function vardannual description

Based on the description of the methodology given in the previous sections, a function vardannual in software R package vardpoor (Breidaks, Liberts and Ivanova, 2019) was created for variance estimation for measuring annual net change or annual for single and multistage stage cluster sampling designs. This function vardannual one of output is table annual_results, which is a data.table containing:

- year - survey years of years for measures of annual;
- year_1 - survey years of years1 for measures of annual net change;
- year_2 - survey years of years2 for measures of annual net change;
- country - survey countries;
- Dom - optional variable of the population domains;

Function vardannual description

- namesY - variable with names of variables of interest;
- namesZ - optional variable with names of denominator for ratio estimation;
- estim_1 - the estimated value for period1 for measures of annual net change;
- estim_2 - the estimated value for period2 for measures of annual net change;
- estim - the estimated value;
- var - the estimated variance;
- se - the estimated standard error;
- rse - the estimated relative standard error (coefficient of variation),
- cv - the estimated relative standard error (coefficient of variation) in percentage,

Function vardannual description

- `absolute_margin_of_error` - the estimated absolute margin of error;
- `relative_margin_of_error` - the estimated relative margin of error in percentage;
- `CI_lower` - the estimated confidence interval lower bound;
- `CI_upper` - the estimated confidence interval upper bound;
- `significant` - is the the difference significant.

vardannual function testing results

Function was tested on Latvia data of LFS 2017 and 2018. The function `vardannual()` is used for the variance estimation of annual net changes with and without calibration.

vardannual function testing results

Calib	Indicator	Difference in p.p.	Variance	Lower CI	Upper CI
Without	Employment rate 20-64, total, in %	2.07	0.25	1.09	3.06
Without	Employment rate 20-64, men, in %	2.29	0.42	0.77	3.52
Without	Employment rate 20-64, women, in %	2.09	0.47	0.74	3.43
With	Employment rate 20-64, total, in %	2.07	0.13	1.36	2.79
With	Employment rate 20-64, men, in %	2.05	0.24	1.89	3.01
With	Employment rate 20-64, women, in %	2.09	0.24	1.12	3.05

The annual net changes of 3 LFS indicators were calculated with the confidence interval, and if the confidence interval for the difference did not include 0, then we were able to conclude that the difference was statistically significant.

Conclusion

- The function *vardannual* in software R package *vardpoor* was implemented in practice
- This function was tested on real data. The function *vardannual* is important to assess quality of LFS estimates and statistical significance of the estimates.
- Variance estimates for annual point estimates and net changes was estimated for 3 LFS indicators using the function *vardannual*.
- When looking at the results with calibration, it can be identified that the confidence interval is narrower than the results without calibration.

Biography

1. BERGER, Y., OSIER, G, GOEDEMÉ, T. (2017). Standard error estimation and related sampling issue, Monitoring social inclusion in Europe (Eurostat), pp. 465 – 480
2. BERGER, Y. G. and PRIAM, R. (2016), A simple variance estimator of change for rotating repeated surveys: an application to the EU-SILC household surveys, University of Southampton, Statistical Sciences Research Institute. Available at <http://eprints.soton.ac.uk/347142>
3. BERGER, Y. G. and OGUZ ALPER, M. (2013), Variance estimation of change of poverty based upon the Turkish EU-SILC survey, paper presented at the NTTS (New Techniques and Technologies for Statistics) Conference, Brussels, 5-7 March 2013

Biography

4. BREIDAKS J., LIBERTS M., IVANOVA, S. (2019). vardpoor: Variance Estimation for Sample Surveys by the Ultimate Cluster, R package version 0.15.0., URL <http://cran.r-project.org/web/packages/vardpoor/index.html>
5. OSIER G. RAYMOND V., (2015) Development of methodology for the estimate of variance of annual net changes for LFS-based indicators. Deliverable 1 - Short document with derivation of the methodology (FINAL), SOGETI
6. OSIER G., PERRAY P., (2016). Variance estimators of annual levels and net changes for a defined set of LFS-based indicators.

Thank you!

[https : //cran.r – project.org/web/packages/vardpoor/index.html](https://cran.r-project.org/web/packages/vardpoor/index.html)

[https : //csblatvia.github.io/vardpoor](https://csblatvia.github.io/vardpoor)