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

Juris Breidaks
(e-mail: Juris.Breidaks@csb.gov.lv)
Central Statistical Bureau of Latvia

ABSTRACT

The paper is devoted to the function *vardannual from R package vardpoor. The Central Statistical Bureau of Latvia in 2017 has developed the function vardannual which is included in the R package vardpoor. In the paper describes the variance estimation of quartely estimates, correlation estimation of two quarter change estimates, and finally it explains how to extend the approach to deal with variance estimation for annual point estimates and net changes for Labour Force Survey (LFS) indicators. Variance estimates for annual point estimates and net changes was estimated for LFS indicators using the function "vardannual". This function was tested on simulated and real data. The function "vardannual" is important to assess quality of LFS estimates and statistical significance of the estimates. The annual net changes of all indicators are calculated with the confidence interval, and if the confidence interval for the difference is not equal to 0, then we are able to conclude that the difference is statistically significant. When looking at the results with calibration, it can be identified that the confidence interval is narrower than the results without calibration. The function "vardannual" in software R package "vardpoor" was implemented in practice, LFS in Latvia.*

Keywords: *Survey sampling, ratio, net changes, annual point estimates, variance estimation.*

JEL Classification: C400, C420

1. 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 applyed for the following LFS indicators:

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

The theoretical basis of *vardannual* was borrowed from the Guillaume Osier and Virginie Raymond (2015) article "Development of methodology for the estimate of variance of annual net changes for LFS-based indicators" and Guillaume Osier and Pauline Perray (2016) article "Variance estimators of annual levels and net changes for a defined set of LFS-based indicators". 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 by Eurostat (2013).

2. METHODOLOGY

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.

In the next sections describe the variance estimation of quartely estimates, correlation estimation of two quarter change estimates, and finally it explains how to extend the approach to deal with variance estimation for annual point estimates and net changes for LFS-based indicators

2.1. VARIANCE ESTIMATION OF QUARTERLY ESTIMATES

The estimate of the ratio is calculated as:

$$\hat{\theta}_t = \frac{\hat{Y}_t}{\hat{Z}_t} = \frac{\sum_{i \in s_t} w_{t,i} \cdot y_{t,i}}{\sum_{i \in s_t} w_{t,i} \cdot z_{t,i}}, \qquad [1]$$

where s_t denotes the sample of individuals at quarter t, the $w_{t,i}$ 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 $\hat{\theta}_t$ 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:

$$\hat{\theta}_t = f(\hat{\tau}_t) \cong \sum_{i \in s_t} w_{t,i} \cdot u_{t,i} + R_{t,n} \qquad [2]$$

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 $\hat{\tau}_t = \sum_{i \in s} w_{t,i} \cdot u_{t,i}$

$$Var(\hat{\theta}_t) \cong Var(\hat{\tau}_t) \quad [3]$$

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

$$\hat{u}_{t,i} = \frac{1}{\hat{Z}_t}\left(y_{t,i} - \hat{\theta}_t z_{t,i}\right). \quad [4]$$

As the weights are calibrated, the calibration residual estimates $\hat{\varphi}_{t,k}$ are calculated by Lundström and Särndal (2001) formula

$$\hat{\varphi}_{t,k} = \sum_{i \in household_k} \hat{u}_{t,i} - x'_{t,k}\hat{B}_t \quad [5]$$

where

$$\hat{B}_t = \left(\sum_s d_{t,k} q_{t,k} x_{t,k} x'_{t,k}\right)^{-1}\left(\sum_s d_{t,k} q_{t,k} x_{t,k} \sum_{i \in household_k} \hat{u}_{t,i}\right). \quad [6]$$

X_t – the matrix of auxiliary variables in quarter t,

$q_{t,k}$ – variable, which is vector of positive values accounting for heteroscedasticity.

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.

$$\check{y}_{t,j} = \sum_{k \in s_t \cap PSU_{t,j}} w_{t,k} \cdot \begin{cases} \hat{\varphi}_{t,k}, & \textit{if calibration is used} \\ \sum_{i \in household_k} \hat{u}_{t,i}, & \textit{if calibration is not used} \end{cases} \quad [7]$$

where $j = \overline{1,m}$, m - denotes total of PSUs at quarter t, PSU_j where the subscript j denotes a PSU at quarter t.

Variance estimation is calculated for linearized variable by ultimate cluster approach:

$$\widehat{var}(\hat{\tau}_t) = \sum_{h=1}^{H_t} \frac{n_{t,h}}{n_{t,h}-1} \sum_{j=1}^{n_{t,h}} \left(\check{y}_{t,j} - \bar{y}_{t,h\bullet}\right)^2, \quad [8]$$

where $\bar{y}_{t,h\bullet} = \frac{\left(\sum_{j=1}^{n_{t,h}} \check{y}_{t,j}\right)}{n_{t,h}}$.

h – the stratum label, with a total of H strata in quarter t.

j – the label of the PSU within stratum h, with a total of $n_{t,h}$ PSUs in quarter t.

Variance estimation is implemented with ultimate cluster approach and regression model, where difference $(\breve{y}_{t,j} - \bar{y}_{t,h\bullet})^2$ is regarded as the estimated residuals of regression model.

Regression model:

$$\breve{y}_{t,j} = \sum_{h=1}^{H} \beta_{t,h} \cdot z_{th,j} + (\varepsilon_{t,j}), \qquad [9]$$

where $\breve{y}_{t,j}$ was defined by (5) and dummy component $z_{th,j}$ is given by:

$$z_{th,j} = \begin{cases} 1, & PSU_j \in s_{t,h} \\ 0, & PSU_j \notin s_{t,h} \end{cases}, \qquad [10]$$

where $s_{t,h}$ is sample for PSU within stratum h at quarter t.

Estimated residuals are used in variance estimator for variable.

$$\widehat{var}(\hat{\tau}_t) = \sum_{h=1}^{H} \frac{n_{t,h}}{n_{t,h}-1} \sum_{j=1}^{n_{t,h}} (\hat{\varepsilon}_{t,j})^2, \qquad [11]$$

2.2. 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 by Särndal et al (1992) and 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.

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

$$\hat{\Delta} = \hat{\theta}_2 - \hat{\theta}_1, \qquad [12]$$

where $\hat{\theta}_t$ are the standard estimator of $\hat{\theta}_t$ 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

$$\hat{\theta}_t = \frac{\sum_{i \in s_t} w_{t,i} \cdot y_{t,i}}{\sum_{i \in s_t} w_{t,i} \cdot z_{t,i}}, \quad t = 1, 2 \qquad [13]$$

Let $\hat{\theta}_t$ be a non-linear statistic and let us assume that $\hat{\theta}_t$ 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:

$$\hat{\theta}_t = f(\hat{\tau}_t) \cong \sum_{k \in s_t} w_{t,k} \cdot \varphi_{t,k} + R_{t,n} \qquad [14]$$

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 $\hat{\theta}_t$ for both t value. We obtain the following formula:

$$\hat{\varphi}_{t,k} = \begin{cases} \sum_{i\in household_k} \hat{u}_{t,i} - x'_{t,k}\ddot{B}_t, & if\ variable\ is\ linearized\ and\ calibration\ is\ used \\ \sum_{i\in household_k} \hat{u}_{t,i}, & if\ variable\ is\ linearized \end{cases} \quad [15]$$

where

$$\hat{u}_{t,i} = \frac{1}{\hat{Z}_t}\left(y_{t,i} - \hat{\theta}_t z_{t,i}\right) \quad [16]$$

$$\hat{B}_t = \left(\sum_s d_{t,k} q_{t,k} x_{t,k}\, x'_{t,k}\right)^{-1}\left(\sum_s d_{t,k} q_{t,k} x_{t,k} \sum_{i\in household_k} \hat{u}_{t,i}\right) \quad [17]$$

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

$$\begin{aligned} &Var(\hat{\Delta}) = Var(\hat{\theta}_2 - \hat{\theta}_1) \cong Var(\hat{\tau}_2 - \hat{\tau}_1) = \\ &= Var(\hat{\tau}_1) + Var(\hat{\tau}_2) - 2 \cdot Cov(\hat{\tau}_2, \hat{\tau}_1) = \\ &= Var(\hat{\tau}_1) + Var(\hat{\tau}_2) - 2 \cdot \sqrt{Var(\hat{\tau}_1)} \cdot \sqrt{Var(\hat{\tau}_2)} \cdot \rho(\hat{\tau}_2, \hat{\tau}_1), \end{aligned} \quad [18]$$

The variance estimates of $Var(\hat{\tau}_1)$ and $Var(\hat{\tau}_2)$ is calculated using the estimator proposed by Net-SILC2 (Berger et al, 2016).

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 (2013) 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:

$$\begin{pmatrix} \varphi_{1j} \\ \varphi_{2j} \end{pmatrix} = \sum_{h=1}^{H} \begin{pmatrix} \beta_h^{(1)} \cdot z_{1h,j} + \beta_h^{(2)} \cdot z_{2h,j} + \beta_h^{(12)} \cdot z_{1h,j} \cdot z_{2h,j} \\ \gamma_h^{(1)} \cdot z_{1h,j} + \gamma_h^{(2)} \cdot u_{2h,j} + \gamma_h^{(12)} \cdot z_{1h,j} \cdot z_{2h,j} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,j} \\ \varepsilon_{2,j} \end{pmatrix}, \quad [19]$$

where $j \in s = s_1 \cup s_2$, and the residuals $\varepsilon_j = \left(\varepsilon_{1,j}, \varepsilon_{2,j}\right)^T$ follow a bivariate distribution with mean zero and an unknown variance-covariance matrix. The The $z_{1h,j}$ and $z_{2h,j}$ are (dummy) design variables which specify the stratification:

$$z_{1h,j} = \begin{cases} 1, if\ j\ \in s_{1h} \\ 0, otherwise' \end{cases} \quad [20]$$

$$z_{2h,j} = \begin{cases} 1, if\ j\ \in s_{2h} \\ 0, otherwise' \end{cases} \quad [21]$$

The quantities $\beta_h^{(1)}$, $\beta_h^{(2)}$, $\gamma_h^{(1)}$, $\gamma_h^{(2)}$, $\beta_h^{(12)}$ and $\gamma_h^{(12)}$ are regression parameters that need to be included into the model. The estimated regression residuals ε_{1j} and ε_{2j} gives an estimate of the correlation $\hat{\rho}(\hat{\tau}_2, \hat{\tau}_1)$ between $\hat{\tau}_1$ and $\hat{\tau}_2$, which can be respectively defined by:

$$\hat{\rho}(\hat{\tau}_2, \hat{\tau}_1) = \frac{\sum_{j\in s_1 \cup s_2} \hat{\varepsilon}_{1;j} \cdot \hat{\varepsilon}_{2;j}}{\sqrt{\sum_{j\in s_1 \cup s_2} (\hat{\varepsilon}_{1;j})^2} \cdot \sqrt{\sum_{j\in s_1 \cup s_2} (\hat{\varepsilon}_{2;j})^2}}, \quad [22]$$

2.3. VARIANCE ESTIMATION FOR ANNUAL POINT ESTIMATES

In order to calculate annual net changes, let

$$\hat{R}_t^{(l)} = \frac{\hat{Y}_t^{(l)}}{\hat{Z}_t^{(l)}} = \frac{\sum_{i\in s_{t,l}} w_{t,i}^{(l)} \cdot y_{t,i}^{(l)}}{\sum_{i\in s_{t,l}} w_{t,i}^{(l)} \cdot z_{t,i}^{(l)}} \quad [23]$$

be the estimated ratio at the quarter l (l Î {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:

$$\hat{\theta}_t = \frac{\hat{R}_t^{(1)}\hat{Z}_t^{(1)} + \hat{R}_t^{(1)}\hat{Z}_t^{(2)} + \hat{R}_t^{(1)}\hat{Z}_t^{(3)} + \hat{R}_t^{(1)}\hat{Z}_t^{(4)}}{\hat{Z}_t^{(1)} + \hat{Z}_t^{(2)} + \hat{Z}_t^{(3)} + \hat{Z}_t^{(4)}} \cong \frac{\hat{R}_t^{(1)} + \hat{R}_t^{(2)} + \hat{R}_t^{(3)} + \hat{R}_t^{(4)}}{4} \quad [24]$$

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

$$Var(\hat{\theta}_t) = \frac{1-f}{16} X_t^T A_t X_t \quad [25]$$

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

$$X_t = \begin{pmatrix} \sqrt{Var(\hat{R}_t^{(1)})} \\ \sqrt{Var(\hat{R}_t^{(2)})} \\ \sqrt{Var(\hat{R}_t^{(3)})} \\ \sqrt{Var(\hat{R}_t^{(4)})} \end{pmatrix} \quad [26]$$

$$A_t = \begin{pmatrix} 1 & \hat{\rho}(\hat{R}_t^{(1)}, \hat{R}_t^{(2)}) & \hat{\rho}(\hat{R}_t^{(1)}, \hat{R}_t^{(3)}) & \hat{\rho}(\hat{R}_t^{(1)}, \hat{R}_t^{(4)}) \\ \hat{\rho}(\hat{R}_t^{(2)}, \hat{R}_t^{(1)}) & 1 & \hat{\rho}(\hat{R}_t^{(2)}, \hat{R}_t^{(4)}) & \hat{\rho}(\hat{R}_t^{(2)}, \hat{R}_t^{(4)}) \\ \hat{\rho}(\hat{R}_t^{(3)}, \hat{R}_t^{(1)}) & \hat{\rho}(\hat{R}_t^{(3)}, \hat{R}_t^{(2)}) & 1 & \hat{\rho}(\hat{R}_t^{(3)}, \hat{R}_t^{(4)}) \\ \hat{\rho}(\hat{R}_t^{(4)}, \hat{R}_t^{(1)}) & \hat{\rho}(\hat{R}_t^{(4)}, \hat{R}_t^{(2)}) & \hat{\rho}(\hat{R}_t^{(4)}, \hat{R}_t^{(3)}) & 1 \end{pmatrix} \quad [27]$$

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

2.4. VARIANCE ESTIMATION FOR ANNUAL NET CHANGE ESTIMATES

In order to calculate annual point estimates, let

$$\hat{R}_t^{(l)} = \frac{\hat{Y}_t^{(l)}}{\hat{Z}_t^{(l)}} = \frac{\sum_{i \in s_{t,l}} w_{t,i}^{(l)} \cdot y_{t,i}^{(l)}}{\sum_{i \in s_{t,l}} w_{t,i}^{(l)} \cdot z_{t,i}^{(l)}} \qquad [28]$$

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

$$\hat{\theta}_1 = \frac{\hat{R}_1^{(1)}\hat{Z}_1^{(1)} + \hat{R}_1^{(1)}\hat{Z}_1^{(2)} + \hat{R}_1^{(1)}\hat{Z}_1^{(3)} + \hat{R}_1^{(1)}\hat{Z}_1^{(4)}}{\hat{Z}_1^{(1)} + \hat{Z}_1^{(2)} + \hat{Z}_1^{(3)} + \hat{Z}_1^{(4)}} \cong \frac{\hat{R}_1^{(1)} + \hat{R}_1^{(2)} + \hat{R}_1^{(3)} + \hat{R}_1^{(4)}}{4} \qquad [29]$$

$$\hat{\theta}_2 = \frac{\hat{R}_2^{(1)}\hat{Z}_2^{(1)} + \hat{R}_2^{(1)}\hat{Z}_2^{(2)} + \hat{R}_2^{(1)}\hat{Z}_2^{(3)} + \hat{R}_1^{(1)}\hat{Z}_1^{(4)}}{\hat{Z}_2^{(1)} + \hat{Z}_2^{(2)} + \hat{Z}_2^{(3)} + \hat{Z}_2^{(4)}} \cong \frac{\hat{R}_1^{(1)} + \hat{R}_1^{(2)} + \hat{R}_1^{(3)} + \hat{R}_1^{(4)}}{4} \qquad [30]$$

The estimate of the annual net changes is written as:

$$\hat{\Delta} = \hat{\theta}_2 - \hat{\theta}_1 \qquad [31]$$

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

$$\mathrm{Var}(\hat{\Delta}) = \frac{1-f}{16} X^T A X \qquad [32]$$

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

$$X = \begin{pmatrix} \sqrt{Var(\hat{R}_1^{(1)})} \\ \sqrt{Var(\hat{R}_1^{(2)})} \\ \sqrt{Var(\hat{R}_1^{(3)})} \\ \sqrt{Var(\hat{R}_1^{(4)})} \\ \sqrt{Var(\hat{R}_2^{(1)})} \\ \sqrt{Var(\hat{R}_2^{(2)})} \\ \sqrt{Var(\hat{R}_2^{(3)})} \\ \sqrt{Var(\hat{R}_2^{(4)})} \end{pmatrix} \qquad [33]$$

$$A = \begin{pmatrix} 1 & \rho(\hat{R}_1^{(1)},\hat{R}_1^{(2)}) & \rho(\hat{R}_1^{(1)},\hat{R}_1^{(3)}) & \rho(\hat{R}_1^{(1)},\hat{R}_1^{(4)}) & -\rho(\hat{R}_1^{(1)},\hat{R}_2^{(1)}) & -\rho(\hat{R}_1^{(1)},\hat{R}_2^{(2)}) & -\rho(\hat{R}_1^{(1)},\hat{R}_2^{(3)}) & -\rho(\hat{R}_1^{(1)},\hat{R}_2^{(4)}) \\ \rho(\hat{R}_1^{(2)},\hat{R}_1^{(1)}) & 1 & \rho(\hat{R}_1^{(2)},\hat{R}_1^{(3)}) & \rho(\hat{R}_1^{(2)},\hat{R}_1^{(4)}) & -\rho(\hat{R}_1^{(2)},\hat{R}_2^{(1)}) & -\rho(\hat{R}_1^{(2)},\hat{R}_2^{(2)}) & -\rho(\hat{R}_1^{(2)},\hat{R}_2^{(3)}) & -\rho(\hat{R}_1^{(2)},\hat{R}_2^{(4)}) \\ \rho(\hat{R}_1^{(3)},\hat{R}_1^{(1)}) & \rho(\hat{R}_1^{(3)},\hat{R}_1^{(2)}) & 1 & \rho(\hat{R}_1^{(3)},\hat{R}_1^{(4)}) & -\rho(\hat{R}_1^{(3)},\hat{R}_2^{(1)}) & -\rho(\hat{R}_1^{(3)},\hat{R}_2^{(2)}) & -\rho(\hat{R}_1^{(3)},\hat{R}_2^{(3)}) & -\rho(\hat{R}_1^{(3)},\hat{R}_2^{(4)}) \\ \rho(\hat{R}_1^{(4)},\hat{R}_1^{(1)}) & \rho(\hat{R}_1^{(4)},\hat{R}_1^{(2)}) & \rho(\hat{R}_1^{(4)},\hat{R}_1^{(3)}) & 1 & -\rho(\hat{R}_1^{(4)},\hat{R}_2^{(1)}) & -\rho(\hat{R}_1^{(4)},\hat{R}_2^{(2)}) & -\rho(\hat{R}_1^{(4)},\hat{R}_2^{(3)}) & -\rho(\hat{R}_1^{(4)},\hat{R}_2^{(4)}) \\ -\rho(\hat{R}_2^{(1)},\hat{R}_1^{(1)}) & -\rho(\hat{R}_2^{(1)},\hat{R}_1^{(2)}) & -\rho(\hat{R}_2^{(1)},\hat{R}_1^{(3)}) & -\rho(\hat{R}_2^{(1)},\hat{R}_1^{(4)}) & 1 & \rho(\hat{R}_2^{(1)},\hat{R}_2^{(2)}) & \rho(\hat{R}_2^{(1)},\hat{R}_2^{(3)}) & \rho(\hat{R}_2^{(1)},\hat{R}_2^{(4)}) \\ -\rho(\hat{R}_2^{(2)},\hat{R}_1^{(1)}) & -\rho(\hat{R}_2^{(2)},\hat{R}_1^{(2)}) & -\rho(\hat{R}_2^{(2)},\hat{R}_1^{(3)}) & -\rho(\hat{R}_2^{(2)},\hat{R}_1^{(4)}) & \rho(\hat{R}_2^{(2)},\hat{R}_2^{(1)}) & 1 & \rho(\hat{R}_2^{(2)},\hat{R}_2^{(3)}) & \rho(\hat{R}_2^{(2)},\hat{R}_2^{(4)}) \\ -\rho(\hat{R}_2^{(3)},\hat{R}_1^{(1)}) & -\rho(\hat{R}_2^{(3)},\hat{R}_1^{(2)}) & -\rho(\hat{R}_2^{(3)},\hat{R}_1^{(3)}) & -\rho(\hat{R}_2^{(3)},\hat{R}_1^{(4)}) & \rho(\hat{R}_2^{(3)},\hat{R}_2^{(1)}) & \rho(\hat{R}_2^{(3)},\hat{R}_2^{(2)}) & 1 & \rho(\hat{R}_2^{(3)},\hat{R}_2^{(4)}) \\ -\rho(\hat{R}_2^{(4)},\hat{R}_1^{(1)}) & -\rho(\hat{R}_2^{(4)},\hat{R}_1^{(2)}) & -\rho(\hat{R}_2^{(4)},\hat{R}_1^{(3)}) & -\rho(\hat{R}_2^{(4)},\hat{R}_1^{(4)}) & \rho(\hat{R}_2^{(4)},\hat{R}_2^{(1)}) & \rho(\hat{R}_2^{(4)},\hat{R}_2^{(2)}) & \rho(\hat{R}_2^{(4)},\hat{R}_2^{(3)}) & 1 \end{pmatrix} \quad [34]$$

3. R FUNCTION "VARDANNUAL"

3.1. R function "vardannual" description

Based on the description of the methodology given in the previous sections, a function "vardannual" in software R package "vardpoor" by Breidaks et al (2018) 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;
- ♦ 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,
- ♦ 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.

3.2. R function "vardannual" results

Function was tested on Latvia data of LFS 2016 and 2017. The function vardannual() is used for the variance estimation of annual net changes without calibration:

```
Rez <- vardannual(Y = c("emp_tot", "emp_men", "emp_women"), H = "STRATA",
                PSU = "PRIMARY_SAMPING_UNIT", w_final = ""
                ID_level1 = "ID_houshold", ID_level2 = "ID_unit", Dom = NULL,
                Z  =  c("tot",  "tot_men",  "tot_women"),  gender  =  NULL,
                country = NULL,
                years = "years", subperiods = "quarters", dataset = LFS,
                year1 = 2016,
                year2 = 2017, X = NULL, countryX = NULL, yearsX = NULL,
                subperiodsX = NULL, X_ID_level1 = NULL,  ind_gr = NULL,
                g = NULL, q = NULL, datasetX = NULL, frate = 18.69,
                percentratio = 1,  use.estVar = FALSE, use.gender = FALSE,
                confidence = 0.95, method = "netchanges")
```

3 LFS indicators calculations without calibration

Table 1

Indicator	2016	2017	Difference in p.p.	Variance of Net changes 2016/2017	Lower CI	Upper CI	Significant
Employment rate 20-64, total, in %	73.23	74.76	1.54	0.29	0.48	2.59	YES
Employment rate 20-64, men, in %	74.71	77.00	2.29	0.56	0.83	3.76	YES
Employment rate 20-64, women, in %	71.84	72.68	0.83	0.50	-0.56	2.23	NO

In table 1 the variance of annual net changes without calibration between 2016 and 2017 was calculated for 3 LFS indicators (the sampling rate *f* is 18.69 %). 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.

The function vardannual() is used for the variance estimation of annual net changes with calibration:

```
Rez <- vardannual(Y = c("emp_tot", "emp_men", "emp_women"), H = "STRATA",
                PSU = "PRIMARY_SAMPING_UNIT", w_final = ""
                ID_level1 = "ID_houshold", ID_level2 = "ID_unit", Dom = NULL,
                Z  =  c("tot",  "tot_men",  "tot_women"),  gender  =  NULL,
                country = NULL,
                years = "years", subperiods = "quarters", dataset = LFS,
                year1 = 2016,
                year2 = 2017, X = calib_names, countryX = NULL, yearsX = "years",
                subperiodsX  =  "quarters",  X_ID_level1  =  "ID_houshold",
                ind_gr = NULL,
                g = "g", q = NULL, datasetX = LFS_cal, frate = 18.69,
                percentratio = 1,  use.estVar = FALSE, use.gender = FALSE,
                confidence = 0.95, method = "netchanges")
```

LFS indicators calculations with calibration

Table 3

Indicator	2016	2017	Difference in p.p.	Variance of Net changes 2016/2017	Lower CI	Upper CI	Significant
Employment rate 20-64, total, in %	73.23	74.76	1.54	0.12	0.86	2.22	YES
Employment rate 20-64, men, in %	74.71	77.00	2.29	0.27	1.28	3.31	YES
Employment rate 20-64, women, in %	71.84	72.68	0.83	0.21	-0.06	1.73	NO

In table 2 the variance of annual net changes with calibration between 2016 and 2017 was calculated for 3 LFS indicators (the sampling rate f is 18.69 %). Looking on results with calibration, it could be seen that confidence interval was narrower than results without calibration.

4. CONCLUSIONS

The function “vardannual” in software R package “vardpoor” was implemented in practice. Variance estimates for annual point estimates and net changes was estimated for 3 LFS indicators using the function “vardannual”. This function was tested on real data. The function “vardannual” is important to assess quality of LFS estimates and statistical significance of the estimates. The annual net changes of 3 LFS indicators are calculated with the confidence interval, and if the confidence interval for the difference is not equal to 0, then we are able to conclude that the difference is statistically significant. When looking at the results with calibration, it can be identified that the confidence interval is narrower than the results without calibration.

References

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.**, 2013, *“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.
4. **BREIDAKS J., LIBERTS M., IVANOVA, S.**, 2018, *vardpoor: Variance Estimation for Sample Surveys by the Ultimate Cluster*, R package version 0.12.0., URL ***http://cran.r-project.org/web/packages/vardpoor/index.html***
5. **DEVILLE J. C.**, 1999, *Variance estimation for complex statistics and estimators: linearization and residual techniques*. Survey Methodology, 25, 193-203, ***http://www5.statcan.gc.ca/bsolc/olc-cel/olc-cel?lang=eng&catno=12-001-X19990024882***

6. **EUROSTAT**, 2013, *"Standard error estimation for the EU-SILC indicators of poverty and social exclusion"*, Eurostat Methodologies and Working papers.
7. **LUNDSTRÖM S., SÄRNDAL C. E.**, 2001, *Estimation in the presence of Nonresponse and Frame Imperfections*. Statistics Sweden, 43-44.
8. **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
9. **OSIER G., PERRAY P.**, 2016, *Variance estimators of annual levels and net changes for a defined set of LFS-based indicators*.
10. **SÄRNDAL, C. E., SWENSSON, B., WRETMAN, J.**, 1992, *Model Assissted Survey Sampling*, 176-181, Springer-Verlag.