# Accuracy investigation of the composite estimators in the case of sample rotation for two-phase sampling scheme

Chadysas Viktoras[1]

[1]Vilnius Gediminas Technical University, e-mail: viktoras.chadysas@vgtu.lt

**Abstract**

In this paper we focus on construction of the combined ratio type estimator of the finite population total and their variances in the case of sample rotation for two-phase sampling scheme. We construct composite estimators of the finite population total without and with the use of auxiliary information known from the previous survey. Two types of sampling design are used for sample selection in each of the phases: simple random sampling without replacement and successive sampling (unequal probability sampling without replacement). A simulation study, based on the real population data, is performed and the proposed estimators are compared.
*Keywords*: Sample rotation, ratio estimator, composite estimator, successive sampling

## 1 Introduction

The Labour Force Survey (LFS) quarterly provides estimates of the number of employed and unemployed individuals. Repeated sampling from a finite population is a sampling procedure used for the LFS.

Let us consider a finite household population $\mathcal{U} = \{1, \ldots, i, \ldots, N\}$ of size $N$. For each household, the number of its members is denoted by $m_i$, $i = 1, 2, \ldots, N$. Then the sum of the household members can be obtained by $M = \sum_{i=1}^{N} m_i$. Let us say that the survey variable $y$ characterizes the number of employed and unemployed individuals in each household. The values $y_i$ of the variable belongs to the set of integers $\{0, 1, \ldots, m_i\}$. We are interested in the estimation of the number of employed (and unemployed) individuals in the population:

$$t_y = \sum_{i=1}^{N} y_i. \tag{1}$$

The previous wave survey data can be used as auxiliary information for estimation of the population total in order to reduce the variance of the estimator. The effectiveness of ratio estimators in the case of any sampling design is described by many authors. If an auxiliary variable is well correlated with the study variable, then it is possible to obtain a more accurate estimate of a parameter.

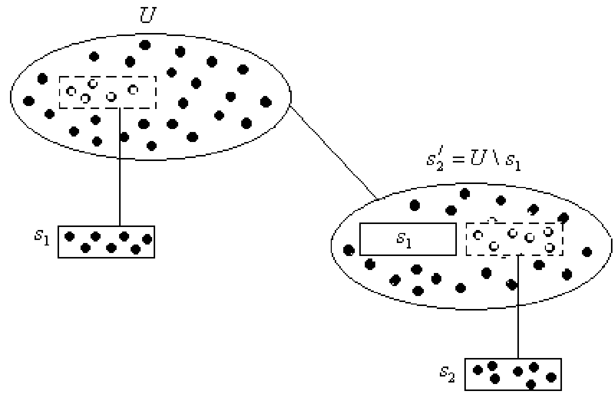
## 2 Sampling rotation and sample selection

LFS is conducted quarterly at Statistics Lithuania. All members of a household are included into the samples for two subsequent quarters, excluded for the next two quarters, and included into the sample once more for another two quarters.

Two-phase sampling scheme as shown in Fig 1. can be used for the estimation of the population total $t_y$ (1).

The sample rotation procedure for the two-phase sampling scheme presented in Fig 1. shows that the whole sample $s$ consists of two samples: $s_1$ and $s_2$. These sample parts are expressed as follows:

Figure 1: The sample selection procedure




$s_1 : \mathcal{U} \longrightarrow s_1$;
$s_2 : \mathcal{U} \longrightarrow s_2' = \mathcal{U} \setminus s_1 \longrightarrow s_2$; (two phases)

## 3 Simple estimators of the population total

Firstly, we construct two separate estimators of the total $t_y$ using data of samples $s_1$ and $s_2$ respectively. Secondly, we propose a combined estimator of the total using the sample rotation scheme (see section 4).

**Step 1**

The sample $s_1$ is selected from the finite population: $\mathcal{U} \longrightarrow s_1$. The corresponding first and second order inclusion probabilities for elements of the sample $s_1$ are respectively:

$$\pi_{1i} = \mathrm{P}(s_1 \subset \mathcal{U} : i \in s_1);$$
$$\pi_{1ij} = \mathrm{P}(s_1 \subset \mathcal{U} : i \in s_1, j \in s_1).$$

Unbiased Horvitz-Thompson (HT) (Horvitz & Thompson, 1952) estimator of the population total:

$$\hat{t}_{1y}^{HT} = \sum_{i \in s_1} \frac{y_i}{\pi_{1i}}. \tag{2}$$

The variance of the estimator $\hat{t}_{1y}^{HT}$ is

$$Var(\hat{t}_{1y}^{HT}) = \sum_{i \in \mathcal{U}} \sum_{j \in \mathcal{U}} (\pi_{1ij} - \pi_{1i}\pi_{1j}) \frac{y_i}{\pi_{1i}} \frac{y_j}{\pi_{1j}}. \tag{3}$$

The variance $Var(\hat{t}_{1y}^{HT})$ is estimated unbiasedly by

$$\widehat{Var}(\hat{t}_{1y}^{HT}) = \sum_{i \in s_1} \sum_{j \in s_1} (1 - \frac{\pi_{1i}\pi_{1j}}{\pi_{1ij}}) \frac{y_i}{\pi_{1i}} \frac{y_j}{\pi_{1j}}. \tag{4}$$

The values of the study variable $y$ in the previous survey can be used as auxiliary information. Let us denote the study variable of the previous wave survey by $x$ with the values $x_i$ and the same variable on the current wave by $y$ with the values $y_i$, $i \in s_1$. We can form the ratio estimator $\hat{t}_{1y}^{rat}$ of the population total $t_y$ by

$$\hat{t}_{1y}^{rat} = \hat{t}_x^{1w} \frac{\hat{t}_{1y}^{HT}}{\hat{t}_{1x}^{HT}} = \hat{t}_x^{1w} \hat{r}, \tag{5}$$

where

$$\hat{t}_{1x}^{HT} = \sum_{i \in s_1} \frac{x_i}{\pi_{1i}}; \quad \hat{r} = \frac{\hat{t}_{1y}^{HT}}{\hat{t}_{1x}^{HT}}.$$

$\hat{t}_x^{1w}$ – an estimate of the population total $t_y$ calculated using all values of the study variable $y$ in the previous survey, $\hat{t}_{1y}^{HT}$ given in (2). The estimator $\hat{t}_{1y}^{rat}$ is non-linear. Usually the Taylor expansion is used to find its properties.

An approximate variance of the ratio estimator $\hat{t}_{1y}^{rat}$ is:

$$AVar(\hat{t}_{1y}^{rat}) = \sum_{i \in \mathcal{U}} \sum_{j \in \mathcal{U}} (\pi_{1ij} - \pi_{1i}\pi_{1j}) \frac{y_i - rx_i}{\pi_{1i}} \frac{y_j - rx_j}{\pi_{1j}}, \tag{6}$$

with $r = \sum_{i \in \mathcal{U}} y_i / \sum_{i \in \mathcal{U}} x_i$.

The variance $AVar(\hat{t}_{1y}^{rat})$ is estimated by

$$\widehat{Var}(\hat{t}_{1y}^{rat}) = \sum_{i \in s_1} \sum_{j \in s_1} (1 - \frac{\pi_{1i}\pi_{1j}}{\pi_{1ij}}) \frac{y_i - \hat{r}x_i}{\pi_{1i}} \frac{y_j - \hat{r}x_j}{\pi_{1j}}, \tag{7}$$

using $\hat{r}$ given in (5)

**Step 2**

The sample $s_2$ is obtained in two-phase sampling: $\mathcal{U} \longrightarrow s_2' = \mathcal{U} \setminus s_1 \longrightarrow s_2$. The corresponding first and second order inclusion probabilities for samples $s_2'$ (first phase) and $s_2$ (second phase) are respectively:

$$\begin{aligned}
\pi_{2i}' &= \mathrm{P}(s_2' \subset \mathcal{U} : i \in s_2') = \mathrm{P}(i \notin s_1) = 1 - \mathrm{P}(i \in s_1); \\
\pi_{2ij}' &= \mathrm{P}(s_2' \subset \mathcal{U} : i \in s_2', j \in s_2') \\
&= 1 - \mathrm{P}(i \in s_1, j \in s_1) - \mathrm{P}(i \in s_1, j \notin s_1) - \mathrm{P}(i \notin s_1, j \in s_1); \\
\pi_{2i|s_2'} &= \mathrm{P}(s_2 \subset s_2' : i \in s_2 \mid s_2'); \\
\pi_{2ij|s_2'} &= \mathrm{P}(s_2 \subset s_2' : i \in s_2, j \in s_2 \mid s_2').
\end{aligned}$$

Under two-phase sampling, using the $\pi^*$ estimator (Särndal *et al.*, 1992), the population total $t_y$ is unbiasedly estimated by

$$\hat{t}_{2y}^{(2)} = \sum_{i \in s_2} \frac{y_i}{\pi_{2i}' \pi_{2i|s_2'}}. \tag{8}$$

In the case of two-phase sampling the variance of the estimator $\hat{t}_{2y}^{(2)}$ may be expressed

$$\begin{aligned}
Var(\hat{t}_{2y}^{(2)}) &= \sum_{i \in \mathcal{U}} \sum_{j \in \mathcal{U}} (\pi_{2ij}' - \pi_{2i}'\pi_{2j}') \frac{y_i}{\pi_{2i}'} \frac{y_j}{\pi_{2j}'} \\
&\quad + \mathrm{E} \sum_{i,j \in s_2'} (\pi_{2ij|s_2'} - \pi_{2i|s_2'}\pi_{2j|s_2'}) \frac{y_i}{\pi_{2i}'\pi_{2i|s_2'}} \frac{y_j}{\pi_{2j}'\pi_{2j|s_2'}}.
\end{aligned} \tag{9}$$

The variance $Var(\hat{t}_{2y}^{(2)})$ is estimated unbiasedly by

$$\begin{aligned}
\widehat{Var}(\hat{t}_{2y}^{(2)}) &= \sum_{i,j \in s_2} \frac{\pi_{2ij}' - \pi_{2i}'\pi_{2j}'}{\pi_{2ij}'\pi_{2ij|s_2'}} \frac{y_i}{\pi_{2i}'} \frac{y_j}{\pi_{2j}'} \\
&\quad + \sum_{i,j \in s_2} \frac{\pi_{2ij|s_2'} - \pi_{2i|s_2'}\pi_{2j|s_2'}}{\pi_{2ij|s_2'}} \frac{y_i}{\pi_{2i}'\pi_{2i|s_2'}} \frac{y_j}{\pi_{2j}'\pi_{2j|s_2'}}
\end{aligned} \tag{10}$$

# 4 Combined estimators of the population total for two-phase sampling scheme

In this section the construction of the combined estimators and their variances of the finite population total (1) in the case of sample rotation for two-phase sampling scheme is presented.

By a linear combination of $\hat{t}_{1y}^{HT}$ and $\hat{t}_{2y}^{(2)}$ we obtain an estimator without the use of auxiliary information of the total

$$\hat{t}_y = \frac{1}{2}\hat{t}_{1y}^{HT} + \frac{1}{2}\hat{t}_{2y}^{(2)}. \tag{11}$$

The variance of estimator (11) of the total $t_y$ can be expressed

$$Var(\hat{t}_y) = \Big(\frac{1}{2}\Big)^2 Var(\hat{t}_{1y}^{HT}) + \Big(\frac{1}{2}\Big)^2 Var(\hat{t}_{2y}^{(2)}) + \frac{1}{2}Cov(\hat{t}_{1y}^{HT}, \hat{t}_{2y}^{(2)}). \tag{12}$$

The variance $Var(\hat{t}_2)$ is estimated unbiasedly by

$$\widehat{Var}(\hat{t}_y) = \Big(\frac{1}{2}\Big)^2 \widehat{Var}(\hat{t}_{1y}^{HT}) + \Big(\frac{1}{2}\Big)^2 \widehat{Var}(\hat{t}_{2y}^{(2)}) + \frac{1}{2}\widehat{Cov}(\hat{t}_{1y}^{HT}, \hat{t}_{2y}^{(2)}). \tag{13}$$

By a linear combination of $\hat{t}_{1y}^{rat}$ and $\hat{t}_{2y}^{(2)}$ we obtain an estimator with the use of auxiliary information of the total

$$\hat{t}_y^{rat} = \frac{1}{2}\hat{t}_{1y}^{rat} + \frac{1}{2}\hat{t}_{2y}^{(2)}. \tag{14}$$

The variance of estimator (14) of the total $t_y$ can be expressed

$$AVar(\hat{t}_y^{rat}) = \Big(\frac{1}{2}\Big)^2 AVar(\hat{t}_{1y}^{rat}) + \Big(\frac{1}{2}\Big)^2 Var(\hat{t}_{2y}^{(2)}) + \frac{1}{2}Cov(\hat{t}_{1y}^{rat}, \hat{t}_{2y}^{(2)}). \tag{15}$$

The variance $Var(\hat{t}_2^{rat})$ is estimated by

$$\widehat{Var}(\hat{t}_y^{rat}) = \Big(\frac{1}{2}\Big)^2 \widehat{Var}(\hat{t}_{1y}^{rat}) + \Big(\frac{1}{2}\Big)^2 \widehat{Var}(\hat{t}_{2y}^{(2)}) + \frac{1}{2}\widehat{Cov}(\hat{t}_{1y}^{rat}, \hat{t}_{2y}^{(2)}). \tag{16}$$

Further we are interested in the estimation of the finite population total $t_y$ using two-phase sampling scheme, when simple random samples of households without replacement and samples with probabilities proportional to household size without replacement are drawn at each of the phases. Data of two quarters is used for the estimation of the population total $t_y$.

## 4.1 Simple random sampling (SRS) of households without replacement

Assume that $s_1$ of size $n_1$ is a simple random sample from the population $\mathcal{U}$ and its complement $s_2' = \mathcal{U} \setminus s_1$ of size $N - n_1$ is also a simple random sample from the population $\mathcal{U}$. $s_2$ of size $n_2$ is a simple random sample from $s_2'$. Then the first and second order inclusion probabilities to be used for (11), (13), (14) and (16) are calculated as follows:

$$\begin{aligned}
\pi_{1i} &= \mathrm{P}(s_1 \subset \mathcal{U} : i \in s_1) = \frac{n_1}{N};\\
\pi_{1ij} &= \mathrm{P}(s_1 \subset \mathcal{U} : i \in s_1, j \in s_1) = \frac{n_1(n_1-1)}{N(N-1)};\\
\pi_{2i}' &= \mathrm{P}(s_2' \subset \mathcal{U} : i \in s_2') = \frac{N-n_1}{N};\\
\pi_{2ij}' &= \mathrm{P}(s_2' \subset \mathcal{U} : i \in s_2', j \in s_2') = 1 - \frac{n_1(n_1-1)}{N(N-1)} - 2\frac{n_1}{N-1}\Big(1-\frac{n_1}{N}\Big);\\
\pi_{2i|s_2'} &= \mathrm{P}(s_2 \subset s_2' : i \in s_2 \mid s_2') = \frac{n_2}{N-n_1};\\
\pi_{2ij|s_2'} &= \mathrm{P}(s_2 \subset s_2' : i \in s_2, j \in s_2 \mid s_2') = \frac{n_2(n_2-1)}{(N-n_1)(N-n_1-1)}.
\end{aligned}$$

We estimate the variances of the estimators $\hat{t}_y$ and $\hat{t}_y^{rat}$ of the total $t_y$ in (13) and (16) respectively with

$$\widehat{Cov}(\hat{t}_{1y}^{HT}, \hat{t}_{2y}^{(2)}) = -\frac{N}{(N-n_1)(n_1-1)}\sum_{i\in s}(y_i-\bar{y})^2 \tag{17}$$

and

$$\widehat{Cov}(\hat{t}_{1y}^{rat}, \hat{t}_{2y}^{(2)}) = -\frac{N^2}{(N-n_1)(n_1-1)}\sum_{i\in s_1}(y_i-\hat{r}x_i)(y_i-\bar{y}) \tag{18}$$

### 4.2 Unequal probability sampling of households without replacement with probability proportional to its size (PPS) - successive sampling

Suppose $\mathcal{U} = \{1,\ldots,i,\ldots,N\}$ is a finite population of size $N$, here $N$ is the number of the households. Let us assume that $m_i$ is the number of the household members, $M = \sum_{i=1}^{N} m_i$ is the sum of the households members.

Assume that sample $s_1$ of size $n_1$, sample $s_2' = \mathcal{U}\setminus s_1$ of size $N-n_1$ and sample $s_2$ of size $n_2$ is drawn according to an order sampling design introduced by Rosen (1996) - successive sampling. To each unit $i$ in the population $\mathcal{U}$, the random number $U_i$ having uniform distribution $U(0,1)$ is generated. Let us assume that $\lambda_i = \frac{n_1 m_i}{M}$, then $Q_i = -ln(1-U_i)/-ln(1-\lambda_i)$, $i\in\mathcal{U}$ are the ranking variables. Values of $Q_i$ are found, the population elements with the $n_1$ smallest $Q$ values constitute the sample $s_1$. Sample $s_2' = \mathcal{U}\setminus s_1$ of size $N-n_1$ is also obtained by a successive order sampling. Now, to each unit $i$ in sample $s_2'$ the random number $U_i$ having uniform distribution $U(0,1)$ is generated, values of ranking variables $Q_i = -ln(1-U_i)/-ln(1-\lambda_i)$, $i\in s_2'$, are recalculated with $\lambda_i = \frac{n_2 m_i}{\sum_{i\in s_2'} m_i}$, and sample $s_2'$ elements with the $n_2$ smallest $Q$ values constitute the sample $s_2$.

Then the first order inclusion probabilities to be used for the estimation of the total $t_y$ in (11) and (14) are calculated as follows:

$$\begin{aligned}
\pi_{1i} &= \mathrm{P}(s_1\subset\mathcal{U} : i\in s_1) \approx \lambda_i = \frac{n_1 m_i}{M};\\
\pi_{2i}' &= \mathrm{P}(s_2'\subset\mathcal{U} : i\in s_2') \approx \lambda_{2i}' = \frac{M-n_1 m_i}{M};\\
\pi_{2i|s_2'} &= \mathrm{P}(s_2\subset s_2' : i\in s_2 \mid s_2') \approx \lambda_{2i} = \frac{n_2 m_i}{M_2}; \quad M_2 = \sum_{i\in s_2'} m_i.
\end{aligned}$$

There are expressions for second order inclusion probabilities $\pi_{1ij}$ in the case of order sampling design presented in (Aires, 1999), but they are too complex and time consuming. Therefore, we approximate $\pi_{1ij}$ for successive sampling design by $\breve{\pi}_{1ij}$ for unequal probability without replacement conditional Poisson sampling design, presented by Aires (1999):

$$\begin{aligned}
\breve{\pi}_{1ij} &= \frac{1}{\gamma_{1i}-\gamma_{1j}}(\gamma_{1i}\breve{\pi}_{1j}-\gamma_{1j}\breve{\pi}_{1i}), && \text{for } \gamma_{1i}\neq\gamma_{1j},\\
\breve{\pi}_{1ij} &= \frac{1}{k_{1i}}\left((n_1-1)\breve{\pi}_{1i} - \sum_{j:\ \gamma_{1j}\neq\gamma_{1i}}\breve{\pi}_{1ij}\right), && \text{for } \gamma_{1i}=\gamma_{1j}, \quad i,j\in\mathcal{U}, \quad i\neq j.
\end{aligned}$$

Here $k_{1i}$ is the number of elements with $j\neq i$ such that $\gamma_{1i}=\gamma_{1j}$ and $\gamma_{1i}=\frac{p_{1i}}{1-p_{1i}}$. The probability $p_{1i}$ is a selection probability of element $i$ for conditional Poisson sampling design, $i\in\mathcal{U}$. Since $p_{1i}$ are unknown, we use approximation result of Bondesson *et al.* (2006):

$$\frac{p_{1i}}{1-p_{1i}} \propto \frac{\lambda_{1i}}{1-\lambda_{1i}} exp\left(\frac{\frac{1}{2}-\lambda_{1i}}{d}\right), \quad d = \sum_{i=1}^{N}\lambda_{1i}(1-\lambda_{1i}).$$

Let us approximate the first order inclusion probabilities:

$$\pi_{1i} = \breve{\pi}_{1i} \approx \lambda_i = \frac{n_1 m_i}{M};$$
$$\pi'_{2i} = \breve{\pi}'_{2i} \approx \lambda'_{2i} = \frac{M - n_1 m_i}{M};$$
$$\pi_{2i|s'_2} = \breve{\pi}_{2i|s'_2} \approx \lambda_{2i} = \frac{n_2 m_i}{M_2}; \quad M_2 = \sum_{i\in s'_2} m_i.$$

The second order inclusion probabilities for samples $s'_2$ and $s_2$ are taken as follows:

$$\begin{aligned}\pi'_{2ij} &= \mathrm{P}(s'_2 \subset \mathcal{U} : i \in s'_2, j \in s'_2)\\ &\cong 1 - \breve{\pi}_{1ij} - \frac{n_1 m_i}{M - m_j}(1-\lambda_j) - \frac{n_1 m_j}{M - m_i}(1-\lambda_i);\end{aligned}$$
$$\breve{\pi}_{2ij|s'_2} = \frac{1}{\gamma_{2i} - \gamma_{2j}}(\gamma_{2i}\breve{\pi}_{2j|s'_2} - \gamma_{2j}\breve{\pi}_{2i|s'_2}), \qquad \text{for } \gamma_{2i} \neq \gamma_{2j},$$
$$\breve{\pi}_{2ij|s'_2} = \frac{1}{k_{2i}}\left((n_2-1)\breve{\pi}_{2i|s'_2} - \sum_{j:\ \gamma_{2j}\neq\gamma_{2i}} \breve{\pi}_{2ij|s'_2}\right), \qquad \text{for } \gamma_{2i} = \gamma_{2j}.$$

Here $k_{2i}$ is the number of elements $j \neq i$ such that $\gamma_{2i} = \gamma_{2j}$ and $\gamma_{2i} = \frac{\lambda_{2i}}{1-\lambda_{2i}} exp\left(\frac{\frac{1}{2}-\lambda_{2i}}{d}\right)$, $d = \sum_{i\in s'_2} \lambda_{2i}(1-\lambda_{2i})$. We have $\breve{\pi}_{2ii|s'_2} = \breve{\pi}_{2i|s'_2}$ in the case $i = j$. After replacing the $\pi$ values in (4), (7) and (10) with the corresponding values presented in this section we estimate the variance of the estimators $\hat{t_2}$ and $\hat{t}_2^{rat}$ of the total $t_y$ in (13) and (16) respectively with

$$\widehat{Cov}(\hat{t}_{1y}^{HT}, \hat{t}_{2y}^{(2)}) = -\frac{M^2}{(N-n_1)n_1}\sum_{i\in s_1}\frac{1}{m_i^2}\left(y_i - \frac{\hat{t}_{1y}^{HT}}{\hat{N}}\right)^2 \tag{19}$$

and

$$\widehat{Cov}(\hat{t}_{1y}^{rat}, \hat{t}_{2y}^{(2)}) = -\frac{M^2}{(N-n_1)n_1}\sum_{i\in s_1}\frac{1}{m_i^2}\left(y_i - \frac{\hat{t}_{1y}^{HT}}{\hat{N}}\right)\left(y_i - \frac{\hat{t}_{1y}^{HT}}{\hat{t}_{1x}^{HT}}x_i\right), \tag{20}$$

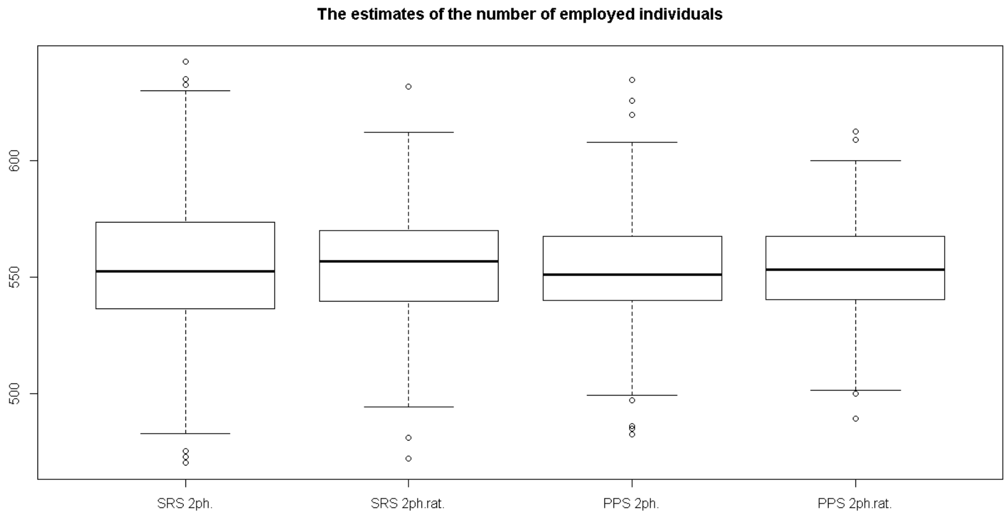
where $\hat{N} = \sum_{i\in s_1} 1/\pi_{1i}$.

# 5 Simulation study

In this section, we present the simulation study for the comparison of the performance of proposed estimators of the total using data of two quarters, with simple random sampling (SRS) and unequal probability sampling (successive order sampling) (PPS) of households without replacement in each of the phase. We study real LFS data of Statistics Lithuania. The study population consists of $N$=500 households. The variables of interest, $y$ and $x$, are the number of employed and unemployed individuals in the population of households. We have selected $B$=500 samples $s_1$ and $s_2$ of size $n_1 = n_2 = 100$ ($n = n_1 + n_2 = 200$) by the simple random sampling and successive sampling scheme.

For each of the estimators $\hat{t}_y$ (11) and $\hat{t}_y^{rat}$ (14), we have calculated the estimates of the population total of the study variable $y$. Also are calculated the averages of the estimates, averages of the estimates of variances. The results of the simulation are presented in Fig 2. and Fig 3.

Figure 2: Box-Plot diagrams of the estimates of the number of employed and the number of unemployed



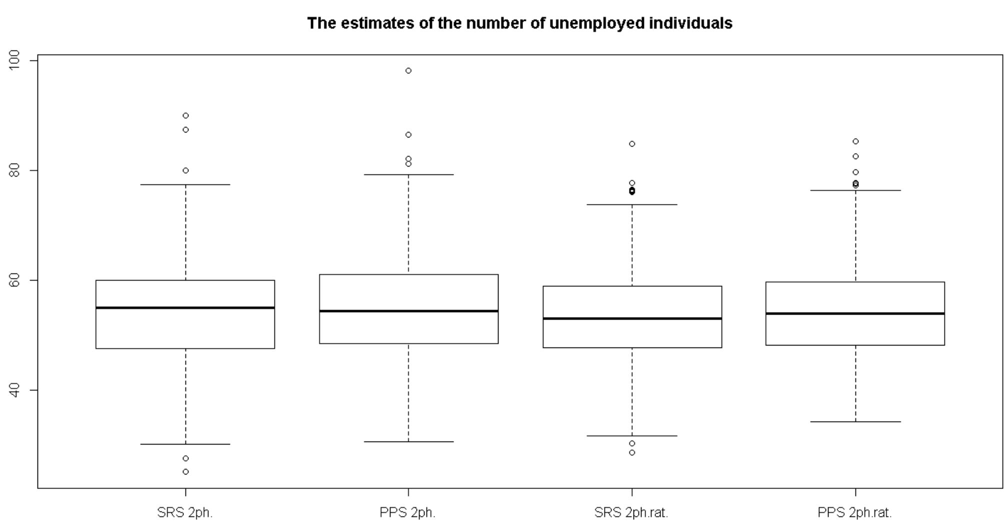
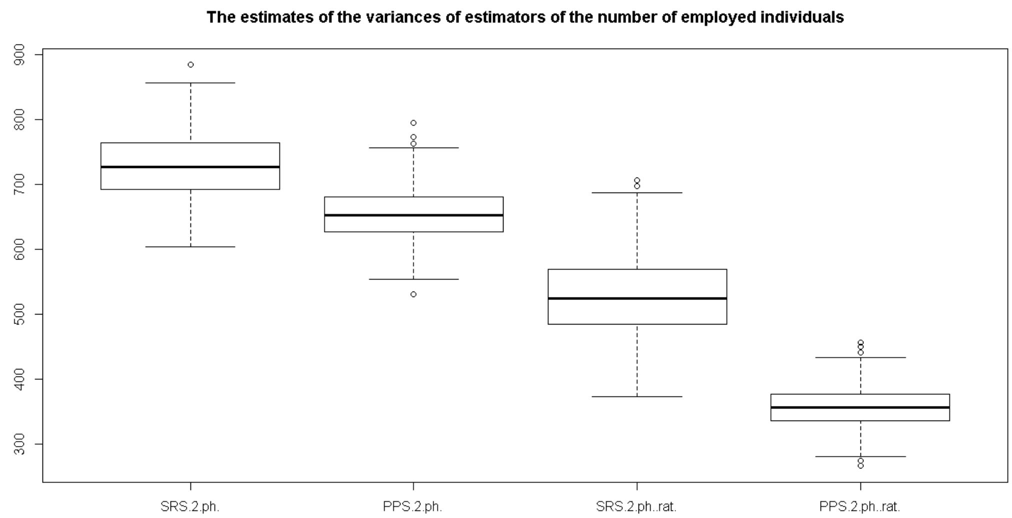
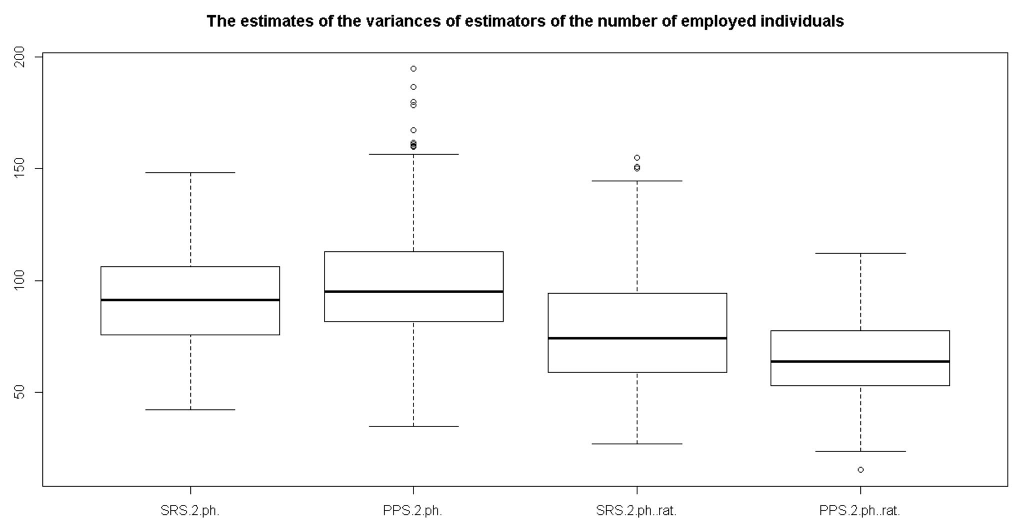



Figure 3: Box-Plot diagrams of estimates of the variances of estimators of the number of employed and the number of unemployed




# 6 Conclusions

The study addresses a practical problem related to the estimation of the number of employed and unemployed persons in two-phase sampling with simple random sampling and successive sampling design in each of the sampling phases. The successive sampling design is effective to estimate the number of employed and its effectiveness is the same as for SRS to estimate the number of the unemployed in the Lithuanian LFS. The ratio type combined estimator in two-phase LFS sampling design gives the smaller variance for the estimate of the number of employed individuals and does not have any effect to the estimates of the number of unemployed in comparison with the corresponding estimators which do not use the auxiliary data.

## References


Aires, N. (1999). Algorithms to find exact inclusion probabilities for conditional poisson sampling and pareto $\pi$ps sampling designs. *Methodology and Computing in Applied Probability* **1:4**, 457 – 469.

Bondesson, L., Traat, I. & Lundqvist, A. (2006). Pareto sampling versus conditional poisson and sampford sampling. *Scandinavian Journal Statitics* **33**, 699 – 720.

Horvitz, D. & Thompson, D. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **78(381)**, 663 – 685.

Rosen, B. (1996). On sampling with probability proportional to size. *R&D Report 1996:1, Statistics Sweden* .

Särndal, C., Swensson, B. & Wretman, J. (1992). *Model assisted survey sampling.* New York: Springer Verlag.