Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 73(1) 2019 19–25

# Calibration Approach for Estimation of Population Ratio under Double Sampling

**Sadikul Islam[1], Hukum Chandra[2], U.C. Sud[2], Saurav Guha[2] and Pradip Basak[2]**
*[1]ICAR-Indian Institute of Soil and Water Conservation, Dehradun*
*[2]ICAR-Indian Agricultural Statistics Research Institute, New Delhi*

*Received 16 July 2018; Revised 30 October 2018; Accepted 14 November 2018*

**SUMMARY**

Ratio estimator is widely used survey estimation method for estimating the finite population mean (or total) using auxiliary variable which is linearly related to study variable. However, this method requires the availability of aggregate level population information for auxiliary variable which may not be always available. As a result, in many practical situations, the ratio method of estimation cannot be applied. Alternatively, the double (or two-phase) sampling approach is often applied in such cases. This paper develops the calibration approach based finite population ratio estimator using the double sampling. It is assumed that the ratio of the total of auxiliary variables is available for the first phase sample only. The expression for variance and estimator of the variance of the proposed estimator is also developed. In addition, optimum sample sizes for the first and second phase samples are also suggested for a fixed cost. Monte Carlo simulations based on real population show that the proposed estimator is efficient than the existing alternative.

*Keywords:* Calibration, Cost function, Double sampling, Population ratio.

## 1. INTRODUCTION

In the sample survey, the auxiliary information is often used to improve the precision of estimators of finite population parameters. The calibration is widely used approach to incorporate auxiliary information in the estimation process to produce efficient estimators of finite population parameters such as finite population mean or total (Deville and Särndal, 1992, Wu and Luan, 2003, Tracy *et al.,* 2003, Rao *et al.,* 2012 and Koyuncu and Kadilar, 2013). However, calibration approach has also been used for estimation of complex parameters like population ratio or product or variance by several researchers including Plikusas and Pumputis (2007, 2010), Kim and Park (2010), Sud*et al.* (2014) and Basak *et al.* (2017) etc. In many practical applications, the estimation of population ratio is often used. For example, if the variable $y$ denotes the number of bullocks on a holding and $z$ its area in acres then the interest may be to estimate the number of bullocks per acre of holding in the population. Similarly, per capita monthly income in socio-economic surveys is obtained by the ratio of the sum of monthly income of households and the size of household, whereas, the productivity of crops is the ratio of total production to the total area where the crop is grown. Likewise, the unemployment rate is obtained by the ratio between the number of unemployed individuals and the number of individuals in the labour force in the country.

If there is a presence of auxiliary information along with the variables under study then the use of the calibration approach may substantially improve the precision of the estimator of population ratio. Raju (2012) used the calibration approach for the estimation of finite population ratio assuming that the information on the auxiliary variable is available at the population level. However, often such aggregate level information on auxiliary variable may not be available at the population level. In such cases, Hidiroglou and Särndal (1998) suggested to use the approach of double (or two-phase) sampling. This article describes calibration approach using double sampling

*Corresponding author:* Sadikul Islam
*E-mail address:* sadikul.islamiasri@gmail.com

to estimate the population ratio when the aggregated population level information on the auxiliary variable is not available. The rest of the article is organized as follows: The following section introduces the calibration estimation of population ratio under double sampling design when the ratio of auxiliary variables is available for the first phase sample. Section 3 presents the variance estimation of the proposed estimator. In addition, optimum two-phase sample sizes for a fixed cost as well as variance expressions under optimum condition are also obtained. Section 4 reports the empirical performance of the proposed estimator using design-based simulations based on real data. Finally, concluding remarks are set out in Section 5.

## 2. METHODOLOGICAL DEVELOPMENT OF ESTIMATOR

Let us assume a finite population $U(1,2...,k,...,N)$ containing $N$ units. Let $Y$ and $Z$ be the two variables belonging to the population $U$ and taking real values $y_1, y_2, ..., y_N$ and $z_1, z_2, ..., z_N$, respectively. Further it is assumed that the population $U$ has two auxiliary variables $X$ and $L$ corresponding to $Y$ and $Z$, respectively and the unit level values of $X$ and $L$ are denoted by $x_1, x_2, ..., x_N$ and $l_1, l_2, ..., l_N$, respectively. It is assumed that the population level aggregate values like population total ($t_x = \sum_{k=1}^{N} x_k$ and $t_l = \sum_{k=1}^{N} l_k$) for both the auxiliary variables are unknown. Under this circumstance, double sampling is often used. In double sampling, the information on the auxiliary character is obtained by selecting a large sample of size $n_1$ denoted by $s_1$, observing the auxiliary variables and further a sub-sample of size $n_2, (n_2 \ll n_1)$ denoted by $s_2$ is selected from the $s_1$ to observe the study variable. Thus, the sampling weight for $k^{\text{th}}$ population unit in $s_1$ is denoted as $d_{1k} = 1/\pi_{1k}$, where $\pi_{1k} = \Pr(k \in s_1)$ is the known first-phase inclusion probability for the $k^{\text{th}}$ population unit. Again, the sampling weight for $k^{\text{th}}$ population unit in $s_2$ is denoted as $d_{2k} = 1/\pi_{k|s_1}$, where $\pi_{k|s_1} = \Pr(k \in s_2 \mid s_1)$ is the known second-phase inclusion probability for $k^{\text{th}}$ population unit. Hence, the total sampling weight for $k^{\text{th}}$ population unit is denoted as $d_k = d_{1k} d_{2k}$, also known as the design weight. Let, $\pi_{1kl}$ and $\pi_{kj|s_1}$ denote the joint inclusion probability of $(k,j)^{\text{th}}$ population unit in $s_1$ and $s_2$, respectively. The population total of $Y$ and Z are given by $t_y = \sum_{k=1}^{N} y_k$ and $t_z = \sum_{k=1}^{N} z_k$, respectively and it is assumed to be unknown. In practice, the actual demand is to estimate of population total instead of population ratio. But, we know that the precision of population total estimation using ratio method is depends on the precision of population ratio estimate. For example, the expression for estimate of population total ($t_y$) of $Y$ using ratio method is denoted as $Est\_t_y = \hat{R}\left(\sum_{k=1}^{N} z_k\right)$ where, $\hat{R}$ is an estimator of population ratio $R = t_y / t_z$. Hence, the precision of $Est\_t_y$ is mostly depends on precision of $\hat{R}$. Following Särndal *et al.*(1992), an estimator of the ratio $R$ without incorporating the auxiliary variables is given by

$$\hat{R} = \frac{\hat{t}_y}{\hat{t}_z} \tag{1}$$

Here $\hat{t}_y = \sum_{k=1}^{n_2} d_k y_k$ and $\hat{t}_z = \sum_{k=1}^{n_2} d_k z_k$. The approximate variance of the $\hat{R}$ (Särndal *et al.*, 1992) is given by

$$AV\left(\hat{R}\right) = \frac{1}{t_z^2}\left[\sum_{k=1}^{N}\sum_{j=1}^{N} \Delta_{kj|s_1} \frac{u_k u_j}{\pi_{k|s_1}\pi_{j|s_1}}\right], \tag{2}$$

where $\Delta_{kj|s_1} = \pi_{kj|s_1} - \pi_{k|s_1}\pi_{j|s_1}$, $u_k = y_k - Rz_k$ and $u_j = y_j - Rz_j$, $k = j = 1, 2, \ldots, N$. The approximate variance estimator of the $\hat{R}$ (Särndal *et al.*, 1992) is then

$$A\hat{V}\left(\hat{R}\right) = \frac{1}{\hat{t}_z^2}\left[\sum_{k=1}^{n_2}\sum_{j=1}^{n_2} \frac{\Delta_{kj|s_1}}{\pi_{kj|s_1}} \frac{\hat{u}_k \hat{u}_j}{\pi_{k|s_1}\pi_{j|s_1}}\right], \tag{3}$$

where, $\hat{u}_k = y_k - \hat{R}z_k$ and $\hat{u}_j = y_j - \hat{R}z_j$, $k = j = 1, 2, \ldots, n_2$.

Let $x_1, x_2, ..., x_{n_1}$ and $l_1, l_2, ..., l_{n_1}$ be the values of the first-phase sample $s_1$ for the auxiliary variable $X$ and $L$, respectively. It is assumed that the unit level auxiliary information is unknown, but, ratio of auxiliary variable total, $R^{(1)} = \sum_{k=1}^{n_1} x_k \Big/ \sum_{k=1}^{n_1} l_k$ is known for $s_1$. Further, it is assumed that for second-phase sample $s_2$ the unit level information for all the four variables ($Y$, $Z$, $X$ and $L$) are known. Now, our aim is to develop calibration estimator of the finite population ratio under double sampling. Following Särndal *et al.* (1992), the expression of calibration estimator of population ratio is given by

$$\hat{R}_{CAL} = \sum_{k=1}^{n_2} w_k y_k \Big/ \sum_{k=1}^{n_2} w_k z_k, \tag{4}$$

where $w_k$ be the calibration weight for both the variable $Y$ and $Z$. Here, we obtain the $w_k$ in such a way that the distance between the original design weight $d_k$ and $w_k$ is minimized by considering the loss function $\sum_{k=1}^{n_2}\frac{(w_k-d_k)^2}{d_k}$ subject to the constraints $\sum_{k=1}^{n_2} w_k x_k - R^{(1)}\sum_{k=1}^{n_2} w_k l_k = 0$. The objective function (denoted as $\phi$) is defined as

$$\phi=\sum_{k=1}^{n_2}\frac{(w_k-d_k)^2}{d_k}+2\lambda\left[\sum_{k=1}^{n_2} w_k x_k - R^{(1)}\sum_{k=1}^{n_2} w_k l_k\right],$$

where, $\lambda$ is a Lagrangian multiplier constant. Now, differentiating the objective function $\phi$ with respect to $w_k$ and equate it to zero leads to calibration weight $w_k$ as

$$w_k = d_k - \left[\sum_{k=1}^{n_2} d_k\left(x_k - R^{(1)}l_k\right)\Big/\sum_{k=1}^{n_2} d_k\left(x_k - R^{(1)}l_k\right)^2\right] d_k\left(x_k - R^{(1)}l_k\right), \tag{5}$$

where, $\lambda=\sum_{k=1}^{n_2} d_k\left(x_k - R^{(1)}l_k\right)\Big/\sum_{k=1}^{n_2} d_k\left(x_k - R^{(1)}l_k\right)^2$.

Replacing the calibrated weights (5) in (4) yields the calibrated estimator of population ratio of form

$$\hat{R}_{CAL}=\frac{\sum_{k=1}^{n_2}\left[d_k-\left\{\sum_{k=1}^{n_2} d_k\left(x_k - R^{(1)}l_k\right)\Big/\sum_{k=1}^{n_2} d_k\left(x_k - R^{(1)}l_k\right)^2\right\}d_k\left(x_k - R^{(1)}l_k\right)\right]y_k}{\sum_{k=1}^{n_2}\left[d_k-\left\{\sum_{k=1}^{n_2} d_k\left(x_k - R^{(1)}l_k\right)\Big/\sum_{k=1}^{n_2} d_k\left(x_k - R^{(1)}l_k\right)^2\right\}d_k\left(x_k - R^{(1)}l_k\right)\right]z_k}. \tag{6}$$

Note that the estimator $\hat{R}_{CAL}$ can also be expressed as

$$\hat{R}_{CAL}=\frac{\hat{t}_y\hat{t}_1-\hat{t}_2\hat{t}_3}{\hat{t}_z\hat{t}_1-\hat{t}_2\hat{t}_4},$$

where, $\hat{t}_y=\sum_{k=1}^{n_2} d_k y_k$, $\hat{t}_1=\sum_{k=1}^{n_2} d_k\left(x_k - R^{(1)}l_k\right)^2$,

$\hat{t}_2=\sum_{k=1}^{n_2} d_k\left(x_k - R^{(1)}l_k\right)$, $\hat{t}_3=\sum_{k=1}^{n_2} d_k\left(x_k - R^{(1)}l_k\right)y_k$,

$\hat{t}_4=\sum_{k=1}^{n_2} d_k\left(x_k - R^{(1)}l_k\right)z_k$ and $\hat{t}_z=\sum_{k=1}^{n_2} d_k z_k$.

## 3. VARIANCE ESTIMATION

The developed calibrated estimator of finite population ratio is non-linear in nature. Hence, to derive the variance expression of the proposed estimator (6) we have to linearize the estimator. Thus, we have used Taylor series linearization technique to derive an approximate variance of the proposed estimator. The approximate variance expression of $\hat{R}_{CAL}$ is as

$$AV\left(\hat{R}_{CAL}\right)=V_1E_2\left[\frac{\hat{t}_y\hat{t}_1-\hat{t}_2\hat{t}_3}{\hat{t}_z\hat{t}_1-\hat{t}_2\hat{t}_4}\right]+E_1V_2\left[\frac{\hat{t}_y\hat{t}_1-\hat{t}_2\hat{t}_3}{\hat{t}_z\hat{t}_1-\hat{t}_2\hat{t}_4}\right],$$
$$=V_1\left(\hat{R}_{CAL}^{(1)}\right)+E_1V_2\left(\hat{R}_{CAL}\right), \tag{7}$$

where, $\hat{R}_{CAL}^{(1)}=E_2\left[\frac{\hat{t}_y\hat{t}_1-\hat{t}_2\hat{t}_3}{\hat{t}_z\hat{t}_1-\hat{t}_2\hat{t}_4}\right]$ and $\hat{R}_{CAL}=\left[\frac{\hat{t}_y\hat{t}_1-\hat{t}_2\hat{t}_3}{\hat{t}_z\hat{t}_1-\hat{t}_2\hat{t}_4}\right]$.

Now, we have to linearized both the term $\hat{R}_{CAL}^{(1)}$ and $\hat{R}_{CAL}$ of (7) separately. The first order Taylor's expansion of the function $\hat{R}_{CAL}^{(1)}$ is

$$\hat{R}_{CAL}^{(1)}\cong\hat{R}_{CAL}^{(1)}(lin)=R+\frac{1}{t_z}\left(t_y^{(1)}-t_y\right)-\frac{R}{t_z}\left(t_z^{(1)}-t_z\right)+\frac{R_c}{t_z}\left(t_2^{(1)}-t_2\right),$$

where, $R_c=\frac{Rt_4-t_3}{t_1}$, $t_y^{(1)}=\sum_{k=1}^{n_1} d_{1k}y_k$, $t_z^{(1)}=\sum_{k=1}^{n_1} d_{1k}z_k$ and $t_2^{(1)}=\sum_{k=1}^{n_1} d_{1k}\left(x_k - R^{(1)}l_k\right)$. Again, the first order Taylor's expansion of the function $\hat{R}_{CAL}$ is given by

$$\hat{R}_{CAL}\cong\hat{R}_{CAL}(lin)=R^{(1)}+\frac{1}{t_z^{(1)}}\left(\hat{t}_y-t_y^{(1)}\right)-\frac{R^{(1)}}{t_z^{(1)}}\left(\hat{t}_z-t_z^{(1)}\right)+\frac{R_c^{(1)}}{t_z^{(1)}}\left(\hat{t}_2-t_2^{(1)}\right),$$

where, $R^{(1)}=\sum_{k=1}^{n_1} x_k\Big/\sum_{k=1}^{n_1} l_k$, $t_y^{(1)}=\sum_{k=1}^{n_1} d_{1k}y_k$,

$t_z^{(1)}=\sum_{k=1}^{n_1} d_{1k}z_k$, $t_1^{(1)}=\sum_{k=1}^{n_1} d_{1k}\left(x_k - R^{(1)}l_k\right)^2$ $R_c^{(1)}=\frac{R^{(1)}t_4^{(1)}-t_3^{(1)}}{t_1^{(1)}}$,

$t_3^{(1)}=\sum_{k=1}^{n_1} d_{1k}\left(x_k - R^{(1)}l_k\right)y_k$ and $t_4^{(1)}=\sum_{k=1}^{n_1} d_{1k}\left(x_k - R^{(1)}l_k\right)z_k$.

Following Särndal *et al.* (1992), the expression of approximate variance of $\hat{R}_{CAL}$ is as

$$AV\left(\hat{R}_{CAL}\right)\cong AV\left[\hat{R}_{CAL}(lin)\right]=\frac{1}{t_z^2}\left[\sum_{k=1}^{N}\sum_{j=1}^{N}\Delta_{kj}^{(1)}\frac{u_k u_j}{\pi_{1k}\pi_{1j}}\right]+$$
$$E_1\left[\frac{1}{(t_z^{(1)})^2}\left(\sum_{k=1}^{n_1}\sum_{j=1}^{n_1}\Delta_{kj|s_1}\frac{v_k}{\pi_{1k}\pi_{k|s_1}}\frac{v_j}{\pi_{1j}\pi_{j|s_1}}\right)\right], \tag{8}$$

where $\Delta_{kj}^{(1)}=\pi_{1kj}-\pi_{1k}\pi_{1j}$, $\Delta_{kj|s_1}=\pi_{kj|s_1}-\pi_{k|s_1}\pi_{j|s_1}$, $u_k=\left[y_k-Rz_k+R_c\left(x_k-R^{(1)}l_k\right)\right]$ and

$$v_k = y_k - R^{(1)}z_k + R_c^{(1)}\left(x_k - R^{(1)}l_k\right).$$

The estimator of the approximate variance of $\hat{R}_{CAL}$ is obtained as

$$A\hat{V}\left(\hat{R}_{CAL}\right) \cong \frac{1}{\hat{t}_z^2}\left[\sum_{k=1}^{n_1}\sum_{j=1}^{n_1}\frac{\Delta_{kj}^{(1)}}{\pi_{1kj}}\frac{\hat{u}_k\hat{u}_j}{\pi_{1k}\pi_{1j}}\right]+$$
$$\left[\frac{1}{\hat{t}_z^2}\left(\sum_{k=1}^{n_1}\sum_{j=1}^{n_1}\Delta_{kj|s_1}\frac{\hat{v}_k}{\pi_{1k}\pi_{2k|s_1}}\frac{\hat{v}_j}{\pi_{1j}\pi_{2j|s_1}}\right)\right], \quad (9)$$

where $\hat{u}_k = \left[y_k - R^{(1)}z_k + R_c^{(1)}\left(x_k - R^{(1)}l_k\right)\right]$,

$\hat{v}_k = y_k - \hat{R}z_k + \hat{R}_c\left(x_k - R^{(1)}l_k\right)$ and $\hat{R}_c = \dfrac{R^{(1)}\hat{t}_4 - \hat{t}_3}{\hat{t}_1}$.

Further, we derive the expression of (8) and (9) under simple random sampling without replacement (SRSWOR). In case of SRSWOR, $\pi_{1k} = \pi_{1l} = \dfrac{n_1}{N}$, $\pi_{1kj} = \dfrac{n_1(n_1-1)}{N(N-1)}$, $\pi_{kj|s_1} = \dfrac{n_2(n_2-1)}{n_1(n_1-1)}$ and $\pi_{k|s_1} = \pi_{j|s_1} = \dfrac{n_2}{n_1}$.

Therefore, the approximate variance of $\hat{R}_{CAL}$ under SRSWOR as

$$AV\left(\hat{R}_{CAL}\right)_{SRS} \cong \left[\frac{N(N-n_1)}{t_z^2 n_1}S_u^2\right]+\left[\frac{1}{t_z^2}\frac{N^2(n_1-n_2)}{n_1n_2}S_v^2\right], \quad (10)$$

where, $S_u^2 = \dfrac{1}{(N-1)}\sum_{k=1}^{N}\left(u_k - \bar{u}\right)^2$,

$S_v^2 = \dfrac{1}{(N-1)}\sum_{k=1}^{N}\left(v_k - \bar{v}\right)^2$, $\bar{u} = N^{-1}\sum_{k=1}^{N}u_k$ and $\bar{v} = N^{-1}\sum_{k=1}^{N}v_k$.

The estimator of approximate variance of $\hat{R}_{CAL}$ under SRSWOR is given by

$$A\hat{V}\left(\hat{R}_{CAL}\right)_{SRS} \cong \left[\frac{N(N-n_1)}{\hat{t}_z^2 n_1}\hat{S}_u^2\right]+\left[\frac{1}{\hat{t}_z^2}\frac{N^2(n_1-n_2)}{n_1n_2}\hat{S}_v^2\right], \quad (11)$$

where $\hat{S}_u^2 = \dfrac{1}{(n_2-1)}\sum_{k=1}^{n_2}\left(\hat{u}_k - \bar{\hat{u}}\right)^2$,

$\hat{S}_v^2 = \dfrac{1}{(n_2-1)}\sum_{k=1}^{n_2}\left(\hat{v}_k - \bar{\hat{v}}\right)^2$, $\bar{\hat{u}} = n_2^{-1}\sum_{k=1}^{n_2}\hat{u}_k$ and $\bar{\hat{v}}\quad n\ \sum\hat{v}$ .

## 3.1 Optimum sample size estimation for a fixed cost

We have also found the optimum sample sizes $n_1$ and $n_2$ for a fixed cost $C_0$ that minimizes the approximate variance (9). The objective function $\psi$ is minimized using Lagrangian multiplier approach subject to the calibration constraint $C_0 = n_1C_1 + n_2C_2$ where $C_1$ and $C_2$ are per unit cost of data collection for $s_1$ and $s_2$, respectively. The objective function used for optimum sample size estimation, is given below:

$$\psi = \left[\frac{N(N-n_1)}{t_z^2 n_1}S_u^2\right]+\left[\frac{1}{t_z^2}\frac{N^2(n_1-n_2)}{n_1n_2}S_v^2\right]+\lambda\left(n_1C_1 + n_2C_2 - C_0\right).$$

First order differentiation of the function $\psi$ with respect to $n_1$ and $n_2$, separately and equating it to zero, we get optimum value of $n_1$ and $n_2$ that minimizes the variance estimator for the fixed cost $C_0$ as

$$n_1^{opt} = \frac{C_0\sqrt{S_u^2}}{C_2\sqrt{S_v^2}+\sqrt{C_2C_1\left(S_u^2 - S_v^2\right)}} \quad (12)$$

$$n_2^{opt} = \frac{C_0\sqrt{\left(S_u^2 - S_v^2\right)}}{\sqrt{C_2C_1S_v^2}+C_1\sqrt{\left(S_u^2 - S_v^2\right)}} \quad (13)$$

### 3.1.1 Variance estimation under two-phase optimum sample size pair

We get expression of the minimum approximate variance of $\hat{R}_{CAL}$ under SRSWOR by substituting the optimum values of $n_1$ and $n_2$ in (10) as

$$AV\left(\hat{R}_{CAL}\right)_{\min} \cong \left[\frac{N\left(N-n_1^{opt}\right)}{t_z^2 n_1^{opt}}S_u^2\right]+\left[\frac{1}{t_z^2}\frac{N^2\left(n_1^{opt}-n_2^{opt}\right)}{n_1^{opt}n_2^{opt}}S_v^2\right]. \quad (14)$$

Similarly, the minimum approximate variance estimator of $\hat{R}_{CAL}$ under SRSWOR is obtained by substituting the optimum values of $n_1$ and $n_2$ in (11) as

$$A\hat{V}\left(\hat{R}_{CAL}\right)_{\min} \cong \left[\frac{N\left(N-n_1^{opt}\right)}{\hat{t}_z^2 n_1^{opt}}\hat{S}_u^2\right]+\left[\frac{1}{\hat{t}_z^2}\frac{N^2\left(n_1^{opt}-n_2^{opt}\right)}{n_1^{opt}n_2^{opt}}\hat{S}_v^2\right],$$

where $\hat{S}_{u\ opt} = \dfrac{}{\left(\quad^{opt}\quad\right)}\sum^{opt}\left(\hat{u}_k - \bar{\hat{u}}\right)$ ,

$\hat{S}_{v,opt}^2 = \dfrac{1}{\left(n_2^{opt}-1\right)}\sum_{k=1}^{n_2^{opt}}\left(\hat{v}_k - \bar{\hat{v}}\right)^2$, $\bar{\hat{u}}_{opt} = (n_2^{opt})^{-1}\sum_{k=1}^{n_2^{opt}}\hat{u}_k$ and

$\bar{\hat{v}}_{opt} = (n_2^{opt})^{-1}\sum_{k=1}^{n_2^{opt}}\hat{v}_k$ .

## 4. EMPIRICAL EVALUATIONS

In this Section, we report the results using the design based simulations that illustrate the performance of the proposed estimators. In particular, we consider two estimators of finite population ratio in our empirical evaluation. These are (i) Ratio estimator given by expression (1), denoted as Est.R, and (ii) Calibration estimator given in (6), denoted as Est.R.CAL. The design based simulations are based on a real dataset of 284 municipalities of Sweden, denoted as the MU284 population. In this data set, the population size is $N$ = 284. From the MU284 population, a sample of large size $n_1 = 150$ is taken using SRSWOR and a second sample of three different size $n_2 = 60,\ 80$ and 100 are selected from the first sample using SRSWOR. Similar to $n_1 = 150$, we select second phase sample of size $n_2 = 60,\ 80$ and 100 using SRSWOR for $n_1 = 175$ and $n_1 = 200$, separately. Here, the aim is to estimate population ratio between variables 1985 population (P85, in thousands) to the variables revenues from the 1985 Municipal taxation (RMT85, measured in millions of kronor). Let us assume that 1975 population (P75, in thousands) as the auxiliary variable to the P85 and the variable number of municipal employees in 1984 (ME84) as the auxiliary variable to the RMT85. The correlations between the variables are presented in Table 1.

**Table 1.** Correlation between different variables in MU284 population

| Variables | RMT85 | P85 | ME84 | P75 |
|---|---|---|---|---|
| RMT85 | 1 | 0.961 | 0.999 | 0.967 |
| P85 | 0.961 | 1 | 0.965 | 0.998 |
| ME84 | 0.999 | 0.965 | 1 | 0.971 |
| P75 | 0.967 | 0.998 | 0.971 | 1 |

The Monte Carlo simulation was run $H$=5000 times. The Simulation study was performed using $R$ software. The performance of the estimators was evaluated by percentage absolute relative bias (ARB), percentage relative root mean squared error (RRMSE) and percentage relative efficiency (RE), defined by

$$ARB(\hat{R}) = \frac{1}{H}\sum_{h=1}^{H}\left|\frac{\hat{R}_h - R}{R}\right| \times 100\,,$$

$$RRMSE(\hat{R}) = \sqrt{H^{-1}\sum_{h=1}^{H}\left(\frac{\hat{R}_h - R}{R}\right)^2} \times 100\,,$$

$$RE(Estimator) = \frac{RRMSE(Est.R)}{RRMSE(Est.R.CAL)} \times 100\,.$$

Here $\hat{R}_h$ denotes the predicted value of the population ratio at simulation $h$, with true value $R$ and $H$ denotes the number of simulations. The values of percentage absolute relative bias (ARB) and the values of percentage relative root mean square error (RRMSE) and percentage relative efficiency (RE) of the proposed estimator are summarized in Table 2. The results in Table 2 show that the values of percentage ARB and RRMSE for the Est.R is more than the Est.R.CAL for all sample sizes. For a fixed first phase sample with the decreases in second phase sample size, the rate of increase of percentage ARB and RRMSE for Est.R is more than the Est.R.CAL. For a fixed second phase sample, the percentage ARB and RRMSE for Est.R.CAL decreases with an increase in first phase sample size whereas the performance of Est.R does not dependent on the first phase sample size. Therefore, it can be concluded that in terms of ARB and RRMSE the estimator Est.R.CAL shows better performance. The performance of the developed estimator is also found to be superior in terms of RE. Figure 1 confirmed the

**Table 2.** Values of percentage absoluterelative biases (ARB), percentage relative root mean squared errors (RRMSE) and percentage relative efficiencies (RE) of the two estimators for different sample sizes

| $n_1$ | $n_2$ | Estimator | ARB | RRMSE | RE |
|---|---|---|---|---|---|
| 150 | 100 | Est.R | 6.41 | 7.37 | 100 |
| | | Est.R.CAL | 5.04 | 5.89 | 125 |
| | 80 | Est.R | 7.39 | 8.48 | 100 |
| | | Est.R.CAL | 5.43 | 6.29 | 135 |
| | 60 | Est.R | 8.82 | 9.87 | 100 |
| | | Est.R.CAL | 5.96 | 6.90 | 143 |
| 175 | 100 | Est.R | 6.42 | 7.37 | 100 |
| | | Est.R.CAL | 4.59 | 5.40 | 137 |
| | 80 | Est.R | 7.33 | 8.43 | 100 |
| | | Est.R.CAL | 4.94 | 5.80 | 145 |
| | 60 | Est.R | 8.72 | 9.77 | 100 |
| | | Est.R.CAL | 5.55 | 6.62 | 148 |
| 200 | 100 | Est.R | 6.48 | 7.44 | 100 |
| | | Est.R.CAL | 4.08 | 4.97 | 150 |
| | 80 | Est.R | 7.34 | 8.42 | 100 |
| | | Est.R.CAL | 4.54 | 5.43 | 155 |
| | 60 | Est.R | 8.69 | 9.75 | 100 |
| | | Est.R.CAL | 5.21 | 6.38 | 153 |

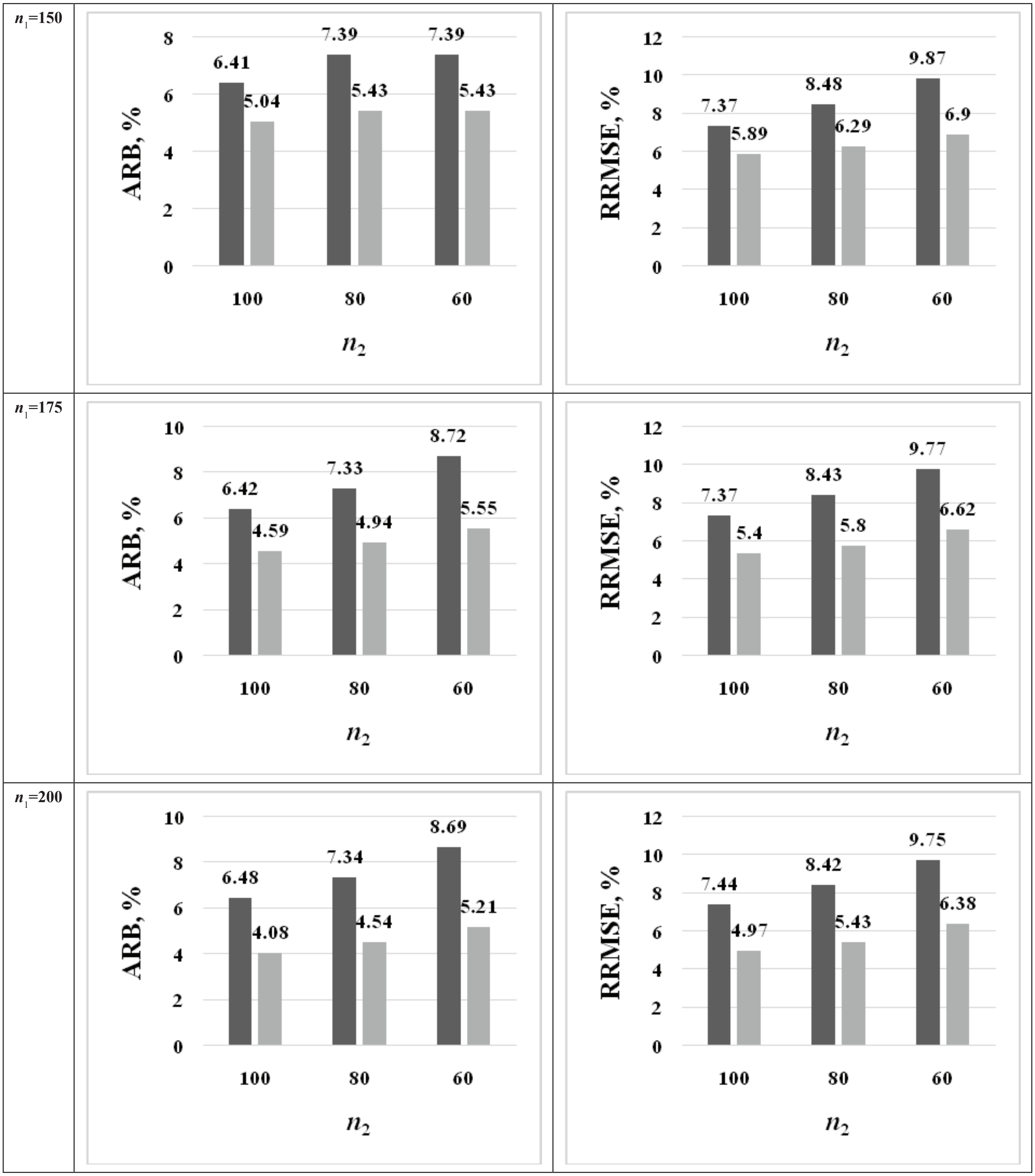


**Fig. 1.** The values of percentage absolute relative bias (ARB) (left side plots) and percentage relative root mean squared error (RRMSE) (right side plots) of Est.R (dark) and Est.R.CAL (light) estimators for different combinations of the first phase ($n_1$) and the second phase ($n_2$) sample sizes

validity of the results of Table 2 in terms of percentage ARB as well as percentage RRMSE.

## 5. CONCLUDING REMARKS

This article discusses the calibration estimation of finite population ratio when population ratio of auxiliary variables is known for the first phase sample only. The developed calibration estimator of finite population ratio gives better performance than the simple ratio estimator. Variance estimation of the calibrated estimator was done through the Taylor series linearization approach. Besides this, optimum sample sizes are also obtained for both first and second phase sample for a fixed cost.

## ACKNOWLEDGMENT

The authors would like to acknowledge the valuable comments and suggestions of the editorial board and an anonymous referee. These led to a considerable improvement in the paper.

## REFERENCES

Basak, P., Sud, U.C. and Chandra, H. (2017). Calibration estimation of regression coefficient for two-stage sampling design. *Journal of the Indian Society of Agricultural Statistics*, **71(1)**, 1-6.

Deville, J.C. and Särndal, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, **87**, 376-382.

Hidiroglou, M.A. and Särndal, C.E. (1998). Use of auxiliary information for two-phase sampling. *Survey Methodology*, **24**, 11-20.

Kim, J.K. and Park, M. (2010). Calibration estimation in survey sampling. *International Statistical Review*,**78**, 21-39.

Koyuncu, N. and Kadilar, C. (2013). Calibration estimator using different distance measures in stratified random sampling. *International Journal of Modern Engineering Research*, **3**, 415-419.

Kumar, R. (2012). An application of calibration approach for estimation of population ratio. *M.Sc. Thesis*, Indian Agricultural Research Institute, New Delhi-12.

Plikusas, A. and Pumputis, D. (2007). Calibrated estimators of the population covariance. *Acta Appllication Mathematicae*, **97**, 177-187.

Plikusas, A. and Pumputis, D. (2010). Estimation of the finite population covariance using calibration. *Lithuanian Mathematical Journal*, **15(3)**, 325-340.

Särndal, C. E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer Verlag, New York.

Sud, U.C., Chandra, H. and Gupta, V.K. (2014). Calibration based product estimator in single and two phase sampling. *Journal of Statistical Theory and Practice*, **8(1)**, 1-11.

Tracy, D.S., Singh, S. and Arnab, R. (2003). Note on calibration in stratified and double sampling. *Survey Methodology,* **29**, 99-104.

Wu, C. and Luan, Y. (2003). Optimal calibration estimators under two-phase sampling. *Journal of Official Statistics*, **19**, 119-131.