STATISTICS IN TRANSITION new series, June 2024
Vol. 25, No. 2, pp. xx–xxx, https://doi.org/10.59170/stattrans-2024-xxx
Received – xx.xx.xxxx; accepted – xx.xx.xxxx

Ratio-type estimator of the population mean in stratified sampling based on the calibration approach

Menakshi Pachori[1], **Neha Garg**[2]

Abstract

An improved ratio-type calibrated estimator was developed using the logarithmic mean in the calibration constraint for the stratified random sampling scheme. The proposed estimator was extended in the case of stratified double sampling and compared with the estimators given by Tracy et al. (2003) together its ratio-type estimators, as well as Nidhi et al. (2017) and Khare et al. (2022). A simulation study was also carried out on both a real and artificial dataset in order to evaluate the performance of the proposed estimator compared to the existing estimators.

Key words: auxiliary information, calibration estimation, stratified random sampling, stratified double sampling, ratio estimator.

1. Introduction

Calibration estimation method provides more precise results by making use of the available auxiliary information. Deville and Sarndal (1992) defined it as a procedure of minimizing a distance function subject to calibration constraints. In this direction, various calibration estimators for various sampling schemes for estimating the population parameters by using different calibration constraints based on the available auxiliary information have been developed by Singh (2003), Kim et al. (2007), Singh and Arnab (2011), Koyuncu and Kadilar (2013 and 2016), Mouhamed et al. (2015), Clement and Enang (2017), Nidhi et al. (2017), Garg and Pachori (2019, 2020 and 2021), Alka et al. (2021), Khare et al. (2022), Pachori et al. (2022), Kumar et al. (2023), etc.

It is a well-known fact in the sample surveys that auxiliary information can be used to increase the precision of the estimators of the population parameters. When study variable Y is highly positively correlated with the available auxiliary variable X, the usual

[1] School of Sciences, Indira Gandhi National Open University, New Delhi-110068, India.
E-mail: pachorim07@gmail.com. ORCID: https://orcid.org/0000-xxxx-xxxx-xxxx.

[2] Corresponding author, School of Sciences, Indira Gandhi National Open University, New Delhi-110068, India. E-mail: nehagarg@ignou.ac.in. ORCID: https://orcid.org/0000-xxxx-xxxx-xxxx.

ratio method of estimation is considered. For example, for estimating total/average yield of a crop (Y) and area of cultivation (X); sale of the departmental stores for current year (Y) and sales recorded for past years (X); volume of the tree (Y) and diameter of the tree (X); sugarcane production in the current year (Y) and sugarcane production in past surveys (X), etc. In the case of heterogeneous population, the stratified random sampling is preferred than the simple random sampling. This article deals with the problem of estimation of population mean of the study variable under stratified random sampling utilizing available auxiliary information with the help of the calibration approach.

In this article, we have suggested a new ratio-type calibration estimator using logarithmic mean of the available auxiliary variable as a calibration constraint under stratified and stratified double sampling for estimating the population mean of the study variable. A simulation study has been conducted on a real as well as an artificial dataset in order to support the study and the result has been compared with the estimator suggested by Tracy et al. (2003) and its ratio-type estimator along with Nidhi et al. (2017) and Khare et al. (2022).

1.1. Calibration estimator under stratified sampling

In stratified random sampling, the population of size N is divided into L homogeneous strata consisting of N_h units in the h[th] stratum such that $\sum_{h=1}^{L} N_h = N$. A sample of size n_h is drawn from the h[th] stratum using Simple Random Sampling Without Replacement (SRSWOR), where *n* is the required sample size such that $\sum_{h=1}^{L} n_h = n$.

Suppose y_{hi} and x_{hi} is the i[th] unit of the study and auxiliary variables, respectively, in the h[th] stratum for $i = 1, 2, ..., n_h$ and $h = 1, 2, ..., L$. Let $W_h = \frac{N_h}{N}$ be the stratum weight and $f_h = \frac{n_h}{N_h}$, the sample fraction.

The calibration estimator for population mean under stratified random sampling given by Tracy et al. (2003) is given as:

$$\bar{y}_{tr} = \sum_{h=1}^{L} \Omega_h \bar{y}_h \tag{1}$$

where Ω_h are the new calibrated weights obtained by minimizing the Chi-square type distance measure, subject to the two calibration constraints:

$$\sum_{h=1}^{L} \Omega_h \bar{x}_h = \sum_{h=1}^{L} W_h \bar{X}_h \tag{2}$$

$$\sum_{h=1}^{L}\Omega_h s_{hx}^2=\sum_{h=1}^{L}W_h S_{hx}^2 \tag{3}$$

where $\bar{x}_h=\frac{\sum_{i=1}^{n_h}x_{hi}}{n_h}$ and $\bar{X}_h=\frac{\sum_{i=1}^{n_h}X_{hi}}{n_h}$ are the h[th] stratum sample and population means of the auxiliary variable, respectively.

$s_{hx}^2=\sum_{i=1}^{n_h}\frac{(x_{hi}-\bar{x}_h)^2}{(n_h-1)}$ and $S_{hx}^2=\sum_{i=1}^{N_h}\frac{(X_{hi}-\bar{X}_h)^2}{(N_h-1)}$ are the h[th] stratum sample and population variances of the auxiliary variable, respectively.

The calibrated weights are obtained as:

$$\Omega_h=W_h+W_hQ_h\bar{x}_h\frac{\left(\sum_{h=1}^{L}W_hQ_hS_{hx}^4\right)\left(\sum_{h=1}^{L}W_h\left(\bar{X}_h-\bar{x}_h\right)-\left(\sum_{h=1}^{L}W_hQ_h\bar{x}_hS_{hx}^2\right)\left(\sum_{h=1}^{L}W_h\left(S_{hx}^2-S_{hx}^2\right)\right)\right)}{\left(\sum_{h=1}^{L}W_hQ_hS_{hx}^4\right)\left(\sum_{h=1}^{L}W_hQ_h\bar{x}_h^2\right)-\left(\sum_{h=1}^{L}W_hQ_h\bar{x}_hS_{hx}^2\right)^2}$$

$$+W_hQ_hs_{hx}^2\frac{\left(\sum_{h=1}^{L}W_h\left(S_{hx}^2-s_{hx}^2\right)\right)\left(\sum_{h=1}^{L}W_hQ_h\bar{x}_h^2\right)-\left(\sum_{h=1}^{L}W_hQ_h\bar{x}_h^2s_{hx}^2\right)\left(\sum_{h=1}^{L}W_h\left(\bar{X}_h-\bar{x}_h\right)\right)}{\left(\sum_{h=1}^{L}W_hQ_hs_{hx}^4\right)\left(\sum_{h=1}^{L}W_hQ_h\bar{x}_h^2\right)-\left(\sum_{h=1}^{L}W_hQ_h\bar{x}_hs_{hx}^2\right)^2}$$

The calibrated estimator given by Tracy et al. (2003) is

$$\bar{y}_{tr}=\sum_{h=1}^{L}W_h\bar{y}_h+\hat{\beta}_{t1}\left(\sum_{h=1}^{L}W_h\left(\bar{X}_h-\bar{x}_h\right)\right)+\hat{\beta}_{t2}\left(\sum_{h=1}^{L}W_h\left(S_{hx}^2-s_{hx}^2\right)\right) \tag{4}$$

where

$$\hat{\beta}_{t1}=\left[\frac{\left(\sum_{h=1}^{L}W_hQ_hs_{hx}^4\right)\left(\sum_{h=1}^{L}W_hQ_h\bar{x}_h\bar{y}_h\right)-\left(\sum_{h=1}^{L}W_hQ_h\bar{x}_hs_{hx}^2\right)\left(\sum_{h=1}^{L}W_hQ_h\bar{y}_hs_{hx}^2\right)}{\left(\sum_{h=1}^{L}W_hQ_hs_{hx}^4\right)\left(\sum_{h=1}^{L}W_hQ_h\bar{x}_h^2\right)-\left(\sum_{h=1}^{L}W_hQ_h\bar{x}_hs_{hx}^2\right)^2}\right]$$

$$\hat{\beta}_{t2}=\left[\frac{\left(\sum_{h=1}^{L}W_hQ_h\bar{x}_h^2\right)\left(\sum_{h=1}^{L}W_hQ_h\bar{y}_hs_{hx}^2\right)-\left(\sum_{h=1}^{L}W_hQ_hs_{hx}^2\right)\left(\sum_{h=1}^{L}W_hQ_h\bar{x}_h\bar{y}_h\right)}{\left(\sum_{h=1}^{L}W_hQ_hs_{hx}^4\right)\left(\sum_{h=1}^{L}W_hQ_h\bar{x}_h^2\right)-\left(\sum_{h=1}^{L}W_hQ_h\bar{x}_hs_{hx}^2\right)^2}\right]$$

To estimate the population mean $\bar{Y}$ under the stratified random sampling, the calibration estimator considered by Nidhi et al. (2017) is:

$$\bar{y}_s=\sum_{h=1}^{L}\Omega_h\bar{y}_h \tag{5}$$

subject to the two calibration constraints:

$$\sum_{h=1}^{L} \Omega_h = 1 \tag{6}$$

$$\sum_{h=1}^{L} \Omega_h \bar{x}_h = \sum_{h=1}^{L} W_h \bar{X}_h \tag{7}$$

The calibrated weight after minimization of Chi-square type distance measure subject to the calibration constraints is given as:

$$\Omega_h = W_h + \left[\frac{(W_h Q_h \bar{x}_h)\left(\sum_{h=1}^{L} W_h Q_h\right) - (W_h Q_h)\left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h\right)}{\left(\sum_{h=1}^{L} W_h Q_h x_h^2\right)\left(\sum_{h=1}^{L} W_h Q_h\right) - \left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h\right)^2} \right] \left(\bar{X} - \sum_{h=1}^{L} W_h \bar{x}_h\right)$$

The estimator given by Nidhi et al. (2017) is:

$$\bar{y}_s = \sum_{h=1}^{L} W_h \bar{y}_h + \hat{\beta}_s \left(\bar{X} - \sum_{h=1}^{L} W_h \bar{x}_h\right) \tag{8}$$

where

$$\hat{\beta}_s = \left[\frac{\left(\sum_{h=1}^{L} W_h Q_h\right)\left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h \bar{y}_h\right) - \left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h\right)\left(\sum_{h=1}^{L} W_h Q_h \bar{y}_h\right)}{\left(\sum_{h=1}^{L} W_h Q_h\right)\left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h^2\right) - \left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h\right)^2} \right]$$

1.2. Calibration Ratio Estimator under Stratified Sampling

If $\bar{y}$ is the sample mean of the study variable Y and $\bar{x}$ is the sample mean of the auxiliary variable X, the usual ratio estimator is given as:

$$\bar{y}_R = \frac{\bar{y}}{\bar{x}} \bar{X}$$

Let $\bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}$ and $\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$ be the sample means of auxiliary variable and study variable, respectively. Now, we can define the ratio-type calibration estimator in stratified random sampling considering the constraints given by Tracy et al. (2003) as:

$$\bar{y}_{r\_tr} = \sum_{h=1}^{L} \Omega_h \bar{y}_{Rh} \tag{9}$$

The calibrated weights are obtained for the estimator defined in equation (9) as:

$$\Omega_h = W_h + W_h Q_h \bar{x}_h \frac{\left(\sum_{h=1}^{L} W_h Q_h S_{hx}^4\right)\left(\sum_{h=1}^{L} W_h \left(\bar{X}_h - \bar{x}_h\right)\right) - \left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h S_{hx}^2\right)\left(\sum_{h=1}^{L} W_h \left(S_{hx}^2 - s_{hx}^2\right)\right)}{\left(\sum_{h=1}^{L} W_h Q_h S_{hx}^4\right)\left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h^2\right) - \left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h S_{hx}^2\right)^2}$$

$$+W_h Q_h s_{hx}^2 \frac{\left(\sum_{h=1}^{L} W_h \left(S_{hx}^2 - s_{hx}^2\right)\right)\left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h^2\right) - \left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h s_{hx}^2\right)\left(\sum_{h=1}^{L} W_h \left(\bar{X}_h - \bar{x}_h\right)\right)}{\left(\sum_{h=1}^{L} W_h Q_h s_{hx}^4\right)\left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h^2\right) - \left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h s_{hx}^2\right)^2}$$

Thus, the ratio-type calibrated estimator considering the constraints given by Tracy et al. (2003) is:

$$\bar{y}_{r\_tr} = \sum_{h=1}^{L} W_h y_{Rh} + \hat{\beta}_{r\_t1}\left(\sum_{h=1}^{L} W_h \left(\bar{X}_h - \bar{x}_h\right)\right) + \hat{\beta}_{r\_t2}\left(\sum_{h=1}^{L} W_h \left(S_{hx}^2 - s_{hx}^2\right)\right) \tag{10}$$

where

$$\hat{\beta}_{r\_t1} = \left[\frac{\left(\sum_{h=1}^{L} W_h Q_h s_{hx}^4\right)\left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h \bar{y}_h\right) - \left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h s_{hx}^2\right)\left(\sum_{h=1}^{L} W_h Q_h \bar{y}_h s_{hx}^2\right)}{\left(\sum_{h=1}^{L} W_h Q_h s_{hx}^4\right)\left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h^2\right) - \left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h s_{hx}^2\right)^2}\right]$$

$$\hat{\beta}_{r\_t2} = \left[\frac{\left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h^2\right)\left(\sum_{h=1}^{L} W_h Q_h \bar{y}_h s_{hx}^2\right) - \left(\sum_{h=1}^{L} W_h Q_h s_{hx}^2\right)\left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h \bar{y}_h\right)}{\left(\sum_{h=1}^{L} W_h Q_h s_{hx}^4\right)\left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h^2\right) - \left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h s_{hx}^2\right)^2}\right]$$

The calibration ratio estimator suggested by Khare et al. (2022) considering the constraints given in Nidhi et al. (2017) is:

$$\bar{y}_{r\_s} = \sum_{h=1}^{L} \Omega_h \bar{y}_{Rh} \tag{11}$$

Minimization of the Chi-square distance measure subject to the calibration constraints leads to the calibrated weight given as:

$$\Omega_h = W_h + \left(W_h Q_h \bar{x}_h\right)\left[\frac{\left(\sum_{h=1}^{L} W_h \bar{X}_h - \sum_{h=1}^{L} W_h \bar{x}_h\right)}{\left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h^2\right)}\right] \tag{12}$$

Thus, the ratio-type calibration estimator given Khare et al. (2022) is

$$\bar{y}_{r\_s} = \sum_{h=1}^{L} W_h \bar{y}_{Rh} + \hat{\beta}_{r\_s}\left(\sum_{h=1}^{L} W_h \bar{X}_h - \sum_{h=1}^{L} W_h \bar{x}_h\right) \tag{13}$$

where $$\hat{\beta}_{r\_s} = \left[\frac{\left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h \bar{y}_{Rh}\right)}{\left(\sum_{h=1}^{L} W_h Q_h \bar{x}_h^2\right)}\right]$$

2. Proposed Calibration Estimator

2.1. Stratified Random Sampling

We propose a new ratio-type calibration estimator for stratified random sampling as:

$$\bar{y}_{r\_L} = \sum_{h=1}^{L} \Omega_h \bar{y}_{Rh} \tag{14}$$

where the calibration weight Ω_h is so chosen in order to minimize the Chi-square type distance measure given as:

$$\sum_{h=1}^{L} \frac{(\Omega_h - W_h)^2}{W_h Q_h} \tag{15}$$

subject to the following calibration constraints using logarithmic mean of the auxiliary variable:

$$\sum_{h=1}^{L} \Omega_h = \sum_{h=1}^{L} W_h \tag{16}$$

$$\sum_{h=1}^{L} \Omega_h \log \bar{x}_h = \sum_{h=1}^{L} W_h \log \bar{X}_h \tag{17}$$

The Lagrange function is given as:

$$L = \sum_{h=1}^{L} \frac{(\Omega_h - W_h)^2}{Q_h W_h} - 2\lambda_1 \left(\sum_{h=1}^{L} \Omega_h - \sum_{h=1}^{L} W_h \right) - 2\lambda_2 \left(\sum_{h=1}^{L} \Omega_h \log \bar{x}_h - \sum_{h=1}^{L} W_h \log \bar{X}_h \right) \tag{18}$$

where λ_1 and λ_2 are the Lagrange multiplier. To find the optimum value of Ω_h, we differentiate the Lagrange function given in equation (18) with respect to Ω_h and equate it to zero. The calibration weight is derived as:

$$\Omega_h = W_h + W_h Q_h \left(\lambda_1 + \lambda_2 \log \bar{x}_h \right) \tag{19}$$

Here, λ_1 and λ_2 are determined by substituting the value of Ω_h from equation (19) to equations (16) and (17), thus we get the calibrated weight as:

$$\begin{aligned} \Omega_h = W_h + W_h Q_h & \left[\frac{-\left(\sum_{h=1}^{L} W_h \left(\log \bar{X}_h - \log \bar{x}_h \right) \right) \left(\sum_{h=1}^{L} W_h Q_h \log \bar{x}_h \right)}{\left(\sum_{h=1}^{L} W_h Q_h \log \bar{x}_h^2 \right) \left(\sum_{h=1}^{L} W_h Q_h \right) - \left(\sum_{h=1}^{L} W_h Q_h \log \bar{x}_h \right)^2} \right] \\ + W_h Q_h \log \bar{x}_h & \left[\frac{\left(\sum_{h=1}^{L} W_h Q_h \right) \left(\sum_{h=1}^{L} W_h \left(\log \bar{X}_h - \log \bar{x}_h \right) \right)}{\left(\sum_{h=1}^{L} W_h Q_h \log \bar{x}_h^2 \right) \left(\sum_{h=1}^{L} W_h Q_h \right) - \left(\sum_{h=1}^{L} W_h Q_h \log \bar{x}_h \right)^2} \right] \end{aligned} \tag{20}$$

On substituting the value of Ω_h from equation (20) in (14), the proposed calibrated estimator is given as:

$$\bar{y}_{r\_L} = \sum_{h=1}^{L} W_h \bar{y}_{Rh} + \hat{\beta}_{r\_L}\left[\sum_{h=1}^{L} W_h \left(\log \bar{X}_h - \log \bar{x}_h\right)\right] \tag{21}$$

where

$$\hat{\beta}_{r\_L} = \left[\frac{\left(\sum_{h=1}^{L} W_h Q_h\right)\left(\sum_{h=1}^{L} W_h Q_h \bar{y}_h \log \bar{x}_h\right) - \left(\sum_{h=1}^{L} W_h Q_h \log \bar{x}_h\right)\left(\sum_{h=1}^{L} W_h Q_h \bar{y}_h\right)}{\left(\sum_{h=1}^{L} W_h Q_h \log \bar{x}_h^2\right)\left(\sum_{h=1}^{L} W_h Q_h\right) - \left(\sum_{h=1}^{L} W_h Q_h \log \bar{x}_h\right)^2}\right]$$

The mean squared error of the estimator $\bar{y}_{r\_L}$ up to the second order of approximation is given as:

$$\mathrm{MSE}\left(\bar{y}_{r\_L}\right) = E\left[\bar{y}_{r\_L} - \bar{Y}\right]^2$$

$$= \sum_{h=1}^{L} W_h^2 f_h' \left[\bar{Y}_h^2 C_{yh}^2 + (\bar{Y}_h + \beta_{r\_L})^2 C_{xh}^2 - 2\rho_h \bar{Y}_h (\bar{Y}_h + \beta_{r\_L}) C_{yh} C_{xh}\right]$$

where $\beta_{r\_L} = \left[\frac{\left(\sum_{h=1}^{L} W_h \bar{Y}_h \log \bar{X}_h\right) - \left(\sum_{h=1}^{L} W_h \log \bar{X}_h\right)\left(\sum_{h=1}^{L} W_h \bar{Y}_h\right)}{\left(\sum_{h=1}^{L} W_h \log \bar{X}_h^2\right) - \left(\sum_{h=1}^{L} W_h \log \bar{X}_h\right)^2}\right]$ and

$$f_h' = \left(\frac{1}{n_h} - \frac{1}{N_h}\right)$$

2.2. Stratified Double Sampling

We have also proposed the calibration estimator in the case of stratified double sampling when the stratum population mean of the auxiliary variable is not known. In this practice, a preliminary sample of size m_h units is drawn as a first phase sample using SRSWOR and a subsample of n_h units is drawn from the preliminary sample of size m_h units by SRSWOR. Let $\bar{x}_h^* = \frac{1}{m_h}\sum_{i=1}^{m_h} x_{hi}$ be the first phase sample mean, while $\bar{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}$ and $\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$ are the second phase sample means of auxiliary and study variables, respectively. The usual ratio estimator in the h^{th} stratum under double sampling is $\bar{y}_{Rh\_d} = \frac{\bar{y}_h}{\bar{x}_h}\bar{x}_h^*$.

Therefore, the proposed calibration estimator in the case of stratified double sampling is given as:

$$\bar{y}_{r\_Ld} = \sum_{h=1}^{L} \Omega_h^* \bar{y}_{Rh\_d} \tag{22}$$

where the calibration weight Ω_h^* is so chosen to minimize the Chi-square type distance measure given as:

$$\sum_{h=1}^{L} \frac{(\Omega_h^* - W_h)^2}{W_h Q_h} \tag{23}$$

subject to the following calibration constraints:

$$\sum_{h=1}^{L} \Omega_h^* = \sum_{h=1}^{L} W_h \tag{24}$$

$$\sum_{h=1}^{L} \Omega_h^* \log \bar{x}_h = \sum_{h=1}^{L} W_h \log \bar{x}_h^* \tag{25}$$

The Lagrange function is given as:

$$L = \sum_{h=1}^{L} \frac{(\Omega_h^* - W_h)^2}{Q_h W_h} - 2\lambda_1 \left(\sum_{h=1}^{L} \Omega_h^* - \sum_{h=1}^{L} W_h \right) - 2\lambda_2 \left(\sum_{h=1}^{L} \Omega_h^* \log \bar{x}_h - \sum_{h=1}^{L} W_h \log \bar{x}_h^* \right) \tag{26}$$

where λ_1 and λ_2 are the Lagrange multipliers. The optimum calibration weight is given as:

$$\Omega_h^* = W_h + W_h Q_h \left[\frac{-\left(\sum_{h=1}^{L} W_h \left(\log \bar{x}_h^* - \log \bar{x}_h \right) \right) \left(\sum_{h=1}^{L} W_h Q_h \log \bar{x}_h \right)}{\left(\sum_{h=1}^{L} W_h Q_h \log \bar{x}_h^2 \right) \left(\sum_{h=1}^{L} W_h Q_h \right) - \left(\sum_{h=1}^{L} W_h Q_h \log \bar{x}_h \right)^2} \right]$$
$$+ W_h Q_h \log \bar{x}_h \left[\frac{\left(\sum_{h=1}^{L} W_h Q_h) \right) \left(\sum_{h=1}^{L} W_h \left(\log \bar{x}_h^* - \log \bar{x}_h \right) \right)}{\left(\sum_{h=1}^{L} W_h Q_h \log x_h^2 \right) \left(\sum_{h=1}^{L} W_h Q_h \right) - \left(\sum_{h=1}^{L} W_h Q_h \log \bar{x}_h \right)^2} \right] \tag{27}$$

Thus, substituting the value of Ω_h^* from equation (27) in (22), the proposed calibrated estimator is obtained as:

$$\bar{y}_{r\_Ld} = \sum_{h=1}^{L} W_h \bar{y}_{Rh\_d} + \left[\hat{\beta}_{r\_Ld} \sum_{h=1}^{L} W_h \left(\log \bar{x}_h^* - \log \bar{x}_h \right) \right] \tag{28}$$

where

$$\hat{\beta}_{r\_Ld} = \left[\frac{\left(\sum_{h=1}^{L} W_h Q_h \right) \left(\sum_{h=1}^{L} W_h Q_h \bar{y}_{Rh\_d} \log \bar{x}_h \right) - \left(\sum_{h=1}^{L} W_h Q_h \log \bar{x}_h \right) \left(\sum_{h=1}^{L} W_h Q_h \bar{y}_{Rh\_d} \right)}{\left(\sum_{h=1}^{L} W_h Q_h \log \bar{x}_h^2 \right) \left(\sum_{h=1}^{L} W_h Q_h \right) - \left(\sum_{h=1}^{L} W_h Q_h \log \bar{x}_h \right)^2} \right]$$

For different values of $Q_{h,}$ we can obtain different forms of calibration estimators.

The mean squared error of the estimator $\bar{y}_{r\_Ld}$ up to the second order of approximation is given as:

$$\mathrm{MSE}\left(\bar{y}_{r\_Ld}\right) = E\left[\bar{y}_{r\_Ld} - \bar{Y}\right]^2$$

$$= \sum_{h=1}^{L} W_h^2 \left[f_h' \bar{Y}_h^2 C_{yh}^2 + f_h''(\bar{Y}_h + \beta_{r\_Ld})^2 C_{xh}^2 - 2 f_h'' \rho_h \bar{Y}_h (\bar{Y}_h + \beta_{r\_Ld}) C_{yh} C_{xh} \right]$$

where $\beta_{r\_Ld} = \left[\dfrac{\left(\sum_{h=1}^{L} W_h \log \bar{X}_h \bar{Y}_h\right) - \left(\sum_{h=1}^{L} W_h \log \bar{X}_h\right)\left(\sum_{h=1}^{L} W_h \bar{Y}_h\right)}{\left(\sum_{h=1}^{L} W_h \log \bar{X}_h^2\right) - \left(\sum_{h=1}^{L} W_h \log \bar{X}_h\right)^2} \right]$ and

$$f_h'' = \left(\frac{1}{n_h} - \frac{1}{m_h} \right)$$

3. Simulation Study

3.1. Real Dataset

The real population used here in order to study the performance of the proposed calibrated estimator is MU284 population given in Appendix B of Sarndal et al. (2003). It comprises 284 units, divided into eight strata of varying sizes. The y and x are the study variable and the auxiliary variable respectively. We have selected simple random samples without replacement (SRSWOR) from each stratum using proportional allocation. A simulated study has been conducted by generating 25,000 samples of different sizes in R-software. The description of the population is given as:

X: REV84, Real state value according to 1984 assessment (in millions)

Y: RMT85, Revenues from 1985 municipal taxation (in millions)

X		**Y**	
Min.	347	Min.	21.00
1st Quartile	1146	1st Quartile	67.75
Median	1854	Median	113.50
Mean	3088	Mean	245.19
3rd Quartile	3345	3rd Quartile	230.25
Max.	59877	Max.	6720.00

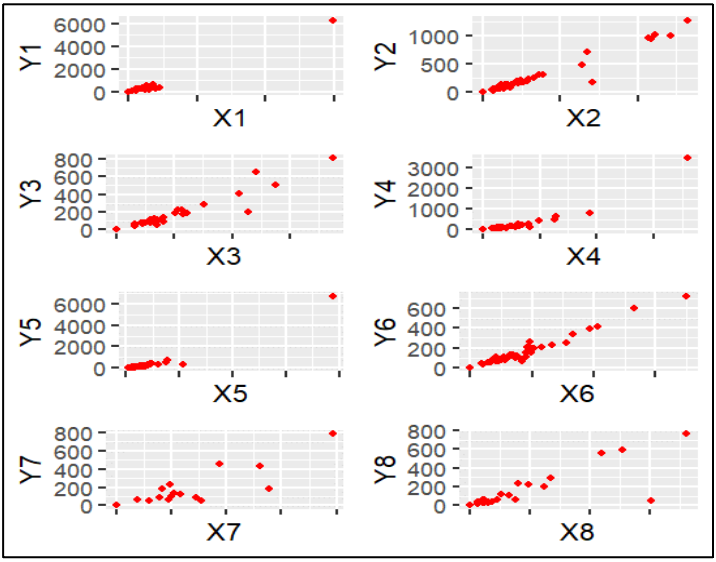

Figure 1: Plot between all strata of the population.

We have computed the empirical percentage absolute relative bias (%ARB) and percentage relative root mean squared error (%RRMSE) of the estimators as:

$$\%ARB(\bar{y}_\alpha) = \left| \frac{(\bar{y}_\alpha - \bar{Y})}{\bar{Y}} \right| \times 100$$

$$\%RRMSE(\bar{y}_\alpha) = \sqrt{\frac{1}{25000} \sum_{i=1}^{25000} \left(\frac{(\bar{y}_\alpha - \bar{Y})}{\bar{Y}} \right)^2} \times 100$$

where $\alpha = tr,\ s,\ r\_tr,\ r\_s,\ r\_L,\ tr.d,\ s.d, r\_tr.d,\ r\_s.d, r\_Ld$

The computed values of the %ARB and %RRMSE are given in the following tables for stratified sampling and stratified double sampling.

Table 1: Percentage Absolute Relative Bias (%ARB) for MU284 Population under Stratified Sampling

Q_h	**Sample Size**	**%ARB** ($\bar{y}_{tr}$)	**%ARB** ($\bar{y}_{r\_tr}$)	**%ARB** ($\bar{y}_s$)	**%ARB** ($\bar{y}_{r\_s}$)	**%ARB** ($\bar{y}_{r\_L}$)
1	**30**	3.004	49.304	5.668	5.969	2.803
	35	3.513	46.094	4.976	5.272	2.596
	40	4.096	42.126	4.505	5.015	2.348
	45	4.937	37.622	3.845	4.320	1.931
	50	5.186	39.019	3.540	3.878	1.821

Table 1: Percentage Absolute Relative Bias (%ARB) for MU284 Population under Stratified Sampling (cont.)

Q_h	Sample Size	%ARB ($\bar{y}_{tr}$)	%ARB ($\bar{y}_{r\_tr}$)	%ARB ($\bar{y}_s$)	%ARB ($\bar{y}_{r\_s}$)	%ARB ($\bar{y}_{r\_L}$)
$\frac{1}{\bar{x}_h}$	**30**	3.374	59.629	4.886	6.367	3.835
	35	3.690	53.318	4.441	5.777	3.564
	40	4.231	47.327	4.077	5.568	3.385
	45	4.762	41.495	3.584	4.918	2.907
	50	4.988	41.971	3.388	4.532	2.795
$\frac{1}{\log \bar{x}_h}$	**30**	3.051	50.355	5.589	6.027	2.952
	35	3.540	46.824	4.922	5.340	2.734
	40	4.125	42.619	4.463	5.094	2.496
	45	4.932	37.982	3.818	4.401	2.067
	50	5.178	39.270	3.525	3.963	1.959

Table 2: Percentage Relative Root Mean Square Error (%RRMSE) for MU284 under Stratified Sampling

Q_h	Sample Size	%RRMSE ($\bar{y}_{tr}$)	%RRMSE ($\bar{y}_{r\_tr}$)	%RRMSE ($\bar{y}_s$)	%RRMSE ($\bar{y}_{r\_s}$)	%RRMSE ($\bar{y}_{r\_L}$)
1	**30**	44.863	105.661	15.356	13.701	12.004
	35	39.069	91.444	14.491	13.389	11.740
	40	37.057	82.211	13.932	12.662	11.082
	45	33.878	74.189	13.424	12.482	10.926
	50	33.319	73.684	13.047	12.098	10.727
$\frac{1}{\bar{x}_h}$	**30**	39.633	106.641	15.124	12.951	11.220
	35	35.028	93.334	14.124	12.709	10.975
	40	33.178	82.878	13.520	11.958	10.318
	45	30.599	73.129	13.003	11.802	10.186
	50	30.383	73.502	12.613	11.513	10.045
$\frac{1}{\log \bar{x}_h}$	**30**	44.124	105.018	15.293	13.575	11.849
	35	38.496	91.088	14.427	13.273	11.596
	40	36.495	81.780	13.866	12.549	10.942
	45	33.411	73.616	13.363	12.371	10.794
	50	32.896	73.266	12.989	11.998	10.605

Table 3: Percentage Absolute Relative Bias (%ARB) for MU284 Population under Stratified Double Sampling

Q_h	Sample Size	%ARB ($\bar{y}_{tr.d}$)	%ARB ($\bar{y}_{r\_tr.d}$)	%ARB ($\bar{y}_{s.d}$)	%ARB ($\bar{y}_{r\_s.d}$)	%ARB ($\bar{y}_{r\_Ld}$)
1	**120;30**	2.418	43.856	4.679	4.847	1.814
	120;35	2.933	40.262	3.881	3.763	1.629
	120;40	2.727	33.652	3.305	3.824	1.405
$\frac{1}{\bar{x}_h}$	**120;30**	2.274	45.533	4.791	5.008	1.941
	120;35	2.749	41.887	4.165	3.917	1.753
	120;40	2.960	36.076	3.505	3.783	1.374
$\frac{1}{\log \bar{x}_h}$	**120;30**	2.834	47.344	4.015	5.123	2.783
	120;35	3.113	41.072	3.441	4.139	2.366
	120;40	2.841	33.352	2.938	4.159	2.197
1	**125;30**	2.707	49.029	4.103	5.260	2.891
	125;35	2.912	43.678	3.734	4.313	2.540
	125;40	3.052	36.093	3.167	4.177	2.288
$\frac{1}{\bar{x}_h}$	**125;30**	2.473	44.109	4.612	4.892	1.948
	125;35	2.962	40.234	3.836	3.815	1.733
	125;40	2.748	33.508	3.268	3.874	1.515
$\frac{1}{\log \bar{x}_h}$	**125;30**	2.329	45.778	4.721	5.050	2.074
	125;35	2.777	41.982	4.122	3.972	1.865
	125;40	2.980	35.963	3.471	3.842	1.499

Table 4: Percentage Relative Root Mean Square Error (%RRMSE) for MU284 under Stratified Double Sampling

Q_h	Sample Size	%RRMSE ($\bar{y}_{tr.d}$)	%RRMSE ($\bar{y}_{r\_tr.d}$)	%RRMSE ($\bar{y}_{s.d}$)	%RRMSE ($\bar{y}_{r\_s.d}$)	%RRMSE ($\bar{y}_{r\_Ld}$)
1	**120;30**	52.709	150.221	21.594	21.904	20.361
	120;35	49.091	133.640	20.869	21.841	19.877
	120;40	43.608	113.620	20.632	21.131	19.725
$\frac{1}{\bar{x}_h}$	**120;30**	47.533	134.851	22.416	20.587	18.853
	120;35	44.435	117.957	21.529	20.658	18.676
	120;40	39.980	97.357	21.084	19.941	18.455
$\frac{1}{\log \bar{x}_h}$	**120;30**	51.970	147.418	21.642	21.704	20.122
	120;35	48.456	131.082	20.906	21.660	19.692
	120;40	43.101	111.061	20.657	20.957	19.536
1	**125;30**	52.495	160.487	20.635	21.585	19.975
	125;35	48.481	135.201	20.205	21.817	19.786
	125;40	43.939	108.609	19.902	20.999	19.510
$\frac{1}{\bar{x}_h}$	**125;30**	47.294	143.024	20.414	20.275	18.495
	125;35	43.954	121.019	19.837	20.549	18.469
	125;40	40.048	93.322	19.502	19.756	18.139
$\frac{1}{\log \bar{x}_h}$	**125;30**	51.742	157.428	20.581	21.383	19.739
	125;35	47.835	132.770	20.147	21.625	19.583
	125;40	43.392	106.078	19.842	20.816	19.305

3.2. Artificial Dataset

A finite population of size N=3000 is generated for 3 strata using SRSWOR within each stratum, where the stratum sizes (N_h) are 1000 each, respectively. The values for the auxiliary variable X are generated considering the exponential distribution with varying values of the parameter for each stratum and the variable of interest Y is generated using the following models:

1st strata: $X_1 = \text{Exponential}(1000;\ 02)$ and $Y_1 = 100 + (\beta_1 * X_1) + \varepsilon_1$

where, $\beta_1 = 0.25$ and $\varepsilon_1 \sim N(0,2)$

2nd strata: $X_2 = \text{Exponential}(1000;\ 03)$ and $Y_2 = 200 + (\beta_2 * X_2) + \varepsilon_2$

where, $\beta_2 = 0.50$ and $\varepsilon_2 \sim N(0,3)$

3rd strata: $X_3 = \text{Exponential}(1000;\ 08)$ and $Y_3 = 300 + (\beta_3 * X_3) + \varepsilon_3$

where, $\beta_3 = 0.75$ and $\varepsilon_3 \sim N(0,4)$

Table 5: Percentage Absolute Relative Bias (%ARB) for Exponential Population under Stratified Sampling

Q_h	**Sample Size**	**%ARB** ($\overline{y}_{tr}$)	**%ARB** ($\overline{y}_{r\_tr}$)	**%ARB** ($\overline{y}_{s}$)	**%ARB** ($\overline{y}_{r\_s}$)	**%ARB** ($\overline{y}_{r\_L}$)
1	**100**	24.082	21.017	1.396	1.690	1.310
	200	12.345	10.586	0.716	0.729	0.593
	300	7.928	6.810	0.415	0.474	0.374
	400	5.867	5.021	0.321	0.322	0.262
	500	4.870	4.294	0.226	0.220	0.180
$\frac{1}{\overline{x}_h}$	**100**	29.162	25.942	1.394	1.851	1.427
	200	14.305	12.355	0.698	0.813	0.645
	300	8.884	7.634	0.401	0.528	0.406
	400	6.437	5.480	0.310	0.361	0.285
	500	5.281	4.617	0.217	0.250	0.196
$\frac{1}{\log \overline{x}_h}$	**100**	20.674	17.744	1.390	1.555	1.247
	200	10.896	9.273	0.730	0.652	0.559
	300	7.196	6.163	0.426	0.424	0.354
	400	5.388	4.618	0.330	0.284	0.246
	500	4.543	4.029	0.234	0.190	0.169

Table 6: Percentage Relative Root Mean Square Error (%RRMSE) for Exponential Population under Stratified Sampling

Q_h	Sample Size	%RRMSE ($\bar{y}_{tr}$)	%RRMSE ($\bar{y}_{r\_tr}$)	%RRMSE ($\bar{y}_s$)	%RRMSE ($\bar{y}_{r\_s}$)	%RRMSE ($\bar{y}_{r\_L}$)
1	**100**	103.650	102.677	9.405	8.880	5.517
	200	56.659	57.615	6.328	5.830	3.661
	300	43.196	43.660	5.040	4.582	2.904
	400	35.456	35.661	4.264	3.886	2.462
	500	30.986	30.966	3.756	3.361	2.134
$\frac{1}{\bar{x}_h}$	**100**	116.032	116.691	9.411	8.930	5.854
	200	63.690	65.117	6.267	5.834	3.869
	300	48.085	48.754	4.977	4.578	3.063
	400	39.427	39.731	4.208	3.878	2.594
	500	34.486	34.523	3.701	3.353	2.248
$\frac{1}{\log \bar{x}_h}$	**100**	97.749	96.804	9.513	8.837	5.249
	200	52.652	53.345	6.449	5.831	3.465
	300	40.404	40.746	5.144	4.590	2.742
	400	33.062	33.191	4.348	3.897	2.324
	500	28.890	28.825	3.832	3.372	2.012

Table 7: Percentage Absolute Relative Bias (%ARB) for Exponential Population under Stratified Double Sampling

Q_h	Sample Size	%ARB ($\bar{y}_{tr.d}$)	%ARB ($\bar{y}_{r\_tr.d}$)	%ARB ($\bar{y}_{s.d}$)	%ARB ($\bar{y}_{r\_s.d}$)	%ARB ($\bar{y}_{r\_Ld}$)
1	**1000;100**	21.484	18.734	1.166	1.671	1.254
	1000; 200	9.858	8.318	0.562	0.717	0.552
	1000; 300	5.971	4.980	0.358	0.402	0.321
$\frac{1}{\bar{x}_h}$	**1000; 100**	25.808	22.927	1.157	1.819	1.351
	1000; 200	11.390	9.689	0.549	0.788	0.597
	1000; 300	6.692	5.579	0.348	0.443	0.348
$\frac{1}{\log \bar{x}_h}$	**1000; 100**	18.584	15.921	1.165	1.547	1.202
	1000; 200	8.721	7.289	0.572	0.652	0.523
	1000; 300	5.433	4.522	0.367	0.364	0.303
1	**1200; 100**	22.256	19.293	1.280	1.622	1.241
	1200; 200	10.769	9.255	0.641	0.672	0.547
	1200; 300	6.640	5.666	0.363	0.397	0.317
$\frac{1}{\bar{x}_h}$	**1200; 100**	26.710	23.547	1.276	1.774	1.347
	1200; 200	12.407	10.743	0.624	0.744	0.592
	1200; 300	7.420	6.329	0.352	0.442	0.344
$\frac{1}{\log \bar{x}_h}$	**1200; 100**	19.249	16.413	1.274	1.494	1.180
	1200; 200	9.574	8.162	0.656	0.607	0.518
	1200; 300	6.042	5.147	0.372	0.354	0.299

Table 8: Percentage Relative Root Mean Square Error (%RRMSE) for Exponential Population under Stratified Double Sampling

Q_h	Sample Size	%RRMSE ($\bar{y}_{tr.d}$)	%RRMSE ($\bar{y}_{r\_tr.d}$)	%RRMSE ($\bar{y}_{s.d}$)	%RRMSE ($\bar{y}_{r\_s.d}$)	%RRMSE ($\bar{y}_{r\_Ld}$)
1	**1000;100**	88.753	92.559	8.995	8.514	5.279
	1000; 200	51.137	52.014	5.825	5.404	3.396
	1000; 300	37.607	37.821	4.412	4.006	2.543
$\frac{1}{\bar{x}_h}$	**1000; 100**	101.307	106.522	8.999	8.567	5.583
	1000; 200	57.689	58.891	5.774	5.411	3.578
	1000; 300	42.086	42.396	4.359	4.000	2.676
$\frac{1}{\log \bar{x}_h}$	**1000; 100**	81.993	85.434	9.101	8.473	5.034
	1000; 200	47.245	47.980	5.932	5.405	3.219
	1000; 300	34.956	35.113	4.502	4.018	2.406
1	**1200; 100**	103.234	106.378	9.108	8.618	5.346
	1200; 200	51.810	52.845	5.941	5.474	3.452
	1200; 300	39.073	39.305	4.590	4.209	2.667
$\frac{1}{\bar{x}_h}$	**1200; 100**	115.116	119.277	9.108	8.670	5.654
	1200; 200	58.494	59.851	5.884	5.474	3.641
	1200; 300	43.586	43.981	4.534	4.204	2.807
$\frac{1}{\log \bar{x}_h}$	**1200; 100**	97.379	99.971	9.223	8.575	5.097
	1200; 200	47.916	48.815	6.055	5.480	3.270
	1200; 300	36.419	36.555	4.684	4.219	2.523

4. Conclusion

The ratio-type calibration estimators have been proposed under the stratified random sampling and stratified double sampling. The simulation study conducted on real as well as on artificial populations has shown that the proposed estimators have less %RRMSE and their value decrease as we increase the sample sizes, which in turn explains its performance better than the existing estimators. It is concluded that the proposed ratio-type calibration estimators using logarithmic stratum mean of the auxiliary variable in calibration constraint for the population mean in the case of stratified random sampling and stratified double sampling are found to be more efficient than the estimators given by Tracy et al. (2003), Nidhi et al. (2017) and Khare et al. (2022).

References

Alka, Ekpenyong, E. J., Rai, P. K., Tiwari, S., (2021). Calibration Estimator of Population Mean under Stratified Systematic Sampling Design. *International Journal of Agricultural and Statistical Science*, 17(Sup.1), pp. 2353–2361.

Clement, E. P., Enang, E. I., (2017). On the efficiency of ratio estimator over the regression estimator. *Communications in Statistics-Theory and Methods*, 46(11), pp. 5357–5367.

Deville, J. C., Sarndal, C. E., (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association,* 87(418), pp. 376–382.

Garg, N., Pachori, M., (2019). Calibration estimation of population mean in stratified sampling using coefficient of skewness. *International Journal of Agricultural and Statistical Sciences*, 15(1), pp. 211–219.

Garg, N., Pachori, M., (2020). Use of coefficient of variation in calibration estimation of population mean in stratified sampling. *Communications in Statistics-Theory and Methods*, 49(23), pp. 5842–5852.

Garg, N., Pachori, M., (2021). A Logarithmic Calibration estimator of Population Mean in Stratified Double Sampling. *International Journal of Agricultural and Statistical Science*, 17(Sup.1), pp. 2019–2025.

Khare, B. B., Singh, S., Mishra, M., (2022). Separate Ratio Estimator Using Calibration Approach for the Population mean in Stratified Random Sampling. *Asian Journal of Probability and Statistics*, 20(3), pp. 64–73.

Kim, J-M., Sungur, E. A., Heo T-Y., (2007). Calibration approach estimators in stratified sampling. *Statistics and Probability Letters,* 77, pp. 99–103.

Koyuncu, N., Kadilar, C., (2013). Calibration estimator using different distance measures in stratified random sampling. *International Journal of Modern Engineering Research (IJMER),* 3(1), pp. 415–419.

Koyuncu, N., Kadilar, C., (2016). Calibration weighting in Stratified Random Sampling. *Communications in Statistics-Simulation and Computation,* 45(7), pp. 2267–2275.

Kumar, R., Biswas, A., Singh, D., (2023). Estimation of Ratio in Finite Population using Calibration Approach under Different Calibrated Weights Systems. *Journal of Indian Society of Agricultural Statistics*, 77(2), pp. 225–232.

Mouhamed, A. M., Ei-Sheikh, A. A., Mohamed, H. A., (2015). A new calibration estimator of Stratified random sampling. *Applied Mathematical Sciences,* 9(35), pp. 1735–1744.

Nidhi, Sisodia, B. V. S., Singh, S., Singh, S. K., (2017). Calibration approach estimation of the mean in stratified sampling and stratified double sampling. *Communications in Statistics-Theory and Methods,* 46(10), pp. 4932–4942.

Pachori, P., Garg, N., Sangal, P. K., Rajesh, (2022). Exponential-type Calibration Estimators of Finite Population Mean in Stratified Sampling. *International Journal of Agricultural and Statistical Science*, 18(1), pp. 431–436.

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

Singh, S., (2003). Golden jubilee year 2003 of the linear regression estimator. Working paper at St. Cloud State University, St. Cloud, MN, USA.

Singh, S., Arnab, R., (2011). On Calibration of design weights. *METRON-International Journal of Statistics,* LXIX(2), pp. 185–205.

Tracy, D. S., Singh, S., Arnab, R., (2003). Note on Calibration in Stratified and Double Sampling. *Survey Methodology,* 29(1), pp. 99–104.