Available online at http://isas.org.in/isa/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 77(2) 2023 225–232

Estimation of Ratio in Finite Population using Calibration Approach under Different Calibrated Weights Systems

Raju Kumar, Ankur Biswas and Deepak Singh

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Received 23 November 2022; Revised 14 June 2023; Accepted 03 August 2023

SUMMARY

The ratio in finite population is one of the most common statistics used in official statistics, demographic studies, agriculture and allied field of agriculture. In this paper, estimators of the ratio/proportion in finite population are developed by incorporating known auxiliary information under the calibration approach. The variance and the estimate of variance for these estimators are obtained. A simulation study is carried out to evaluate the performance of proposed estimators comparing them with a simple estimator of the Population ratio that does not incorporate auxiliary information.

Keywords: Calibration estimator; Population ratio; Probability sample; Auxiliary information.

1. INTRODUCTION

The ratio in a finite population, $R = Y/X$, where Y and X are random variables, is one of the most common statistics used in official statistics, agriculture and allied fields of agriculture. It has wide applications like age or sex ratio of animals in wildlife population, the number of bullocks per acre of holding, the unemployment rate in agriculture or other sectors, average salary etc. For the real life example we can consider the study taken by Smith *et al.* (1995) for assessing the sampling for species of waterfowl. Their evolution was based on a simulation experiment, and the samples collected from a count of ring-necked ducks (Aythyacollaris), blue-winged teals (Anasdiscors), and green-winged teals (Anascrecca) in a 5,000 km^2 area of central Florida. Ring-necked Ducks prefer shallow, freshwater wetlands with stable water levels and abundant emergent and submerged or floating plants. The Blue-winged Teal inhabits seasonal wetlands and wet meadows as well as shallow semi-permanent marshes (BNA). Green-winged Teal prefers shallow ponds with lots of emergent vegetation. On the basis of this, they defined available habit as open water and wetlands with herbaceous emergent vegetation. In environmental and ecological studies, such as the work by Smith *et al.* (1995), researchers often go beyond studying the total number or density of a particular bird or animal species in a given area. They also investigate the ratio of major bird species as an important measure to assess resource availability in different habitats. This approach aids in the effective management and conservation of bird species. In this scenario, the ratio of bird species serves as the parameter of interest, with the number of individuals for each species. Here number of ring-necked ducks and blue-winged tealscan be considered as variables X and Y, respectively. The habitat in which these birds reside plays a crucial role in determining their population numbers. Therefore, the corresponding habitat areas can be selected as auxiliary information for estimating the ratio of these bird species. For variable X, the auxiliary information can be the areas of open water, while for variable Y, the areas of wetlands can be used as auxiliary information. Each of Y and X are assumed to be estimated from the sample i.e. $r = y/x$. Commonly y and x are both simple totals in the sample of the "y" and "x" variate. When

Corresponding author: Deepak Singh
E-mail address: deepaksingh2112@gmail.com

a survey is conducted for the estimation of population parameters, survey experts are always concerned about the improvement of the precision of estimators. For this, the auxiliary information is the most important tool in sample surveys to improve the precision of estimators when measuring population parameters like mean, total, ratio etc. Using auxiliary information Deville and Särndal (1992) developed the calibration method which is used for increasing the precision of estimators and is now widely used to develop estimates of important population parameters.

Kish *et al.* (1962) discussed the ratio mean and ratio bias of two random variables in surveys. Chang and Huang (2013) proposed an improved estimator of the ratio of the population mean in the survey sampling when some observations are missing.

The problem of estimating the population ratio of two totals using the calibration method was attempted by Plikusas (2001) and Krapavickaite and Plikusas (2005). The theory for estimating the population ratio of two totals under a stratified random sampling design was developed by Barktus and Pumputis (2010) using the calibration approach. Sadikul (2019) *et al.* developed a calibration approach for the estimation of population ratio under double sampling when the availability of aggregate level population information for auxiliary variables is available. The calibration approach has been widely applied to estimate complex parameters such as population ratio, product, or variance. (Kim and Park (2010), Sud *et al.* (2014), Basak *et al.* (2017) and ozgul (2021)).

The aim of this paper is, to develop a calibration estimator of finite population ratio depending on the extent of availability of auxiliary information. By considering the different weights systems, variance and estimate of variance for the various calibration estimators of population ratio are obtained. A simulation study is used to compare the various estimators empirically.

2. THEORETICAL DEVELOPMENTS

Let a finite population U consisting of N distinguishable units $(1,2,...,N)$ labelled units. Let Y_i the value of Y and X_i the value of X are two variables defined understudy for i^{th} unit in the population and are unknown but observable in population U and take values $y_1, y_2, \dots, y_N$ and $x_1, x_2, \dots, x_N$, where $t_y = \sum_{i\in U} y_i$ and $t_x = \sum_{i\in U} x_i$. The objective is to estimate the population ratio denoted by $R = t_y / t_x$. Suppose, from population U of size N a sample $s (s \subset U)$ of size n, be drawn with any design. Let the first and second-order inclusion probabilities are $\pi_i = p(i \in s)$ and $\pi_{ij} = p(i \text{ and } j \in s)$ with the assumption that $\pi_i = p(i \in s)$ and $\pi_{ij} = p(i \text{ and } j \in s)$ strictly positive and known which is common in the literature. For the elements $i \in s$, observe (y_i, x_i). Horvitz Thompson estimator of the total is defined as $\hat{t}_{y\pi} = \sum_{i\in s} \frac{y_i}{\pi_i}$, $\hat{t}_{x\pi} = \sum_{i\in s} \frac{x_i}{\pi_i}$ and $\hat{t}_{a\pi} = \sum_{i\in s} \frac{a_i}{\pi_i}$ and $\hat{t}_{b\pi} = \sum_{i\in s} \frac{b_i}{\pi_i}$.

2.1 Simple estimator of population ratio

Consider a finite population $U = \{1, 2, ..., N\}$, consisting of N units. Let Y and X be two variables defined on the population U and take values $y_1, y_2, \dots, y_N$ and $x_1, x_2, \dots, x_N$ such that $t_y = \sum_{i\in U} y_i$ and $t_x = \sum_{i\in U} x_i$ are unknown. We are interested in the estimation of the ratio of the two totals $R = t_y / t_x$.

An estimator of the ratio R which does not incorporate the auxiliary information is given by

$$\hat{R} = \frac{\hat{t}_{y\pi}}{\hat{t}_{x\pi}} = \frac{\sum_{i\in s} \frac{y_i}{\pi_i}}{\sum_{i\in s} \frac{x_i}{\pi_i}} \tag{2.1.1}$$

Since ratio between two unknown totals is not a simple structure as population total, estimation of a population ratio can be obtained by Taylor linearization technique (see Särndal *et al.* (1992), page 177-179)

$$\hat{R} = \frac{\hat{t}_{y\pi}}{\hat{t}_{x\pi}} = f(\hat{t}_{y\pi}, \hat{t}_{x\pi})$$

The first-order Taylor expansion of the function $\hat{R}$ is given by $\hat{R}(l)$

$$\hat{R}(l) \cong R + \frac{1}{t_x}\left[\hat{t}_y - t_y\right] - \frac{R}{t_x}\left[\hat{t}_x - t_x\right] \cong R + \frac{1}{t_x}\left[\hat{t}_y - R\hat{t}_x\right].$$

The variance $\hat{R}(l)$ in the first order of approximation is given by $AVar(\hat{R}(l))$

$$AVar\left(\hat{R}(l)\right) = v(\hat{R}_0(l) = \frac{1}{t_x^2} Var\left[\hat{t}_y - R\hat{t}_x\right]$$

$$\text{or, } v(\hat{R}_0(l) = \frac{1}{t_x^2} \sum_U \sum_U \Delta_{ij} \frac{v_i}{\pi_i} \frac{v_j}{\pi_j}. \tag{2.1.2}$$

where

$v_i = y_i - Rx_i; \Delta_{ij} = \pi_{ij} - \pi_i\pi_j$.

An estimator of variance $\hat{R}$ is given by

$$\hat{v}(\hat{R}_{(l)}) = \frac{1}{\hat{t}_x^2}\sum_s\sum_s \frac{\Delta_{ij}}{\pi_{ij}}\frac{\hat{v}_i}{\pi_i}\frac{\hat{v}_j}{\pi_j}, \tag{2.1.3}$$

where

$\hat{v}_i = y_i - \hat{R}x_i$ and the quantity $1/t_x^2$ is estimated by $1/\hat{t}_x^2$.

2.2 Calibrated Estimators of the Population Ratio under different weights system

Situation-I: Same weight system for numerator and denominator

Suppose variable *b* is known and serves as an auxiliary variable for the variable *Y* and similarly a is known and serves as an auxiliary variable for the variable *X*, their values are $b_1, b_2, \ldots, b_N$ and $a_1, a_2, \ldots, a_N$ respectively.

Suppose population totals of the auxiliary variables are known i.e. $A = t_a = \sum_{i\in U} a_i$ and $B = t_b = \sum_{i\in U} b_i$ are known.

If w_i's are calibrated weights then the proposed estimator is

$$\hat{R}_I = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i x_i}$$

where the w_i's are obtained by minimizing the distance between the design weights $d_i = \frac{1}{\pi_i}; i = 1,2,\ldots,n$ and the calibrated weights $w_i;\ i = 1,2,\ldots,n$ subject to the constraints $\sum_{i=1}^n w_i a_i = A$ and $\sum_{i=1}^n w_i b_i = B$ (Deville & Särndal (1992)). Specifically, we minimize the Chi-square distance function $\phi = \sum_{i=1}^n \frac{(w_i - d_i)^2}{d_i q_i}$ subject to the constraints $\sum_{i=1}^n w_i a_i = A$ and $\sum_{i=1}^n w_i b_i = B$. Here, the q_i's are known positive individual's weight unrelated to d_i and used to generalize the distance function and the nature of estimator depends upon. For, the simplification we have taken $q_i = 1$ which is most common in application (Jaiswal *et al.*(2023).

We use the Lagrange multiplier technique for minimization. Thus, we consider the function

$$\psi = \frac{1}{2}\sum_{i=1}^n \frac{(w_i - d_i)^2}{d_i} + \lambda_1\left(\sum_{i=1}^n w_i a_i - A\right) + \lambda_2\left(\sum_{i=1}^n w_i b_i - B\right).$$

Differentiation of function ψ w.r.to w_i and equating the resultant expression to zero gives

$$w_i = d_i - \lambda_1 a_i d_i - \lambda_2 b_i d_i \tag{2.2.1}$$

Multiplying the expression (2.2.1) with a_i and summing over the sample values give

$$\sum_{i=1}^n w_i a_i = \sum_{i=1}^n a_i d_i - \lambda_1\sum_{i=1}^n a_i^2 d_i - \lambda_2\sum_{i=1}^n a_i b_i d_i$$

Multiplying the expression (2.2.1) with b_i and summing over the sample values give

$$\sum_{i=1}^n w_i b_i = \sum_{i=1}^n b_i d_i - \lambda_1\sum_{i=1}^n a_i b_i d_i - \lambda_2\sum_{i=1}^n b_i^2 d_i.$$

Solving the above equations gives

$$\lambda_1 = \frac{\sum_{i=1}^n a_i b_i d_i (B - \sum_{i=1}^n b_i d_i) + \left[\sum_{i=1}^n d_i b_i^2 (A - \sum_{i=1}^n a_i d_i)\right]}{\left(\sum_{i=1}^n a_i^2 d_i\right)\left(\sum_{i=1}^n b_i^2 d_i\right) - \left(\sum_{i=1}^n a_i b_i d_i\right)^2}$$

$$\lambda_2 = \frac{\left[\sum_{i=1}^n a_i b_i d_i (A - \sum_{i=1}^n a_i d_i) + \sum_{i=1}^n d_i a_i^2 (B - \sum_{i=1}^n b_i d_i)\right]}{\left(\sum_{i=1}^n a_i^2 d_i\right)\left(\sum_{i=1}^n b_i^2 d_i\right) - \left(\sum_{i=1}^n a_i b_i d_i\right)^2}.$$

Putting the value of λ_1 and λ_2 in equation (2.2.1) and after simplification we get

$$w_i = d_i + \frac{a_i d_i\left[\sum_{i=1}^n d_i a_i^2 (B - \sum_{i=1}^n b_i d_i) - \sum_{i=1}^n a_i b_i d_i (A - \sum_{i=1}^n a_i d_i)\right]}{\left(\sum_{i=1}^n a_i^2 d_i\right)\left(\sum_{i=1}^n b_i^2 d_i\right) - \left(\sum_{i=1}^n a_i b_i d_i\right)^2}$$

$$+ \frac{b_i d_i\left[\sum_{i=1}^n d_i b_i^2 (A - \sum_{i=1}^n a_i d_i) - \sum_{i=1}^n a_i b_i d_i (B - \sum_{i=1}^n b_i d_i)\right]}{\left(\sum_{i=1}^n a_i^2 d_i\right)\left(\sum_{i=1}^n b_i^2 d_i\right) - \left(\sum_{i=1}^n a_i b_i d_i\right)^2}.$$

Thus the proposed calibrated estimator of the population ratio is

$$\hat{R}_I = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i x_i}$$

$$= \frac{\sum_{i=1}^{n} d_i y_i + (B - \sum_{i=1}^{n} b_i d_i)\hat{L}_1 + (A - \sum_{i=1}^{n} a_i d_i)\hat{L}_2}{\sum_{i=1}^{n} d_i x_i + (B - \sum_{i=1}^{n} b_i d_i)\hat{L}_3 + (A - \sum_{i=1}^{n} a_i d_i)\hat{L}_4}$$

$$= \frac{\hat{t}_y + (t_b - \hat{t}_b)\hat{L}_1 + (t_a - \hat{t}_a)\hat{L}_2}{\hat{t}_x + (t_b - \hat{t}_b)\hat{L}_3 + (t_a - \hat{t}_a)\hat{L}_4} \quad (2.2.2)$$

where

$$\hat{L}_1 = \frac{\sum_{i=1}^{n} a_i d_i y_i \sum_{i=1}^{n} d_i a_i^2 - \sum_{i=1}^{n} b_i d_i y_i \sum_{i=1}^{n} a_i b_i d_i}{\left(\sum_{i=1}^{n} a_i^2 d_i\right)\left(\sum_{i=1}^{n} b_i^2 d_i\right) - \left(\sum_{i=1}^{n} a_i b_i d_i\right)^2}$$

$$\hat{L}_2 = \frac{\sum_{i=1}^{n} b_i d_i y_i \sum_{i=1}^{n} d_i b_i^2 - \sum_{i=1}^{n} a_i d_i y_i \sum_{i=1}^{n} a_i b_i d_i}{\left(\sum_{i=1}^{n} a_i^2 d_i\right)\left(\sum_{i=1}^{n} b_i^2 d_i\right) - \left(\sum_{i=1}^{n} a_i b_i d_i\right)^2}$$

$$\hat{L}_3 = \frac{\sum_{i=1}^{n} a_i d_i x_i \sum_{i=1}^{n} d_i a_i^2 - \sum_{i=1}^{n} b_i d_i x_i \sum_{i=1}^{n} a_i b_i d_i}{\left(\sum_{i=1}^{n} a_i^2 d_i\right)\left(\sum_{i=1}^{n} b_i^2 d_i\right) - \left(\sum_{i=1}^{n} a_i b_i d_i\right)^2}$$

$$\hat{L}_4 = \frac{\sum_{i=1}^{n} b_i d_i x_i \sum_{i=1}^{n} d_i b_i^2 - \sum_{i=1}^{n} a_i d_i x_i \sum_{i=1}^{n} a_i b_i d_i}{\left(\sum_{i=1}^{n} a_i^2 d_i\right)\left(\sum_{i=1}^{n} b_i^2 d_i\right) - \left(\sum_{i=1}^{n} a_i b_i d_i\right)^2}$$

Following Särndal *et al.* (1992), using the Taylor Linearization technique, we obtained the first-order Taylor expansion of the $\hat{R}_I$

$$\hat{R}_I(l) \cong R + \frac{1}{t_x}\left[\hat{t}_y - t_y\right] - \frac{R}{t_x}\left[\hat{t}_x - t_x\right] + \frac{L_3 R - L_1}{t_x}\left[\hat{t}_b - t_b\right] + \frac{L_4 R - L_2}{t_x}\left(\hat{t}_a - t_a\right),$$

where quantity L_1, L_2, L_3 and L_4 are estimate by $\hat{L}_1, \hat{L}_2, \hat{L}_3, \hat{L}_4$.

The approximate variance of $\hat{R}_I(l)$ in the first order of approximation is obtained by

$$AVar\left(\hat{R}_I(l)\right) = v(\hat{R}_{Io}(l) = \frac{1}{t_x^2} Var\left[\hat{t}_y - R\hat{t}_x + \left(L_3 R - L_1\right)\hat{t}_b + \left(L_4 R - L_2\right)\hat{t}_a\right]$$

$$\text{or, } v(\hat{R}_{Io}(l) = \frac{1}{t_x^2}\sum_U \sum_U \Delta_{ij} \frac{v_i}{\pi_i}\frac{v_j}{\pi_j}, \quad (2.2.3)$$

where

$$v_i = y_i - Rx_i + \left(L_3 R - L_1\right)b_i + \left(L_4 R - L_2\right)a_i.$$

An estimator of variance $\hat{R}$ is given by

$$\hat{v}(\hat{R}_I(l)) = \frac{1}{\hat{t}_x^2}\sum_s \sum_s \frac{\Delta_{ij}}{\pi_{ij}}\frac{\hat{v}_i}{\pi_i}\frac{\hat{v}_j}{\pi_j},$$

where

$$v_i = y_i - \hat{R}x_i + \left(\hat{L}_3 \hat{R} - \hat{L}_1\right)b_i + \left(\hat{L}_4 \hat{R} - \hat{L}_2\right)a_i.$$

Situation-II: Different weights system for numerator and denominator

Like situation-I, the totals *A* and *B* are known individually. We use two systems of calibrated weights for the estimation of the population ratio of the two totals. If w_{3i} and w_{4i} (i=1,2,…,n) are the calibrated weights then the proposed estimator is

$$\hat{R}_{II} = \frac{\sum_{i \in s} w_{3i} y_i}{\sum_{i \in s} w_{4i} x_i}.$$

Where w_{3i} and w_{4i} are obtained by minimizing the distance between the design weights d_i and calibrated weight w_{3i} and w_{4i} subject to the constraints $\sum_{i=1}^{n} w_{3i} a_i = A$ and $\sum_{i=1}^{n} w_{4i} b_i = B$. Specifically, we minimized the loss functions $L_1 = \sum_{i=1}^{n} \frac{(w_{3i} - d_i)^2}{d_i q_{1i}}$ and $L_2 = \sum_{i=1}^{n} \frac{(w_{4i} - d_i)^2}{d_i q_{2i}}$ with $q_{1i} = 1$ and $q_{2i} = 1$ subject to the constraints $\sum_{i=1}^{n} w_{3i} a_i = A$ and $\sum_{i=1}^{n} w_{4i} b_i = B$.

We use the Lagrange multiplier technique for minimization. Thus, we consider the function

$$\psi = \frac{1}{2}\sum_{i=1}^{n} \frac{(w_{3i} - d_i)^2}{d_i} + \frac{1}{2}\sum_{i=1}^{n} \frac{(w_{4i} - d_i)^2}{d_i} + \lambda_1\left(\sum_{i=1}^{n} w_{3i} a_i - A\right) + \lambda_2\left(\sum_{i=1}^{n} w_{4i} b_i - B\right).$$

Differentiation of function, ψ w.r.to w_{3i} and w_{4i} equating to zero gives

$$w_{3i} = d_i - \lambda_1 a_i d_i, \quad (2.2.4)$$

$$w_{4i} = d_i - \lambda_2 b_i d_i. \quad (2.2.5)$$

Multiplying 2.2.4 by a_i and summing and similarly multiplying (2.2.5) by b_i and summing gives

$$\lambda_1 = \frac{\sum_{i=1}^{n} a_i d_i - a_i w_{3i}}{\sum_{i=1}^{n} a_i^2 d_i}, \quad \text{and} \quad \lambda_2 = \frac{\sum_{i=1}^{n} b_i d_i - b_i w_{4i}}{\sum_{i=1}^{n} b_i^2 d_i}.$$

Putting the value λ_1 and λ_2 in equations (2.2.4) and (2.2.5) gives

$$w_{3i} = d_i + \frac{d_i a_i}{\sum_{i=1}^{n} d_i a_i^2}\left(A - \sum_{i=1}^{n} d_i a_i\right),$$

$$w_{4i} = d_i + \frac{d_i b_i}{\sum_{i=1}^{n} d_i b_i^2}\left(B - \sum_{i=1}^{n} d_i b_i\right).$$

Hence the proposed calibrated estimator of the ratio of two population totals is

$$\hat{R}_{II} = \frac{\sum_{i=1}^{n} w_{4i} y_i}{\sum_{i=1}^{n} w_{3i} x_i}$$

$$= \frac{\sum_{i=1}^{n}\left\{d_i + \frac{d_i b_i}{\sum_{i=1}^{n} d_i b_i^2}\left(B - \sum_{i=1}^{n} d_i b_i\right)\right\} y_i}{\sum_{i=1}^{n}\left\{d_i + \frac{d_i a_i}{\sum_{i=1}^{n} d_i a_i^2}\left(B - \sum_{i=1}^{n} d_i a_i\right)\right\} x_i}$$

$$= \frac{\hat{t}_y + \left(t_b - \hat{t}_b\right)\hat{\beta}_y}{\hat{t}_x + \left(t_a - \hat{t}_a\right)\hat{\beta}_x}, \qquad (2.2.6)$$

where

$$\hat{\beta}_x = \frac{\sum_{i=1}^{n} d_i b_i x_i}{\sum_{i=1}^{n} d_i b_i^2},$$

$$\hat{\beta}_y = \frac{\sum_{i=1}^{n} d_i b_i y_i}{\sum_{i=1}^{n} d_i b_i^2}.$$

The first-order Taylor expansion of the function $\hat{R}_{II}$ is given by

$$\hat{R}_{II}\left(l\right) \cong R + \frac{1}{t_x}\left[\hat{t}_y - t_y\right] - \frac{R}{t_x}\left[\hat{t}_x - t_x\right] + \frac{R\beta_x}{t_x}\left(\hat{t}_a - t_a\right) - \frac{\beta_y}{t_x}\left[\hat{t}_b - t_b\right].$$

The variance $\hat{R}_{II}\left(l\right)$ in the first order of approximation is given by

$$AVar\left(\hat{R}_{II}(l)\right) = v\left(\hat{R}_{II0}(l)\right) = \frac{1}{t_x^2} Var\left[\left[\hat{t}_y - t_y\right] - R\left[\hat{t}_x - t_x\right] + R\beta_x\left(\hat{t}_a - t_a\right) - \beta_y\left[\hat{t}_b - t_b\right]\right]$$

$$v\left(\hat{R}_{II0}(l)\right) = \frac{1}{t_x^2} Var\left[\left(\hat{t}_y - R\hat{t}_x\right) - \left(\beta_y \hat{t}_b - R\hat{t}_a \beta_x\right)\right]$$

$$= \frac{1}{t_x^2}\sum_U \sum_U \Delta_{ij}\frac{v_{3i}}{\pi_i}\frac{v_{3j}}{\pi_j}, \qquad (2.2.7)$$

where

$$v_{3i} = y_i - Rx_i - \beta_y b_i + Ra_i\beta_x.$$

The estimate of the variance $\hat{R}_{II}\left(l\right)$ to the first order of approximation is given by

$$AVar\left(\hat{R}_{II}(l)\right) = \frac{1}{t_x^2} Var\left[\left[\hat{t}_y - t_y\right] - R\left[\hat{t}_x - t_x\right] + R\beta_x\left(\hat{t}_a - t_a\right) - \beta_y\left[\hat{t}_b - t_b\right]\right]$$

$$= \frac{1}{t_x^2} Var\left[\left(\hat{t}_y - R\hat{t}_x\right) - \left(\beta_y \hat{t}_b - R\hat{t}_a \beta_x\right)\right]$$

$$= \frac{1}{t_x^2}\sum_U \sum_U \Delta_{ij}\frac{v_{3i}}{\pi_i}\frac{v_{3j}}{\pi_j}, \qquad (2.2.8)$$

where

$$v_{3i} = y_i - Rx_i - \beta_y b_i + Ra_i\beta_x.$$

An estimator of variance $\hat{R}$ is given by

$$\hat{v}(\hat{R}_{II}(l)) = \frac{1}{\hat{t}_x^2}\sum_s \sum_s \frac{\Delta_{ij}}{\pi_{ij}}\frac{\hat{v}_i}{\pi_i}\frac{\hat{v}_j}{\pi_j},$$

$$\hat{v}_i = y_i - \hat{R}x_i - \hat{\beta}_y b_i + \hat{R}a_i\hat{\beta}_x.$$

Situation-III: The population ratio of auxiliary variables is used under calibration.

The population ratio of auxiliary variables is available. For example, suppose that we are interested in estimation of productivity of a crop and from previous year data on fertilizer consumption per hectare is available. i.e. population ratio of auxiliary variables is known.

$R_0 = \sum_{i\in U} b_i \Big/ \sum_{i\in U} a_i$ is known.

The proposed estimator using the calibrated weights w_i is given by $\hat{R}_{III} = \sum_{i=1}^{n} w_i y_i \Big/ \sum_{i=1}^{n} w_i x_i$

where the w_{2i}'s are obtained by minimizing the distance between the design weights d_i and calibrated weight w_{2i} subject to constraints $R_0 = \sum_{i=1}^{n} w_{2i} b_i \Big/ \sum_{i=1}^{n} w_{2i} a_i$. Specifically, we minimize the function $\phi = \sum_{i=1}^{n} \frac{(w_{2i}-d_i)^2}{d_i q_i}$ with $q_i = 1$ subject to constraints $R_0 = \sum_{i=1}^{n} w_{2i} b_i \Big/ \sum_{i=1}^{n} w_{2i} a_i$.

We use the Lagrange multiplier technique for minimization. Thus we consider the function

$$\psi = \sum_{i=1}^{n} \frac{(w_{2i}-d_i)^2}{d_i} + \lambda_1 \left(\sum_{i=1}^{n} R_0 w_{2i} a_i - \sum_{i=1}^{n} w_{2i} b_i \right).$$

Differentiation of function ψ w.r.to w_{2i} and equating to zero gives

$$w_{2i} = d_i - \lambda d_i (b - R_0 a_i) \qquad (2.2.9)$$

where $\lambda = \lambda_1/2$.

Multiplying (2.2.9) by $(b_i - Ra_i)$ and summing over sample values give

$$\lambda = \frac{\sum_{i=1}^{n} d_i (b_i - R_o a_i)}{\sum_{i=1}^{n} d_i (b_i - R_o a_i)^2}.$$

Putting the value λ in equation (2.2.9) give

$$w_{2i} = d_i - \left[\frac{\sum_{i=1}^{n} d_i (b_i - R_o a_i)}{\sum_{i=1}^{n} d_i (b_i - R_o a_i)^2} \right] d_i (b_i - R_o a_i).$$

Hence the proposed calibrated estimator of the population ratio is

$$\hat{R}_{III} = \frac{\sum_{i=1}^{n} w_{2i} y_i}{\sum_{i=1}^{n} w_{2i} x_i}$$

$$= \frac{\sum_{i=1}^{n} \left\{ d_i - \left[\sum_{i=1}^{n} d_i (b_i - R_o a_i) \Big/ \sum_{i=1}^{n} d_i (b_i - R_o a_i)^2 \right] d_i (b_i - R_o a_i) \right\} y_i}{\sum_{i=1}^{n} \left\{ d_i - \left[\sum_{i=1}^{n} d_i (b_i - R_o a_i) \Big/ \sum_{i=1}^{n} d_i (b_i - R_o a_i)^2 \right] d_i (b_i - R_o a_i) \right\} x_i}$$

$$= \frac{\sum_{i=1}^{n} d_i y_i \sum_{i=1}^{n} d_i (b_i - R_o a_i)^2 - \sum_{i=1}^{n} d_i (b_i - R_o a_i) y_i \sum_{i=1}^{n} d_i (b_i - R_o a_i)}{\sum_{i=1}^{n} d_i x_i \sum_{i=1}^{n} d_i (b_i - R_o a_i)^2 - \sum_{i=1}^{n} d_i (b_i - R_o a_i) y_i \sum_{i=1}^{n} d_i (b_i - R_o a_i) x_i}.$$

$$= \frac{\hat{t}_y \hat{t}_1 - \hat{t}_2 \hat{t}_3}{\hat{t}_x \hat{t}_1 - \hat{t}_2 \hat{t}_4}, \qquad (2.2.10)$$

where

$$t_1 = \sum_{i=1}^{N} d_i (b_i - R_o a_i)^2,$$

$$t_2 = \sum_{i=1}^{N} d_i (b_i - R_o a_i) y_i,$$

$$t_3 = \sum_{i=1}^{N} d_i (b_i - R_o a_i),$$

$$t_4 = \sum_{i=1}^{N} d_i (b_i - R_o a_i) x_i.$$

The first-order Taylor expansion of the function $\hat{R}_{II}$ is given by

$$\hat{R}_{III}(l) \cong R + \frac{1}{t_x}(\hat{t}_y - t_y) - \frac{R}{t_x}(t_x - t_y) + \frac{t_y t_4 - t_x t_3}{t_x^2 t_1}(\hat{t}_2 - t_2)$$

$$= R + \frac{\hat{t}_y - R\hat{t}_x}{\hat{t}_x} + \frac{1}{t_x} \frac{Rt_4 - t_3}{t_4}(\hat{t}_b - R_0 \hat{t}_a) + \frac{R_0 t_a - t_b}{t_x}.$$

The variance $\hat{R}_{III}(l)$ in the first order of approximation is given by

$$AVar(\hat{R}_{III}(l)) = v(\hat{R}_{III0}(l)) = \frac{1}{t_x^2} Var\left[(\hat{t}_y - R\hat{t}_x) + \frac{Rt_4 - t_3}{t_4}(\hat{t}_b - R_0 \hat{t}_a) \right]$$

$$= \frac{1}{t_x^2} Var\left[(\hat{t}_y - R\hat{t}_x) + L(\hat{t}_b - R_0 \hat{t}_a) \right]$$

$$v(\hat{R}_{III0}(l)) = \frac{1}{t_x^2} \sum_U \sum_U \Delta_{ij} \frac{v_{2i}}{\pi_i} \frac{v_{2j}}{\pi_j}. \qquad (2.2.11)$$

where

$$v_{2i} = y_i - Rx_i + L(b_i - R_0 a_i).$$

An estimator of variance $\hat{R}$ is given by

$$\hat{v}(\hat{R}_{III}(l)) = \frac{1}{\hat{t}_x^2} \sum_s \sum_s \frac{\Delta_{ij}}{\pi_{ij}} \frac{\hat{v}_i}{\pi_i} \frac{\hat{v}_j}{\pi_j},$$

where

$$\hat{v}_{2i} = y_i - \hat{R}x_i + \hat{L}(b_i - R_0 a_i).$$

3. SIMULATION STUDY

To study the performance of the proposed estimator, several multivariate normal populations of size 1000 were generated for different values of coefficient of variation (CV) c_x, c_y, c_a, c_b with different intensity levels of correlation among the study and the auxiliary variables i.e. low, moderate and high correlation. Hence, from each multivariate normal population, repeated samples of sizes 20 and 50 were drawn 2000 times by SRSWOR using the PROC SURVEYSELECT procedure in the SAS package. The results of the simulation are presented in table 1 for all estimators $\hat{\theta} = \hat{R}_{\mathrm{I}}(l),\ \hat{R}_{\mathrm{II}}(l),\ \hat{R}_{\mathrm{III}}(l)$. The average mean square error of 2000 samples has been considered for efficiency comparison between the developed estimator $(\hat{R}_j)$, j=I, II, III and Simple estimator of population ratio $(\hat{R}_0)$. We calculated empirical mean square error and relative efficiency using the following formulas

$$R.E. = \frac{\frac{1}{K}\sum_{i=1}^{K}\hat{M}SE(\hat{R}_0)}{\frac{1}{K}\sum_{i=1}^{K}\hat{M}SE(\hat{R}_j)},\ \forall\, j=I,\ II,\ III;\ k=2000,$$

where for all estimators estimated bias $B\hat{i}as(\hat{R}_i) = \bar{\hat{R}}_i - R_i$ and estimated mean square errors $\hat{M}SE(\hat{R}_i) = A\,\mathrm{var}(\hat{R}_i) + \left(\hat{B}ias(\hat{R}_i)\right)^2 \quad \forall\ i=0,\ I,\ II,\ III.$

Three different cases of correlation between the study and the auxiliary variables were taken.

Case-I. Uncorrelated variables.

A positive definite covariance matrix with the following combination of correlation coefficients was taken.

$\rho(x,y) = 0.08$, $\rho(y,b) = 0.12$, $\rho(x,a) = 0.11$, $\rho(a,b) = 0.1$, $\rho(x,b) = 0.13$, $\rho(y,a) = 0.1$.

Case-II. Uncorrelated study variables but correlated auxiliary variables

A positive definite covariance matrix with the following combination of correlation coefficients was taken.

$\rho(x,y) = 0.09$, $\rho(y,b) = 0.83$, $\rho(x,a) = 0.83$, $\rho(a,b) = 0.04$, $\rho(x,b) = 0.03$, $\rho(y,a) = 0.04$.

Case-III Highly correlated study and auxiliary variables

A positive definite covariance matrix with the following combination of correlation coefficients was taken.

$\rho(x,y) = 0.8$, $\rho(y,b) = 0.83$, $\rho(x,a) = 0.83$, $\rho(a,b) = 0.5$, $\rho(x,b) = 0.5$, $\rho(y,a) = 0.5$.

The relative efficiencies of the proposed estimators are worked out by fixing the CV values of the variables ‘x’, ‘y’, ‘a’ and ‘b’. For fixing the CV, we were taken different combination of mean and variance for generating population. Considering different CV values and correlation between the study and the auxiliary variables, Table 1 is prepared through simulation runs.

For all estimators, the relative efficiency decreases as population dispersion (in terms of CV) increases, and it so happens across all levels of correlation between the study and auxiliary variables. $\hat{R}_{III}$ has the best performance, particularly when population is less diverse (i.e., has low CV), the variables are well correlated and the weights involves ratio from the auxiliary variables. Furthermore, the efficiency increases with correlation increasing, irrespective of the population CV.

Table 1. Relative efficiency of different calibrated estimators

Estimator	Sample Size	R.E. ($c_x=c_y=c_a=c_b=10\%$.) Low C.V			R.E. ($c_x=c_y=c_a=c_b=25\%$) Moderate C.V			R.E. ($c_x=c_y=c_a=c_b=40\%$) High C.V		
		Case I	Case II	Case III	Case I	Case II	Case III	Case I	Case II	Case III
$\hat{R}_I$	20	1.24	2.09	2.82	1.08	1.30	1.67	1.39	1.69	1.61
	50	1.31	3.12	3.32	1.19	2.29	2.57	1.65	1.89	2.25
$\hat{R}_{II}$	20	2.10	3.41	3.62	2.05	2.39	2.15	1.45	2.27	2.15
	50	3.15	3.53	4.26	2.32	3.14	2.99	1.91	2.74	2.68
$\hat{R}_{III}$	20	2.53	2.73	4.35	2.65	2.67	2.63	1.58	2.54	2.52
	50	2.70	3.52	4.53	3.18	3.83	3.47	1.98	2.82	3.40

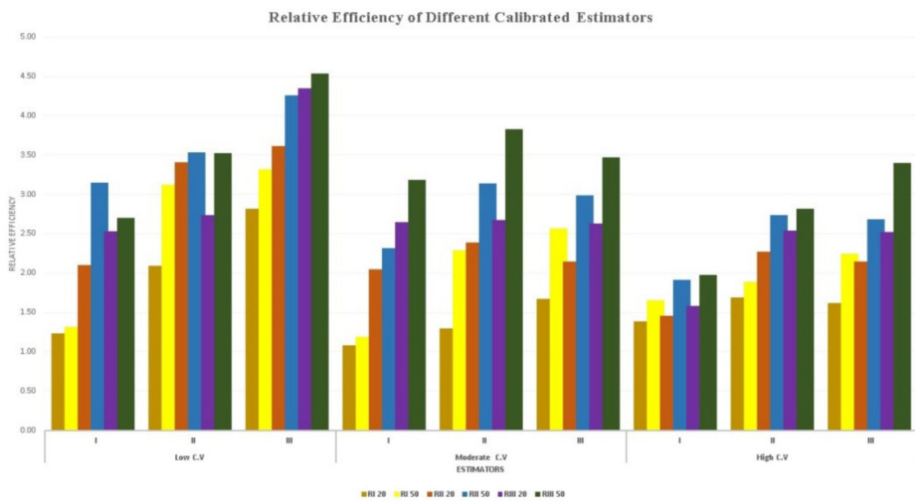

Fig. 1. Relative Efficiency of Different Calibrated Estimators

4. CONCLUSION

In the proposed study we have developed the calibrated estimators utilizing auxiliary information. It has been observed that with the increase in correlation between study and auxiliary variables improves the efficiency of calibrated estimators at constant CV. At constant correlation, the efficiency of calibrated estimators is inversely proportional to the CV. Calibrated estimator developed under the situation in which population ratio of auxiliary variables is used with single weight system under calibration has highest relative efficiency. So we concluded that the population ratio of auxiliary variables with a single weight system is preferred over a different weights systems.

ACKNOWLEDGEMENTS

The authors thank the reviewers for their constructive comments and suggestions to improve the earlier version of the manuscript.

REFERENCES

Bartkus, I. and Pumputis, D. (2010). Estimation of finite population ratio using the calibration of strata weight. *Workshop on Survey Sampling Theory and Methodology*, Lithuania. 23-27.

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.

Chang, H.J. and Huang, K.C. (2013). On estimation of ratio of population mean in survey sampling when some observations are missing. *Journal of information and Optimizations Science,* **3**, 429-436.

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

Estevao, V.M. and Särndal, C.E. (2002). The ten cases of auxiliary information for calibration in two-phase sampling. *Journal of Official Statistics*, **18**, 233-255.

Islam, S, Chandra , H, Sud, U.C., Guha, S. and Basak, P. (2019). Calibration Approach for Estimation of Population Ratio under Double Sampling. *Journal of Indian Society of Agricultural Statistics*, **73(1)**, 19-25.

Jaiswal, A.K., Usman, M. and Singh, G.N. (2023). New calibration estimation procedure in the presence of unit nonresponse, *Ain Shams Engineering Journal*, **14**, 101910.

Kim J.K. and Park M. (2010). Calibration estimation in survey sampling. *International Statistical Review*, **78(1)**, 21-39.

Kish, L., Namboodiri, N.K. and Pillai, R.K. (1962). The Ratio Bias in Surveys. *Journal of the American Statistical Association,* **57(300)**, 863-876.

Krapavickaite, D. and Plikusas, A. (2005). Estimation of a Ratio in the Finite Population. *Informatics,* **16**, 347-364.

Ozgul, N. (2021). New improved calibration estimator based on two auxiliary variables in stratified two-phase sampling. *Journal of Statistical Computation and Simulation,* **91(6)**, 1243-1256.

Plikusas, A. (2001). Calibrated estimators of the ratio. *Lithuanian Mathematical Journal*, (special issue) **41**, 457-462.

Särndal, C.E., Swensson, B and Wretman J.H. (1992). *Model assisted survey sampling*,Springer-Verlag, New York.

Smith, D.R., Conroy, M.J., and Brakhage, D.H. (1995). Efficiency of Adaptive Cluster Sampling for Estimating Density of Wintering Waterfowl. *Biometrics*, **51(2),** 777-788.

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.