**International Journal of Science Academic Research**
Vol. 04, Issue 12, pp.6700-6707, December, 2023
Available online at http://www.scienceijsar.com

ISSN: 2582-6425

**Research Article**

# CALIBRATION APPROACH ESTIMATION OF THE POPULATION MEAN IN STRATIFIED SAMPLING

**Anupama Bhadauriya and *Namita Srivastava**

Department of Statistics, St. John's College Agra, UP India

**Received** 19th October 2023; **Accepted** 10th November 2023; **Published online** 29th December 2023

**Abstract**

Calibration approach is most important in survey sampling. Calibration estimation improves the precision of the estimates of population parameters by using auxiliary information. We proposed two new calibration estimators for estimating the population mean in stratified random sampling with a set of new calibration constraints. The proposed estimator has been compared with Tracy *et al.* (2003) estimator with the help of simulation study. The resultant new estimators are more efficient than Tracy *et al.* (2003) estimator in stratified random sampling.

**Keywords:** Auxiliary Information, Calibration Estimation, Stratified Sampling.

## 1. INTRODUCTION

Calibration is commonly used in survey sampling to include auxiliary information to increase the precision of the estimators of population parameter, Deville and Sarndal (1992) first proposed calibration estimators in survey sample and calibration estimator has been studied by many survey statisticians. A few key references are Kim *et al.* (2007), Singh and Arnab (2011), Koyuncu and Kadilar (2013), Mouhamed *et al.* (2015), Koyuncu and Kadilar (2016), Clement and Enang (2017), Nidhi *et al.* (2017), Neha Garg & Meenakshi Pachori (2020 etc., have contributed in the direction of developing some calibrated estimators for different population parameters using different calibration constraints under various sampling schemes. Thus the term calibration estimation as introduced by Deville and Sarndal (1992) is a procedure of minimizing a distance function subject to calibration constraints. In this paper, two new calibration estimators for population mean under stratified random sampling is suggested using new calibration constraints, which includes mean and coefficient of variation, coefficient of variation and correlation coefficient. The use of these constraints makes the estimator more efficient. In Section 3, the proposed estimators has been determined using calibration approach. Section 4 .Simulation study has been carried out (using R software) in Section 5 to check the performance of the proposed estimators with the other existing estimators.

### 2. Procedure, notations, and definitions

Consider a finite population $U$ of size $N$ with unit labels $1,2 \ldots\ldots N$. let $y_{i,}\ i = 1,2 \ldots.N$ be the study variable and $x_i$, $i = 1,2 \ldots.N$ be the auxiliary variable linearly related to $y$. The population total of the auxiliary variable $X = \sum_{i\in U} x_i$ is assumed to be known. The objective is to estimate the population total for the study variable $Y = \sum_{i\in U} y_i$.

A sample S is drawn such that $s = \{1,2, \ldots.n\} \subset U$ using a probability sampling design $P$.

The first and second order inclusion probability *i.e.,* $\pi_i = P_r(i\ \epsilon\ s)$ and $\pi_{ij}$: the second order inclusion probability$i.e.,\pi_{ij} = P_r(i, j\ \epsilon\ s)$.

To estimate the population total $t_y = \sum_{i=1}^{n} y_i$ Deville (1988) used calibration technique on known population totals to modify the basic sampling design weights, $d_i = 1/\pi_i$ defined as the inverse of the inclusion probability for unit $\boldsymbol{i}$. The weights $d_i = 1/\pi_i$ appear in Horvitz-Thompson estimator.

Then,

$$\hat{Y}_{HT} = \sum_{i\epsilon s} y_i / \pi_i$$
$$\hat{Y}_{HT} = \sum_{i\epsilon s} d_i y_i \qquad (2.1)$$

The calibration estimator of $t_y$ is defined as

$$\hat{Y}_c = \sum_{i\epsilon s} w_i y_i$$

***Corresponding Author:** ***Namita Srivastava***
Department of Statistics, St. John's College Agra, UP India

Where the weight $w_i$ is chosen as possible is an average sense for given metric to the $d_i$ based on a distance function given that the calibration equation is

$$\sum_{i\epsilon s} w_i x_i = X.$$

The distance measure most commonly chosen is the Chi square distance function given as $D = \sum_{i\epsilon s} \frac{(w_i - d_i)^2}{d_i q_i}$, where $q_i$, $i\epsilon s$ are suitable chosen positive scale factor which decide the form of estimator. Using the method of Lagrange's multiplier, the calibration weight is obtained as

$$w_i = d_i + (\lambda d_i x_i q_i / \sum_{i\epsilon s} d_i q_i\, x_i^2)(X - \sum_{i\epsilon s} d_i x_i) \qquad (2.2)$$

The resulting estimator of $Y$ using calibration weights was obtained as the generalized regression estimator of $Y$ by Deville and Sarndal (1992) as

$$\hat{Y}_{GREG} = \sum_{i\epsilon s} d_i y_i + \hat{\beta}_{DS}\left(X - \hat{X}_{HT}\right)$$
$$\hat{Y}_{HT} + \hat{\beta}_{DS}\left(X - \hat{X}_{HT}\right) \qquad (2.3)$$

Where $\hat{\beta}_{DS} = \sum_{i\epsilon s} d_i x_i\, q_i / \sum_{i\epsilon s} d_i q_i\, x_i^2$ is a weighted estimator of the multiple regression coefficients.
Variance of $\hat{Y}_{GREG}$ for a large sample size provided by Deville and Sarndal (1992) is

$$V\left(\hat{Y}_{GREG}\right) = \frac{1}{2}\sum\sum_{i\neq j\in U} D_{ij}\, \pi_{ij}\left(d_i E_i - d_j E_j\right)^2$$
Where $D_{ij} = \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}}, E_i = y_i - Bx_i \text{ and } B = \frac{\sum_{i\epsilon U} y_i x_i q_i}{\sum_{i\epsilon U} q_i x_i^2}$

A consistent and approximate unbiased estimator of variance is given as

$$V\left(\hat{Y}_{GREG}\right) = \frac{1}{2}\sum\sum_{i\neq j\in U} D_{ij}\left(w_i e_i - w_j e_j\right)^2 \qquad (2.4)$$
where $e_i = y_i - \hat{\beta}_{DS} x_i$

The calibration estimator under the stratified random sampling for population mean $\bar{Y}$defined by Tracy *et al.* (2003) is given as

$$\bar{y}_{st}(t) = \sum_{h=1}^{L} \Omega_h \bar{y}_h \qquad (2.5)$$

Where the weight $\Omega_h$ are chosen such that the chi square distance function

$$\sum_{h=1}^{L} \frac{(\Omega_h - W_h)^2}{W_h Q_h} \qquad (2.6)$$

Where $Q_h$ denotes suitable weight to different forms of estimator

Subject to the following calibration constraints

$$\sum_{h=1}^{L} \Omega_h \bar{x}_h = \sum_{h=1}^{L} W_h \bar{X}_h \qquad (2.7)$$

$$\sum_{h=1}^{L} \Omega_h s_{hx}^2 = \sum_{h=1}^{L} W_h S_{hx}^2 \qquad (2.8)$$

Where $\bar{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}$ and $\bar{X}_h = \frac{1}{N_h}\sum_{i=1}^{n_h} x_{hi}$ are the $h^{th}$stratum sample and population means of the auxiliary variable, respectively. $S^2{}_{xh} = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(x_{hi} - \bar{X}_h)^2$ denotes the $h^{th}$ stratum population variance, $s^2{}_{xh} = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(x_{hi} - \bar{x}_h)^2$ denotes the $h^{th}$ stratum sample variance,

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

Thus, the estimator given by Tracy *et al.* (2003) is

$$\bar{y}_{st}(t) = \sum_{h=1}^{L} W_h \left(\bar{y}_h + \beta_{1(T)}(\bar{X}_h - \bar{x}_h) + \beta_{2(T)}(S_{hx}^2 - s_{hx}^2)\right)$$

Where

$\beta_{1(T)} =$ $W_h Q_h \bar{x}_h \bar{y}_h [\sum_{h=1}^{L} W_h Q_h s_{hx}{}^4 \sum_{h=1}^{L} W_h(\bar{X}_h - \bar{x}_h) - \sum_{h=1}^{L} W_h Q_h \bar{x}_h s_{hx}^2 \sum_{h=1}^{L} W_h(S_{hx}^2 - s_{hx}^2) / \sum_{h=1}^{L} W_h Q_h s_{hx}{}^4 \sum_{h=1}^{L} W_h Q_h \bar{x}_h{}^2 - (\sum_{h=1}^{L} W_h Q_h s_{hx}^2)^2]$

$\beta_{2(T)}=$ $W_h Q_h s_{hx}^2 \bar{y}_h [\sum_{h=1}^{L} W_h Q_h \bar{x}_h{}^2 \sum_{h=1}^{L} W_h(S_{hx}^2 - s_{hx}^2) - \sum_{h=1}^{L} W_h Q_h s_{hx}^2 \bar{x}_h \sum_{h=1}^{L} W_h(\bar{X}_h - \bar{x}_h) / \sum_{h=1}^{L} W_h Q_h s_{hx}{}^4 \sum_{h=1}^{L} W_h Q_h \bar{x}_h{}^2 - (\sum_{h=1}^{L} W_h Q_h s_{hx}^2)^2 ]$

## 3. Proposed Calibration estimators in stratified sampling

Suppose we have population $N$ units that is first sub divided in to $L$ homogeneous sub group called strata. Such that the $h^{th}$ stratum Consist of $N_h$ units $h = 1,2 \dots L$ and $\sum_{h=1}^{L} N_{h=} N$. Suppose further that a$n_h$ be the size of the sample for $h^{th}$ stratum such that $\sum_{h=1}^{L} n_h = n$. Let $y$ and $x$ be the study and auxiliary variables taking values $y_{hi}$ and $x_{hi}$, respectively for $i^{th}$unit $(i = 1,2, \dots, N)$ and $W_h = \frac{N_h}{N}$ is the known proportion population units falling in the $h^{th}$ stratum.The unbiased estimator of population mean $\bar{Y}$ is given by

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_{h,} \qquad (3.1)$$

Where $\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$ denotes the $h^{th}$ stratum sample mean. Under SRSWOR sampling, the variance of the estimator $\bar{y}_{st}$ is give by

$$V(\bar{y}_{st}) = \frac{1}{\bar{Y}^2} \sum_{h=1}^{l} W_h{}^2 \left(\frac{1-f_h}{n_h}\right) S^2{}_{hy} \qquad (3.2)$$

Where $S^2{}_{yh} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2$ denotes the $h^{th}$ stratum population variance, $\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{n_h} Y_{hi}$ denotes the $h^{th}$ stratum population mean and $f_h = n_h / N_h$,

**Case1:** Let $X_{hi}$, $i = 1,2 \dots, N_h$; $h = 1,2 \dots L$. denote the value of the $i^{th}$ unit of the auxiliary variable in the $h^{th}$ stratum about which information may be known at the stratum level consider a new alternative (calibration) estimator for stratified sampling of the form

$$\bar{y}_{st}(n) = \sum_{h=1}^{L} \Omega_h \bar{y}_h \qquad (3.3)$$

Where the weight $\Omega_h$ are chosen such that the chi square distance function

$$\sum_{h=1}^{L} \frac{(\Omega_h - W_h)^2}{W_h Q_h} \qquad (3.4)$$

Where $Q_h$ denotes suitable weight to different forms of estimator

Subject to the following calibration constraints

$$\sum_{h=1}^{L} \Omega_h \bar{x}_h = \sum_{h=1}^{L} W_h \bar{X}_h \qquad (3.5)$$
$$\sum_{h=1}^{L} \Omega_h c_{hx}^2 = \sum_{h=1}^{L} W_h C_{hx}^2 \qquad (3.6)$$

Where $\bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}$ and $\bar{X}_h = \frac{1}{N_h} \sum_{i=1}^{n_h} x_{hi}$ are the $h^{th}$stratum sample and population means of the auxiliary variable, respectively. $S_{xh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)^2$ denotes the $h^{th}$ stratum population variance, $s_{xh}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2$ denotes the $h^{th}$ stratum sample variance, $c_{xh}^2 = \frac{s^2{}_{xh}}{\bar{x}^2}$, $C^2{}_{xh} = \frac{S^2{}_{xh}}{\bar{X}^2}$ be the coefficient of variation of $h^{th}$ stratum.

Thus the calibration estimation problem reduce to an optimization problem where

$$\phi = \sum_{h=1}^{L} \frac{(\Omega_h - W_h)^2}{W_h Q_h} - 2\lambda_1 (\sum_{h=1}^{L} \Omega_h \bar{x}_h - \sum_{h=1}^{L} W_h \bar{X}_h) - 2\lambda_2 (\sum_{h=1}^{L} \Omega_h c_{hx}^2 - \sum_{h=1}^{L} W_h C_{hx}^2)$$

Where $\lambda_1$ and $\lambda_2$ are the Lagrange's multipliers Differentiate above equation w.r.to $\Omega_h$ and equating to zero for obtaining the value of $\Omega_h$

$$\Omega_h = W_h + W_h Q_h (\lambda_1 \bar{x}_h + \lambda_2 c_{hx}^2) \qquad (3.7)$$

From equation (3.5)

$$\sum_{h=1}^{L} \{W_h + W_h Q_h (\lambda_1 \bar{x}_h + \lambda_2 c_{hx}^2)\} \bar{x}_h = \sum_{h=1}^{L} W_h \bar{X}_h$$
$$\sum_{h=1}^{L} \lambda_1 W_h Q_h \bar{x}_h{}^2 + \sum_{h=1}^{L} \lambda_2 W_h Q_h c_{hx}^2 \bar{x}_h = \sum_{h=1}^{L} W_h \bar{X}_h - \sum_{h=1}^{L} W_h \bar{x}_h \qquad (3.8)$$

From equation (3.6)

$$\sum_{h=1}^{L}\{W_h + W_hQ_h(\lambda_1\bar{x}_h + \lambda_2 c_{hx}^2)\}c_{hx}^2 = \sum_{h=1}^{L} W_h C_{hx}^2$$
$$\sum_{h=1}^{L}\lambda_1 W_hQ_hc_{hx}^2\bar{x}_h + \sum_{h=1}^{L}\lambda_2 W_hQ_h c_{hx}^4 = \sum_{h=1}^{L} W_hC_{hx}^2 - \sum_{h=1}^{L} W_hc_{hx}^2 \quad (3.9)$$

On solving equations (3.8) and (3.9) we get

$$\lambda_1 = \{\sum_{h=1}^{L} W_hQ_h\, c_{hx}{}^4 \sum_{h=1}^{L} W_h(\bar{X}_h - \bar{x}_h) - \sum_{h=1}^{L} W_hQ_hc_{hx}^2 \sum_{h=1}^{L} W_h(C_{hx}^2 - c_{hx}^2)\}/\sum_{h=1}^{L} W_hQ_h\, c_{hx}{}^4 \sum_{h=1}^{L} W_hQ_h\,\bar{x}_h{}^2 - (\sum_{h=1}^{L} W_hQ_hc_{hx}^2)^2$$

$$\lambda_2 = \sum_{h=1}^{L} W_hQ_hc_{hx}^2\bar{x}_h \sum_{h=1}^{L} W_h(\bar{X}_h - \bar{x}_h) + \sum_{h=1}^{L} W_hQ_h\,\bar{x}_h{}^2 \sum_{h=1}^{L} W_h(C_{hx}^2 - c_{hx}^2)/\sum_{h=1}^{L} W_hQ_h\, c_{hx}{}^4 \sum_{h=1}^{L} W_hQ_h\,\bar{x}_h{}^2 - (\sum_{h=1}^{L} W_hQ_hc_{hx}^2)^2$$

On substuting $\lambda_1, \lambda_2$ in equation (3.7) we get

$$\Omega_h = W_h + W_hQ_h\bar{x}_h\left[\sum_{h=1}^{L} W_hQ_h\,c_{hx}{}^4\sum_{h=1}^{L} W_h(\bar{X}_h - \bar{x}_h) - \sum_{h=1}^{L} W_hQ_hc_{hx}^2\sum_{h=1}^{L} W_h(C_{hx}^2 - c_{hx}^2)\Big/\sum_{h=1}^{L} W_hQ_h\,c_{hx}{}^4\sum_{h=1}^{L} W_hQ_h\,\bar{x}_h{}^2 - \left(\sum_{h=1}^{L} W_hQ_hc_{hx}^2\right)^2\right]$$
$$+ W_hQ_hc_{hx}^2\left[\sum_{h=1}^{L} W_hQ_h\,\bar{x}_h{}^2\sum_{h=1}^{L} W_h(C_{hx}^2 - c_{hx}^2) - \sum_{h=1}^{L} W_hQ_hc_{hx}^2\bar{x}_h\sum_{h=1}^{L} W_h(\bar{X}_h - \bar{x}_h)\Big/\sum_{h=1}^{L} W_hQ_h\,c_{hx}{}^4\sum_{h=1}^{L} W_hQ_h\,\bar{x}_h{}^2 - \left(\sum_{h=1}^{L} W_hQ_hc_{hx}^2\right)^2\right]$$

On substuting the $\Omega_h$ in (3.3) we get

$$\bar{y}_{st}(n) = \sum_{h=1}^{L} W_h\left(\bar{y}_h + \beta_1(\bar{X}_h - \bar{x}_h) + \beta_2(C_{hx}^2 - c_{hx}^2)\right)$$

Where

$$\beta_1 = W_hQ_h\bar{x}_h\bar{y}_h\left[\sum_{h=1}^{L} W_hQ_h\,c_{hx}{}^4\sum_{h=1}^{L} W_h(\bar{X}_h - \bar{x}_h) - \sum_{h=1}^{L} W_hQ_h\bar{x}_hc_{hx}^2\sum_{h=1}^{L} W_h(C_{hx}^2 - c_{hx}^2)\Big/\sum_{h=1}^{L} W_hQ_h\,c_{hx}{}^4\sum_{h=1}^{L} W_hQ_h\,\bar{x}_h{}^2 - \left(\sum_{h=1}^{L} W_hQ_hc_{hx}^2\right)^2\right]$$

$$\beta_2 = W_hQ_hc_{hx}^2\bar{y}_h[\sum_{h=1}^{L} W_hQ_h\,\bar{x}_h{}^2 \sum_{h=1}^{L} W_h(C_{hx}^2 - c_{hx}^2) - \sum_{h=1}^{L} W_hQ_hc_{hx}^2\bar{x}_h \sum_{h=1}^{L} W_h(\bar{X}_h - \bar{x}_h)/\sum_{h=1}^{L} W_hQ_h\, c_{hx}{}^4 \sum_{h=1}^{L} W_hQ_h\,\bar{x}_h{}^2 - (\sum_{h=1}^{L} W_hQ_hc_{hx}^2)^2\,]$$

**Case2:** The calibration estimator under the stratified random sampling for population mean defined by Tracy *et al.* (2003) is given as

$$\bar{y}_{st}(cv) = \sum_{h=1}^{L} \Omega_h\bar{y}_h \quad (3.10)$$

Where $\Omega_h$ is the weight minimize the distance measure In this study we consider following distance function
Where the weight $\Omega_h$ are chosen such that the chi square distance function

$$\sum_{h=1}^{L}\frac{(\Omega_h - W_h)^2}{W_hQ_h} \quad (3.11)$$

Where $Q_h$ denotes suitable weight to different forms of estimator

Subject to the following calibration constraints

$$\sum_{h=1}^{L}\Omega_hc_{hx} = \sum_{h=1}^{L} W_h\, C_{hx} \quad (3.12)$$
$$\sum_{h=1}^{L}\Omega_h\rho_{hx}^2 = \sum_{h=1}^{L} W_h\,\boldsymbol{\rho_{hx}^2} \quad (3.13)$$

Where $s^2{}_{yh} = \frac{1}{n_h-1}\sum_{i=1}^{n_h}(y_{hi} - \bar{y}_h)^2$ and $s^2{}_{xh} = \frac{1}{n_h-1}\sum_{i=1}^{n_h}(x_{hi} - \bar{x}_h)^2$ be the sample variance of $y$ and $x$ respectively in $h^{th}$ stratum corresponding to the population variances$S^2{}_{yh} = \frac{1}{N_h-1}\sum_{i=1}^{N_h}(y_{hi} - \bar{Y}_h)^2$and $S^2{}_{xh} = \frac{1}{N_h-1}\sum_{i=1}^{N_h}(x_{hi} - \bar{X}_h)^2$, $S_{yxh} = \frac{1}{N_h-1}\sum_{i=1}^{N_h}(x_{hi} - \bar{X}_h)\,(y_{hi} - \bar{Y}_h)$ Covariance between $y\ and\ x$ then $\boldsymbol{\rho_{hx}} = {S_{yxh}}\big/{S_{yh} * S_{xh}}$ , $\rho_{hx} = {s_{yxh}}/{s_{yh} * s_{xh}}$ is the correlation coefficient of $h^{th}$ stratum and $c_{xh} = \frac{s_{xh}}{\bar{X}}$ , $C_{xh} = \frac{S_{xh}}{\bar{X}}$ be the coefficient of variation of $h^{th}$ stratum.

Thus the calibration estimation problem reduce to an optimization problem where

$$\phi = \sum_{h=1}^{L} \frac{(\Omega_h - W_h)^2}{W_h Q_h} - 2\lambda_1 \left(\sum_{h=1}^{L} \Omega_h c_{hx} - \sum_{h=1}^{L} W_h C_{hx}\right) - 2\lambda_2 \left(\sum_{h=1}^{L} \Omega_h \rho_{hx}^2 - \sum_{h=1}^{L} W_h \boldsymbol{\rho}_{\boldsymbol{hx}}^{\boldsymbol{2}}\right)$$

where $\lambda_1$ and $\lambda_2$ are the Lagrange's multipliers Differentiate above equation w.r.to $\Omega_h$ and equating to zero for obtaining the value of $\Omega_h$

$$\Omega_h = W_h + W_h Q_h (\lambda_1 c_{hx} + \lambda_2 \rho_{hx}^2) \qquad (3.14)$$

From equation (3.12)

$$\sum_{h=1}^{L} \{W_h + W_h Q_h (\lambda_1 c_{hx} + \lambda_2 \rho_{hx}^2)\} c_{hx} = \sum_{h=1}^{L} W_h C_{hx}$$
$$\sum_{h=1}^{L} \lambda_1 W_h Q_h {c_{hx}}^2 + \sum_{h=1}^{L} \lambda_2 W_h Q_h \rho_{hx}^2 \bar{x}_h = \sum_{h=1}^{L} W_h C_{hx} - \sum_{h=1}^{L} W_h c_{hx} \qquad (3.15)$$

From equation (3.13)

$$\sum_{h=1}^{L} \{W_h + W_h Q_h (\lambda_1 c_{hx} + \lambda_2 \rho_{hx}^2)\} \rho_{hx}^2 = \sum_{h=1}^{L} W_h \boldsymbol{\rho}_{\boldsymbol{hx}}^{\boldsymbol{2}}$$
$$\sum_{h=1}^{L} \lambda_1 W_h Q_h \rho_{hx}^2 c_{hx} + \sum_{h=1}^{L} \lambda_2 W_h Q_h {\rho_{hx}}^4 = \sum_{h=1}^{L} W_h \boldsymbol{\rho}_{\boldsymbol{hx}}^{\boldsymbol{2}} - \sum_{h=1}^{L} W_h \rho_{hx}^2 \qquad (3.16)$$

On solving equations (3.15) and (3.16) we get

$$\lambda_1 = \left\{\sum_{h=1}^{L} W_h Q_h {\rho_{hx}}^4 \sum_{h=1}^{L} W_h (C_{hx} - c_{hx}) - \sum_{h=1}^{L} W_h Q_h \rho_{hx}^2 \sum_{h=1}^{L} W_h (\boldsymbol{\rho}_{\boldsymbol{hx}}^{\boldsymbol{2}} - \rho_{hx}^2)\right\} / \sum_{h=1}^{L} W_h Q_h {\rho_{hx}}^4 \sum_{h=1}^{L} W_h Q_h {c_{hx}}^2 - \left(\sum_{h=1}^{L} W_h Q_h \rho_{hx}^2\right)^2$$

$$\lambda_2 = \sum_{h=1}^{L} W_h Q_h \rho_{hx}^2 c_{hx} \sum_{h=1}^{L} W_h (C_{hx} - c_{hx}) + \sum_{h=1}^{L} W_h Q_h {c_{hx}}^2 \sum_{h=1}^{L} W_h (\boldsymbol{\rho}_{\boldsymbol{hx}}^{\boldsymbol{2}} - \rho_{hx}^2) / \left\{\sum_{h=1}^{L} W_h Q_h {\rho_{hx}}^4 \sum_{h=1}^{L} W_h (C_{hx} - c_{hx}) - \sum_{h}^{L} \right. \left(\sum_{h=1}^{L} W_h Q_h \rho_{hx}^2\right)^2$$

On substuting $\lambda_1, \lambda_2$ in equation (3.14) we get

$$\Omega_h = W_h + W_h Q_h c_{hx} \left[\sum_{h=1}^{L} W_h Q_h {\rho_{hx}}^4 \sum_{h=1}^{L} W_h (C_{hx} - c_{hx}) - \sum_{h=1}^{L} W_h Q_h \rho_{hx}^2 \sum_{h=1}^{L} W_h (\boldsymbol{\rho}_{\boldsymbol{hx}}^{\boldsymbol{2}} - \rho_{hx}^2) \Big/ \sum_{h=1}^{L} W_h Q_h {\rho_{hx}}^4 \sum_{h=1}^{L} W_h Q_h c_{hx}^2 - \left(\sum_{h=1}^{L} W_h Q_h \rho_{hx}^2\right)^2\right]$$
$$+ W_h Q_h c_{hx}^2 \left[\sum_{h=1}^{L} W_h Q_h c_{hx}^2 \sum_{h=1}^{L} W_h (\boldsymbol{\rho}_{\boldsymbol{hx}}^{\boldsymbol{2}} - \rho_{hx}^2) - \sum_{h=1}^{L} W_h Q_h \rho_{hx}^2 c_{hx} \sum_{h=1}^{L} W_h (C_{hx} - c_{hx}) \Big/ \sum_{h=1}^{L} W_h Q_h {\rho_{hx}}^4 \sum_{h=1}^{L} W_h Q_h c_{hx}^2 - \left(\sum_{h=1}^{L} W_h Q_h \rho_{hx}^2\right)^2\right]$$

On substuting the $\Omega_h$ in (3.10) we get

$$\bar{y}_{st}(cv) = \sum_{h=1}^{L} W_h \left(\bar{y}_h + \beta_1 (C_{hx} - c_{hx}) + \beta_2 \left(\boldsymbol{\rho}_{\boldsymbol{hx}}^{\boldsymbol{2}} - \rho_{hx}^2\right)\right)$$

Where

$$\beta_1 = W_h Q_h c_{hx} \bar{y}_h \left[\sum_{h=1}^{L} W_h Q_h \rho_{hx}^4 \sum_{h=1}^{L} W_h ((C_{hx} - c_{hx})) - \sum_{h=1}^{L} W_h Q_h c_{hx} \rho_{hx}^2 \sum_{h=1}^{L} W_h (\boldsymbol{\rho}_{\boldsymbol{hx}}^{\boldsymbol{2}} - \rho_{hx}^2) \Big/ \sum_{h=1}^{L} W_h Q_h {\rho_{hx}}^4 \sum_{h=1}^{L} W_h Q_h {c_{hx}}^2 - \left( \right.\right.$$

$$\beta_2 = W_h Q_h \rho_{hx}^2 \bar{y}_h \left[\sum_{h=1}^{L} W_h Q_h c_{hx}^2 \sum_{h=1}^{L} W_h (\boldsymbol{\rho}_{\boldsymbol{hx}}^{\boldsymbol{2}} - \rho_{hx}^2) - \sum_{h=1}^{L} W_h Q_h c_{hx}^2 \bar{x}_h \sum_{h=1}^{L} W_h (C_{hx} - c_{hx}) / \sum_{h=1}^{L} W_h Q_h \rho_{hx}^4 \sum_{h=1}^{L} W_h Q_h c_{hx}^2 - \left(\sum_{h=1}^{L} W_h Q_h \rho_{hx}^2\right)^2\right]$$

Since the ratio $\Omega_h / W_h \rightarrow 1$in probability, as the sample size in each stratum tends to infinity, the proposed estimator f the population mean is consistent.

**4. Simulation Study**

To study the performance of the proposed estimator we use the following data set

**Source of Data**

Population, is taken from the census of India 2011(Uttar Pradesh , Series 10, Part 12B and District census hand book, AGRA).

**Population: 1** The considered data relates to total area of 45 villages of Khandauli block at Agra districts (U.P). We consider the numbers of agricultural laborers in villages as study variable $y$ and the total area of villages as auxiliary variable $x$

We divided the whole population of 45 villages is divided in to 5 strata according to the area. Accordingly we have:

| Strata | Area in Hectare |
|---|---|
| 1 | (1-4400) (21 Villages) |
| 2 | (4400-8400) (10 Villages) |
| 3 | (8400-12400) (6 Villages) |
| 4 | (12400-16500) (5Villages) |
| 5 | (16500-20900) (3Villages) |

**Table 1. Parametric values of the population (1)**

| Population | Stratum 1 | Stratum 2 | Stratum 3 | Stratum 4 | Stratum 5 |
|---|---|---|---|---|---|
| $N = 45$ | $N_1$ =21 | $N_2$ =10 | $N_3$ =6 | $N_4$ =5 | $N_5$ =3 |
| $n = 23$ | $n_1$ =10 | $n_2$ =5 | $n_2$ =3 | $n_2$ =3 | $n_2$=2 |
| $\bar{Y} = 173.508$ | $\bar{Y}_1$=112.09 | $\bar{Y}_2$=175.9 | $\bar{Y}_3$=149.83 | $\bar{Y}_4$=232.2 | $\bar{Y}_5$=545 |
| $\bar{X} = 463.37$ | $\bar{X}_1$=196.4 | $\bar{X}_2$=413.411 | $\bar{X}_3$=672.33 | $\bar{X}_4$=844.26 | $\bar{X}_5$=1445.97 |

**Population:2** For this we considered the 2011census data which is relates to the total number of agricultural laborers, total area total population, and total numbers of cultivators of 55 villages of Etmadpur block of Agra districts (U.P). We take the numbers of Agricultural laborers in villages as y, the total area of villages as $x_1$, the total population of villages as $x_2$, and the total numbers of Cultivators as $x_3$.

The whole population of 55 villages stratified in to 5 strata according to the Area. So the strata become as under

| Strata | Area in Hectare |
|---|---|
| 1 | (1-3647) (31 Villages) |
| 2 | (3647-7222) (11 Villages) |
| 3 | (7222-9973) (5 Villages) |
| 4 | (9973-12307) (3Villages) |
| 5 | (12307-18173) (5Villages) |

**Table 2. Parametric values of the population (2)**

| Population | Stratum 1 | Stratum 2 | Stratum 3 | Stratum 4 | Stratum 5 |
|---|---|---|---|---|---|
| $N = 55$ | $N_1$ =31 | $N_2$ =11 | $N_3$ =5 | $N_4$ =3 | $N_5$ =5 |
| $n = 31$ | $n_1$ =15 | $n_2$ =5 | $n_2$ =3 | $n_2$ =2 | $n_2$=3 |
| $\bar{Y} = 147.4$ | $\bar{Y}_1$=72.38 | $\bar{Y}_2$=158.81 | $\bar{Y}_3$=194.4 | $\bar{Y}_4$=250.6 | $\bar{Y}_5$=478.2 |
| $\bar{X} = 330.43$ | $\bar{X}_1$=117.67 | $\bar{X}_2$=324.96 | $\bar{X}_3$=550.23 | $\bar{X}_4$=777.97 | $\bar{X}_5$=1173.2 |

We calculated empirical mean square error and relative efficiency using following formulas

The MSE of Tracy et.al (2003) estimator

$$MSE\bar{y}_{st}(t) = \frac{1}{50}\sum_{j=1}^{50}\left[\sum_{h=1}^{3} W_h\left(\bar{y}_h + \beta_{1(T)}(\bar{X}_h - \bar{x}_h) + \beta_{2(T)}(S_{hx}^2 - s_{hx}^2)\right) - \bar{Y}\right]^2$$

Similarly the empirical mean squared error of the proposed estimators is given by

$$MSE\bar{y}_{st}(n) = \frac{1}{50}\sum_{j=1}^{50}\left[\sum_{h=1}^{3} W_h\left(\bar{y}_h + \beta_1(\bar{X}_h - \bar{x}_h) + \beta_2(C_{hx}^2 - c_{hx}^2)\right) - \bar{Y}\right]^2$$

$$MSE\bar{y}_{st}(cv) = \frac{1}{50}\sum_{j=1}^{50}\left[\sum_{h=1}^{3} W_h\left(\bar{y}_h + \beta_1(C_{hx} - c_{hx}) + \beta_2(\boldsymbol{\rho_{hx}^2} - \rho_{hx}^2)\right) - \bar{Y}\right]^2$$

The percent R.E of the proposed estimators with respect Tracy (2003) estimator to is given by

$$RE(n) = \frac{MSE\bar{y}_{st}(t)}{MSE\bar{y}_{st}(n)} \times 100$$

$$RE(cv) = \frac{MSE\bar{y}_{st}(t)}{MSE\bar{y}_{st}(cv)} \times 100$$

**The MSE and RE for population 1,2**

| **Population** | **Estimator** | **MSE** | **RE** |
|---|---|---|---|
| 1 | Tracy($t$) | 51.192 | 100 |
| | Proposed($n$) | 36.734 | 141.34 |
| | Proposed($cv$) | 39.59 | 131.14 |
| 2 | Tracy($t$) | 53.34 | 100 |
| | Proposed($n$) | 34.308 | 155.47 |
| | Proposed($cv$) | 45.57 | 117.05 |

**Conclusion**

The new calibration estimators proposed in this paper to estimate the population mean in case of stratified random sampling, utilizing mean and coefficient of variation, coefficient of variation and correlation coefficient of auxiliary variable in calibration constraint are found to be more efficient than Tracy et al. (2003) estimators. A simulation study has been carried using R Software to compare the efficiency of the proposed estimators with the existing estimator through which it is observed that the proposed estimators are having less MSE and so more efficient for stratified random sampling.

## REFERENCES


Clement, E. P., and E. I. Enang, on the efficiency of ratio estimator over the regression estimator, communications in Statistics: Theory and Methods, 46 (11), 2017, 5357–67. doi:10.1080/ 03610926.2015.1100741.

Deville, J. C., and C. E. Sarndal, Calibration estimators in survey sampling, Journal of the American Statistical Association, 87 (418), 1992, 376–82.

Kim, J.-M., E. A. Sungur, and T.Y. Heo, calibration approach estimators in stratified sampling, statistics and Probability Letters, 77 (1), 2007, 99–103.

Koyuncu, N., and C. Kadilar, Calibration estimator using different distance measures in stratified random sampling, International Journal of Modern Engineering Research, 3 (1), 2013,415–9.

Koyuncu, N., and C. Kadilar, Calibration weighting in stratified random sampling. Communications in Statistics: Simulation and Computation, 45 (7), 2016, 2267–75.

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

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

Sarndal, C. E., B. Swensson, and J. Wretman, Model assisted survey sampling, New York, 1992, Springer Verlag.

Singh, S., and R. Arnab, On Calibration of design weights, METRON: International Journal of Statistics, LXIX 69 (2), 2011,185–205.

Tracy, D. S., S. Singh, and R. Arnab.(2003), Note on calibration in stratified and double sampling, Survey Methodology, 29 (1),2003, 99–104.


**Appendix**

**R Calculations**

```
install.packages("xlsx")
library(xlxs)
library(readxl)
N1<-31;N2<-11;N3<-5;N4<-3;N5<-5
N<-55
n1<-15;n2<-5;n3<-3;n4<-2;n5<-3
n<-28
Q<-1;W1<-N1/N;W2<-N2/N;W3<-N3/N;W4<-N4/N;W5<-N5/N
data<-read_excel("C:/Users/dell pc/Desktop/Book2.xlsx")
head(data)
data1<-as.data.frame(data)
Book2<-data.frame(data1,g=c(rep(1,31),rep(2,11),rep(3,5),rep(4,3),rep(5,5)))
dim(data1)
head(Book2)
X<-split(Book2,Book2$g)
y<-lapply(seq_along(X), function(x)as.data.frame(X[[x]])[,1:2])
A1<-y[[1]]
A2<-y[[2]]
A3<-y[[3]]
A4<-y[[4]]
A5<-y[[5]]
X1<-mean(A1$X);X2<-mean(A2$X);X3<-mean(A3$X);X4<-mean(A4$X);X5<-mean(A5$X)
S1X<-sd(A1$X);S2X<-sd(A2$X);S3X<-sd(A3$X);S4X<-sd(A4$X);S5X<-sd(A5$X)
y1<-mean(A1$Y);y2<-mean(A2$Y);y3<-mean(A3$Y);y4<-mean(A4$Y);y5<-mean(A5$Y)
S1y<-sd(A1$Y);S2y<-sd(A2$Y);S3y<-sd(A3$Y);S4y<-sd(A4$Y);S5y<-sd(A5$Y)
Mys<-W1*y1+W2*y2+W3*y3+W4*y4+W5*y5
S1XY<-cov(A1$X,A1$Y);S2XY<-cov(A2$X,A2$Y);S3XY<-cov(A3$X,A3$Y);S4XY<-cov(A4$X,A4$Y);S5XY<-
cov(A5$X,A5$Y)
r1<-cor(A1$X,A1$Y);r2<-cor(A2$X,A2$Y);r3<-cor(A3$X,A3$Y);r4<-cor(A4$X,A4$Y);r5<-cor(A5$X,A5$Y)
ybart<-NA;MSEt<-NA;ybartcv<-NA;MSEcv<-NA;ybartn<-NA;MSEn<-NA
for(i in 1:50){
```

```
sam<-sample(1:31,15,replace=F)
sam1<-sample(1:11,5,replace=F)
sam2<-sample(1:5,3,replace=F)
sam3<-sample(1:3,2,replace=F)
sam4<-sample(1:5,3,replace=F)
sam11<-sample(A1$X,15);sam12<-sample(A2$X,5);sam13<-sample(A3$X,3);sam14<-sample(A4$X,2);sam15<-
sample(A5$X,3)
x11<-mean(sam11);x12<-mean(sam12);x13<-mean(sam13);x14<-mean(sam14);x15<-mean(sam15)
x11s<-sd(sam11);x12s<-sd(sam12);x13s<-sd(sam13);x14s<-sd(sam14);x15<-sd(sam15)
sam21<-sample(A1$Y,15);sam22<-sample(A2$Y,5);sam23<-sample(A3$Y,3);sam24<-sample(A4$Y,2);sam25<-
sample(A5$Y,3)
y11<-mean(sam21);y12<-mean(sam22);y13<-mean(sam23);y14<-mean(sam24);y15<-mean(sam25)
r11<-cor(sam11,sam21);r12<-cor(sam12,sam22);r13<-cor(sam13,sam23);r14<-cor(sam14,sam24);r15<-cor(sam15,sam25)
Q1<-Q*W1*x11s^4+Q*W2*x12s^4+Q*W3*x13s^4+Q*W4*x14s^4+Q*W5*x15s^4
Q2<-Q*W1*y11*x11+Q*W2*y12*x12+Q*W3*y13*x13+Q*W4*y14*x14+Q*W5*y15*x15
Q3<-Q*W1*x11s^2*x11+Q*W2*x12s^2*x12+Q*W3*x13s^2*x13+Q*W4*x14s^2*x14+Q*W5*x15s^2*x15
Q4<-Q*W1*x11s^2*y11+Q*W2*x12s^2*y12+Q*W3*x13s^2*y13+Q*W4*x14s^2*y14+Q*W5*x15s^2*y15
Q5<-Q*W1*x11^2+Q*W2*x12^2+Q*W3*x13^2+Q*W4*x14^2+Q*W5*x15^2
Q6<-Q*W1*x11s^2+Q*W2*x12s^2*+Q*W3*x13s^2+Q*W4*x14s^2+Q*W5*x15s^2
b1nk[i]<-Q1*Q2-Q3*Q2/Q1*Q5-Q6^2
b2nk[i]<-Q5*Q4-Q3*Q4/Q1*Q5-Q6^2
ybartnk[i]<-W1*y11+W2*y12+W3*y13+W4*y14+W5*y15+b1nk[i]*(W1*(X1-x11)+(W2*(X2-x12))+(W3*(X3-
x13))+(W4*(X4-x14))+(W5*(X5-x15)))+b2nk[i]*W1*(S1X^2-x11s^2)                    +W2*(S2X^2-x12s^2)+W3*(S3X^2-
x13s^2)+W4*(S4X^2-x14s^2)+W5*(S5X^2-x15s^2)
MSEt[i]<-(ybartnk[i]-Mys)^2
Q7<-Q*W1*r11^4+Q*W2*r12^4+Q*W3*r13^4+Q*W4*r14^4+Q*W5*r15^4
Q8<-Q*W1*cv11*y11+Q*W2*cv12*y12+Q*W3*cv13*y13+Q*W4*cv14*y14+Q*W5*cv15*y15
Q9<-Q*W1*r11^2*cv11+Q*W2*r12^2*cv12+Q*W3*r13^2*cv13+Q*W4*r14^2*cv14+Q*W5*r15^2*cv15
Q10<-Q*W1*r11^2*y11+Q*W2*r12^2*y12+Q*W3*r13^2*y13+Q*W4*r14^2*y14+Q*W5*r15^2*y15
Q11<-Q*W1*cv11^2+Q*W2*cv12^2+Q*W3*cv13^2+Q*W4*cv14^2+Q*W5*cv15^2
Q12<-Q*W1*r11^2+Q*W2*r12^2+Q*W3*r13^2+Q*W4*r14^2+Q*W5*r15^2
b1[i]<-Q7*Q8-Q9*Q8/Q11*Q7-Q12^2
b2[i]<-Q11*Q10-Q10*Q9/Q11*Q7-Q12^2
ybartcv[i]<-W1*y11+W2*y12+W3*y13+W4*y14+W5*y15+b1[i]*(W1*(Cx1-cv11)+W2*(Cx2-cv12)+W3*(Cx3-
cv13)+W4*(Cx4-cv14)+W5*(Cx5-cv15))+b2[i]*(W1*(r1^2-r11^2)+W2*(r2^2-r12^2)+W3*(r3^2-r13^2)+W4*(r4^2-
r14^2)+W5*(r5^2-r15^2))
MSEcv[i]<-(ybartcv[i]-Mys)^2
Q13<-Q*W1*cv11^4+Q*W2*cv12^4+Q*W3*cv13^4+Q*W4*cv14^4+Q*W5*cv15^4
Q14<-Q*W1*y11*x11+Q*W2*y12*x12+Q*W3*y13*x13+Q*W4*y14*x14+Q*W5*y15*x15
Q15<-Q*W1*cv11^2*x11+Q*W2*cv12^2*x12+Q*W3*cv13^2*x13+Q*W4*cv14^2*x14+Q*W5*cv15^2*x15
Q16<-Q*W1*cv11^2*y11+Q*W2*cv12^2*y12+Q*W3*cv13^2*y13+Q*W4*cv14^2*y14+Q*W5*cv15^2*y15
Q17<-Q*W1*x11^2+Q*W2*x12^2+Q*W3*x13^2+Q*W4*x14^2+Q*W5*x15^2
Q18<-Q*W1*cv11^2+Q*W2*cv12^2*+Q*W3*cv13^2+Q*W4*cv14^2+Q*W5*cv15^2
b1n[i]<-Q13*Q14-Q15*Q14/Q13*Q17-Q18^2
b2n[i]<-Q17*Q16-Q15*Q16/Q13*Q17-Q18^2
ybartn[i]<-W1*y11+W2*y12+W3*y13+W4*y14+W5*y15+b1n[i]*(W1*(X1-x11)+(W2*(X2-x12))+(W3*(X3-x13))+(W4*(X4-
x14))+(W5*(X5-x15)))+b2n[i]*W1*(S1X^2-x11s^2)                    +W2*(S2X^2-x12s^2)+W3*(S3X^2-x13s^2)+W4*(S4X^2-
x14s^2)+W5*(S5X^2-x15s^2)
MSEn[i]<-(ybartn[i]-Mys)^2
}
Pre<-MSEt[i]/ MSEn [i]*100
Pre<-MSEt[i]/ MSEcv [i]*100
```

*******