INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

Check for updates

OPEN ACCESS

Received: 05/08/2025

Accepted: 29/10/2025

Published: 23/11/2025

Citation: Patel A, Garg N, Ray BK (2025) Traditional and Calibration-based Estimation of the Population Mean for Stratified Random Sampling under Non-Response. Indian Journal of Science and Technology 18(42): 3357-3369. https://doi.org/10.17485/IJST/v18i42.1366

* **Corresponding author**.

anantpatel880@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Traditional and Calibration-based Estimation of the Population Mean for Stratified Random Sampling under Non-Response

Anant Patel[1]*, Neha Garg[1], Basant Kumar Ray[2]

1 School of Sciences, Indira Gandhi National Open University, New Delhi- 110068, India
2 Department of Statistics, Banaras Hindu University, Varanasi- 221005, Uttar Pradesh, India

Abstract

Objectives: To propose estimators for the population mean considering the non-response presented in the data, utilizing an auxiliary variable with the help of both traditional and calibration estimation methods. **Methods:** To investigate how the suggested calibration estimators perform in the case of non-response, a simulation study has been conducted using the R software. **Findings:** As per the percentage relative efficiencies (%PRE) using Population-I and II, the suggested estimators outperformed the existing estimators considered here, The percentage relative efficiencies (%PRE) of the proposed estimators $\bar{y}_{TST}, \bar{y}_{TY}, \bar{y}_{CST} and \bar{y}_{CY}$ vary from 200.513% to 215.641%, 200.524% to 215.654%, 217.910% to 247.667% and 217.904% to 247.662%, respectively, while 158.931% to 164.414% for the existing estimator $\bar{y}_{Lee}$ with respect to the estimator $\bar{y}_{HH}$, as shown in Table 4 for Population I. Similarly, the percentage relative efficiencies (%PRE) of the estimators with respect to $\bar{y}_{HH}$ those depicted in Table 4 vary from 144.9% to 148.3%, 237.2% to 264.7 %, 237.3 % to 264.7 %, 279.6% to 324.5% and 279.6% to 324.5% for the existing estimator $\bar{y}_{Lee}$, and the suggested estimators $\bar{y}_{TST}, \bar{y}_{TY}, \bar{y}_{CST} and \bar{y}_{CY}$, respectively, for Population II. **Novelty:** The proposed method has efficiently improved the traditional existing estimator's precision in estimation. Using this method, we can derive more new calibrated estimators.

Keywords: Calibration approach; Stratified random sampling; Chi-square type distance function; Non-response

1 Introduction

Non-response presents a major hurdle in survey research, as the absence of participation from certain individuals or units can lead to biased results and compromise the study's credibility. To address this, researchers employ several techniques, with calibration being a particularly effective method. In this study, we employed the calibration approach to address non-response. This method, along with a traditional approach, integrates auxiliary variables into the estimation process. The issue of non-response was first addressed in [1], which were pioneers in the development of the calibration estimation method.

Later, many authors utilized the calibration approach for parameter estimation [2], [3]. Many researchers have made significant contributions to address the non-response issue in the sample surveys as elaborated in [4], [5] and [6].

The challenge of estimating the population mean utilizing auxiliary information in the case of non-response has been examined in [7]. Two cases were used to examine this situation: (i) neither the study variable nor the auxiliary variable responds, and the population mean of the auxiliary variable is known; and (ii) full data on the auxiliary variable and the population mean of the auxiliary variable are available, while non-response occurs only on the study variable. The estimator considered in [1] and a few other existing estimators have been conceptually compared with the suggested estimators.

Even though the data may not be normally distributed, a novel approach was suggested for estimating the population mean in successive sampling across two occasions. In contrast to ordinary least squares (OLS) approaches, this approach was resilient to outliers, meaning that extreme values had less impact on it. The suggested method used estimators of the quantile regression ratio type, which took advantage of available information on the auxiliary variables [4]. A two-phase sampling technique was employed, and six classes of estimators were proposed for the population mean in stratified random sampling using an auxiliary variable, assuming that non-response occurs only in the study variable [8]. In case of mixed response and non-response patterns are detected, a new sampling strategy was considered in [9] to overcome the difficulties that frequently arise in estimating problems. In order to lower the impact of non-response on the data, calibration estimators of the population mean were introduced in the proposed setup.

The difficulty of estimating the finite population proportion under the Probability Proportional to Size (PPS) sampling technique was discussed when complete information is unavailable because of non-response [10]. Based on the availability of auxiliary data, the calibrated estimators of the population proportion were developed under PPS sampling in the presence of non-response. In addition to the design-based Hansen and Hurwitz type estimator in the presence of non-response, the developed estimators of the population proportion were compared with the design-based Horvitz-Thompson estimator and Horvitz-Thompson type calibration estimator that were obtained on the complete response units.

In this paper, both conventional and calibration techniques of estimation are considered to develop improved estimators for the population mean under stratified random sampling, utilizing an auxiliary variable in the presence of non-response. First, the conventional technique is applied by incorporating the given auxiliary information, and then the calibration technique is used to further enhance the estimators of the population mean. A chi-square-type distance function is considered, which is designed explicitly for non-response situations. The researchers typically calibrated stratum weights corresponding to non-responding units to obtain estimators of the population mean for a stratified population using a calibration approach. The proposed calibration-based estimators were derived in the current study by calibrating the weights assigned to the stratum means of the responding units and the sub-sampled units drawn from the non-responding sample. The performance of the proposed calibration estimator in the presence of non-response has also been examined through a simulation analysis conducted in R software.

2 Methodology

2.1 Sampling Design and Notations

Let $P = \{(y_1, x_1), (y_2, x_2), \ldots, (y_N, x_N)\}$ be a set of finite population of N well-defined and identifiable elements in a mail survey design with the study variable Y and auxiliary variable X. Given a population exhibiting heterogeneity, we divided it into L homogeneous groups guided by prior knowledge about the population. Let the h^{th} group be made up of N_h units such that $\sum_{h=1}^{L} N_h = N$. Assume that out of N_h units, N_{h1} units provide the responses on the surveyor's first attempt, while not being able to get responses on N_{h2}. Using the Simple Random Sampling Without Replacement (SRSWOR) technique, select $n_h \in s_h$ units randomly from the h^{th}stratum such that $\sum_{h=1}^{L} n_h = n$. Suppose we do not receive responses on each sample unit drawn from the h^{th} stratum. It is assumed that out of n_h units, $n_{h1} \in s_{rh}$ units respond, while the remaining $n_{h2} \in s_{nh}$ units do not respond in the h^{th} stratum. We also suppose that the n_{h1} and n_{h2} sample units, respectively, are drawn from the N_{h1} responding and N_{h2} non-responding units. Let us choose a sub-sample of $n_{sh} \in ss_h$ units utilizing the SRSWOR scheme from the sample s_{nh} of n_{h2} non-responding units using the method [1], such that $n_{sh} = \frac{n_{h2}}{g_h}; g_h > 1$.

To escape from notational ambiguity, we consider the following notations:

Y: Study variable.

X: Auxiliary variable.

(y_{hi}, x_{hi}): Pair of the i^{th} unit in the h^{th} stratum.

(y_{rhi}, x_{rhi}): Pair of the i^{th} responding units in the h^{th} stratum.

(y_{nhi}, x_{nhi}): Pair of the i^{th} non-responding units in the h^{th} stratum.
ρ: Correlation coefficient between study variable Y and auxiliary variable X.
ρ_h : Correlation coefficient between the study variable and auxiliary variable in the h^{th} stratum.
$R = \frac{\bar{Y}}{\bar{X}}$ and $R_h = \frac{\bar{Y}_h}{\bar{X}_h}$.
The combined means, mean-squares, and covariance of Y and X variables at the population level are defined as:
$\overline{Y} = \frac{\sum_{h=1}^{L} N_h \overline{Y}_h}{N}, \overline{X} = \frac{\sum_{h=1}^{L} N_h \overline{X}_h}{N}$,
$S_Y^2 = \frac{1}{N-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(y_{hi}-\overline{Y}\right)^2, S_X^2 = \frac{1}{N-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(x_{hi}-\overline{X}\right)^2$,
and $S_{XY} = \frac{1}{N-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(y_{hi}-\overline{Y}\right)\left(x_{hi}-\overline{X}\right)$.
The means, mean-squares and covariance of Y and X variables for the h^{th} stratum at the population level are given as:
$\overline{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}, \overline{X}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} x_{hi}$,
$S_{hY}^2 = \frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(y_{hi}-\overline{Y}_h\right)^2, S_{hX}^2 = \frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(x_{hi}-\overline{X}_h\right)^2$,
and $S_{hXY} = \frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(y_{hi}-\overline{Y}_h\right)\left(x_{hi}-\overline{X}_h\right)$.
The population means, mean-squares and covariance of Y and X variables of the N_{h2} non-responding units for the h^{th} stratum are
$\overline{Y}_{N_{h2}} = \frac{1}{N_{h2}}\sum_{i=1}^{N_{h2}} y_{nhi}, \overline{X}_{N_{h2}} = \frac{1}{N_{h2}}\sum_{i=1}^{N_{h2}} x_{nhi}$,
$S_{hY(2)}^2 = \frac{1}{N_{h2}-1}\sum_{i=1}^{N_{h2}}\left(y_{nhi}-\overline{Y}_{N_{h2}}\right)^2, S_{hX(2)}^2 = \frac{1}{N_{h2}-1}\sum_{i=1}^{N_{h2}}\left(x_{nhi}-\overline{X}_{N_{h2}}\right)^2$,
and $S_{hXY(2)} = \frac{1}{N_{h2}-1}\sum_{i=1}^{N_{h2}}\left(y_{nhi}-\overline{Y}_{N_{h2}}\right)\left(x_{nhi}-\overline{X}_{N_{h2}}\right)$.
Sample means, mean-squares and covariance of Y and X variables for the h^{th} stratum are
$\bar{y}_h = \frac{1}{n_h}\sum_{i\in s_h} y_{hi}, \bar{x}_h = \frac{1}{n_h}\sum_{i\in s_h} x_{hi}, s_{hy}^2 = \frac{1}{n_h-1}\sum_{i\in s_h}\left(y_{hi}-\bar{y}_h\right)^2, s_{hx}^2 = \frac{1}{n_h-1}\sum_{i\in s_h}\left(x_{hi}-\bar{x}_h\right)^2, s_{hxy} = \frac{1}{n_h-1}\sum_{i\in s_h}\left(y_{hi}-\bar{y}_h\right)\left(x_{hi}-\bar{x}_h\right)$.
Sample means of Y and X variables for the n_{h1} responding units drawn from the N_{h1} units are
$\overline{y}_{rh} = \frac{1}{n_{h1}}\sum_{i\in s_{rh}} y_{rhi}$ and $\overline{x}_{rh} = \frac{1}{n_{h1}}\sum_{i\in s_{rh}} x_{rhi}$.
Sub-sample means, mean-squares and covariance of Y and X variables for the n_{sh} responding units drawn from the n_{h2} non-responding units are
$\overline{y}_{sh} = \frac{1}{n_{sh}}\sum_{i\in ss_h} y_{nhi}, \quad \overline{x}_{sh} = \frac{1}{n_{sh}}\sum_{i\in ss_h} x_{nhi}, \quad s_{hy(2)}^2 = \frac{1}{n_{sh}-1}\sum_{i\in ss_h}\left(y_{nhi}-\overline{y}_{sh}\right)^2, \quad s_{hx(2)}^2 = \frac{1}{n_{sh}-1}\sum_{i\in ss_h}\left(x_{nhi}-\overline{x}_{sh}\right)^2$ and $s_{hxy(2)} = \frac{1}{n_{sh}-1}\sum_{i\in ss_h}\left(y_{nhi}-\overline{y}_{sh}\right)\left(x_{nhi}-\overline{x}_{sh}\right)$.

2.2 Existing Traditional and Calibration-based Estimators

The estimator of the population mean proposed by[11] in SRSWOR is given as:

$$T_{ST} = \bar{y}\frac{\overline{X}+\rho}{\overline{x}+\rho} \tag{1}$$

The *MSE* of the estimator T_{ST} is expressed as

$$MSE\left[T_{ST}\right] = \frac{(1-f)}{n}\left[S_Y^2 + R^2\left(\frac{\overline{X}}{\overline{X}+\rho}\right)^2 S_X^2 - 2R\left(\frac{\overline{X}}{\overline{X}+\rho}\right)S_{XY}\right] \tag{2}$$

where $f = \frac{n}{N}$ and $R = \frac{\overline{Y}}{\overline{X}}$.
The estimator of the population mean proposed by[11] in SRSWOR is given as

$$T_Y = \bar{y}\frac{n\overline{X}+\rho}{n\bar{x}+\rho} \tag{3}$$

The *MSE* of the estimator T_Y is

$$MSE\left[T_Y\right] = \frac{(1-f)}{n}\left[S_Y^2 + R^2\left(\frac{n\overline{X}}{n\overline{X}+\rho}\right)^2 S_X^2 - 2R\left(\frac{n\overline{X}}{n\overline{X}+\rho}\right)S_{XY}\right] \tag{4}$$

If we consider the stratified random sampling, the usual estimator of the population mean $\overline{Y}$ is given as

$$\bar{y}_{st} = \sum_{h=1}^{L} w_h \bar{y}_h \tag{5}$$

where $w_h = \frac{N_h}{N}$.

In the cases where information is incomplete for certain units, the conventional estimator for the mean of a stratified population, as prescribed in[1], is expressed as

$$\bar{y}_{HH} = \sum_{h=1}^{L} w_h \bar{y}_h^* \tag{6}$$

where $\bar{y}_h^* = \frac{n_{h1}\bar{y}_{rh} + n_{h2}\bar{y}_{sh}}{n_h} = w_{h1}\bar{y}_{rh} + w_{h2}\bar{y}_{sh}$.

The variance expression for the estimator $\bar{y}_{HH}$ is given as

$$V\left[\bar{y}_{HH}\right] = \sum_{h=1}^{L} \frac{w_h^2 \left(1 - f_h\right)}{n_h} S_{hY}^2 + \sum_{h=1}^{L} \frac{w_h^2 \left(g_h - 1\right) W_{h2}}{n_h} S_{hY(2)}^2 \tag{7}$$

where $W_{h2} = \frac{N_{h2}}{N_h}$ is the non-response rate in the population for the h^{th} stratum.

When some units do not respond, the following calibration estimator is proposed[3] for the population mean of a stratified population in the presence of non-response:

$$\bar{y}_{Lee} = \sum_{h=1}^{L} \delta_h^* \bar{y}_h^* \tag{8}$$

where δ_h^* is the calibration weight for the h^{th} stratum.

The chi-square type distance function, as used in[3], is given as:

$$D_{Lee} = \frac{1}{2} \sum_{h=1}^{L} \frac{\left(\delta_h^* - w_h\right)^2}{w_h q_h} \tag{9}$$

Two calibration constraints considered for the derivation of the suggested calibration estimator are

$$\sum_{h=1}^{L} \delta_h^* \bar{x}_h^* = \overline{X} \tag{10}$$

$$\sum_{h=1}^{L} \delta_h^* = \sum_{h=1}^{L} w_h \tag{11}$$

The final form of the estimator $\bar{y}_{Lee}$ is given by

$$\bar{y}_{Lee} = \sum_{h=1}^{L} w_h \bar{y}_h^* + \hat{\beta}_{Lee} \left[\overline{X} - \sum_{h=1}^{L} w_h \bar{x}_h^*\right] \tag{12}$$

where $\hat{\beta}_{Lee} = \frac{\left(\sum_{h=1}^{L} q_h w_h \bar{y}_h^* \bar{x}_h^*\right)\left(\sum_{h=1}^{L} q_h w_h\right) - \left(\sum_{h=1}^{L} q_h w_h \bar{y}_h^*\right)\left(\sum_{h=1}^{L} q_h w_h \bar{x}_h^*\right)}{\left(\sum_{h=1}^{L} q_h w_h \bar{x}_h^{*2}\right)\left(\sum_{h=1}^{L} q_h w_h\right) - \left(\sum_{h=1}^{L} q_h w_h \bar{x}_h^*\right)^2}$ and $\bar{x}_h^* = \frac{n_{h1}\bar{x}_{rh} + n_{h2}\bar{x}_{sh}}{n_h}$.

The variance expression for the calibration estimator $\bar{y}_{Lee}$ under the proportional allocation is given as

$$V\left(\bar{y}_{Lee}\right) = \sum_{h=1}^{L} \left(\frac{1}{n} - \frac{1}{N}\right) w_h \sigma_{he}^2 + \sum_{h=1}^{L} \frac{\left(g_h - 1\right) w_h W_{h2}}{n} \sigma_{he(2)}^2 \tag{13}$$

where $\sigma_{he}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} e_{hi}^{*2}$; $e_{hi}^* = \left(y_{hi} - \overline{Y}_h\right) - \hat{\beta}_{Lee}\left(x_{hi} - \overline{X}_h\right)$,

$\sigma_{he(2)}^2 = \frac{1}{N_{h2} - 1} \sum_{k=1}^{N_{h2}} e_{nhk}^{*2}$; $e_{nhk}^* = \left(y_{nhk} - \overline{Y}_{N_{h2}}\right) - \hat{\beta}_{Lee}\left(x_{nhk} - \overline{X}_{N_{h2}}\right)$ and $\hat{\beta}_{Lee}$ is the estimator of population level value β_{Lee}.

2.3 Proposed Calibration Estimators in Stratified Sampling

Before defining the suggested estimators, let us recall[1], the usual estimator $\bar{y}_{HH}$ given in **equation (6)** as:

$$\bar{y}_{HH} = \sum_{h=1}^{L} w_h \{w_{h1}\bar{y}_{rh} + w_{h2}\bar{y}_{sh}\} \tag{14}$$

Let us define an estimator of the population mean by incorporating the auxiliary information into the estimation process using the traditional method.

$$\bar{y}_{TM} = \sum_{h=1}^{L} w_h T_h \{w_{h1}\bar{y}_{rh} + w_{h2}\bar{y}_{sh}\} \tag{15}$$

where T_h is a function of the sample mean $(\bar{x}_h)$ and population mean $(\bar{X}_h)$ utilizing a known parameter (g_h) of the auxiliary variable X defined for the h^{th} stratum, i.e., $T_h = f(\bar{x}_h, \bar{X}_h, g_h)$.

In this paper, we will consider two functions, and you can also generate many more estimators using different forms of the function T_h.

We now suggest new estimators of the population mean by incorporating auxiliary information using the calibration approach when some units do not respond, as follows:

$$\bar{y}_{CT} = \sum_{h=1}^{L} \{\delta_{h1}\bar{y}_{rh} + \delta_{h2}\bar{y}_{sh}\} T_h$$

$$\text{or} \qquad \bar{y}_{CT} = \sum_{h=1}^{L} \delta_{h1}\bar{y}_{rh} T_h + \sum_{h=1}^{L} \delta_{h2}\bar{y}_{sh} T_h \tag{16}$$

Here, δ_{h1} and δ_{h2} are the new weights assigned to the sample mean $\bar{y}_{rh}$ and subsample mean $\bar{y}_{sh}$ of the responding and subsample from the non-responding groups, respectively. Let us define a new chi-square type distance function suitable for our situation.

$$D = \sum_{h=1}^{L} \frac{(\delta_{h1} - w_h w_{h1})^2}{w_h w_{h1} q_{h1}} + \sum_{h=1}^{L} \frac{(\delta_{h2} - w_h w_{h2})^2}{w_h w_{h2} q_{h2}} \tag{17}$$

Subject to the following calibration conditions:

$$\sum_{h=1}^{L} \{\delta_{h1}\bar{x}_{rh} + \delta_{h2}\bar{x}_{sh}\} = \overline{X} \tag{18}$$

$$\text{and} \qquad \sum_{h=1}^{L} \{\delta_{h1} + \delta_{h2}\} = 1 \tag{19}$$

Note that q_{h1} and q_{h2} quantities in the aforementioned **equation (17)** are appropriately selected constants that determine the various estimators' forms. Thus, we obtain a family of calibration estimators for various values of q_{h1} and q_{h2}.

Let us define the Lagrange function to minimize the chi-square type distance function **(17)** subject to calibration conditions **(18)** and **(19)**.

$$\begin{aligned} L \; = \; & \sum_{h=1}^{L} \frac{(\delta_{h1} - w_h w_{h1})^2}{w_h w_{h1} q_{h1}} + \sum_{h=1}^{L} \frac{(\delta_{h2} - w_h w_{h2})^2}{w_h w_{h2} q_{h2}} - 2\lambda_1 \left\{ \sum_{h=1}^{L} \{\delta_{h1}\bar{x}_{rh} + \delta_{h2}\bar{x}_{sh}\} - \overline{X} \right\} \\ & -2\lambda_2 \left\{ \sum_{h=1}^{L} \{\delta_{h1} + \delta_{h2}\} - 1 \right\} \end{aligned} \tag{20}$$

Let us now differentiate the Lagrange function in **equation (20)** with respect to δ_{h1} and equate the derivative to zero.

$$\begin{aligned} \frac{\partial L}{\partial \delta_{h1}} \; = \; & \frac{2(\delta_{h1} - w_h w_{h1})}{w_h w_{h1} q_{h1}} - 2\lambda_1 \bar{x}_{rh} - 2\lambda_2 = 0 \\ & \Rightarrow \frac{(\delta_{h1} - w_h w_{h1})}{w_h w_{h1} q_{h1}} - \lambda_1 \bar{x}_{rh} - \lambda_2 = 0 \\ & \Rightarrow \delta_{h1} = w_h w_{h1} (1 + \lambda_1 q_{h1} \bar{x}_{rh} + \lambda_2 q_{h1}) \end{aligned} \tag{21}$$

Let us again differentiate the Lagrange function in **equation (20)** with respect to δ_{h2} and equate the derivative to zero.

$$\begin{aligned} \frac{\partial L}{\partial \delta_{h2}} \; = \; & \frac{2(\delta_{h2} - w_h w_{h2})}{w_h w_{h2} q_{h2}} - 2\lambda_1 \bar{x}_{sh} - 2\lambda_2 = 0 \\ & \Rightarrow \frac{(\delta_{h2} - w_h w_{h2})}{w_h w_{h2} q_{h2}} - \lambda_1 \bar{x}_{sh} - \lambda_2 = 0 \\ & \Rightarrow \delta_{h2} = w_h w_{h2} (1 + \lambda_1 q_{h2} \bar{x}_{sh} + \lambda_2 q_{h2}) \end{aligned} \tag{22}$$

Putting the values of δ_{h1} and δ_{h2} from **equations (21)** and **(22)** into **equation (18)**, we get

$$\sum_{h=1}^{L}\{w_h w_{h1}(1+\lambda_1 q_{h1}\bar{x}_{rh}+\lambda_2 q_{h1})\}\bar{x}_{rh}+\sum_{h=1}^{L}\{w_h w_{h2}(1+\lambda_1 q_{h2}\bar{x}_{sh}+\lambda_2 q_{h2})\}\bar{x}_{sh}=\overline{X}$$

$$\lambda_1\textstyle\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}\bar{x}_{rh}^2+w_{h2}q_{h2}\bar{x}_{sh}^2\}+\lambda_2\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh}\}=\overline{X}-\bar{x}_{HH} \tag{23}$$

Substituting the values of δ_{h1} and δ_{h2} from **equations (21)** and **(22)** into **equation (19)**, we obtain

$$\sum_{h=1}^{L}\{w_h w_{h1}(1+\lambda_1 q_{h1}\bar{x}_{rh}+\lambda_2 q_{h1})\}+\sum_{h=1}^{L}\{w_h w_{h2}(1+\lambda_1 q_{h2}\bar{x}_{sh}+\lambda_2 q_{h2})\}=1$$

$$\lambda_1\textstyle\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh}\}+\lambda_2\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}+w_{h2}q_{h2}\}=0 \tag{24}$$

After solving the system of **equations (23) and (24)**,

we obtain $\lambda_1=\frac{\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}+w_{h2}q_{h2}\}[\overline{X}-\bar{x}_{HH}]}{\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}\bar{x}_{rh}^2+w_{h2}q_{h2}\bar{x}_{sh}^2\}\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}+w_{h2}q_{h2}\}-\left[\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh}\}\right]^2}$

$$\lambda_2=-\frac{\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh}\}[\overline{X}-\bar{x}_{HH}]}{\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}\bar{x}_{rh}^2+w_{h2}q_{h2}\bar{x}_{sh}^2\}\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}+w_{h2}q_{h2}\}-\left[\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh}\}\right]^2}$$

The calibrated weights after substituting the values λ_1 and $\lambda_2$are determined as:

$$\delta_{h1}=w_h w_{h1}+w_h w_{h1}q_{h1}\frac{\begin{bmatrix}\bar{x}_{rh}\sum_{h=1}^{L}\{w_h w_{h1}q_{h1}+w_h w_{h2}q_{h2}\}\\-\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh}\}\end{bmatrix}[\overline{X}-\bar{x}_{HH}]}{\begin{bmatrix}\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}\bar{x}_{rh}^2+w_{h2}q_{h2}\bar{x}_{sh}^2\}\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}+w_{h2}q_{h2}\}\\-\left\{\sum_{h=1}^{L} w_h(w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh})\right\}^2\end{bmatrix}}$$

and

$$\delta_{h2}=w_h w_{h2}+w_h w_{h2}q_{h2}\frac{\begin{bmatrix}\bar{x}_{sh}\sum_{h=1}^{L}\{w_h w_{h1}q_{h1}+w_h w_{h2}q_{h2}\}\\-\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh}\}\end{bmatrix}[\overline{X}-\bar{x}_{HH}]}{\begin{bmatrix}\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}\bar{x}_{rh}^2+w_{h2}q_{h2}\bar{x}_{sh}^2\}\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}+w_{h2}q_{h2}\}\\-\left\{\sum_{h=1}^{L} w_h(w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh})\right\}^2\end{bmatrix}}$$

The final form of the calibration estimator given in **equation (16)** is established as:

$$\begin{aligned}
\bar{y}_{CT} &= \sum_{h=1}^{L} w_h\left(w_{h1}\bar{y}_{rh}+w_{h2}\bar{y}_{sh}\right)T_h \\
&+\frac{\left[\begin{array}{l}\displaystyle\sum_{h=1}^{L} w_h T_h\left\{w_{h1}q_{h1}\bar{y}_{rh}\bar{x}_{rh}+w_{h2}q_{h2}\bar{y}_{sh}\bar{x}_{sh}\right\}\sum_{h=1}^{L}\left\{w_h w_{h1}q_{h1}+w_h w_{h2}q_{h2}\right\}\\ \displaystyle-\sum_{h=1}^{L} w_h\left\{w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh}\right\}\sum_{h=1}^{L} w_h T_h\left\{w_{h1}q_{h1}\bar{y}_{rh}+w_{h2}q_{h2}\bar{y}_{sh}\right\}\end{array}\right]\left[\overline{X}-\bar{x}_{HH}\right]}{\left[\begin{array}{l}\displaystyle\sum_{h=1}^{L} w_h\left\{w_{h1}q_{h1}\bar{x}_{rh}^2+w_{h2}q_{h2}\bar{x}_{sh}^2\right\}\sum_{h=1}^{L} w_h\left\{w_{h1}q_{h1}+w_{h2}q_{h2}\right\}\\ \displaystyle-\left\{\sum_{h=1}^{L} w_h\left(w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh}\right)\right\}^2\end{array}\right]} \\
\text{or } \bar{y}_{CT} &= \sum_{h=1}^{L} w_h\bar{y}_h^{*}T_h+\hat{\beta}\left[\overline{X}-\bar{x}_{HH}\right]
\end{aligned} \tag{25}$$

where

$$\hat{\beta}=\frac{\left[\begin{array}{l}\displaystyle\sum_{h=1}^{L} w_h T_h\left\{w_{h1}q_{h1}\bar{y}_{rh}\bar{x}_{rh}+w_{h2}q_{h2}\bar{y}_{sh}\bar{x}_{sh}\right\}\sum_{h=1}^{L}\left\{w_h w_{h1}q_{h1}+w_h w_{h2}q_{h2}\right\}\\ \displaystyle-\sum_{h=1}^{L} w_h\left\{w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh}\right\}\sum_{h=1}^{L} w_h T_h\left\{w_{h1}q_{h1}\bar{y}_{rh}+w_{h2}q_{h2}\bar{y}_{sh}\right\}\end{array}\right]}{\left[\begin{array}{l}\displaystyle\sum_{h=1}^{L} w_h\left\{w_{h1}q_{h1}\bar{x}_{rh}^2+w_{h2}q_{h2}\bar{x}_{sh}^2\right\}\sum_{h=1}^{L} w_h\left\{w_{h1}q_{h1}+w_{h2}q_{h2}\right\}\\ \displaystyle-\left\{\sum_{h=1}^{L} w_h\left(w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh}\right)\right\}^2\end{array}\right]}$$

Let us consider two cases of function T_h to define the proposed traditional and calibration estimators of the population mean.

Case 1: Let us first take the functional form of T_h inspired by the estimator proposed[11] for the h^{th} stratum, *i.e.*,$T_h = \frac{\overline{X}_h+\rho_h}{\bar{x}_h^{*}+\rho_h}$. The traditional estimator of the population mean suggested in **equation (15)** in the presence of the non-response will reduce to

$$\bar{y}_{TST}=\sum_{h=1}^{L} w_h\left(\frac{\overline{X}_h+\rho_h}{\bar{x}_h^{*}+\rho_h}\right)\bar{y}_h^{*} \tag{26}$$

The suggested calibration estimator recommended in **equation (25)** for $T_h = \frac{\overline{X}_h+\rho_h}{\bar{x}_h^{*}+\rho_h}$ is given as:

$$\bar{y}_{CST}=\sum_{h=1}^{L} w_h\bar{y}_h^{*}\left(\frac{\overline{X}_h+\rho_h}{\bar{x}_h^{*}+\rho_h}\right)+\hat{\beta}_{ST}\left[\overline{X}-\bar{x}_{HH}\right] \tag{27}$$

where

$$\hat{\beta}_{ST}=\frac{\left[\begin{array}{l}\displaystyle\sum_{h=1}^{L} w_h\left(\frac{\overline{X}_h+\rho_h}{\bar{x}_h^{*}+\rho_h}\right)\left\{w_{h1}q_{h1}\bar{y}_{rh}\bar{x}_{rh}+w_{h2}q_{h2}\bar{y}_{sh}\bar{x}_{sh}\right\}\sum_{h=1}^{L}\left\{w_h w_{h1}q_{h1}+w_h w_{h2}q_{h2}\right\}\\ \displaystyle-\sum_{h=1}^{L} w_h\left\{w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh}\right\}\sum_{h=1}^{L} w_h\left(\frac{\overline{X}_h+\rho_h}{\bar{x}_h^{*}+\rho_h}\right)\left\{w_{h1}q_{h1}\bar{y}_{rh}+w_{h2}q_{h2}\bar{y}_{sh}\right\}\end{array}\right]}{\left[\begin{array}{l}\displaystyle\sum_{h=1}^{L} w_h\left\{w_{h1}q_{h1}\bar{x}_{rh}^2+w_{h2}q_{h2}\bar{x}_{sh}^2\right\}\sum_{h=1}^{L} w_h\left\{w_{h1}q_{h1}+w_{h2}q_{h2}\right\}\\ \displaystyle-\left\{\sum_{h=1}^{L} w_h\left(w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh}\right)\right\}^2\end{array}\right]}$$

Case 2: In the same way, if the functional form of T$_h$ is encouraged in [11] for the h^{th} stratum, *i.e.* $T_h = \frac{n\overline{X}_h+\rho_h}{n\bar{x}_h^*+\rho_h}$, the traditional non-response version of the estimator proposed in **equation (15)** is obtained as:

$$\bar{y}_{TY} = \sum_{h=1}^{L} w_h \left(\frac{n\overline{X}_h+\rho_h}{n\bar{x}_h^*+\rho_h}\right)\bar{y}_h^* \tag{28}$$

And the suggested calibration estimator in **equation (25)** can be written for $T_h = \frac{n\overline{X}_h+\rho_h}{n\bar{x}_h^*+\rho_h}$ as:

$$\bar{y}_{CY} = \sum_{h=1}^{L} w_h \bar{y}_h^* \left(\frac{n\overline{X}_h+\rho_h}{n\bar{x}_h^*+\rho_h}\right) + \hat{\beta}_Y\left[\overline{X}-\bar{x}_{HH}\right] \tag{29}$$

where

$$\hat{\beta}_Y = \frac{\begin{bmatrix}\sum_{h=1}^{L} w_h \left(\frac{n\overline{X}_h+\rho_h}{n\bar{x}_h^*+\rho_h}\right)\{w_{h1}q_{h1}\bar{y}_{rh}\bar{x}_{rh}+w_{h2}q_{h2}\bar{y}_{sh}\bar{x}_{sh}\}\sum_{h=1}^{L}\{w_h w_{h1}q_{h1}+w_h w_{h2}q_{h2}\} \\ -\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh}\}\sum_{h=1}^{L} w_h\left(\frac{n\overline{X}_h+\rho_h}{n\bar{x}_h^*+\rho_h}\right)\{w_{h1}q_{h1}\bar{y}_{rh}+w_{h2}q_{h2}\bar{y}_{sh}\}\end{bmatrix}}{\begin{bmatrix}\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}\bar{x}_{rh}^2+w_{h2}q_{h2}\bar{x}_{sh}^2\}\sum_{h=1}^{L} w_h\{w_{h1}q_{h1}+w_{h2}q_{h2}\} \\ -\left\{\sum_{h=1}^{L} w_h\left(w_{h1}q_{h1}\bar{x}_{rh}+w_{h2}q_{h2}\bar{x}_{sh}\right)\right\}^2\end{bmatrix}}$$

Similarly, many more traditional and calibration forms of the estimators can be derived using different forms of function T_h. Due to the length of the research paper, we have considered only two functions here.

2.4 Mean Squared Errors of the Proposed Estimators

Let us derive the expressions of the variances for both cases of the proposed traditional and calibration estimators considered in the presence of non-response in **equations (26) to (29)** one by one.

Case 1: We first consider the proposed traditional estimator defined in **equation (26)** as:

$\bar{y}_{TST} = \sum_{h=1}^{L} w_h \left(\frac{\overline{X}_h+\rho_h}{\bar{x}_h^*+\rho_h}\right)\bar{y}_h^*$

For the large sample approximation, we define $\bar{y}_h^* = \overline{Y}_h(1+e_{0h})$ and $\bar{x}_h^* = \overline{X}_h(1+e_{1h})$.

where $E(e_{0h}) = E(e_{1h}) = 0, E(e_{0h}^2) = \frac{1}{\overline{Y}_h^2}\left[\frac{(1-f_h)}{n_h}S_{hY}^2 + \frac{(g_h-1)W_{h2}}{n_h}S_{hY(2)}^2\right], E(e_{1h}^2) = \frac{1}{\overline{X}_h^2}\left[\frac{(1-f_h)}{n_h}S_{hX}^2 + \frac{(g_h-1)W_{h2}}{n_h}S_{hX(2)}^2\right]$,

and $E(e_{0h}e_{1h}) = \frac{1}{\overline{X}_h\overline{Y}_h}\left\{\frac{(1-f_h)}{n_h}S_{hXY} + \frac{(g_h-1)W_{h2}}{n_h}S_{hXY(2)}\right\}$.

After putting these approximations in the estimator $\bar{y}_{TST}$, we obtain

$$\bar{y}_{TST} = \sum_{h=1}^{L} w_h\left(\frac{\overline{X}_h+\rho_h}{\overline{X}_h(1+e_{1h})+\rho_h}\right)\overline{Y}_h(1+e_{0h})$$

$\bar{y}_{TST} = \sum_{h=1}^{L} w_h\left(1+\frac{\overline{X}_h e_{1h}}{\overline{X}_h+\rho_h}\right)^{-1}\overline{Y}_h(1+e_{0h})$

The expression of the estimator in terms of e_{0h} and e_{1h} up to the first order of approximation is

$\bar{y}_{TST} = \sum_{h=1}^{L} w_h\overline{Y}_h\left(1+e_{0h}-\frac{\overline{X}_h e_{1h}}{\overline{X}_h+\rho_h}\right)+\ldots$

The *MSE* of the estimator $\overline{y}_{TST}$ is derived as

$$\begin{aligned}
MSE\left(\bar{y}_{TST}\right) &= E\left[\left\{\sum_{h=1}^{L} w_h \overline{Y}_h\left(1+e_{0h}-\frac{\overline{X}_h e_{1h}}{\overline{X}_h+\rho_h}\right)-\sum_{h=1}^{L} w_h \overline{Y}_h\right\}^2\right] \\
MSE\left(\bar{y}_{TST}\right) &= \sum_{h=1}^{L} w_h^2 \overline{Y}_h^2\left\{E\left(e_{0h}^2\right)+\left(\frac{\overline{X}_h}{\overline{X}_h+\rho_h}\right)^2 E\left(e_{1h}^2\right)-2\left(\frac{\overline{X}_h}{\overline{X}_h+\rho_h}\right) E\left(e_{0h}e_{1h}\right)\right\} \\
MSE\left(\bar{y}_{TST}\right) &= \sum_{h=1}^{L} w_h^2\left[\frac{(1-f_h)}{n_h} S_{hY}^2+\frac{(g_h-1) W_{h2}}{n_h} S_{hY(2)}^2\right. \\
&\quad +R_h^2\left(\frac{\overline{X}_h}{\overline{X}_h+\rho_h}\right)^2\left\{\frac{(1-f_h)}{n_h} S_{hX}^2+\frac{(g_h-1) W_{h2}}{n_h} S_{hX(2)}^2\right\} \\
&\quad \left.-2R_h\left(\frac{\overline{X}_h}{\overline{X}_h+\rho_h}\right)\left\{\frac{(1-f_h)}{n_h} S_{hXY}+\frac{(g_h-1) W_{h2}}{n_h} S_{hXY(2)}\right\}\right]
\end{aligned} \tag{30}$$

Let us consider the proposed calibration estimator $\overline{y}_{CST}$ given in **equation (28)**.

$$\bar{y}_{CST}=\sum_{h=1}^{L} w_h \bar{y}_h^*\left(\frac{\overline{X}_h+\rho_h}{\bar{x}_h^*+\rho_h}\right)+\hat{\beta}_{ST}\left[\overline{X}-\bar{x}_{HH}\right]$$

Let us take $q_{h1}=1$ and $q_{h2}=1$. The proposed calibration estimator for this choice of q_{h1} and q_{h2} reduces to

$$\bar{y}_{CST}=\sum_{h=1}^{L} w_h \bar{y}_h^*\left(\frac{\overline{X}_h+\rho_h}{\bar{x}_h^*+\rho_h}\right)+\frac{\left[\begin{array}{l}\sum_{h=1}^{L} w_h\left(\frac{\overline{X}_h+\rho_h}{\bar{x}_h^*+\rho_h}\right)\left\{w_{h1}\bar{y}_{rh}\bar{x}_{rh}+w_{h2}\bar{y}_{sh}\bar{x}_{sh}\right\}\cdot 1 \\ -\bar{x}_{HH}\sum_{h=1}^{L} w_h\left(\frac{\overline{X}_h+\rho_h}{\bar{x}_h^*+\rho_h}\right)\bar{y}_h^*\end{array}\right]}{\left[\sum_{h=1}^{L} w_h\left\{w_{h1}\bar{x}_{rh}^2+w_{h2}\bar{x}_{sh}^2\right\}\cdot 1-\bar{x}_{HH}^2\right]}\left[\overline{X}-\bar{x}_{HH}\right]$$

To obtain the MSE of the suggested estimator, the proposed calibration estimator can be expressed as:

$$\bar{y}_{CST}=\phi\left(\tau_1,\tau_2,\tau_3,\tau_4\right)=\tau_1+\frac{\tau_3-\tau_2\tau_1}{\tau_4-\left(\tau_2\right)^2}\left[\overline{X}-\tau_2\right]$$

where
$\tau_1=\sum_{h=1}^{L} w_h \bar{y}_h^*\left(\frac{\overline{X}_h+\rho_h}{\bar{x}_h^*+\rho_h}\right)$, $\tau_2=\bar{x}_{HH}$, $\tau_3=\sum_{h=1}^{L} w_h\left(\frac{\overline{X}_h+\rho_h}{\bar{x}_h^*+\rho_h}\right)\left(w_{h1}\bar{x}_{rh}\bar{y}_{rh}+w_{h2}\bar{x}_{sh}\bar{y}_{sh}\right)$ and $\tau_4=\sum_{h=1}^{L} w_h\left(w_{h1}\bar{x}_{rh}^2+w_{h2}\bar{x}_{sh}^2\right)$.

The first-order partial derivatives of the function $\overline{y}_{CST}=\phi\left(\tau_1,\tau_2,\tau_3,\tau_4\right)$ with respect to component estimators τ_1,τ_2,τ_3 and τ_4 at the parametric values $\xi_1=\overline{Y}$, $\xi_2=\overline{X}$, $\xi_3=\sum_{h=1}^{L} w_h\overline{Y}_h\overline{X}_h$, and $\xi_4=\sum_{h=1}^{L} w_h\overline{X}_h^2$, respectively, are

$\phi_1=\left.\frac{\partial\phi}{\partial\tau_1}\right|_{(\tau_i=\xi_i;\, i=1,2,3,4)}=1$, $\phi_2=\left.\frac{\partial\phi}{\partial\tau_2}\right|_{(\tau_i=\xi_i;\, i=1,2,3,4)}=-\frac{\xi_3-\xi_2\xi_1}{\xi_4-\left(\xi_2\right)^2}$,

$\phi_3=\left.\frac{\partial\phi}{\partial\tau_3}\right|_{(\tau_i=\xi_i;\, i=1,2,3,4)}=0$,

and $\phi_4=\left.\frac{\partial\phi}{\partial\tau_4}\right|_{(\tau_i=\xi_i;\, i=1,2,3,4)}=0$.

The linear form of the estimator $\bar{y}_{CST}$ using the Taylor Linearization technique for large *n* is

$$\bar{y}_{CST}=\overline{Y}+1\cdot\left[\tau_1-\xi_1\right]+\phi_2\left[\tau_2-\xi_2\right]$$

or $\bar{y}_{CST}=\overline{Y}+\left[\tau_1-\xi_1\right]+\phi_2\left[\tau_2-\xi_2\right]$

or $\bar{y}_{CST}-\overline{Y}=\left[\tau_1-\xi_1\right]+\phi_2\left[\tau_2-\xi_2\right]$

On squaring both sides and taking the expectation in both sides of the resulting expression, we get

$$\begin{aligned}
E\left[\bar{y}_{CST}-\overline{Y}\right]^2 &= E\left[\left(\tau_1-\xi_1\right)+\phi_2\left(\tau_2-\xi_2\right)\right]^2 \\
MSE\left[\bar{y}_{CST}\right] &= E\left[\left(\tau_1-\xi_1\right)^2+\phi_2^2\left(\tau_2-\xi_2\right)^2+2\phi_2\left(\tau_1-\xi_1\right)\left(\tau_2-\xi_2\right)\right] \\
MSE\left[\bar{y}_{CST}\right] &= E\left(\tau_1-\xi_1\right)^2+\phi_2^2 E\left(\tau_2-\xi_2\right)^2+2\phi_2 E\left(\tau_1-\xi_1\right)\left(\tau_2-\xi_2\right) \\
MSE\left[\bar{y}_{CST}\right] &= MSE\left[\tau_1\right]+\phi_2^2 V\left[\tau_2\right]+2\phi_2 E\left(\tau_1-\xi_1\right)\left(\tau_2-\xi_2\right)
\end{aligned} \tag{28}$$

where

$$\begin{aligned} MSE[\tau_1] &= \textstyle\sum_{h=1}^{L} w_h^2 \left[\frac{(1-f_h)}{n_h} S_{hY}^2 + \frac{(g_h-1)W_{h2}}{n_h} S_{hY(2)}^2 \right. \\ &\quad + R_h^2 \left(\frac{\overline{X}_h}{\overline{X}_h+\rho_h} \right)^2 \left\{ \frac{(1-f_h)}{n_h} S_{hX}^2 + \frac{(g_h-1)W_{h2}}{n_h} S_{hX(2)}^2 \right\} \\ &\quad \left. -2R_h \left(\frac{\overline{X}_h}{\overline{X}_h+\rho_h} \right) \left\{ \frac{(1-f_h)}{n_h} S_{hXY} + \frac{(g_h-1)W_{h2}}{n_h} S_{hXY(2)} \right\} \right] \\ V[\tau_2] = V[\bar{x}_{HH}] &= \textstyle\sum_{h=1}^{L} w_h^2 \left[\frac{(1-f_h)}{n_h} S_{hX}^2 + \frac{(g_h-1)W_{h2}}{n_h} S_{hX(2)}^2 \right] \\ E\left[\left\{\tau_1-\overline{Y}\right\}\left\{\tau_2-\overline{X}\right\}\right] &= \textstyle\sum_{h=1}^{L} w_h^2 \left[\left\{ \frac{(1-f_h)}{n_h} S_{hXY} + \frac{(g_h-1)W_{h2}}{n_h} S_{hXY(2)} \right\} \right. \\ &\quad \left. -R_h \left(\frac{\overline{X}_h}{\overline{X}_h+\rho_h} \right) \left\{ \frac{(1-f_h)}{n_h} S_{hX}^2 + \frac{(g_h-1)W_{h2}}{n_h} S_{hX(2)}^2 \right\} \right] \end{aligned}$$

Case 2: Let us consider the proposed traditional estimator given in **equation (25)**.

$$\bar{y}_{TY} = \sum_{h=1}^{L} w_h \left(\frac{n\overline{X}_h+\rho_h}{n\bar{x}_h^*+\rho_h} \right) \bar{y}_h^*$$

Let us proceed as in Case 1, the MSE of the estimator $\bar{y}_{TY}$ is given as

$$\begin{aligned} MSE(\bar{y}_{TY}) &= \textstyle\sum_{h=1}^{L} w_h^2 \left[\dfrac{(1-f_h)}{n_h} S_{hY}^2 + \dfrac{(g_h-1)W_{h2}}{n_h} S_{hY(2)}^2 \right. \\ &\quad + R_h^2 \left(\dfrac{n\overline{X}_h}{n\overline{X}_h+\rho_h} \right)^2 \left\{ \dfrac{(1-f_h)}{n_h} S_{hX}^2 + \dfrac{(g_h-1)W_{h2}}{n_h} S_{hX(2)}^2 \right\} \\ &\quad \left. -2R_h \left(\dfrac{n\overline{X}_h}{n\overline{X}_h+\rho_h} \right) \left\{ \dfrac{(1-f_h)}{n_h} S_{hXY} + \dfrac{(g_h-1)W_{h2}}{n_h} S_{hXY(2)} \right\} \right] \end{aligned} \tag{29}$$

Let us consider the proposed calibration estimator $\bar{y}_{CY}$ in **equation (26)**.

$$\bar{y}_{CY} = \sum_{h=1}^{L} w_h \bar{y}_h^* \left(\frac{n\overline{X}_h+\rho_h}{n\bar{x}_h^*+\rho_h} \right) + \hat{\beta}_Y \left[\overline{X} - \bar{x}_{HH}\right]$$

Let us take $q_{h1} = 1$ and $q_{h2} = 1$. The proposed estimator $\bar{y}_{CY}$ for these choices of q_{h1} and q_{h2} is

$$\bar{y}_{CY} = \sum_{h=1}^{L} w_h \bar{x}_h^* \left(\frac{n\overline{X}_h+\rho_h}{n\bar{x}_h^*+\rho_h} \right) + \frac{\left[\begin{array}{l} \displaystyle\sum_{h=1}^{L} w_h \left(\frac{n\overline{X}_h+\rho_h}{n\bar{x}_h^*+\rho_h} \right) \left\{ w_{h1}\bar{y}_{rh}\bar{x}_{rh} + w_{h2}\bar{y}_{sh}\bar{x}_{sh} \right\} \cdot 1 \\ \displaystyle -\bar{x}_{HH} \sum_{h=1}^{L} w_h \left(\frac{n\overline{X}_h+\rho_h}{n\bar{x}_h^*+\rho_h} \right) \bar{y}_h^* \end{array} \right]}{\left[\sum_{h=1}^{L} w_h \left\{ w_{h1}\bar{x}_{rh}^2 + w_{h2}\bar{x}_{sh}^2 \right\} \cdot 1 - \bar{x}_{HH}^2 \right]} \left[\overline{X} - \bar{x}_{HH}\right]$$

Let us express the proposed calibration estimator to obtain the MSE of the suggested estimator.

$$\bar{y}_{CY} = \varphi(\eta_1, \eta_2, \eta_3, \eta_4) = \eta_1 + \frac{\eta_3 - \eta_2\eta_1}{\eta_4 - (\eta_2)^2} \left[\overline{X} - \eta_2\right]$$

where $\eta_1 = \sum_{h=1}^{L} w_h \bar{y}_h^* \left(\frac{n\overline{X}_h+\rho_h}{n\bar{x}_h^*+\rho_h} \right)$, $\eta_2 = \bar{x}_{HH}$, $\eta_3 = \sum_{h=1}^{L} w_h \left(\frac{n\overline{X}_h+\rho_h}{n\bar{x}_h^*+\rho_h} \right) (w_{h1}\bar{x}_{rh}\bar{y}_{rh} + w_{h2}\bar{x}_{sh}\bar{y}_{sh})$ and $\eta_4 = \sum_{h=1}^{L} w_h (w_{h1}\bar{x}_{rh}^2 + w_{h2}\bar{x}_{sh}^2)$ are the estimators corresponding to parameters $\zeta_1 = \overline{Y}$, $\zeta_2 = \overline{X}$, $\zeta_3 = \sum_{h=1}^{L} w_h \overline{Y}_h \overline{X}_h$, and $\zeta_4 = \sum_{h=1}^{L} w_h \overline{X}_h^2$ respectively.

Proceeding in a similar manner as Case 1, the *MSE* of the estimator $\bar{y}_{CY}$ is derived as:

$$MSE[\bar{y}_{CY}] = MSE[\eta_1] + \varphi_2^2 V[\bar{x}_{HH}] + 2\varphi_2 E\left[\left\{\eta_1 - \overline{Y}\right\}\left\{\bar{x}_{HH} - \overline{X}\right\}\right] \tag{30}$$

where

$$\begin{aligned}
MSE[\eta_1] &= \sum_{h=1}^{L} w_h^2 \left[\frac{(1-f_h)}{n_h} S_{hY}^2 + \frac{(g_h-1)W_{h2}}{n_h} S_{hY(2)}^2 \right. \\
&\quad + R_h^2 \left(\frac{n\overline{X}_h}{n\overline{X}_h+\rho_h} \right)^2 \left\{ \frac{(1-f_h)}{n_h} S_{hX}^2 + \frac{(g_h-1)W_{h2}}{n_h} S_{hX(2)}^2 \right\} \\
&\quad \left. -2R_h \left(\frac{n\overline{X}_h}{n\overline{X}_h+\rho_h} \right) \left\{ \frac{(1-f_h)}{n_h} S_{hXY} + \frac{(g_h-1)W_{h2}}{n_h} S_{hXY(2)} \right\} \right] \\
V[\bar{x}_{HH}] &= \sum_{h=1}^{L} w_h^2 \left[\frac{(1-f_h)}{n_h} S_{hX}^2 + \frac{(g_h-1)W_{h2}}{n_h} S_{hX(2)}^2 \right] \\
E\left[\left\{ \eta_1 - \overline{Y} \right\} \left\{ \overline{x}_{HH} - \overline{X} \right\} \right] &= \sum_{h=1}^{L} w_h^2 \left[\left\{ \frac{(1-f_h)}{n_h} S_{hXY} + \frac{(g_h-1)W_{h2}}{n_h} S_{hXY(2)} \right\} \right. \\
&\quad \left. -R_h \left(\frac{n\overline{X}_h}{n\overline{X}_h+\rho_h} \right) \left\{ \frac{(1-f_h)}{n_h} S_{hX}^2 + \frac{(g_h-1)W_{h2}}{n_h} S_{hX(2)}^2 \right\} \right] \\
\varphi_2 &= -\frac{\zeta_3-\zeta_2\zeta_1}{\zeta_4-(\zeta_2)^2}
\end{aligned}$$

3 Result and Discussion

We conducted a simulation-based numerical study using hypothetically generated data to assess the effectiveness of the suggested calibration estimators. We follow the five steps considered in [12]. Here, we have generated two populations based on different characteristics in the R software.

3.1 Population I

We have generated a population comprising 26,000 individuals for four strata. A stratified sample of size $n = 5200$ is drawn. The characteristics of the generated population are shown in Table 1.

Table 1. Description of Generated Population

Stratum No.	Stratum Size(N_h)	Distribution of Study Variable(y_{hi}) $N(\mu_{hy},\sigma_h^2)$	Distribution of auxiliary variable(x_{hi}) $N(\mu_{hx},\sigma_h^2)$	Correlation between variablesy_{hi}andx_{hi}
1	12000	$N(110, 81)$	$N(200, 81)$	0.85
2	5000	$N(150,\ 100)$	$N(240,\ 100)$	0.75
3	2000	$N(180,\ 121)$	$N(280,\ 121)$	0.60
4	7000	$N(130,\ 64)$	$N(320,\ 64)$	0.75

3.2 Population II

We generated a second population of size 14,400 individuals across four strata. The drawn sample size is $n = 2880$. The characteristics of the population are shown in Table 2.

Table 2. Description of Generated Population

Stratum No.	Stratum Size(N_h)	Distribution of Study Variable(y_{hi}) $N(\mu_{hy},\sigma_h^2)$	Distribution of auxiliary variable(x_{hi}) $N(\mu_{hx},\sigma_h^2)$	Correlation between variablesy_{hi}andx_{hi}
1	6000	$N(170, 12)$	$N(300, 12)$	0.84
2	4500	$N(150,\ 10)$	$N(240,\ 10)$	0.78
3	1200	$N(130,\ 7)$	$N(270,\ 7)$	0.70
4	2700	$N(190,\ 9)$	$N(210,\ 9)$	0.85

We draw 1000 samples from both populations. Using the observations on n_{h1} responding and the n_{sh} sub-sampled units, calculate the $\left[\bar{y}_{HH}-\bar{Y}\right]^2$, $\left[\bar{y}_{Lee}-\bar{Y}\right]^2$, $\left[\bar{y}_{TST}-\bar{Y}\right]^2$, $\left[\bar{y}_{TY}-\bar{Y}\right]^2$, $\left[\bar{y}_{CST}-\bar{Y}\right]^2$ and $\left[\bar{y}_{CY}-\bar{Y}\right]^2$ corresponding to each sample. We average out the results obtained from each sample. The mean squared errors (*MSEs)* of estimators $\bar{y}_{HH}$, $\bar{y}_{Lee}$, $\bar{y}_{TST}$, $\bar{y}_{TY}$, $\bar{y}_{CST}$ and $\bar{y}_{CY}$ and the percentage relative efficiencies (*PREs)* of estimators $\bar{y}_{Lee}$, $\bar{y}_{CST}$, $\bar{y}_{TST}$, $\bar{y}_{TY}$ and $\bar{y}_{CY}$ with respect to $\bar{y}_{HH}$ for various values of $g_h = \frac{n_{h2}}{s_h}$ and non-response rates are shown in Table 3 and Table 4, respectively, for both populations.

Table 3. *MSEs of* estimators $\bar{y}_{HH}, \bar{y}_{Lee}, \bar{y}_{TST}, \bar{y}_{TY}, \bar{y}_{CST}$ and $\bar{y}_{CY}$

Population	g_h	Non-Response Rate (in %)	$MSE(\bar{y}_{HH})$	$MSE(\bar{y}_{Lee})$	$MSE(\bar{y}_{TST})$	$MSE(\bar{y}_{TY})$	$MSE(\bar{y}_{CST})$	$MSE(\bar{y}_{CY})$
		10	0.0147	0.0092	0.0071	0.0071	0.0061	0.0061
	2	20	0.0157	0.0097	0.0074	0.0074	0.0066	0.0066
		30	0.0170	0.0104	0.0079	0.0079	0.0069	0.0069
		10	0.0144	0.0091	0.0072	0.0072	0.0066	0.0066
I	3	20	0.0189	0.0116	0.0090	0.0090	0.0081	0.0081
		30	0.0245	0.0150	0.0114	0.0114	0.0096	0.0096
		10	0.0171	0.0106	0.0083	0.0083	0.0074	0.0074
	4	20	0.0218	0.0134	0.0102	0.0102	0.0089	0.0089
		30	0.0284	0.0174	0.0132	0.0132	0.0112	0.0112
		10	0.0340	0.0227	0.01325	0.01325	0.01157	0.01157
	2	20	0.0412	0.0275	0.01563	0.01563	0.01308	0.01308
		30	0.0464	0.0313	0.01753	0.01753	0.01430	0.01430
		10	0.0381	0.0257	0.01507	0.01507	0.01283	0.01283
II	3	20	0.0458	0.0305	0.01743	0.01743	0.01473	0.01473
		30	0.0557	0.0379	0.02221	0.02221	0.01859	0.01859
		10	0.0423	0.0290	0.01724	0.01724	0.01465	0.01465
	4	20	0.0514	0.0348	0.02103	0.02103	0.01838	0.01838
		30	0.0662	0.0457	0.02791	0.02791	0.02368	0.02368

Table 4. *PREs* of $\bar{y}_{Lee}, \bar{y}_{CST}, \bar{y}_{TST}, \bar{y}_{TY}$ and $\bar{y}_{CY}$ with respect to $\bar{y}_{HH}$

Population	g_h	Non-Response Rate (in %)	$PRE(\bar{y}_{Lee})$	$PRE(\bar{y}_{TST})$	$PRE(\bar{y}_{TY})$	$PRE(\bar{y}_{CST})$	$PRE(\bar{y}_{CY})$
		10	160.2	208.1	208.1	243.3	243.3
	2	20	161.9	211.6	211.6	237.2	237.2
		30	164.4	215.6	215.7	247.7	247.7
		10	158.9	200.5	200.5	217.9	217.9
I	3	20	162.2	209.9	209.9	231.4	231.4
		30	163.4	214.8	214.8	254.4	254.4
		10	161.6	206.4	206.4	232.1	232.1
	4	20	163.3	214.6	214.6	244.1	244.1
		30	163.6	214.7	214.7	253.3	253.3
		10	149.7	256.5	256.5	293.9	293.9
	2	20	149.6	263.3	263.3	314.7	314.7
		30	148.3	264.7	264.7	324.5	324.5
		10	147.9	252.5	252.5	296.6	296.6
II	3	20	150.0	262.6	262.6	310.7	310.7
		30	147.1	250.9	250.9	299.8	299.8
		10	145.8	245.4	245.4	288.8	288.8
	4	20	147.7	244.3	244.3	279.7	279.7
		30	144.9	237.2	237.2	279.6	279.6

This study was conducted with different non-response rates (10%, 20%, and 30%) and varying values of g_h (2, 3, and 4). From Table 3, you can observe that the values of MSEs decrease as the non-response rate and g_h decrease for both populations. The percentage relative efficiencies (%PRE) of the proposed estimators $\bar{y}_{TST}, \bar{y}_{TY}, \bar{y}_{CST} and \bar{y}_{CY}$ vary from 200.513% to 215.641%, 200.524% to 215.654%, 217.910% to 247.667% and 217.904% to 247.662%, respectively, while 158.931% to 164.414% for the existing estimator $\bar{y}_{Lee}$ with respect to the estimator $\bar{y}_{HH}$, as shown in Table 4 for Population I. Similarly, the percentage relative efficiencies (%PRE) of the estimators with respect to $\bar{y}_{HH}$ those depicted in Table 4 vary from 144.9% to 148.3%, 237.2% to 264.7 %, 237.3 % to 264.7 %, 279.6% to 324.5% and 279.6% to 324.5% for the existing estimator $\bar{y}_{Lee}$, and the suggested estimators $\bar{y}_{TST}, \bar{y}_{TY}, \bar{y}_{CST} and \bar{y}_{CY}$, respectively, for Population II.

4 Conclusion

This research introduces a set of estimators for the population mean that address non-response by leveraging an auxiliary variable, employing both conventional and calibration-based approaches. Variance expressions for these estimators were derived using Taylor linearization. Furthermore, a simulation study was carried out to evaluate their performance and validate their efficiency.

Hence, it demonstrates that the proposed estimators using the traditional method ($\bar{y}_{TST}$ and $\bar{y}_{TY}$) outperform the $\bar{y}_{HH}$ and $\bar{y}_{Lee}$. The estimators ($\bar{y}_{CST}$ and $\bar{y}_{CY}$) obtained by applying the calibration approach to traditional estimators also perform better than the existing estimators $\bar{y}_{HH}$ and $\bar{y}_{Lee}$ for Populations I and II in the presence of non-response under stratified random sampling. The proposed estimators can be generalized to a multi-auxiliary setup under non-response as well as under different sampling designs. This study can be further extended in the case of two-phase sampling when the population mean of the auxiliary variable is not known in advance. Thus, we recommend the proposed study for its use in practice.

5 Data availability statement

We have generated a population-I comprising 26,000 individuals for four strata. A stratified sample of size $n = 5200$ is drawn. The characteristics of the generated population are shown in Table 1. And generated a second population-II of size 14,400 individuals across four strata. The drawn sample size is $n = 2880$. The characteristics of the population are shown in Table 2.

6 Details of Ethical Clearance

We understand the importance of ethical clearance. Since in this study we do not involve any living being or societally entity. This is a pure fundamental development in the area of parameter estimation in a sample survey. And related to data, we have mentioned in the data availability statement section.

References

1) Hansen MH, Hurwitz WN. The problem of non-response in sample surveys. *Journal of the American Statistical Association*. 1946;41(236):517–529. https://doi.org/10.1080/01621459.1946.10501894.
2) Audu A, Aphane M, Ahmad J, Singh RV. On the calibration estimators of finite population proportion under remainder systematic sampling. *Quality & Quantity*. 2025;p. 1–7. https://doi.org/10.1007/s11135-025-02091-0.
3) Dykes L, Singh S, Sedory SA, Louis V. Calibrated estimators of population mean for a mail survey design. *Communications in Statistics-Theory and Methods*. 2015;44(16):3403–3427. https://doi.org/10.1080/03610926.2013.841932.
4) Pandey MK, Singh GN, Bandyopadhyay A. Quantile regression-ratio-type estimator of population mean in stratified successive sampling using calibrated weights. *Mathematical Population Studies*. 2025;p. 1–43. https://doi.org/10.1080/08898480.2025.2536773.
5) team P. Integrated Development Environment for R. *Posit Software, PBC, Boston, MA*. 2024. Available from: http://www.posit.co/.
6) Team RC. A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria*. 2020. Available from: https://www.R-project.org/.
7) Yadav SK, Dixit MK, Dungana HN, Mishra SS. Improved estimators for estimating average yield using auxiliary variable. *International Journal of Mathematical, Engineering and Management Sciences*. 2019;4(5):1228. https://doi.org/10.33889/IJMEMS.2019.4.5-097.
8) Singh R, Khare S, Khare BB, Jha PS. The General Classes of Estimators for Population Mean Under Stratified Two Phase Random Sampling in The Presence of Non-Response. *International Journal of Agricultural & Statistical Sciences*. 2020;16(2). Available from: https://connectjournals.com/03899.2020.16.557.
9) Singh GN, Pandey MK, Bandyopadhyay A, Meetei MZ, Zaagan AA, Mahnashi AM, et al. Estimation of Population Mean Using Calibrated Weights in Stratified Random Successive Sampling in Presence of Incomplete Data. *Journal of Mathematics*. 2025;2025(1):6778010. https://doi.org/10.1155/jom/6778010.
10) Garg N, Patel A, Pachori M. Calibration estimation of population proportion in probability proportional to size sampling in the presence of non-response. *Indian Journal of Science and Technology*. 2024;17(9):841–851. https://doi.org/10.17485/IJST/v17i9.2938.
11) Singh HP, Tailor R. Use of known correlation coefficient in estimating the finite population mean. *Statistics in transition*. 2003;6(4):555–560. https://doi.org/10.6092/issn.1973-2201/1300.
12) Chaudhary MK, Ray BK, Vishwakarma GK, Kadilar C. A calibration-based approach on estimation of mean of a stratified population in the presence of non response. *Communications in Statistics-Theory and Methods*. 2024;53(19):7054–7068. https://doi.org/10.1080/03610926.2023.2257818.