STATISTICS IN TRANSITION new series, December 2026
Vol. 27, No. 4, pp. xx–XX, https://doi.org/10.59139/stattrans-2026-xxx
Received – xx.xx.xxxx; accepted – xx.xx.xxxx

Estimation of Finite Population Mean in Survey Sampling with Improved Ratio Type Imputation Strategies for Non-Response

Deepjan Gohain[1], **Tarun Kumari**[2], **B. K. Singh**[3]

Abstract

In the cases where the research variable shows non-response, this work presents four efficient chain ratio-type imputation techniques and their corresponding point estimators of finite population mean under simple random sampling and double sampling schemes. The bias of the proposed estimators, MSE, and minimum MSE are calculated in terms of population parameters up to first-order approximations. It is shown that under the above two schemes, the proposed estimators outperform the conventional ones in terms of efficiency. The theoretical results are supported by empirical study as well.

Key words: chain ratio estimator, imputation, bias, mean square error (MSE), simple random sampling, double sampling.

1. Introduction

Population parameters are generally calculated by using a set of suitable statistics, known as an estimator. For example, the best-suited and unbiased estimator of the population mean is the sample mean. The sample mean, however, has a large amount of variation, leading to the search for an estimator with a lesser MSE than the sample mean. Such cases are dealt with by the use of the auxiliary variable X which exhibits a high correlation with the variable Y considered for study. Cochran (1940) introduced an efficient population mean estimator utilizing data from a strongly correlated population variable which is biased but has lesser MSE than the sample mean known as the classical ratio estimator. Using Simple Random Sampling, Srivenkataramana (1980), Bahl and Tuteja (1991), and many others extended the technique further. The population mean of the auxiliary variable is assumed to be known in advance, in each of these scenarios. In situations where this information is not available, the double sampling method also known as the two-phase method suggested by Neyman (1938) comes into the picture. Two-phase sampling method has been utilized in the works by Kumar and Bahl (2006), Singh and Vishwakarma (2007), Kalita et al. (2013), Shukla (2019) and many others. Sometimes it is seen that the variable considered for the study

[1]Dept of Statistics, Lakhimpur Commerce College, Assam, India. E-mail: deepjangohain1982@gmail.com ORCID: https://orcid.org/0000-0003-2211-268X.
[2]Dept of Mathematics, Lakhimpur Commerce College, Assam, India. E-mail: tarun.nlp@gmail.com. ORCID: https://orcid.org/0009-0000-6322-6494.
[3]Dept of Mathematics, NERIST, Arunachal Pradesh, India. E-mail: bksinghnerist@gmail.com.

© D. Gohain, T. Kumari, B. K. Singh. Article available under the CC BY-SA 4.0 licence

is correlated with more than one variable and in scenarios like this, two different auxiliary variables, viz., X and Z, are considered for the estimation of the finite population mean of the study variable. Chand (1975) introduced a method of chaining the information of two auxiliary variables with the study variable under the double sampling method. In similar lines work has been done by Singh and Choudhury (2012), Kalita and Singh (2013), Singh and Majhi (2014), Singh et al. (2015) and others. The methods described above are used when the information of all the variables X, Y and Z is available in the sample. Occasionally, in some real-world scenarios, unit information is unavailable because the units within the sample(s) did not respond.

The technique used for handling missing data in sample surveys is called Imputation. It is one of the most commonly used scientific techniques to replace such missing information with the information available in the sample. In 1976, Rubin explained that missing data can follow different patterns. One common type is "missing at random" (MAR), where the chance of data being missing depends on the information we already have, but not on the actual missing values themselves. Later, Heitjan and Basu (1996) clearly explained the difference between MAR and "missing completely at random" (MCAR), where missing data is unrelated to either the observed or missing values. Researchers have also developed methods to handle missing data by using extra information called auxiliary variables. For example, Lee et al. (1994, 1995) used these variables to improve estimates, and Singh and Horn (2000) suggested combining them with average-based methods. Ahmed et al. (2006) tested different imputation methods and compared how well they worked. Similarly, Rao and Sitter (1995) applied these methods to two-phase sampling, and Kadilar and Cingi (2008) as well as Diana and Perri (2010) proposed other useful techniques.

In this study, we assume that the missing data is MCAR. Some well-known imputation techniques such as the mean, ratio, and compromised methods are used to estimate the finite population mean of the variable under study. Researchers have developed various other improved methods of imputation in Upadhyaya and Singh (1999), Ahmed et al. (2006), Shukla and Thakur (2008), Kadilar and Cingi (2008), Diana and Perri (2010) and Gogoi and Singh (2017), Gogoi and Singh (2018), Gogoi and Singh (2019), Nath and Singh (2018), Nath et al. (2020). More specifically we consider the imputation methods suggested by Pandey et al. (2021), Pandey et al. (2025) and Thakur and Shukla (2022).

In the context of simple random sampling without replacement setup and two-phase sampling setup, the goal of this article is to propose more effective and superior estimators than those that already exist.

2. Notations

Let $\Omega = \{P_1, P_2, \ldots P_N\}$ be a population of size N where Y is the variable considered for the study and X & Z are the auxiliary variables. It is assumed that the information on the study variable is missing at random in the samples and the information on the auxiliary variables is available in the samples. The variables Y, X and Z have population means $\overline{Y}$, $\overline{X}$ and $\overline{Z}$ respectively where $\overline{X}$ is unknown.

2.1. Simple Random Sampling Scheme

A sample S of size $n(n<N)$ is drawn from the population Ω. Let the number of responding units of Y in S be $r(r<n)$, denoted by subset A and the number of non-responding units be $(n-r)$, denoted by A^c.
Let us consider the following list of notations as described in Table 1, to be used henceforth.

Table 1. List of Notations

Notation	Description
$\bar{y}_r$ and $\bar{x}_r$	sample mean of the responding units of Y and X respectively in A
$\bar{x}_n$ and $\bar{z}_n$	sample mean of the units of X and Z respectively in S
S_Y^2, S_X^2, S_Z^2	population mean square of the units of Y, X and Z respectively
ρ_{YX}, ρ_{YZ} & ρ_{XZ}	population correlation coefficient of Y and X, Y and Z & X and Z respectively
C_Y, C_X and C_Z	$\frac{S_Y}{\bar{Y}}$, $\frac{S_X}{\bar{X}}$ and $\frac{S_Z}{\bar{Z}}$ respectively

2.2. Double Sampling Scheme

A sample S_1 of size n_1 is drawn from the population Ω by using the SRSWOR method, where the unknown population mean of the auxiliary variable X is estimated. Secondly, a sample S of size $n(n<n_1)$ is selected from S_1 by using the SRSWOR method which has r responding units and $(n-r)$ non-responding units of the study variable Y only denoted by A and A^c such that $A \cup A^c = S$ and $A \cap A^c = \phi$.

Table 2 describes the list of notations to be used under double sampling.

Table 2. List of Notations

Notation	Description
$\bar{y}_r$ and $\bar{x}_r$	sample means of the responding units of Y and X respectively, in A
$\bar{x}_n$ and $\bar{z}_n$	sample means of the units of X and Z respectively, in S
$\bar{x}_1$ and $\bar{z}_1$	sample means of the units of X and Z respectively, in S_1
S_Y^2, S_X^2, S_Z^2	population mean square of the units of Y, X and Z respectively
ρ_{YX}	population correlation coefficient of Y and X
ρ_{YZ}	population correlation coefficient of Y and Z
ρ_{XZ}	population correlation coefficient of X and Z
C_Y, C_X and C_Z	$\frac{S_Y}{\bar{Y}}$, $\frac{S_X}{\bar{X}}$ and $\frac{S_Z}{\bar{Z}}$ respectively

3. Proposed method of Imputation in Simple Random Sampling

Ahmed et al. (2006) proposed a new imputation approach under simple random sampling that does not employ $\overline{X}$, as explained below.

$$y_{.i} = \begin{cases} y_i & \text{if } i \in A \\ \frac{\bar{y}_r}{n-r}\left\{n\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\kappa} - r\right\} & \text{if } i \in A^c \end{cases}$$

The corresponding point estimator is

$$t = \overline{y}_r \left(\frac{\overline{x}_n}{\overline{x}_r} \right)^{\kappa}$$

where κ, the constant is considered in a way that yields a minimum MSE of the point estimator.

Inspired by the works of Lee et al. (1994) and Ahmed et al. (2006), we proposed the following ratio chain type imputation method under simple random sampling.

$$y_{.i}^{*} = \begin{cases} y_i & \text{if } i \in A \\ \frac{\overline{y}_r}{n-r} \left\{ n(\frac{\overline{x}_n}{\overline{x}_r})^{\alpha} (\frac{\overline{Z}}{\overline{z}_n})^{\beta} - r \right\} & \text{if } i \in A^c \end{cases}$$

The point estimator of the population mean $\overline{Y}$ obtained by this method is

$$t^{*} = \overline{y}_r \left(\frac{\overline{x}_n}{\overline{x}_r} \right)^{\alpha} \left(\frac{\overline{Z}}{\overline{z}_n} \right)^{\beta}$$

where α and β are constants, considered in a way that yields a minimum mean square error of the corresponding point estimator.

4. Properties of the proposed point estimator t^{*}

To determine the characteristics of the proposed point estimator, we utilize the following transformations.

$\overline{y}_r = \overline{Y}(1+e_0) \qquad \overline{x}_r = \overline{X}(1+e_1)$
$\overline{x}_n = \overline{X}(1+e_2) \qquad \overline{z}_n = \overline{Z}(1+e_3)$, where $e_i \in (0,1)$ for $i = 1,2,3$
$E(e_i) = 0$ for $i = 0,1,2,3$
$E(e_0^2) = \delta_1 C_Y^2 \qquad E(e_1^2) = \delta_1 C_X^2 \qquad E(e_2^2) = \delta_2 C_X^2 \qquad E(e_3^2) = \delta_2 C_Z^2$
$E(e_0 e_1) = \delta_1 \rho_{YX} C_Y C_X \qquad E(e_0 e_2) = \delta_2 \rho_{YX} C_Y C_X \qquad E(e_0 e_3) = \delta_2 \rho_{YZ} C_Y C_Z$
$E(e_1 e_2) = \delta_2 C_X^2 \qquad E(e_1 e_3) = \delta_2 \rho_{ZX} C_Z C_X \qquad E(e_2 e_3) = \delta_2 \rho_{ZX} C_Z C_X$
$\delta_1 = \frac{1}{r} - \frac{1}{N} \qquad \delta_2 = \frac{1}{n} - \frac{1}{N} \qquad \delta_4 = \frac{1}{r} - \frac{1}{n}$

Theorem 4.1. *The point estimator t^{*} represented in terms of e_i's $(i \in \{0,1,2,3\})$ is*

$$\begin{aligned} t^{*} - \overline{Y} = \overline{Y} \Bigg[& e_0 - \alpha(e_1 - e_2 + e_0 e_1 - e_0 e_2) - \beta(e_0 e_3 + e_3) + \frac{\alpha^2}{2}(e_1 - e_2)^2 \\ & + \frac{\alpha}{2}(e_1^2 - e_2^2) + \frac{(\beta^2 + \beta)}{2} e_3^2 + \alpha\beta(e_1 e_3 - e_2 e_3) \Bigg] \end{aligned} \tag{1}$$

Proof.

$$\begin{aligned}
t^* &= \bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\alpha}\left(\frac{\overline{Z}}{\bar{z}_n}\right)^{\beta} \\
&= \overline{Y}(1+e_0)(1+e_2)^{\alpha}(1+e_1)^{-\alpha}(1+e_3)^{-\beta} \\
&= \overline{Y}\Big\{1+e_0-\alpha(e_1-e_2)-\beta e_3+\frac{\alpha(\alpha+1)}{2}e_1^2+\frac{\beta(\beta+1)}{2}e_3^2 \\
&\quad +\frac{\alpha(\alpha-1)}{2}e_2^2+\alpha\beta e_1e_3-\alpha\beta e_2e_3-\alpha^2e_1e_2-\beta e_0e_3-\alpha e_0e_1+\alpha e_0e_2\Big\}
\end{aligned}$$

$$\begin{aligned}
t^*-\overline{Y} = \overline{Y}\Big\{&e_0-\alpha(e_1-e_2+e_0e_1-e_0e_2)-\beta(e_0e_3+e_3)+\frac{\alpha^2}{2}(e_1-e_2)^2+\frac{\alpha}{2}(e_1^2-e_2^2) \\
&+\frac{(\beta^2+\beta)}{2}e_3^2+\alpha\beta(e_1e_3-e_2e_3)\Big\}
\end{aligned}$$

□

Theorem 4.2. *The Bias of t^* is*

$$B(t^*) = \overline{Y}\Big[\frac{\alpha^2+\alpha}{2}\delta_4C_X^2+\frac{\beta^2+\beta}{2}\delta_2C_Z^2-\alpha\delta_4\rho_{YX}C_YC_X-\beta\delta_2\rho_{YZ}C_YC_Z\Big] \qquad (2)$$

Proof.

$$\begin{aligned}
B(t^*) &= E(t^*-\overline{Y}) \\
&= \overline{Y}\Big\{\alpha E(e_0e_2-e_0e_1)-\beta E(e_0e_3)+\frac{\alpha^2}{2}E(e_1-e_2)^2 \\
&\quad +\frac{\alpha}{2}E(e_1^2-e_2^2)+\frac{\alpha}{2}E(e_1^2-e_2^2)+\frac{\beta^2+\beta}{2}E(e_3^2)+\alpha\beta E(e_1e_3)\Big\} \\
&= \overline{Y}\Big\{\frac{\alpha^2+\alpha}{2}\delta_1C_X^2+\frac{\alpha^2-\alpha}{2}\delta_2C_X^2+\frac{\beta^2+\beta}{2}\delta_2C_Z^2-\alpha^2\delta_2C_X^2 \\
&\quad -\alpha\delta_1\rho_{YX}C_YC_X+\alpha\delta_2\rho_{YX}C_YC_X-\beta\delta_2\rho_{YZ}C_YC_Z\Big\} \\
&= \overline{Y}\Big\{\frac{\alpha^2+\alpha}{2}(\delta_1-\delta_2)C_X^2+\frac{\beta^2+\beta}{2}\delta_2C_Z^2-\alpha(\delta_1-\delta_2)\rho_{YX}C_YC_X-\beta\delta_2\rho_{YZ}C_YC_Z\Big\} \\
&= \overline{Y}\Big\{\frac{\alpha^2+\alpha}{2}\delta_4C_X^2+\frac{\beta^2+\beta}{2}\delta_2C_Z^2-\alpha\delta_4\rho_{YX}C_YC_X-\beta\delta_2\rho_{YZ}C_YC_Z\Big\}
\end{aligned}$$

□

Theorem 4.3. *The MSE of t^* is*

$$M(t^*) = \overline{Y}^2\Big[\delta_1C_Y^2+\alpha^2\delta_4C_X^2+\beta^2\delta_2C_Z^2-2\alpha\delta_4\rho_{YX}C_YC_X-2\beta\delta_2\rho_{YZ}C_YC_Z\Big] \qquad (3)$$

Proof. Squaring both sides of Equation (1), and taking expectations we have

$$\begin{aligned} M(t^*) &= \overline{Y}^2 E\{e_0^2 + \alpha^2(e_1^2 + e_2^2 - e_1e_2) + \beta^2 e_3^2 - 2\alpha(e_0e_1 - e_0e_2) - 2\beta e_0e_3 \\ &\qquad + 2\alpha\beta(e_1e_3 - e_2e_3)\} \\ &= \overline{Y}^2\{\delta_1 C_Y^2 + \alpha^2(\delta_1 C_X^2 + \delta_2 C_X^2 - 2\delta_2 C_X^2) + \beta^2\delta_2 C_Z^2 - 2\alpha(\delta_1 - \delta_2)\rho_{YX}C_YC_X \\ &\qquad - 2\beta\delta_2\rho_{YZ}C_YC_Z + 2\alpha\beta(\delta_2\rho_{XZ}C_XC_Z - \delta_2\rho_{XZ}C_XC_Z)\} \\ &= \overline{Y}^2\left\{\delta_1 C_Y^2 + \alpha^2\delta_4 C_X^2 + \beta^2\delta_2 C_Z^2 - 2\alpha\delta_4\rho_{YX}C_YC_X - 2\beta\delta_2\rho_{YZ}C_YC_Z\right\} \end{aligned}$$

□

Theorem 4.4. *The minimum MSE of t^* is*

$$M(t^*)_{min} = \overline{Y}^2\left[\delta_1 C_Y^2 - \delta_4\rho_{YX}^2 C_Y^2 - \delta_2\rho_{YZ}^2 C_Y^2\right] \tag{4}$$

Proof. Using the principle of maxima and minima in Equation (3), we get

$$\begin{aligned} \alpha C_X - \rho_{YX}C_Y &= 0 \\ \beta C_Z - \rho_{YZ}C_Y &= 0 \end{aligned}$$

Solving the above two equations we get $\alpha_{opt} = \frac{\rho_{YX}C_Y}{C_X}$ and $\beta_{opt} = \frac{\rho_{YZ}C_Y}{C_Z}$.

Thus, putting $\alpha = \alpha_{opt}$ and $\beta = \beta_{opt}$ in Equation (3) we get

$$M(t^*)_{min} = \overline{Y}^2\left\{\delta_1 C_Y^2 - \delta_4\rho_{YX}^2 C_Y^2 - \delta_2\rho_{YZ}^2 C_Y^2\right\}$$

□

5. Remarks on Practicability of the Estimator t^*

For practical use of the estimator t^* proposed above, the population parameters $\rho_{YX}, \rho_{YZ}, C_Y$ and C_X are assumed to be known. In the case of unavailability of the above data, we have to replace α_{opt} and β_{opt} by $\hat{\alpha}_{opt}$ and $\hat{\beta}_{opt}$ respectively, where, $\hat{\alpha}_{opt} = \frac{\rho_{yx}C_y}{C_x}$ and $\hat{\beta}_{opt} = \frac{\rho_{yz}C_y}{C_z}$. Note that, $\rho_{yx}, \rho_{yz}, C_y, C_x$ and C_z are calculated from the sample units of the responding unit A of the sample S.

6. Particular Cases

A. For $\alpha = \kappa$ and $\beta = 0$, the point estimator t^* transforms into the point estimator suggested by Ahmed et al. (2006).

$$t^* = \overline{y}_r\left(\frac{\overline{x}_n}{\overline{x}_r}\right)^{\kappa}\left(\frac{\overline{Z}}{\overline{z}_n}\right)^{0} = \overline{y}_r\left(\frac{\overline{x}_n}{\overline{x}_r}\right)^{\kappa} = t$$

B. For $\alpha = 1$ and $\beta = 0$, the point estimator t^* transforms into $\overline{y}_{RAT}$.

$$t^* = \overline{y}_r \left(\frac{\overline{x}_n}{\overline{x}_r}\right)^1 \left(\frac{\overline{Z}}{\overline{z}_n}\right)^0 = \overline{y}_r \left(\frac{\overline{x}_n}{\overline{x}_r}\right) = \overline{y}_{RAT}$$

C. For $\alpha = \alpha^*$ and $\beta = 1$, the point estimator t^* transforms into the following ratio-type point estimator

$$t^* = \overline{y}_r \left(\frac{\overline{x}_n}{\overline{x}_r}\right)^{\alpha^*} \left(\frac{\overline{Z}}{\overline{z}_n}\right) = t_0^* \text{ (say)}$$

D. For $\alpha = 0$ and $\beta = \beta^*$, the point estimator t^* transforms into the following ratio-type point estimator

$$t^* = \overline{y}_r \left(\frac{\overline{x}_n}{\overline{x}_r}\right) \left(\frac{\overline{Z}}{\overline{z}_n}\right)^{\beta^*} = t_0^{**} \text{ (say)}$$

E. For $\alpha = 0$ and $\beta = 0$, the point estimator t^* transforms into $\overline{y}_m$

$$t^* = \overline{y}_r \left(\frac{\overline{x}_n}{\overline{x}_r}\right)^0 \left(\frac{\overline{Z}}{\overline{z}_n}\right)^0 = \overline{y}_r = \overline{y}_m$$

Theorem 6.1. *The Bias and MSE of t are*

$$B(t) = \overline{Y}\left\{ (\frac{\kappa^2 + \kappa}{2})\delta_4 C_X^2 - \kappa\delta_4 \rho_{YX} C_Y C_X \right\} \tag{5}$$

$$M(t) = \overline{Y}^2\left\{ \delta_1 C_Y^2 + \kappa^2\delta_4 C_X^2 - 2\kappa\delta_4 \rho_{YX} C_Y C_X \right\} \tag{6}$$

$$M(t)_{min} = \overline{Y}^2\left\{ \delta_1 C_Y^2 - \delta_4 \rho_{YX}^2 C_Y^2 \right\} \tag{7}$$

Proof. $B(t)$ and $M(t)$ can be obtained by putting $\alpha = \kappa$ and $\beta = 0$ in Equations (2) and (3).

Again, equating the differentiation of Equation (6) w.r.t κ to zero we get $\kappa_{opt} = \rho_{YX}\frac{C_Y}{C_X}$. Putting $\kappa = \kappa_{opt}$ in Equation (6) we get

$$M(t)_{min} = \overline{Y}^2\left\{ \delta_1 C_Y^2 - \delta_4 \rho_{YX}^2 C_Y^2 \right\}$$

□

7. Efficiency Comparison: Theoretical

7.1. Comparison with the mean method of imputation

In this method, the sample units in S are defined as

$$y_{.i} = \begin{cases} y_i, & \text{if } i \in A \\ \overline{y}_r, & \text{if } i \in A^c \end{cases}$$

Using the above description, the point estimator of $\overline{Y}$ is given by

$$\overline{y}_m = \frac{1}{n}\sum_{i\in S} y_{.i} = \overline{y}_r$$

where

$$\overline{y}_r = \frac{1}{r}\sum_{i\in A} y_{.i}$$

The Bias and the variance of $\overline{y}_m$ are given by

$$B(\overline{y}_m) = 0 \tag{8}$$

$$V(\overline{y}_m) = \overline{Y}^2\delta_1 C_Y^2 \tag{9}$$

From Equations (4) and (9) we have

$$V(\overline{y}_m) - M(t^*)_{min} = \overline{Y}^2\{\delta_4\rho_{YX}^2 C_Y^2 + \delta_2\rho_{YZ}^2 C_Y^2\} > 0$$

Thus, t^* is more efficient than $\overline{y}_m$.

7.2. Comparison with the ratio method of imputation

In this method, the sample units in S are defined as

$$y_{.i} = \begin{cases} y_i, & \text{if } i \in A \\ \hat{b}x_i, & \text{if } i \in A^c \end{cases}$$

where

$$\hat{b} = \frac{\sum_{i\in A} y_i}{\sum_{i\in A} x_i}$$

Using the above description, the point estimator of $\overline{Y}$ is given by,

$$\overline{y}_{RAT} = \overline{y}_r\frac{\overline{x}_n}{\overline{x}_r}$$

The Bias and the MSE of $\overline{y}_{RAT}$ are given by

$$B(\overline{y}_{RAT}) = \overline{Y}\delta_4(C_X^2 - \rho_{YX}C_Y C_X) \tag{10}$$

$$M(\overline{y}_{RAT}) = \overline{Y}^2[\delta_1 C_Y^2 + \delta_4(C_X^2 - 2\rho_{YX}C_Y C_X)] \tag{11}$$

From Equations (4) and (11) we have

$$M(\overline{y}_{RAT}) - M(t^*)_{min} = \overline{Y}^2\{\delta_4(C_X - \rho_{YX}C_Y)^2 + \delta_2\rho_{YZ}^2 C_Y^2\} > 0$$

7.3. Comparison with the compromised method of imputation

In this method, the sample units of S are defined as

$$y_{.i} = \begin{cases} \alpha \frac{n}{r} y_i + (1-\alpha)\hat{b}x_i, & \text{if } i \in A \\ (1-\alpha)\hat{b}x_i, & \text{if } i \in A^c \end{cases}$$

where, α is a constant, chosen in a way that gives the minimum MSE of the point estimator .

The corresponding point estimator of the population mean here is given by

$$\bar{y}_{Comp} = \alpha \bar{y}_r + (1-\alpha)\bar{y}_r \frac{\bar{x}_n}{\bar{x}_r}$$

The Bias and the MSE of $\bar{y}_{Comp}$ are given by

$$B(\bar{y}_{Comp}) = \overline{Y}\delta_4(C_X - \rho_{YX}C_Y)\rho_{YX}C_Y \tag{12}$$

$$M(\bar{y}_{Comp}) = \overline{Y}^2(\delta_1 C_Y^2 - \delta_4 \rho_{YX}^2 C_Y^2) \tag{13}$$

From Equations (4) and (13) we have

$$M(\bar{y}_{Comp}) - M(t^*)_{min} = \overline{Y}^2\{\delta_2 \rho_{YZ}^2 C_Y^2\} > 0$$

7.4. Comparison with the Ahmed et al. (2006) method of imputation

From Equations (4) and (7) we have

$$M(t)_{min} - M(t^*)_{min} = \overline{Y}^2\{\delta_2 \rho_{YZ}^2 C_Y^2\} > 0$$

7.5. Comparison with the Pandey et al. (2021) method of imputation

They suggested three logarithmic type imputation techniques in Simple Random Sampling to estimate finite population mean of the study variable. The point estimators obtained through the suggested imputation methods are

$$\psi(\alpha) = \bar{y}_r + \alpha \log\left(\frac{\bar{x}_n}{\overline{\overline{X}}}\right)$$

$$\psi(\beta) = \bar{y}_r + \beta \log\left(\frac{\bar{x}_r}{\overline{\overline{X}}}\right)$$

$$\psi(\gamma) = \bar{y}_r + \gamma \log\left(\frac{\bar{x}_r}{\bar{x}_n}\right)$$

where α, β and γ are the constants to be determined such that the point estimators $\psi(\alpha)$, $\psi(\beta)$ and $\psi(\gamma)$ respectively has the minimum MSE.

We consider the point estimator $\psi(\gamma)$ only, as we assume that $\overline{X}$ is not known in advance. The bias and MSE at the optimum value of $\gamma_{opt} = \rho_{YX}\frac{S_Y}{C_X}$ are

$$B[\psi(\gamma)] = \frac{1}{2}\delta_4\rho_{YX}S_Y C_X \tag{14}$$

$$M[\psi(\gamma)]_{min} = S_Y^2\left\{\delta_1 - \delta_4\rho_{YX}^2\right\} \tag{15}$$

From Equation (4) and (15), we have

$$M[\psi(\gamma)]_{min} - M(t^*)_{min} = \delta_2\rho_{YX}^2 S_Y^2 > 0$$

7.6. Comparison with the Panday et al. (2025) suggested methods of imputation

They suggested three imputation methods in Simple Random Sampling to estimate finite population mean of the study variable. The point estimators obtained through the suggested imputation methods are

$$\psi(a_1) = \overline{y}_r + a_1\log\left\{1+\sin\left(\frac{\overline{x}_n}{\overline{X}} - 1\right)\right\}$$

$$\psi(a_2) = \overline{y}_r + a_2\log\left\{1+\sin\left(\frac{\overline{x}_r}{\overline{X}} - 1\right)\right\}$$

$$\psi(a_3) = \overline{y}_r + a_3\log\left\{1+\sin\left(\frac{\overline{x}_r}{\overline{x}_n} - 1\right)\right\}$$

where a_1, a_2 and a_3 are constants to be determined such that the point estimators $\psi(a_1)$, $\psi(a_2)$ and $\psi(a_3)$ have minimum MSE respectively.

We consider the point estimator $\psi(a_3)$ only for comparison, as we assume that $\overline{X}$ is not known in advance. The bias and MSE at the optimum value of $a_{3(opt)} = -\rho_{YX}\frac{S_Y}{C_X}$ are

$$B[\psi(a_3)] = -\frac{1}{2}\overline{Y}\delta_4\rho_{YX}C_Y C_X \tag{16}$$

$$M[\psi(a_3)]_{min} = \overline{Y}^2\left\{\delta_1 C_Y^2 - \delta_4\rho_{YX}^2 C_Y^2\right\} \tag{17}$$

From Equation (4) and (17), we have

$$M[\psi(a_3)]_{min} - M(t^*)_{min} = \delta_2\left(\overline{Y}\rho_{YX}C_Y\right)^2 > 0$$

Thus, the proposed point estimator t^* is more efficient than the existing point estimators $\overline{y}_m$, $\overline{y}_{RAT}$, $\overline{y}_{Comp}$, Ahmed et al. (2006) suggested point estimator t, Panday et al. (2021) suggested point estimator $\psi(\gamma)$ and Panday et al. (2025) suggested point estimator $\psi(a_3)$.

8. Proposed Methods of Imputation under Double Sampling

Thakur and Yadav (2011) proposed three imputation methods under two-phase sampling as described below:

(1)

$$y_{1i} = \begin{cases} y_i & \text{if } i \in A \\ \frac{\bar{y}_r}{n-r}\left\{n\left(\frac{\bar{x}_1}{\bar{x}_n}\right)^{\gamma_1} - r\right\} & \text{if } i \in A^c \end{cases}$$

The corresponding point estimator of $\overline{Y}$ is

$$t_1 = \bar{y}_r \left(\frac{\bar{x}_1}{\bar{x}_n}\right)^{\gamma_1}$$

(2)

$$y_{2i} = \begin{cases} y_i & \text{if } i \in A \\ \frac{\bar{y}_r}{n-r}\left\{n\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\gamma_2} - r\right\} & \text{if } i \in A^c \end{cases}$$

The corresponding point estimator of $\overline{Y}$ is

$$t_2 = \bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\gamma_2}$$

(3)

$$y_{3i} = \begin{cases} y_i & \text{if } i \in A \\ \frac{\bar{y}_r}{n-r}\left\{n\left(\frac{\bar{x}_1}{\bar{x}_r}\right)^{\gamma_3} - r\right\} & \text{if } i \in A^c \end{cases}$$

The corresponding point estimator of $\overline{Y}$ is

$$t_3 = \bar{y}_r \left(\frac{\bar{x}_1}{\bar{x}_r}\right)^{\gamma_3}$$

where γ_1, γ_2, and γ_3 are constants, chosen in a way that yields minimum mean square errors for the estimators t_1, t_2 and t_3 .

Motivated by Ahmed et al. (2006) and Thakur and Yadav (2011), we have proposed three Ratio chain-type imputation methods under the double sampling method.

(1)

$$y^*_{1i} = \begin{cases} y_i & \text{if } i \in A \\ \frac{\bar{y}_r}{n-r}\left\{n\left(\frac{\bar{x}_1}{\bar{x}_n}\right)^{\alpha_1}\left(\frac{\overline{Z}}{\bar{z}_1}\right)^{\beta_1} - r\right\} & \text{if } i \in A^c \end{cases}$$

(2)

$$y^*_{2i} = \begin{cases} y_i & \text{if } i \in A \\ \frac{\bar{y}_r}{n-r}\left\{n\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\alpha_2}\left(\frac{\bar{z}_1}{\bar{z}_n}\right)^{\beta_2} - r\right\} & \text{if } i \in A^c \end{cases}$$

(3)

$$y_{3i}^{*}=\begin{cases} y_i & \text{if } i\in A \\ \frac{\bar{y}_r}{n-r}\left\{n(\frac{\bar{x}_1}{\bar{x}_r})^{\alpha_3}(\frac{\overline{Z}}{\bar{z}_1})^{\beta_3}-r\right\} & \text{if } i\in A^c \end{cases}$$

Using the imputation methods described above, the corresponding point estimators of the population mean $\overline{Y}$ for methods (1), (2) and (3) respectively are:

$$t_1^{*}=\bar{y}_r\left(\frac{\bar{x}_1}{\bar{x}_n}\right)^{\alpha_1}\left(\frac{\overline{Z}}{\bar{z}_1}\right)^{\beta_1}$$

$$t_2^{*}=\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\alpha_2}\left(\frac{\bar{z}_1}{\bar{z}_n}\right)^{\beta_2}$$

$$t_3^{*}=\bar{y}_r\left(\frac{\bar{x}_1}{\bar{x}_r}\right)^{\alpha_3}\left(\frac{\overline{Z}}{\bar{z}_1}\right)^{\beta_3}$$

where α_i and β_i $(i=1,2,3)$ are constants, chosen in a way that yields the minimum MSE of the estimator t_1^{*}, t_2^{*} and t_3^{*} respectively.

9. Properties of t_1^{*}, t_2^{*} and t_3^{*}

The following transformations are used to determine the characteristics of the proposed point estimators.

$\bar{y}_r=\overline{Y}(1+e_0)$ $\bar{x}_1=\overline{X}(1+e_1)$ $\bar{x}_n=\overline{X}(1+e_2)$ $\bar{z}_1=\overline{Z}(1+e_3)$ such that $e_i\in(0,1)$ for $i=1,2,3$

$\bar{x}_r=\overline{X}(1+e_1^{*})$ $\bar{z}_n=\overline{Z}(1+e_3^{*})$ such that $e_j^{*}\in(0,1)$ for $j=1,3$

$E(e_i)=0\ \ \forall i=0,1,2,3$ and $E(e_j^{*})=0\ \ \forall j=1,3$

$E(e_0^2)=\delta_1C_Y^2$ $E(e_1^2)=\delta_3C_X^2$ $E(e_2^2)=\delta_2C_X^2$

$E(e_3^2)=\delta_3C_X^2$ $E(e_1^{*2})=\delta_1C_X^2$ $E(e_3^{*2})=\delta_2C_Z^2$

$E(e_0e_1)=\delta_3\rho_{YX}C_YC_X$ $E(e_0e_2)=\delta_2\rho_{YX}C_YC_X$ $E(e_0e_3)=\delta_3\rho_{YZ}C_YC_Z$

$E(e_0e_1^{*})=\delta_1\rho_{YX}C_YC_X$ $E(e_0e_3^{*})=\delta_2\rho_{YZ}C_YC_Z$ $E(e_1e_2)=\delta_3C_X^2$

$E(e_1e_3)=\delta_3\rho_{XZ}C_XC_Z$ $E(e_1e_1^{*})=\delta_3C_X^2$ $E(e_1e_3^{*})=\delta_3\rho_{XZ}C_XC_Z$

$E(e_2e_3)=\delta_3\rho_{XZ}C_XC_Z$ $E(e_2e_3^{*})=\delta_2\rho_{XZ}C_XC_Z$ $E(e_1^{*}e_2)=\delta_2C_X^2$

$$E(e_1^* e_3) = \delta_3 \rho_{XZ} C_X C_Z \qquad E(e_1^* e_3^*) = \delta_2 \rho_{XZ} C_X C_Z \qquad E(e_3 e_3^*) = \delta_3 C_Z^2$$

$$\delta_1 = \tfrac{1}{r} - \tfrac{1}{N} \qquad \delta_2 = \tfrac{1}{n} - \tfrac{1}{N} \qquad \delta_3 = \tfrac{1}{n_1} - \tfrac{1}{N} \qquad \delta_4 = \tfrac{1}{r} - \tfrac{1}{n} \qquad \delta_5 = \tfrac{1}{r} - \tfrac{1}{n_1} \qquad \delta_6 = \tfrac{1}{n} - \tfrac{1}{n_1}$$

Theorem 9.1. *Biases of t_1^*, t_2^* and t_3^* are*

$$i)\ B(t_1^*) = \frac{\overline{Y}}{2}\{\delta_5(\alpha_1^2+\alpha_1)C_X^2 + \delta_3(\beta_1^2+\beta)C_Z^2 - 2\alpha_1\delta_5\rho_{YX}C_Y C_X - 2\beta_1\delta_3\rho_{YZ}C_Y C_Z\} \quad (18)$$

$$ii)\ B(t_2^*) = \overline{Y}\{\delta_4(\frac{\alpha_2^2+\alpha_2}{2})C_X^2 + \delta_6(\frac{\beta_2^2+\beta_2}{2})C_Z^2 - \delta_4\alpha_2\rho_{YX}C_Y C_X - \delta_6\beta_2\rho_{YZ}C_Y C_Z\} \quad (19)$$

$$iii)\ B(t_3^*) = \frac{\overline{Y}}{2}\{\delta_6(\alpha_3^2+\alpha_3)C_X^2 + \delta_3(\beta_3^2+\beta)C_Z^2 - 2\alpha_3\delta_6\rho_{YX}C_Y C_X - 2\beta_3\delta_3\rho_{YZ}C_Y C_Z\} \quad (20)$$

Proof. The point estimators t_1^*, t_2^* and t_3^*, expressed in terms e_i's, e_1^* and e_3^* $(i = 0,1,2,3)$ and approximated up to the first order are:

$$\begin{aligned} t_1^* - \overline{Y} = \overline{Y}\{&e_0 + \alpha_1(e_1 - e_2 + e_2^2 - e_1e_2 + e_0e_1 - e_0e_2) - \beta_1(e_3 + e_0e_3) \\ &+ \frac{\alpha_1(\alpha_1-1)}{2}(e_1^2 + e_2^2 - 2e_1e_2) + \frac{\beta_1(\beta_1+1)}{2}e_3^2 - \alpha_1\beta_1(e_1e_3 - e_2e_3)\} \end{aligned} \quad (21)$$

$$\begin{aligned} t_2^* - \overline{Y} = \overline{Y}\{&e_0 + \alpha_2(e_2 - e_1^* + e_1^{*2} - e_1^* - e_1^*e_2 + e_0e_2 - e_0e_1^*) + \beta_2(e_3 - e_3^{*2} - e_3e_3^* + e_0e_3 - e_0e_3^*) \\ &+ \frac{\alpha_2(\alpha_2-1)}{2}(e_2^2 + e_1^{*2} - 2e_3e_3^*) + \alpha\beta(e_2e_3 - e_2e_3^* - e_1^*e_3 + e_1^*e_3^*)\} \end{aligned} \quad (22)$$

$$\begin{aligned} t_3^* - \overline{Y} = \overline{Y}\{&e_0 + \alpha_3(e_1 - e_1^* + e_1^{*2} - e_1e_1^* + e_0e_1 - e_0e_1^*) + \beta_3(e_3 + e_0e_3) \\ &+ \frac{\alpha_3(\alpha_3-1)}{2}(e_1^2 + e_1^{*2} - 2e_1e_1^*) + \frac{\beta_3(\beta_3+1)}{2}e_3^2 - \alpha_3\beta_3(e_1e_3 - e_1^*e_3)\} \end{aligned} \quad (23)$$

Taking expectations on both sides of the Equations (21), (22) and (23), we get

$$i)\ B(t_1^*) = \frac{\overline{Y}}{2}\{\delta_5(\alpha_1^2+\alpha_1)C_X^2 + \delta_3(\beta_1^2+\beta)C_Z^2 - 2\alpha_1\delta_5\rho_{YX}C_Y C_X - 2\beta_1\delta_3\rho_{YZ}C_Y C_Z\}$$

$$ii)\ B(t_2^*) = \overline{Y}\{\delta_4(\frac{\alpha_2^2+\alpha_2}{2})C_X^2 + \delta_6(\frac{\beta_2^2+\beta_2}{2})C_Z^2 - \delta_4\alpha_2\rho_{YX}C_Y C_X - \delta_6\beta_2\rho_{YZ}C_Y C_Z\}$$

$$\begin{aligned} iii)\ B(t_3^*) = \overline{Y}E\{&e_0 + \alpha_3(e_1 - e_1^* + e_1^{*2} - e_1e_1^* + e_0e_1 - e_0e_1^*) + \beta_3(e_3 + e_0e_3) + \\ &\frac{\alpha_3(\alpha_3-1)}{2}(e_1^2 + e_1^{*2} - 2e_1e_1^*) + \frac{\beta_3(\beta_3+1)}{2}e_3^2 - \alpha_3\beta_3(e_1e_3 - e_1^*e_3)\} \end{aligned}$$

□

Theorem 9.2. *MSEs of t_1^*, t_2^* and t_3^* are*

$$i)\ M(t_1^*) = \overline{Y}^2\{\delta_1 C_Y^2 + \alpha_1^2\delta_6 C_X^2 + \beta_1^2\delta_3 C_Z^2 - 2\alpha_1\delta_6\rho_{YX}C_Y C_X - 2 - 2\beta_1\delta_3\rho_{YZ}C_Y C_Z\} \quad (24)$$

$$ii)\ M(t_2^*) = \overline{Y}^2\{\delta_1 C_Y^2 + \alpha_2^2\delta_4 C_X^2 + \beta_2^2\delta_6 C_Z^2 - 2\alpha_2\delta_4\rho_{YX}C_Y C_X - 2\beta_2\delta_6\rho_{YZ}C_Y C_Z\} \quad (25)$$

$$iii)\ M(t_3^*) = \overline{Y}^2\{\delta_1 C_Y^2 + \alpha_3^2\delta_5 C_X^2 + \beta_3^2\delta_3 C_Z^2 - 2\alpha_3\delta_5\rho_{YX}C_Y C_X - 2\beta_3\delta_3\rho_{YZ}C_Y C_Z\} \quad (26)$$

Proof. i) Taking expectations after squaring both sides of Equation (21), we get

$$\begin{aligned} M(t_1^*) &= \overline{Y}^2 E\{e_0^2 + \beta_1^2 e_3^2 + \alpha_1^2(e_1 - e_2)^2 - 2\beta_1 e_0 e_3 + 2\alpha_1(e_0 e_1 - e_0 e_2) - 2\alpha_1\beta_1(e_1 e_3 - e_2 e_3)\} \\ &= \overline{Y}^2\{\delta_1 C_Y^2 + \alpha_1^2(\delta_2 - \delta_3)C_X^2 + \beta_1^2\delta_3 C_Z^2 - 2\alpha_1(\delta_2 - \delta_3)\rho_{YX}C_Y C_X - 2\beta_1\delta_3\rho_{YZ}C_Y C_Z\} \\ &= \overline{Y}^2\{\delta_1 C_Y^2 + \alpha_1^2\delta_6 C_X^2 + \beta_1^2\delta_3 C_Z^2 - 2\alpha_1\delta_6\rho_{YX}C_Y C_X - 2 - 2\beta_1\delta_3\rho_{YZ}C_Y C_Z\} \end{aligned}$$

ii) Similarly, taking expectations after squaring both sides of the Equation (22), we get

$$\begin{aligned} M(t_2^*) &= \overline{Y}^2\{\delta_1 C_Y^2 + \alpha_2^2(\delta_1 - \delta_2)C_X^2 + \beta_2^2(\delta_2 - \delta_3)C_Z^2 - 2\alpha_2(\delta_1 - \delta_2)\rho_{YX}C_Y C_X \\ &\quad - 2\beta_2(\delta_2 - \delta_3)\rho_{YZ}C_Y C_Z\} \\ &= \overline{Y}^2\{\delta_1 C_Y^2 + \alpha_2^2\delta_4 C_X^2 + \beta_2^2\delta_6 C_Z^2 - 2\alpha_2\delta_4\rho_{YX}C_Y C_X - 2\beta_2\delta_6\rho_{YZ}C_Y C_Z\} \end{aligned}$$

iii) And taking expectations after squaring both sides of Equation (23), we get

$$\begin{aligned} M(t_3^*) &= \overline{Y}^2\{\delta_1 C_Y^2 + \beta_3^2\delta_3 C_Z^2 + \alpha_3^2(\delta_1 - \delta_3)C_X^2 - 2\alpha_3(\delta_1 - \delta_3)\rho_{YX}C_Y C_X - 2\beta_3\delta_3\rho_{YZ}C_Y C_Z\} \\ &= \overline{Y}^2\{\delta_1 C_Y^2 + \alpha_3^2\delta_5 C_X^2 + \beta_3^2\delta_3 C_Z^2 - 2\alpha_3\delta_5\rho_{YX}C_Y C_X - 2\beta_3\delta_3\rho_{YZ}C_Y C_Z\} \end{aligned}$$

□

Theorem 9.3. *Minimum MSEs of t_1^*, t_2^* and t_3^* are*

$$i)\quad M(t_1^*)_{min} = \overline{Y}^2\{\delta_1 C_Y^2 - \delta_6\rho_{YX}^2 C_Y^2 - \delta_3\rho_{YZ}^2 C_Y^2\} \quad (27)$$

$$ii)\quad M(t_2^*)_{min} = \overline{Y}^2\{\delta_1 C_Y^2 - \delta_4\rho_{YX}^2 C_Y^2 - \delta_6\rho_{YZ}^2 C_Y^2\} \quad (28)$$

$$iii)\quad M(t_3^*)_{min} = \overline{Y}^2\{\delta_1 C_Y^2 - \delta_5\rho_{YX}^2 C_Y^2 - \delta_3\rho_{YZ}^2 C_Y^2\} \quad (29)$$

Proof. i) Equating the differentiation of Equation (24) w.r.t α_1 and β_1 to zero, we get

$$\alpha_{1\,opt} = \frac{\rho_{YX}C_Y}{C_X} \text{ and } \beta_{1\,opt} = \frac{\rho_{YZ}C_Y}{C_Z}$$

Now, putting the values of $\alpha_1 = \alpha_{1opt}$ and $\beta_1 = \beta_{1opt}$ in Equation (24), we obtain

$$M(t_1^*)_{min} = \overline{Y}^2\{\delta_1 C_Y^2 - \delta_6 \rho_{YX}^2 C_Y^2 - \delta_3 \rho_{YZ}^2 C_Y^2\}$$

ii) Similarly, proceeding for Equation (25), we have

$$\alpha_{2opt} = \frac{\rho_{YX} C_Y}{C_X} \text{ and } \beta_{2opt} = \frac{\rho_{YZ} C_Y}{C_Z}$$

$$M(t_2^*)_{min} = \overline{Y}^2\{\delta_1 C_Y^2 - \delta_4 \rho_{YX}^2 C_Y^2 - \delta_6 \rho_{YZ}^2 C_Y^2\}$$

iii) And for Equation (26), we have

$$\alpha_{3opt} = \frac{\rho_{YX} C_Y}{C_X} \text{ and } \beta_{3opt} = \frac{\rho_{YZ} C_Y}{C_Z}$$

$$M(t_3^*)_{min} = \overline{Y}^2\{\delta_1 C_Y^2 - \delta_5 \rho_{YX}^2 C_Y^2 - \delta_3 \rho_{YZ}^2 C_Y^2\}$$

□

9.1. Remarks on Practicability of the Proposed estimators

To use the proposed estimator in practical situations, $\rho_{YX}, \rho_{YZ}, C_X, C_Y$ and C_Z are known in advance. In the case of unavailability, we would replace α_{iopt} and β_{iopt} by $\hat{\alpha}_{iopt}$ and $\hat{\beta}_{iopt}$ $(i = 1,2,3)$ respectively, where, $\hat{\alpha}_{1opt} = \frac{\rho_{yx} C_y}{C_x}, \hat{\beta}_{1opt} = \frac{\rho_{yz} C_y}{C_z}$; $\hat{\alpha}_{2opt} = \frac{\rho_{yx} C_y}{C_x}, \hat{\beta}_{2opt} = \frac{\rho_{yz} C_y}{C_z}$; $\hat{\alpha}_{3opt} = \frac{\rho_{yx} C_y}{C_x}, \hat{\beta}_{3opt} = \frac{\rho_{yz} C_y}{C_z}$.

It may be noted that $\rho_{yx}, \rho_{yz}, C_y, C_x$, and C_z are calculated from the responding units of S_1 .

9.2. Particular Cases

For $\alpha_1 = \gamma_1$ and $\beta_1 = 0$; $\alpha_2 = \gamma_2$ and $\beta_2 = 0$; $\alpha_3 = \gamma_3$ and $\beta_3 = 0$, the point estimators t_1^*, t_2^* and t_3^* transform into the ratio-type point estimator suggested by Thakur and Yadav (2011).

$$t_1^* \to \overline{y}_r \left(\frac{\overline{x}_1}{\overline{x}_n}\right)^{\gamma_1} \left(\frac{\overline{Z}}{\overline{z}_1}\right)^0 = \overline{y}_r \left(\frac{\overline{x}_1}{\overline{x}_n}\right)^{\gamma_1} = t_1$$

$$t_2^* \to \overline{y}_r \left(\frac{\overline{x}_n}{\overline{x}_r}\right)^{\gamma_2} \left(\frac{\overline{z}_1}{\overline{z}_n}\right)^0 = \overline{y}_r \left(\frac{\overline{x}_n}{\overline{x}_r}\right)^{\gamma_2} = t_2$$

$$t_3^* \to \overline{y}_r \left(\frac{\overline{x}_1}{\overline{x}_r}\right)^{\gamma_3} \left(\frac{\overline{Z}}{\overline{z}_1}\right)^0 = \overline{y}_r \left(\frac{\overline{x}_1}{\overline{x}_r}\right)^{\gamma_3} = t_3$$

Theorem 9.4. *The Bias and the MSE of* t_1 *are*

$$i) \quad B(t_1) = \overline{Y}\{\delta_5(\gamma_1^2 + \gamma_1) C_X^2 - 2\gamma_1 \delta_5 \rho_{YX} C_Y C_X\} \tag{30}$$

$$ii) \quad M(t_1) = \overline{Y}^2\{\delta_1 C_Y^2 + \gamma_1^2 \delta_6 C_X^2 - 2\gamma_1 \delta_6 \rho_{YX} C_Y C_X\} \tag{31}$$

$$iii) \quad M(t_1)_{min} = \overline{Y}^2\{\delta_1 C_Y^2 - \delta_6 \rho_{YX}^2 C_Y^2\} \tag{32}$$

Proof. Putting $\alpha_1 = \gamma_1$ and $\beta_1 = 0$ in Equation (18) and Equation (24) we get the above $B(t_1)$ and $M(t_1)$.

Equating to zero, the differentiation of Equation (31) w.r.t. γ_1, we get $\gamma_{1\,opt} = \rho_{YX}\frac{C_Y}{C_X}$.

Putting $\gamma_1 = \gamma_{1\,opt}$ in Equation (31) we have

$$M(t_1)_{min} = \overline{Y}^2\{\delta_1 C_Y^2 - \delta_6 \rho_{YX}^2 C_Y^2\}$$

□

Theorem 9.5. *The Bias and the MSE of t_2 are*

$$i) \quad B(t_2) = \frac{\overline{Y}}{2}\{\delta_4(\gamma_2^2 + \gamma_2)C_X^2 - 2\gamma_2 \delta_4 \rho_{YX} C_Y C_X\} \tag{33}$$

$$ii) \quad M(t_2) = \overline{Y}^2\{\delta_1 C_Y^2 + \gamma_2^2 \delta_4 C_X^2 - 2\gamma_2 \delta_4 \rho_{YX} C_Y C_X\} \tag{34}$$

$$iii) \quad M(t_2)_{min} = \overline{Y}^2\{\delta_1 C_Y^2 - \delta_4 \rho_{YX}^2 C_Y^2\} \tag{35}$$

Proof. $B(t_2)$ and $M(t_2)$ can be obtained by putting $\alpha_2 = \gamma_2$ and $\beta_2 = 0$ in Equation (19) and Equation (25).

Equating to zero, the differentiation of Equation (34) w.r.t. γ_2, we get $\hat{\gamma}_2 = \rho_{YX}\frac{C_Y}{C_X}$.

Putting $\gamma_2 = \hat{\gamma}_2$ in Equation (34) we have

$$M(t_2)_{min} = \overline{Y}^2\{\delta_1 C_Y^2 - \delta_4 \rho_{YX}^2 C_Y^2\}$$

□

Theorem 9.6. *The Bias and the MSE of the point estimator t_3 are*

$$i) \quad B(t_3) = \frac{\overline{Y}}{2}\{\delta_6(\gamma_3^2 + \gamma_3)C_X^2 - 2\gamma_3 \delta_6 \rho_{YX} C_Y C_X\} \tag{36}$$

$$ii) \quad M(t_3) = \overline{Y}^2\{\delta_1 C_Y^2 + \gamma_3^2 \delta_5 C_X^2 - 2\gamma_3 \delta_5 \rho_{YX} C_Y C_X\} \tag{37}$$

$$iii) \quad M(t_3)_{min} = \overline{Y}^2\{\delta_1 C_Y^2 - \delta_5 \rho_{YX}^2 C_Y^2\} \tag{38}$$

Proof. $B(t_3)$ and $M(t_3)$ can be obtained by putting $\alpha_3 = \gamma_3$ and $\beta_3 = 0$ in Equation (20) and Equation (26).

Equating to zero, the differentiation of Equation (37) w.r.t. γ_3, we get $\hat{\gamma}_3 = \rho_{YX}\frac{C_Y}{C_X}$.

Putting $\gamma_3 = \hat{\gamma}_3$ in Equation (37) we have

$$M(t_3)_{min} = \overline{Y}^2\{\delta_1 C_Y^2 - \delta_5 \rho_{YX}^2 C_Y^2\}$$

□

10. Efficiency comparison: Theoretical

In this section we compare the estimators t_1^*, t_2^* and t_3^* with the point estimators t_1, t_2 and t_3 suggested by Thakur and Yadav (2011).
From Equations (32) and (27), we get

$$M(t_1)_{min} - M(t_1^*)_{min} = \overline{Y}^2(\delta_3 \rho_{YZ}^2 C_Y^2) > 0$$

From Equations (35) and (28), we get

$$M(t_2)_{min} - M(t_2^*)_{min} = \overline{Y}^2(\delta_6 \rho_{YZ}^2 C_Y^2) > 0$$

From Equations (38) and (29), we get

$$M(t_3)_{min} - M(t_3^*)_{min} = \overline{Y}^2(\delta_3 \rho_{YZ}^2 C_Y^2) > 0$$

Thus, the proposed estimators outperform the point estimators suggested by Thakur and Yadav (2011)

11. Empirical Study I

To evaluate the efficiency of the proposed imputation methods of the finite population mean employing two auxiliary variables in simple random sampling, we have tested the following two populations. Let the Percent Relative Efficiencies (PRE) of the suggested imputation techniques in comparison with the mean, ratio, and compromised imputation methods be respectively denoted by

$$PRE_I = \frac{M(\bar{y}_m)}{M(\cdot)} \times 100; \quad PRE_{II} = \frac{M(\bar{y}_{RAT})}{M(\cdot)} \times 100; \quad PRE_{III} = \frac{M(\bar{y}_{Comp})}{M(\cdot)} \times 100$$

Population I:

{Source: Cochran (1977), Page No. 182}

Y: Number of cases in the placebo group
X: Number of cases in the placebo group with paralytic polio
Z: Number of cases in the paralytic polio group without inoculation.
Here, $N = 34 \quad n_1 = 20 \quad n = 12 \quad r = 8$

Population II:
This data was artificially generated for three variables Y, X, and Z by Shukla et al. (2011).
Here, $N = 200 \quad n_1 = 80 \quad n = 30 \quad r = 22$

Values of various parameters of the variables corresponding to Population I and Population II are given in Table 3. The MSEs and PREs of the existing and the proposed estimator(s) in SRS are shown in Table 4 and Table 5 respectively.

Table 3. Parameters of the Population I and II

Parameters	Population I	Population II
$\overline{Y}$	4.92	42.485
$\overline{X}$	2.59	18.515
$\overline{Z}$	2.91	20.52
ρ_{YX}	0.7326	0.8734
ρ_{YZ}	0.6430	0.8667
ρ_{XZ}	0.6837	0.9943
C_Y^2	1.0248	0.10804
C_X^2	1.5175	0.1410
C_Z^2	1.1492	0.10864

Table 4. The MSE of the existing estimators and the suggested estimator t^* in SRS

Estimator	Population I	Population II
$\overline{y}_m$	2.37127	7.88902
$\overline{y}_{RAT}$	2.05891	6.25695
$\overline{y}_{COMP}$	1.81650	6.08591
t^*	1.26357	1.93526

Table 5. The PRE of the suggested estimator t^* in SRS

PRE	Population I	Population II
PRE_I	187.66	407.65
PRE_{II}	162.94	323.31
PRE_{III}	143.76	314.48

The MSEs and the PREs of the proposed estimator(s) in double sampling are given in Table 6 and Table 7 respectively.

Table 6. The MSE of the suggested estimators in double sampling

Estimator	Population I	Population II
t_1^*	1.71635	3.69085
t_2^*	1.47464	3.03389
t_3^*	1.16156	1.88789

Table 7. The PRE of the suggested estimators in double sampling

Estimator	PRE	Population I	Population II
t_1^*	PRE_I	138.16	213.75
	PRE_{II}	119.96	169.18
	PRE_{III}	105.84	164.55
t_2^*	PRE_I	160.80	260.03
	PRE_{II}	139.62	206.24
	PRE_{III}	123.18	200.59
t_3^*	PRE_I	204.15	417.87
	PRE_{II}	177.25	331.43
	PRE_{III}	156.38	322.37

12. Empirical Study II (Simulation)

For the simulation study consider the data which is an artificially generated population of size $N = 200$ containing values of the study variable Y and the auxiliary variables X and Z (Population II), as attached in Appendix A. In SRS the different steps followed are:

Step 1: Draw a random sample of size $n = 80$ from the population.

Step 2: Drop away 50 units randomly from each sample corresponding to the variable Y.

Step 3: Compute and impute the dropped units of Y with the help of existing and proposed imputation methods.

Step 4: Obtain the estimates of the population mean for existing and proposed imputation methods.

Step 5: Repeat the step 1 to step 4, $k = 500000$ times, which provides multiple sample-based estimates $\hat{T}_1, \hat{T}_2, \hat{T}_3, \ldots \hat{T}_{500000}$.

Step 6: Calculate Bias and MSE of $\hat{T}$:

$$B(\hat{T}) = \frac{1}{k}\sum_{i=1}^{k}(\hat{T}_i - \overline{Y})$$

$$M(\hat{T}) = \frac{1}{k}\sum_{i=1}^{k}(\hat{T}_i - \overline{Y})^2$$

For Double Sampling, the different steps for the simulation procedure are as follows:

Step 1: Draw a preliminary random S_1, of size $n_1 = 80$ from the population of size $N = 200$.

Step 2: Again, draw a random sample of size $n = 30$ from S_1.

Step 3: Drop away 8 units randomly from each sample corresponding to the variable Y.

Step 4: Compute and impute the dropped units of Y with the help of existing and proposed imputation methods.

Step 5: Obtain the estimates of the population mean for existing and proposed imputation methods.

Step 6: Repeat the step 1 to step 4, $k = 500000$ times, which provides multiple sample-based estimates $\hat{T}_1, \hat{T}_2, \hat{T}_3, \ldots \hat{T}_{500000}$.

Step 7: Calculate Bias and MSE of $\hat{T}$:

$$B(\hat{T}) = \frac{1}{k}\sum_{i=1}^{k}(\hat{T}_i - \overline{Y})$$

$$M(\hat{T}) = \frac{1}{k}\sum_{i=1}^{k}(\hat{T}_i - \overline{Y})^2$$

The well-known software *Wolfram Mathematica* is used to run the program.

The Bias, MSEs, and PREs of the existing and the proposed estimator(s) in SRS and double sampling obtained from simulation are given in Table 8 and Table 9 respectively.

Table 8. Bias, MSE and PRE of the existing and the suggested estimator(s) t^* in SRS

Estimator	Bias	MSE	PRE_I	PRE_{II}	PRE_{III}
$\bar{y}_m$	0.000465	7.88478	100	-	-
$\bar{y}_{RAT}$	0.017715	6.24919	126.1728	100	-
$\bar{y}_{Comp}$	0.013653	6.07536	129.8346	102.8612	100
t	0.007075	6.07294	129.8346	102.9022	100.0398
t^*	0.001515	1.93044	408.4447	323.7184	314.7137

Table 9. Bias, MSE and PRE of the existing and the suggested estimator(s) t^* in Double Sampling

Estimator	Bias	MSE	PRE_I	PRE_{II}	PRE_{III}
$\bar{y}_m$	-0.00060	7.82456	100	-	-
$\bar{y}_{RAT}$	0.00615	6.18800	126.4473	100	-
$\bar{y}_{Comp}$	0.00126	6.01266	130.1347	102.9162	100
t_1	0.00361	4.82623	162.1257	128.2160	124.5829
t_2	-0.00533	6.01022	130.1875	102.9579	100.0406
t_3	0.01267	2.99276	262.7896	206.7698	200.9069
t_1^*	0.01026	3.69628	212.7696	167.4115	162.6679
t_2^*	0.00659	3.03264	259.3305	204.0466	198.2649
t_3^*	0.01941	1.87155	420.2164	330.6350	321.2663

13. Conclusion

The proposed exponential ratio-type imputation techniques outperform the existing techniques such as the mean, ratio, and compromised ones in SRS and double sampling. The suggested point estimators' MSEs are computed in terms of the parameters, and are compared to the MSEs of the existing estimators $\bar{y}_m, \bar{y}_{RAT}, \bar{y}_{Comp}$. It has been demonstrated both theoretically and empirically, that the *MSE*s of the suggested estimators are lower than that of the existing ones under both sampling schemes.

References

Ahmed, M. S., Al-Titi, O., Al-Rawi, Z., and Abu-Dayyeh, W., (2006). Estimation of a Population Mean using Different Imputation Methods. *Statistics in Transition*, 7(6), pp.1247–1264.

Bahl, S., Tuteja, R., (1991). Ratio and Product type Exponential Estimator. *Information and Optimization Sciences*, 12, pp. 159–163.

Chand, L., (1975). Some Ratio type Estimator based on Two or More Auxiliary Variables. *Unpublished PhD dissertation*, Iowa State University, Ames, Iowa.

Cochran, W., (1940). The Estimation of the Yields of the Cereal Experiments by Sampling for the Ratio of Grain to Total Produce. *Journal of Agricultural Science*, 30(2), pp. 262–275.

Cochran, W., (1977). Sampling Techniques. *Wiley Eastern Limited*, New York.

Diana, G., Perri, P., (2010). Improved Estimators of the Population mean for Missing data. *Communication in Statistics: Theory and Methods*, 39, pp. 3245–3251.

Gogoi, U., Singh, B., (2017). Estimation of Population Mean using Exponential Dual to Ratio type Compromised Imputation of Missing Data in Survey Sampling. *Journal of*

Statistics Application and Probability, 6(3), pp. 515–522.

Gogoi, U., Singh, B., (2018). Estimation of Population Mean using Imputation Techniques in Sample Survey. *International Journal of Statistics and Economics*, 19(3), pp. 90–103.

Gogoi, U., Singh, B., (2019). Study of a One Parameter Class of Ratio cum Product type Compromised Imputation Technique in Survey Sampling. *International Journal of Mathematics and Statistics*, 20(2), pp. 16–30.

Kadilar, C., Cingi, H., (2008). Estimators for the Population Mean in the case of Missing Data. *Communication in Statistics: Theory and Methods*, 37, pp. 2226–2236.

Kalita, D. and Singh, B., (2013). Exponential Dual to Ratio and Dual to Product-type estimators of finite population mean in double sampling. *Elixir Statistics*, 59, pp. 15458–15470.

Kalita, D., Singh, B., and Choudhury, S., (2013). Improved Exponential Chain Ratio and Product type Estimator for Finite Population Mean in Double Sampling. *International Journal of Statistical Science*, 13, pp. 21–37.

Kumar, M., Bahl, S., (2006). Class of Dual to Ratio Estimators for Double Sampling. *Statistical Papers*, (47), pp. 319–326.

Lee, H., Rancourt, E., and Sarndal, C., (1994). Experiments with Variance Estimation from Survey Data with Imputed Values. *Journal of Official Statistics*, 10(3), pp. 231–243.

Lee, H., Rancourt, E., and Sarndal, C., (1995). Variance Estimation in the presence of Imputed Data for the generalized Estimation System. *Proceedings of the American Statistical Association (Social Survey Research Methods Section)*, pp. 384–389.

Nath, K., Gohain, D., Gogoi, N., and Singh, B., (2020). Study of Imputation based Generalized Class of Dual to Product cum Dual to Ratio Estimators for Missing Data in Two-phase Sampling. *Journal of Mathematical and Computational Science*, 11(6), pp. 7921–7938.

Nath, K., Singh, B., (2018). Population Mean Estimation using Ratio cum Product Compromised Method of Imputation in Two-phase Sampling Scheme. *Asian Journal Mathematics and Statistics*, 11(1), pp. 27–39.

Neyman, J., (1938). Contribution to the Theory of Sampling Human Population. *Journal of American Statistical Association*, 33, pp. 101–116.

Pandey, A., Singh, G., Bhattacharyya, D., Ali, A., Al-Thubaiti, S., and H.A., Y., (2021). Some Classes of Logarithmic type Imputation Techniques for handling Missing Data. *Computational Intelligence and Neuroscience*, 2021(1), pp. 1–11.

Pandey, M., Singh, G., and Zaman, T., (2025). Estimation of Population Mean using Some Improved Imputation methods for Missing Data in Sample Surveys. *Communication in Statistics-Theory and Methods*, 54(8), pp. 2378–2392.

Shukla, A. K., (2019). Efficient Classes of Modified Ratio and Product Estimators for Estimating Population Mean under Two-phase Sampling. *International Journal of Engineering Research and Technology (IJERT)*, 8(12), pp. 659–663.

Shukla, D., Thakur, N., (2008). Estimation of Mean Imputation of Missing Data using Factor-type Estimators. *Statistics in Transition*, 9(1), pp. 33–48.

Shukla, D., Thakur, N., Thakur, D., and Pathak, S., (2011). Linear combination based Imputation method for Missing Data in Sample. *International Journal of Modern Engineering Research*, 1(2), pp. 580– 596.

Singh, B. K., Choudhury, S., (2012). Exponential Chain Ratio and Product type Estimator for Finite Population Mean under Double Sampling Scheme. *Global Journal of Science Frontier Research: Mathematics and Decision Science*, 12(6), pp. 13–24.

Singh, G., Majhi, D., (2014). Some Chain type Exponential Estimators of Population Mean in two-phase Sampling. Statistics in Transition, 15(2), pp. 221–230.

Singh, H., Vishwakarma, G., (2007). Modified Exponential Ratio and Product Estimators for Finite Population Mean in Two-phase Sampling. *Austrian Journal of Statistics*, 30(3), pp. 217–225.

Singh, P., Singh, A., and Singh, V., (2015). On the Use of Compromised Imputation for Missing Data using Factor-type Estimators. *Journal of Statistics Applications and Probability Letters*, 2(2), pp. 105–113.

Singh, S., Horn, S., (2000). Compromised Imputation in Survey Sampling. *Metrika*, 51(3), pp. 267–276.

Srivenkankataramana, T., (1980). Dual to Ratio Estimator in Sample Survey. *Biometrika*, 67(1), pp. 199–204.

Thakur, N., Shukla, D., (2022). Missing Data Estimation based on the Chaining Technique in Survey Sampling. Statistics in Transition New Series, 23(4), pp. 91–111.

Thakur, N., Yadav, K., (2011). A New Imputation Method for Missing Attribute values in Data Mining. *Journal of Applied Computer Science and Mathematics*, 10(5), pp. 14–19.

Upadhyaya, L., Singh, H., (1999). Use of Transformed Auxiliary Variable in Estimating the Finite Population Mean. *Biometrical Journal*, 41(5), pp. 627–636.

Appendix

Table. Population N = 200

Y_i	45	50	39	60	42	38	28	42	38	35	40	55	45	36	40	58	56	62	58	46
X_i	15	20	23	35	18	12	8	15	17	13	29	35	20	14	18	25	28	21	19	18
Z_i	16	22	26	37	19	14	11	17	18	15	30	37	23	15	19	27	30	22	21	21
Y_i	36	43	68	70	50	56	45	32	30	38	35	41	45	65	30	28	32	38	61	58
X_i	15	20	38	42	23	25	18	11	09	17	13	15	18	25	09	08	11	13	23	21
Z_i	18	22	39	44	25	26	19	13	12	20	16	17	19	27	12	10	13	14	24	23
Y_i	65	62	68	85	40	32	60	57	47	55	67	70	60	40	35	30	25	38	23	55
X_i	27	25	30	45	15	12	22	19	17	21	25	30	27	21	15	17	09	15	11	21
Z_i	28	26	33	46	17	15	23	20	19	23	26	32	30	3	7	18	12	18	14	24
Y_i	50	69	53	55	71	74	55	39	43	45	61	72	65	39	43	57	37	71	71	70
X_i	15	23	29	30	33	31	17	14	17	19	25	31	30	19	21	23	15	30	32	29
Z_i	17	24	30	33	35	32	19	16	19	21	27	33	32	21	23	25	17	32	33	32
Y_i	73	63	67	47	53	51	54	57	59	39	23	25	35	30	38	60	60	40	47	30
X_i	28	23	23	17	19	17	18	21	23	20	07	09	15	11	13	25	27	15	17	11
Z_i	30	25	24	20	22	20	21	23	26	22	10	11	18	14	14	26	29	18	20	14
Y_i	57	54	60	51	26	32	30	45	55	54	33	33	20	25	28	40	33	38	41	33
X_i	31	23	25	17	09	11	13	19	25	27	13	11	07	09	13	15	13	17	15	13
Z_i	32	25	27	19	12	13	14	20	27	28	16	14	09	10	14	17	14	20	17	15
Y_i	30	35	20	18	20	27	23	42	37	45	37	37	37	34	41	35	39	45	24	27
X_i	11	15	08	07	09	13	12	25	21	22	15	16	17	13	20	15	21	25	11	13
Z_i	13	18	11	08	12	16	14	26	24	23	16	18	19	16	22	18	23	26	14	14
Y_i	23	20	26	26	40	56	41	47	43	33	37	27	21	23	24	21	39	33	25	35
X_i	09	08	11	12	15	25	15	25	21	15	17	13	11	11	09	08	15	17	11	19
Z_i	11	10	14	15	17	26	17	27	22	17	19	16	13	12	12	11	17	20	13	20
Y_i	45	40	31	20	40	50	45	35	30	35	32	27	30	33	31	47	43	35	30	40
X_i	21	23	15	11	20	25	23	17	16	18	15	13	14	17	15	25	23	17	16	19
Z_i	22	25	18	13	21	27	26	19	17	19	17	16	16	14	17	28	25	18	18	22
Y_i	35	35	46	39	35	30	31	53	63	41	52	43	39	37	20	23	35	39	45	37
X_i	19	19	23	15	17	13	19	25	35	21	25	19	18	17	11	09	15	17	19	19
Z_i	22	21	24	17	20	15	22	26	36	23	26	20	20	19	13	12	17	18	21	22