Alexandria Engineering Journal 95 (2024) 236–246

Contents lists available at ScienceDirect

Alexandria Engineering Journal

journal homepage: www.elsevier.com/locate/aej

Original Article

Novel imputation methods under stratified simple random sampling

Anoop Kumar [a], Shashi Bhushan [b,*], Manahil SidAhmed Mustafa [c], Ramy Aldallal [d], Hassan M. Aljohani [e], Fatimah A. Almulhim [f]

[a] Department of Statistics, Central University of Haryana, Mahendergarh, 123031, India
[b] Department of Statistics, University of Lucknow, U.P., 226007, India
[c] Department of Statistics, Faculty of Science, University of Tabuk, Tabuk, Saudi Arabia
[d] Department of Management, College of Business Administration in Hawtat Bani Tamim, Prince Sattam bin Abdulaziz University, Saudi Arabia
[e] Department of Mathematics and Statistics, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia
[f] Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

ARTICLE INFO

MSC:
62D05
62D10

Keywords:
Missing values
Imputation
Stratified simple random sampling

ABSTRACT

This paper addresses some classes of combined and separate imputation methods (CSIMs) of the population mean under stratified simple random sampling (SSRS) along with their characteristics. To the best of our knowledge, these imputation methods (IMs) have yet not been studied by any author under SSRS, hence these IMs are called 'novel'. In addition, the existing CSIMs are distinguished as the members of the suggested CSIMs, respectively. The theoretical conditions under which the proposed IMs perform better are obtained by comparing the proposed IMs with the existing IMs. To validate the theoretical findings, the numerical and simulation studies are conducted on real and artificial populations, respectively.

1. Introduction

In a sample survey, it is widely recognized that the complete information concerning any situation or episode is essential to make inferences. Incomplete information or missing values in the data set may impair the whole inference. The most familiar method which is used till date to tackle the issue of missing values is imputation. Many articles have been published so far to estimate the population mean in the presence of missing values utilizing simple random sampling (SRS). [1] mooted three key strategies of missing values, namely, missing at random (MAR), observed at random (OAR), and parameter distribution (PD). [2] marked a difference between missing at random and missing completely at random (MCAR) strategy. Subsequently, [3] suggested ratio category of estimators in case of missing values. In the presence of missing data, [4] developed an improved population mean imputation method. [5] suggested some optimal imputations for estimating population mean in case of missing data. [6] utilized moments of higher order of an auxiliary variable and suggested regression type IMs under SRS, whereas [7] considered higher order moments and suggested improved form of regression type IMs. [8] developed a generalized class of estimators utilizing the dual of the supplementary variable in the case of non-response. [9] introduced composite imputation consisting of mean estimators utilizing robust quantile regression. [10] suggested logarithmic IMs using MCAR strategy. Recently, utilizing MCAR strategy, [11] and [12] suggested some IMs based on single and multi-auxiliary information under ranked set sampling. The basic concentration of this article is to discuss the MCAR strategy in the area of socio-economic investigations where sample units can be distinguished very cheaply.

It is well known that the stratified simple random sampling (SSRS) is better representative of a heterogeneous population than SRS. The SSRS enhances the efficiency of the estimators by separating the population into homogeneous strata over the sampling units. A new estimator for mean under SSRS was suggested by [13]. [14] considered SSRS to estimate the population mean using auxiliary characters. An effective exponential mean estimator for SSRS was studied by [15]. Utilizing SSRS, [16] developed an exponential ratio estimator of the median which is an alternative to the regression estimator of the median. [17] suggested dual utilization of auxiliary information for the estimation of finite population mean under SSRS, while [18] carried out a simulation study by utilizing dual auxiliary variable to estimate population mean under SSRS. In stratified two-phase sampling, [19] recommended using exponential ratio and product type estimators of the mean, whereas

* Corresponding author.
E-mail address: bhushan_s@lkouniv.ac.in (S. Bhushan).

https://doi.org/10.1016/j.aej.2024.03.088
Received 23 September 2023; Received in revised form 23 February 2024; Accepted 25 March 2024
Available online 4 April 2024

[20] suggested a generalized estimator for population mean utilizing auxiliary attribute. In SSRS, [21] investigated a few efficient types of estimators, however, [22] utilized bivariate auxiliary information and suggested a few improved types of estimators. [23] produced several enhanced classes of estimators by utilizing stratified ranked set sampling (SRSS), however, an efficient estimation of population mean under SRSS was presented by [24]. The issue of mean estimation under SRSS was discussed by [25]. A miniscule work has been done till date for estimating the population mean in case of missing data under SSRS. The estimate of the population mean under SSRS was taken into consideration by [26] using a few ratio type imputation approaches. Motivated by [27], [28] suggested a separate regression type estimator under SSRS, which is best linear unbiased estimator (BLUE). Under SSRS, there is no imputation method which compete with the BLUE. Apart from this, no study is available which considers CSIMs simultaneously. In the present study, these issues are taken into consideration, and the following objectives have been set:

- To propose novel CSIMs for the population mean under SSRS which compete with the BLUE.
- To compare theoretically the CSIMs with the corresponding conventional CSIMs.
- To perform an empirical study utilizing real and artificially rendered symmetric and skewed populations, respectively.

1.1. Methodology and notation

Consider a population $\Theta = (\Theta_1, \Theta_2, ..., \Theta_N)$ consisting of N units which is divided into L mutually exclusive and exhaustive strata with N_h units in the h^{th} stratum. Let a sample of size n_h units be chosen from the h^{th} stratum containing N_h elements to estimate the population mean. Let r be the number of units providing a response out of n selected units. The set of responding units is referred by R_u and the set of non-responding units is referred by $\bar{R}_u$. The notation of symbols employed throughout this study are defined below.

$(N,\ n)$; population and sample size,
$(N_h,\ n_h)$; population and sample size in stratum h,
$W_h = N_h/N$; h^{th} stratum's weight,
$(\bar{y}_h = \sum_{i=1}^{n_h} y_{hi}/n_h,\ \bar{x}_h = \sum_{i=1}^{n_h} x_{hi}/n_h)$; sample mean of variables $(y,\ x)$ in stratum h,
$(\bar{y}_{st} = \sum_{h=1}^{L} W_h y_h,\ \bar{x}_{st} = \sum_{h=1}^{L} W_h x_h)$; sample mean of variables $(y,\ x)$,
$(\bar{Y}_h = \sum_{i=1}^{N_h} y_{hi}/N_h,\ \sum_{i=1}^{N_h} x_{hi}/N_h)$; population mean of variables $(y,\ x)$ in stratum h,
$(\bar{Y} = \bar{Y}_{st} = \sum_{h=1}^{L} W_h Y_h,\ \bar{X} = \bar{X}_{st} = \sum_{h=1}^{L} W_h X_h)$; population mean of variables $(y,\ x)$,
$R = \bar{Y}/\bar{X}$; population ratio,
$R_h = \bar{Y}_h/\bar{X}_h$; h^{th} stratum's population ratio,
$(S^2_{y_h} = (N_h-1)^{-1}\sum_{i=1}^{N_h}(y_{hi}-\bar{Y}_h)^2,\ S^2_{x_h} = (N_h-1)^{-1}\sum_{i=1}^{N_h}(x_{hi}-\bar{X}_h)^2)$; population variance of variables $(y,\ x)$ in stratum h,
$S_{xy_h} = (N_h-1)^{-1}\sum_{i=1}^{N_h}(x_{hi}-\bar{X}_h)(y_{hi}-\bar{Y}_h)$; population covariance between variables x and y in stratum h,
$\rho_{xy_h} = S_{xy_h}/S_{x_h}S_{y_h}$; population correlation coefficient for the variables x and y in stratum h,
$(C_{y_h},\ C_{x_h})$; population variation coefficients for the variables $(y,\ x)$ in stratum h.

In this article, the undermentioned notation are considered to establish the attributes of the combined point estimators.
Let $\bar{y}_{r_{st}} = \bar{Y} + \varepsilon_{0_{st}}$, $\bar{x}_{r_{st}} = \bar{X} + \varepsilon_{1_{st}}$ and $\bar{x}_{n_{st}} = \bar{X} + \varepsilon_{2_{st}}$ such that $E(\varepsilon_{0_{st}}) = E(\varepsilon_{1_{st}}) = E(\varepsilon_{2_{st}}) = 0$, $E(\varepsilon^2_{0_{st}}) = \sum_{h=1}^{L} W_h^2 \gamma_h^* S^2_{y_h} = I_0^*$, $E(\varepsilon^2_{1_{st}}) = \sum_{h=1}^{L} W_h^2 \gamma_h^* S^2_{x_h} = I_1^*$, $E(\varepsilon^2_{2_{st}}) = E(\varepsilon_{1_{st}}, \varepsilon_{2_{st}}) = \sum_{h=1}^{L} W_h^2 \gamma_h S^2_{x_h} = I_1$, $E(\varepsilon_{0_{st}}, \varepsilon_{1_{st}}) = \sum_{h=1}^{L} W_h^2 \gamma_h^* \rho_{xy_h} S_{x_h} S_{y_h} = I_{01}^*$ and $E(\varepsilon_{0_{st}}, \varepsilon_{2_{st}}) = \sum_{h=1}^{L} W_h^2 \gamma_h\ \rho_{xy_h} S_{x_h} S_{y_h} = I_{01}$, where $\gamma_h^* = (1/r_h) - (1/N_h)$ and $\gamma_h = (1/n_h) - (1/N_h)$.

The undermentioned notation are considered to determine the attributes of the separate point estimators.
Let $\bar{y}_{r_h} = \bar{Y}_h + e_{0_h}$, $\bar{x}_{r_h} = \bar{X}_h + e_{1_h}$, $\bar{x}_{n_h} = \bar{X}_h + e_{2_h}$ such that $E(e_{0_h}) = E(e_{1_h}) = E(e_{2_h}) = 0$, $E(e^2_{0_h}) = \gamma_h^* S^2_{y_h} = J_0^*$, $E(e^2_{1_h}) = \gamma_h^* S^2_{x_h} = J_1^*$, $E(e^2_{2_h}) = E(e_{1_h}, e_{2_h}) = \gamma_h S^2_{x_h} = J_1$, $E(e_{0_h}, e_{1_h}) = \gamma_h^* \rho_{xy_h} S_{x_h} S_{y_h} = J_{01}^*$ and $E(e_{0_h}, e_{2_h}) = \gamma_h \rho_{xy_h} S_{x_h} S_{y_h} = J_{01}$.

The following sections make up the article's structure. We examine the commonly used CSIMs in Section 2. We propose some CSIMs in Section 3. Section 4 provides a theoretical comparison between existing and suggested IMs. Section 5 provides the simulation study along with the discussion of simulation results. The illustration of the proposed methods is shown in Section 6. Section 7 provides the conclusion of the study.

2. Existing imputation methods

The mean IM under SSRS is

$$y_{.i_m} = \begin{cases} y_i & \text{if } i \in R_u \\ \bar{y}_{r_{st}} & \text{if } i \in \bar{R}_u \end{cases}$$

The sequent estimator is

$$T_m = \bar{y}_{r_{st}}$$

where the stratified sample mean of study variable y is given by $\bar{y}_{r_{st}} = \sum_{h=1}^{L} W_h \bar{y}_h$.
Further, we consider some prominent commonly used CSIMs.

2.1. Combined imputation methods

The IMs are divided into following strategies in the availability of auxiliary informations.
Strategy I: If $\bar{X}$ is known and $\bar{x}_{n_{st}}$ is used.
Strategy II: If $\bar{X}$ is known and $\bar{x}_{r_{st}}$ is used.
Strategy III: If $\bar{X}$ is unknown and $\bar{x}_{n_{st}}$, and $\bar{x}_{r_{st}}$ are used.
The conventional combined ratio IMs are prescribed under SSRS as
Strategy I

$$y^c_{.i_{R_1}} = \begin{cases} y_i & \text{if } i \in R_u \\ \frac{1}{n-r}\left(n\bar{y}_{r_{st}}\frac{\bar{X}}{\bar{x}_{n_{st}}} - r\bar{y}_{r_{st}}\right) & \text{if } i \in \bar{R}_u \end{cases}$$

Strategy II

$$y^c_{.i_{R_2}} = \begin{cases} y_i & \text{if } i \in R_u \\ \frac{1}{n-r}\left(n\bar{y}_{r_{st}}\frac{\bar{X}}{\bar{x}_{r_{st}}} - r\bar{y}_{r_{st}}\right) & \text{if } i \in \bar{R}_u \end{cases}$$

Strategy III

$$y^c_{.i_{R_3}} = \begin{cases} y_i & \text{if } i \in R_u \\ \frac{1}{n-r}\left(n\bar{y}_{r_{st}}\frac{\bar{x}_{n_{st}}}{\bar{x}_{r_{st}}} - r\bar{y}_{r_{st}}\right) & \text{if } i \in \bar{R}_u \end{cases}$$

The sequent estimators are given as follows:

$$T^c_{R_1} = \bar{y}_{r_{st}}\frac{\bar{X}}{\bar{x}_{n_{st}}}$$

$$T^c_{R_2} = \bar{y}_{r_{st}}\frac{\bar{X}}{\bar{x}_{r_{st}}}$$

$$T^c_{R_3} = \bar{y}_{r_{st}}\frac{\bar{x}_{n_{st}}}{\bar{x}_{r_{st}}}$$

where $\bar{x}_{n_{st}} = \sum_{h=1}^{L} W_h \bar{x}_h$ is the sample mean of the auxiliary variable x under SSRS.
Motivated by [4], we define the classical regression IM under SSRS as
Strategy I

$$y^c_{.iDP_1} = \begin{cases} \bar{y}_i & \text{if } i \in R_u \\ \bar{y}_{r_{st}} + \frac{n}{n-r} b_1(\bar{X} - \bar{x}_{n_{st}}) & \text{if } i \in \bar{R}_u \end{cases}$$

Strategy II

$$y^c_{.iDP_2} = \begin{cases} \bar{y}_i & \text{if } i \in R_u \\ \bar{y}_{r_{st}} + \frac{n}{n-r} b_2(\bar{X} - \bar{x}_{r_{st}}) & \text{if } i \in \bar{R}_u \end{cases}$$

Strategy III

$$y^c_{.iDP_3} = \begin{cases} \bar{y}_i & \text{if } i \in R_u \\ \bar{y}_{r_{st}} + \frac{n}{n-r} b_3(\bar{x}_{n_{st}} - \bar{x}_{r_{st}}) & \text{if } i \in \bar{R}_u \end{cases}$$

The sequent estimators are given as follows:

$$T^c_{DP_1} = \bar{y}_{r_{st}} + b_1(\bar{X} - \bar{x}_{n_{st}})$$

$$T^c_{DP_2} = \bar{y}_{r_{st}} + b_2(\bar{X} - \bar{x}_{r_{st}})$$

$$T^c_{DP_3} = \bar{y}_{r_{st}} + b_3(\bar{x}_{n_{st}} - \bar{x}_{r_{st}})$$

where b_j, $j = 1, 2, 3$ are the regression coefficients for the respective strategies.
On the lines of [27], [26], and [28], we investigate the following combined ratio categories of IMs based on SSRS as
Strategy I

$$y^c_{.iS_1} = \begin{cases} y_i & \text{if } i \in R_u \\ \frac{n}{n-r}\left[\bar{y}_{r_{st}}\left(\frac{\bar{X}}{\bar{x}_{n_{st}}}\right)^{\beta_1} - \bar{y}_{r_{st}}\right] & \text{if } i \in \bar{R}_u \end{cases}$$

$$y^c_{.iS_4} = \begin{cases} y_i & \text{if } i \in R_u \\ \frac{n}{n-r}\left[\bar{y}_{r_{st}}\left(\frac{\bar{X}}{\beta_4\bar{x}_{n_{st}} + (1-\beta_4)\bar{X}}\right) - \bar{y}_{r_{st}}\right] & \text{if } i \in \bar{R}_u \end{cases}$$

Strategy II

$$y^c_{.iS_2} = \begin{cases} y_i & \text{if } i \in R_u \\ \frac{n}{n-r}\left[\bar{y}_{r_{st}}\left(\frac{\bar{X}}{\bar{x}_{r_{st}}}\right)^{\beta_2} - \bar{y}_{r_{st}}\right] & \text{if } i \in \bar{R}_u \end{cases}$$

$$y^c_{.iS_5} = \begin{cases} y_i & \text{if } i \in R_u \\ \frac{n}{n-r}\left[\bar{y}_{r_{st}}\left(\frac{\bar{X}}{\beta_5\bar{x}_{r_{st}} + (1-\beta_5)\bar{X}}\right) - \bar{y}_{r_{st}}\right] & \text{if } i \in \bar{R}_u \end{cases}$$

Strategy III

$$y^c_{.iS_3} = \begin{cases} y_i & \text{if } i \in R_u \\ \frac{n}{n-r}\left[\bar{y}_{r_{st}}\left(\frac{\bar{x}_{n_{st}}}{\bar{x}_{r_{st}}}\right)^{\beta_3} - \bar{y}_{r_{st}}\right] & \text{if } i \in \bar{R}_u \end{cases}$$

$$y^c_{.iS_6} = \begin{cases} y_i & \text{if } i \in R_u \\ \frac{n}{n-r}\left[\bar{y}_{r_{st}}\left(\frac{\bar{X}}{\beta_6\bar{x}_{r_{st}} + (1-\beta_6)\bar{x}_{n_{st}}}\right) - \bar{y}_{r_{st}}\right] & \text{if } i \in \bar{R}_u \end{cases}$$

The sequent estimators are given by

$$T^c_{S_1} = \bar{y}_{r_{st}}\left(\frac{\bar{X}}{\bar{x}_{n_{st}}}\right)^{\beta_1}$$

$$T^c_{S_2} = \bar{y}_{r_{st}}\left(\frac{\bar{X}}{\bar{x}_{r_{st}}}\right)^{\beta_2}$$

$$T^c_{S_3} = \bar{y}_{r_{st}}\left(\frac{\bar{x}_{n_{st}}}{\bar{x}_{r_{st}}}\right)^{\beta_3}$$

$$T^c_{S_4} = \bar{y}_{r_{st}}\left(\frac{\bar{X}}{\beta_4\bar{x}_{n_{st}} + (1-\beta_4)\bar{X}}\right)$$

$$T^c_{S_5} = \bar{y}_{r_{st}}\left(\frac{\bar{X}}{\beta_5\bar{x}_{r_{st}} + (1-\beta_5)\bar{X}}\right)$$

$$T^c_{S_6} = \bar{y}_{r_{st}}\left(\frac{\bar{X}}{\beta_6\bar{x}_{r_{st}} + (1-\beta_6)\bar{x}_{n_{st}}}\right)$$

where β_i; $i = 1, 2, \ldots, 6$ are properly chosen scalars.
In Appendix A, the mean square error (MSE) of the subsequent estimators derived from various IMs is provided.

2.2. Separate imputation methods

The separate methods of imputation are classified into following strategies in the accessibility of auxiliary informations.
Strategy I: If $\bar{X}_h$ is known and $\bar{x}_{n_h}$ is used.
Strategy II: If $\bar{X}_h$ is known and $\bar{x}_{r_h}$ is used.
Strategy III: If $\bar{X}_h$ is unknown and $\bar{x}_{n_h}$, $\bar{x}_{r_h}$ are used.
The classical separate ratio type methods of imputation under SSRS is defined as
Strategy I

$$y^s_{.iR_1} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \frac{1}{n-r}\left(n\bar{y}_{r_h}\frac{\bar{X}_h}{\bar{x}_{n_h}} - r\bar{y}_{r_h}\right) & \text{if } i \in \bar{R}_{uh} \end{cases}$$

Strategy II

$$y^s_{.iR_2} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \frac{1}{n-r}\left(n\bar{y}_{r_h}\frac{\bar{X}_h}{\bar{x}_{r_h}} - r\bar{y}_{r_h}\right) & \text{if } i \in \bar{R}_{uh} \end{cases}$$

Strategy III

$$y^s_{.iR_3} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \frac{1}{n-r}\left(n\bar{y}_{r_h}\frac{\bar{x}_{n_h}}{\bar{x}_{r_h}} - r\bar{y}_{r_h}\right) & \text{if } i \in \bar{R}_{uh} \end{cases}$$

The sequent estimators are given by

$$T^s_{R_1} = \sum_{h=1}^{L} W_h \bar{y}_{r_h}\frac{\bar{X}_h}{\bar{x}_{n_h}}$$

$$T^s_{R_2} = \sum_{h=1}^{L} W_h \bar{y}_{r_h}\frac{\bar{X}_h}{\bar{x}_{r_h}}$$

$$T^s_{R_3} = \sum_{h=1}^{L} W_h \bar{y}_{r_h}\frac{\bar{x}_{n_h}}{\bar{x}_{r_h}}$$

Following [4], we define the separate regression IMs under SSRS as
Strategy I

$$y^s_{.iDP_1} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \bar{y}_{r_h} + \frac{n}{n-r} b_{1_h}(\bar{X}_h - \bar{x}_{n_h}) & \text{if } i \in \bar{R}_{uh} \end{cases}$$

Strategy II

$$y^s_{.iDP_2} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \bar{y}_{r_h} + \frac{n}{n-r} b_{2_h}(\bar{X}_h - \bar{x}_{r_h}) & \text{if } i \in \bar{R}_{uh} \end{cases}$$

Strategy III

$$y^s_{.iDP_3} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \bar{y}_{r_h} + \frac{n}{n-r} b_{3_h}(\bar{x}_{n_h} - \bar{x}_{r_h}) & \text{if } i \in \bar{R}_{uh} \end{cases}$$

The sequent separate estimators are given by

$$T^s_{DP_1} = \sum_{h=1}^{L} W_h[\bar{y}_{r_h} + b_{1_h}(\bar{X}_h - \bar{x}_{n_h})]$$

$$T^s_{DP_2} = \sum_{h=1}^{L} W_h[\bar{y}_{r_h} + b_{2_h}(\bar{X}_h - \bar{x}_{r_h})]$$

$$T^s_{DP_3} = \sum_{h=1}^{L} W_h[\bar{y}_{r_h} + b_{3_h}(\bar{x}_{n_h} - \bar{x}_{r_h})]$$

On the lines of [27], we investigate the following separate ratio categories of IMs based on SSRS as

Strategy I

$$y^s_{.is_1} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \frac{n}{n-r}\left[\bar{y}_{r_h}\left(\frac{\bar{X}_h}{\bar{x}_{n_h}}\right)^{\beta_{1_h}} - \bar{y}_{r_h}\right] & \text{if } i \in \bar{R}_{uh} \end{cases}$$

$$y^s_{.is_4} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \frac{n}{n-r}\left[\bar{y}_{r_h}\left(\frac{\bar{X}_h}{\beta_{4_h}\bar{x}_{n_h}+(1-\beta_{4_h})\bar{X}_h}\right) - \bar{y}_{r_h}\right] & \text{if } i \in \bar{R}_{uh} \end{cases}$$

Strategy II

$$y^s_{.is_2} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \frac{n}{n-r}\left[\bar{y}_{r_h}\left(\frac{\bar{X}_h}{\bar{x}_{r_h}}\right)^{\beta_{2_h}} - \bar{y}_{r_h}\right] & \text{if } i \in \bar{R}_{uh} \end{cases}$$

$$y^s_{.is_5} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \frac{n}{n-r}\left[\bar{y}_{r_h}\left(\frac{\bar{X}_h}{\beta_{5_h}\bar{x}_{r_h}+(1-\beta_{5_h})\bar{X}_h}\right) - \bar{y}_{r_h}\right] & \text{if } i \in \bar{R}_{uh} \end{cases}$$

Strategy III

$$y_{.is_3} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \frac{n}{n-r}\left[\bar{y}_{r_h}\left(\frac{\bar{x}_{n_h}}{\bar{x}_{r_h}}\right)^{\beta_{3_h}} - \bar{y}_{r_h}\right] & \text{if } i \in \bar{R}_{uh} \end{cases}$$

$$y^s_{.is_6} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \frac{n}{n-r}\left[\bar{y}_{r_h}\left(\frac{\bar{X}_h}{\beta_{6_h}\bar{x}_{r_h}+(1-\beta_{6_h})\bar{x}_{n_h}}\right) - \bar{y}_{r_h}\right] & \text{if } i \in \bar{R}_{uh} \end{cases}$$

The sequent estimators are given by

$$T^s_{S_1} = \sum_{h=1}^{L} W_h\bar{y}_{r_h}\left(\frac{\bar{X}_h}{\bar{x}_{n_h}}\right)^{\beta_{1_h}}$$

$$T^s_{S_2} = \sum_{h=1}^{L} W_h\bar{y}_{r_h}\left(\frac{\bar{X}_h}{\bar{x}_{r_h}}\right)^{\beta_{2_h}}$$

$$T^s_{S_3} = \sum_{h=1}^{L} W_h\bar{y}_{r_h}\left(\frac{\bar{x}_{n_h}}{\bar{x}_{r_h}}\right)^{\beta_{3_h}}$$

$$T^s_{S_4} = \sum_{h=1}^{L} W_h\bar{y}_{r_h}\left(\frac{\bar{X}_h}{\beta_{4_h}\bar{x}_{n_h} + (1-\beta_{4_h})\bar{X}_h}\right)$$

$$T^s_{S_5} = \sum_{h=1}^{L} W_h\bar{y}_{r_h}\left(\frac{\bar{X}_h}{\beta_{5_h}\bar{x}_{r_h} + (1-\beta_{5_h})\bar{X}_h}\right)$$

$$T^s_{S_6} = \sum_{h=1}^{L} W_h\bar{y}_{r_h}\left(\frac{\bar{X}_h}{\beta_{6_h}\bar{x}_{r_h} + (1-\beta_{6_h})\bar{x}_{n_h}}\right)$$

where β_{i_h}; $i = 1, 2, ..., 6$ are properly chosen scalars.

In Appendix B, the minimum MSE of the sequent estimators made up of several separate IMs is listed.

3. Suggested imputation methods

In the previous section, the IMs do not compete with the regression IM which is BLUE. The nub of this section is to suggest a few novel combined and separate population mean methods of imputation in case of missing data in SSRS which can compete with the BLUE. Motivated by [5], [29], [21], [10], and [30], we propose the following CSIMs based on SSRS.

3.1. Combined imputation methods

We suggest nine novel combined methods of imputation in the strategies described in the former section as

Strategy I

$$y_{.i_{ak_1}} = \begin{cases} \alpha_1 y_i & \text{if } i \in R_u \\ \alpha_1\bar{y}_{r_{st}} + \frac{n\theta_1}{n-r}(\bar{x}_{n_{st}} - \bar{X}) & \text{if } i \in \bar{R}_u \end{cases}$$

$$y_{.i_{ak_4}} = \begin{cases} y_i & \text{if } i \in R_u \\ \frac{1}{n-r}\left[n\alpha_4\bar{y}_{r_{st}}\left(\frac{\bar{X}}{\bar{x}_{n_{st}}}\right)^{\theta_4} - r\bar{y}_{r_{st}}\right] & \text{if } i \in \bar{R}_u \end{cases}$$

$$y_{.i_{ak_7}} = \begin{cases} y_i & \text{if } i \in R_u \\ \frac{1}{n-r}\left[n\alpha_7\bar{y}_{r_{st}}\left(\frac{\bar{X}}{\bar{X}+\theta_7(\bar{x}_{n_{st}}-\bar{X})}\right) - r\bar{y}_{r_{st}}\right] & \text{if } i \in \bar{R}_u \end{cases}$$

Strategy II

$$y_{.i_{ak_2}} = \begin{cases} \alpha_2 y_i & \text{if } i \in R_u \\ \alpha_2\bar{y}_{r_{st}} + \frac{n\theta_2}{n-r}(\bar{x}_{r_{st}} - \bar{X}) & \text{if } i \in \bar{R}_u \end{cases}$$

$$y_{.i_{ak_5}} = \begin{cases} y_i & \text{if } i \in R_u \\ \frac{1}{n-r}\left[n\alpha_5\bar{y}_{r_{st}}\left(\frac{\bar{X}}{\bar{x}_{r_{st}}}\right)^{\theta_5} - r\bar{y}_{r_{st}}\right] & \text{if } i \in \bar{R}_u \end{cases}$$

$$y_{.i_{ak_8}} = \begin{cases} y_i & \text{if } i \in R_u \\ \frac{1}{n-r}\left[n\alpha_8\bar{y}_{r_{st}}\left(\frac{\bar{X}}{\bar{X}+\theta_8(\bar{x}_{r_{st}}-\bar{X})}\right) - r\bar{y}_{r_{st}}\right] & \text{if } i \in \bar{R}_u \end{cases}$$

Strategy III

$$y_{.i_{ak_3}} = \begin{cases} \alpha_3 y_i & \text{if } i \in R_u \\ \alpha_3\bar{y}_{r_{st}} + \frac{n\theta_3}{n-r}(\bar{x}_{r_{st}} - \bar{x}_{n_{st}}) & \text{if } i \in \bar{R}_u \end{cases}$$

$$y_{.i_{ak_6}} = \begin{cases} y_i & \text{if } i \in R_u \\ \frac{1}{n-r}\left[n\alpha_6\bar{y}_{r_{st}}\left(\frac{\bar{x}_{n_{st}}}{\bar{x}_{r_{st}}}\right)^{\theta_6} - r\bar{y}_{r_{st}}\right] & \text{if } i \in \bar{R}_u \end{cases}$$

$$y_{.i_{ak_9}} = \begin{cases} y_i & \text{if } i \in R_u \\ \frac{1}{n-r}\left[n\alpha_9\bar{y}_{r_{st}}\left(\frac{\bar{x}_{n_{st}}}{\bar{x}_{n_{st}}+\theta_9(\bar{x}_{n_{st}}-\bar{x}_{r_{st}})}\right) - r\bar{y}_{r_{st}}\right] & \text{if } i \in \bar{R}_u \end{cases}$$

Under the above strategies, the point estimators are given as follows:

$$T^c_{ak_1} = \alpha_1\bar{y}_{r_{st}} + \theta_1(\bar{x}_{n_{st}} - \bar{X})$$

$$T^c_{ak_2} = \alpha_2\bar{y}_{r_{st}} + \theta_2(\bar{x}_{r_{st}} - \bar{X})$$

$$T^c_{ak_3} = \alpha_3\bar{y}_{r_{st}} + \theta_3(\bar{x}_{r_{st}} - \bar{x}_{n_{st}})$$

$$T^c_{ak_4} = \alpha_4\bar{y}_{r_{st}}\left(\frac{\bar{X}}{\bar{x}_{n_{st}}}\right)^{\theta_4}$$

$$T^c_{ak_5} = \alpha_5\bar{y}_{r_{st}}\left(\frac{\bar{X}}{\bar{x}_{r_{st}}}\right)^{\theta_5}$$

$$T^c_{ak_6} = \alpha_6\bar{y}_{r_{st}}\left(\frac{\bar{x}_{n_{st}}}{\bar{x}_{r_{st}}}\right)^{\theta_6}$$

$$T^c_{ak_7} = \alpha_7 \bar{y}_{r_{st}} \left[\frac{\bar{X}}{\bar{X} + \theta_7(\bar{x}_{n_{st}} - \bar{X})} \right]$$

$$T^c_{ak_8} = \alpha_8 \bar{y}_{r_{st}} \left[\frac{\bar{X}}{\bar{X} + \theta_8(\bar{x}_{r_{st}} - \bar{X})} \right]$$

$$T^c_{ak_9} = \alpha_9 \bar{y}_{r_{st}} \left[\frac{\bar{x}_{n_{st}}}{\bar{x}_{n_{st}} + \theta_9(\bar{x}_{r_{st}} - \bar{x}_{n_{st}})} \right]$$

where α_j and θ_j; $j = 1, 2, ..., 9$ have been properly chosen scalars. Notably, the suggested combined IMs $y^c_{.i_{ak_j}}$, $j = 1, 2, ..., 9$ reduce to the current combined IMs for known values of scalars as

1. Conventional mean IM $y^c_{.i_m}$ for $(\alpha_j, \theta_j;\ j = 4, 5, 6) = (1, 0)$.
2. Classical ratio IM $y^c_{.i_{R_j}}$, $j = 1, 2, 3$ for $(\alpha_j, \theta_j;\ j = 4, 5, 6) = (1, 1)$.
3. Imputation techniques envisaged on the lines of [27] $y^c_{.i_{S_j}}$, $j = 1, 2, ..., 6$ for $(\alpha_j, \theta_j;\ j = 4, 5, ..., 9) = (\alpha_j, 1)$.
4. Imputation techniques envisaged on the lines of [4] $y^c_{.i_{DP_j}}$, $j = 1, 2, 3$ for $(\alpha_j, \theta_j;\ j = 1, 2, 3) = (1, b_j)$.

3.1.1. MSE and minimum MSE of the proposed combined estimators

The MSE of the sequent estimators consisting of the suggested methods imputation is provided as

$$MSE(T^c_{ak_1}) = (\alpha_1 - 1)^2 \bar{Y}^2 + \alpha_1^2 \bar{Y}^2 I_0^* + \theta_1^2 \bar{X}^2 I_1 + 2\alpha_1 \theta_1 \bar{X}\bar{Y} I_{01}$$

$$MSE(T^c_{ak_2}) = (\alpha_2 - 1)^2 \bar{Y}^2 + \alpha_2^2 \bar{Y}^2 I_0^* + \theta_2^2 \bar{X}^2 I_1^* + 2\alpha_2 \theta_2 \bar{X}\bar{Y} I_{01}^*$$

$$MSE(T^c_{ak_3}) = \left[\begin{array}{l} (\alpha_3 - 1)^2 \bar{Y}^2 + \alpha_3^2 \bar{Y}^2 I_0^* + \theta_3^2 \bar{X}^2 \left\{ I_1^* - I_1 \right\} \\ +2\alpha_3 \theta_3 \bar{X}\bar{Y} \left\{ I_{01}^* - I_{01} \right\} \end{array} \right]$$

$$MSE(T^c_{ak_4}) = \bar{Y}^2 \left[\begin{array}{l} 1 + \alpha_4^2 \left\{ 1 + I_0^* + \theta_4(2\theta_4 + 1) I_1 - 4\theta_4 I_{01} \right\} \\ -2\alpha_4 \left\{ 1 - \theta_4 I_{01} + \frac{\theta_4(\theta_4+1)}{2} I_1 \right\} \end{array} \right]$$

$$MSE(T^c_{ak_5}) = \bar{Y}^2 \left[\begin{array}{l} 1 + \alpha_5^2 \left\{ 1 + I_0^* + \theta_5(2\theta_5 + 1) I_1^* - 4\theta_5 I_{01}^* \right\} \\ -2\alpha_5 \left\{ 1 - \theta_5 I_{01}^* + \frac{\theta_5(\theta_5+1)}{2} I_1^* \right\} \end{array} \right]$$

$$MSE(T^c_{ak_6}) = \bar{Y}^2 \left[\begin{array}{l} 1 + \alpha_6^2 \left\{ 1 + I_0^* + \theta_6(2\theta_6 + 1)(I_1^* - I_1) - 4\theta_6 (I_{01}^* - I_{01}) \right\} \\ -2\alpha_6 \left\{ 1 - \theta_6 (I_{01}^* - I_{01}) + \frac{\theta_6(\theta_6+1)}{2} (I_1^* - I_1) \right\} \end{array} \right]$$

$$MSE(T^c_{ak_7}) = \bar{Y}^2 \left[\begin{array}{l} 1 + \alpha_7^2 \left\{ 1 + I_0^* + 3\theta_7^2 I_1 - 4\theta_7 I_{01} \right\} \\ -2\alpha_7 \left\{ 1 + \theta_7^2 I_1 - \theta_7 I_{01} \right\} \end{array} \right]$$

$$MSE(T^c_{ak_8}) = \bar{Y}^2 \left[\begin{array}{l} 1 + \alpha_8^2 \left\{ 1 + I_0^* + 3\theta_8^2 I_1^* - 4\theta_8 I_{01}^* \right\} \\ -2\alpha_8 \left\{ 1 + \theta_8^2 I_1^* - \theta_8 I_{01}^* \right\} \end{array} \right]$$

$$MSE(T^c_{ak_9}) = \bar{Y}^2 \left[\begin{array}{l} 1 + \alpha_9^2 \left\{ 1 + I_0^* + 3\theta_9^2 (I_1^* - I_1) - 4\theta_9 (I_{01}^* - I_{01}) \right\} \\ -2\alpha_9 \left\{ 1 + \theta_9^2 (I_1^* - I_1) - \theta_9 (I_{01}^* - I_{01}) \right\} \end{array} \right]$$

The minimum MSE of the sequent estimators consisting of the suggested IMs is provided by

$$minMSE(T^c_{ak_j}) = \bar{Y}^2 (1 - \alpha_{j(opt)}) = \bar{Y}^2 \left(1 - \frac{A_j^2}{B_j} \right);\ j = 1, 2, 3 \tag{3.1}$$

$$minMSE(T^c_{ak_j}) = \bar{Y}^2 \left(1 - \frac{A_j^2}{B_j} \right);\ j = 4, 5, 6, 7, 8, 9 \tag{3.2}$$

Appendix C contains a the derivations of these MSE expressions along with brief annotations.

3.2. Separate imputation methods

We suggest the following new separate IMs under the strategies described primarily as

Strategy I

$$y^s_{.i_{ak_1}} = \begin{cases} \alpha_{1_h} y_i & \text{if } i \in R_{uh} \\ \alpha_{1_h} \bar{y}_{r_h} + \frac{n\theta_{1_h}}{n-r} (\bar{x}_{n_h} - \bar{X}_h) & \text{if } i \in \bar{R}_{uh} \end{cases}$$

$$y^s_{.i_{ak_4}} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \frac{1}{n-r} \left[n\alpha_{4_h} \bar{y}_{r_h} \left(\frac{\bar{X}_h}{\bar{x}_{n_h}} \right)^{\theta_{4_h}} - r\bar{y}_{r_h} \right] & \text{if } i \in \bar{R}_{uh} \end{cases}$$

$$y^s_{.i_{ak_7}} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \frac{1}{n-r} \left[n\alpha_{7_h} \bar{y}_{r_h} \left(\frac{\bar{X}_h}{\bar{X}_h + \theta_{7_h}(\bar{x}_{n_h} - \bar{X}_h)} \right) - r\bar{y}_{r_h} \right] & \text{if } i \in \bar{R}_{uh} \end{cases}$$

Strategy II

$$y^s_{.i_{ak_2}} = \begin{cases} \alpha_{2_h} y_i & \text{if } i \in R_{uh} \\ \alpha_{2_h} \bar{y}_{r_h} + \frac{n\theta_{2_h}}{n-r} (\bar{x}_{r_h} - \bar{X}_h) & \text{if } i \in \bar{R}_{uh} \end{cases}$$

$$y^s_{.i_{ak_5}} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \frac{1}{n-r} \left[n\alpha_{5_h} \bar{y}_{r_h} \left(\frac{\bar{X}_h}{\bar{x}_{r_h}} \right)^{\theta_{5_h}} - r\bar{y}_{r_h} \right] & \text{if } i \in \bar{R}_{uh} \end{cases}$$

$$y^s_{.i_{ak_8}} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \frac{1}{n-r} \left[n\alpha_{8_h} \bar{y}_{r_h} \left(\frac{\bar{X}_h}{\bar{X}_h + \theta_{8_h}(\bar{x}_{r_h} - \bar{X}_h)} \right) - r\bar{y}_{r_h} \right] & \text{if } i \in \bar{R}_{uh} \end{cases}$$

Strategy III

$$y^s_{.i_{ak_3}} = \begin{cases} \alpha_{3_h} y_i & \text{if } i \in R_{uh} \\ \alpha_{3_h} \bar{y}_{r_h} + \frac{n\theta_{3_h}}{n-r} (\bar{x}_{r_h} - \bar{x}_{n_h}) & \text{if } i \in \bar{R}_{uh} \end{cases}$$

$$y^s_{.i_{ak_6}} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \frac{1}{n-r} \left[n\alpha_{6_h} \bar{y}_{r_h} \left(\frac{\bar{x}_{n_h}}{\bar{x}_{r_h}} \right)^{\theta_{6_h}} - r\bar{y}_{r_h} \right] & \text{if } i \in \bar{R}_{uh} \end{cases}$$

$$y^s_{.i_{ak_9}} = \begin{cases} y_i & \text{if } i \in R_{uh} \\ \frac{1}{n-r} \left[n\alpha_{9_h} \bar{y}_{r_h} \left(\frac{\bar{x}_{n_h}}{\bar{x}_{n_h} + \theta_{9_h}(\bar{x}_{n_h} - \bar{x}_{r_h})} \right) - r\bar{y}_{r_h} \right] & \text{if } i \in \bar{R}_{uh} \end{cases}$$

The sequent estimators are given by

$$T^s_{ak_1} = \sum_{h=1}^{L} W_h [\alpha_{1_h} \bar{y}_{r_h} + \theta_{1_h} (\bar{x}_{n_h} - \bar{X}_h)]$$

$$T^s_{ak_2} = \sum_{h=1}^{L} W_h [\alpha_{2_h} \bar{y}_{r_h} + \theta_{2_h} (\bar{x}_{r_h} - \bar{X}_h)]$$

$$T^s_{ak_3} = \sum_{h=1}^{L} W_h [\alpha_{3_h} \bar{y}_{r_h} + \theta_{3_h} (\bar{x}_{r_h} - \bar{x}_{n_h})]$$

$$T^s_{ak_4} = \sum_{h=1}^{L} W_h \alpha_{4_h} \bar{y}_{r_h} \left(\frac{\bar{X}_h}{\bar{x}_{n_h}} \right)^{\theta_{4_h}}$$

$$T^s_{ak_5} = \sum_{h=1}^{L} W_h \alpha_{5_h} \bar{y}_{r_h} \left(\frac{\bar{X}_h}{\bar{x}_{r_h}} \right)^{\theta_{5_h}}$$

$$T^s_{ak_6} = \sum_{h=1}^{L} W_h \alpha_{6_h} \bar{y}_{r_h} \left(\frac{\bar{x}_{n_h}}{\bar{x}_{r_h}} \right)^{\theta_{6_h}}$$

$$T^s_{ak_7} = \sum_{h=1}^{L} W_h \alpha_{7_h} \bar{y}_{r_h} \left[\frac{\bar{X}_h}{\bar{X}_h + \theta_{7_h} (\bar{x}_{n_h} - \bar{X}_h)} \right]$$

$$T^s_{ak_8} = \sum_{h=1}^{L} W_h \alpha_{8_h} \bar{y}_{r_h} \left[\frac{\bar{X}_h}{\bar{X}_h + \theta_{8_h} (\bar{x}_{r_h} - \bar{X}_h)} \right]$$

$$T^s_{ak_9} = \sum_{h=1}^{L} W_h \alpha_{9_h} \bar{y}_{r_h} \left[\frac{\bar{x}_{n_h}}{\bar{x}_{n_h} + \theta_{9_h} (\bar{x}_{r_h} - \bar{x}_{n_h})} \right]$$

where α_{j_h} and θ_{j_h}; $j=1,2,\dots 9$ are the properly chosen scalars.
It is important to note that the suggested separate IMs $y^{s}_{.i_{ak_j}}$, $j=1,2,\dots,9$ reduce to the existing separate IMs for known values of scalars as

1. Conventional mean IM $y^{c}_{.i_m}$ for $(\alpha_{j_h},\theta_{j_h};\ j=4,5,6)=(1,0)$.
2. Classical ratio IM $y^{c}_{.i_{R_{j_h}}}$, $j=1,2,3$ for $(\alpha_{j_h},\theta_{j_h};\ j=4,5,6)=(1,1)$.
3. Imputation techniques envisaged on the lines of [27] $y^{c}_{.i_{S_{j_h}}}$, $j=1,2,\dots,6$ for $(\alpha_{j_h},\theta_{j_h};\ j=4,5,\dots,9)=(\alpha_{j_h},1)$.
4. Imputation techniques envisaged on the lines of [4] $y^{c}_{.i_{DP_{j_h}}}$, $j=1,2,3$ for $(\alpha_{j_h},\theta_{j_h};\ j=1,2,3)=(1,b_{j_h})$.

3.2.1. MSE and minimum MSE of the proposed separate estimators

The MSE of the sequent estimators based on the suggested IMs is given by

$$MSE(T^s_{ak_1}) = \sum_{h=1}^{L} W_h^2\left\{(\alpha_{1_h}-1)^2\bar{Y}_h^2+\alpha_{1_h}^2\bar{Y}_h^2J_0^*+\theta_{1_h}^2\bar{X}_h^2J_1+2\alpha_{1_h}\theta_{1_h}\bar{X}_h\bar{Y}_hJ_{01}\right\}$$

$$MSE(T^s_{ak_2}) = \sum_{h=1}^{L} W_h^2\left\{(\alpha_{2_h}-1)^2\bar{Y}_h^2+\alpha_{2_h}^2\bar{Y}_h^2J_0^*+\theta_{2_h}^2\bar{X}_h^2J_1^*+2\alpha_{2_h}\theta_{2_h}\bar{X}_h\bar{Y}_hJ_{01}^*\right\}$$

$$MSE(T^s_{ak_3}) = \sum_{h=1}^{L} W_h^2\left[\begin{array}{l}(\alpha_{3_h}-1)^2\bar{Y}_h^2+\alpha_{3_h}^2\bar{Y}_h^2J_0^*+\theta_{3_h}^2\bar{X}_h^2\left\{J_1^*-J_1\right\}\\ +2\alpha_{3_h}\theta_{3_h}\bar{X}_h\bar{Y}_h\left\{J_{01}^*-J_{01}\right\}\end{array}\right]$$

$$MSE(T^s_{ak_4}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[\begin{array}{l}1+\alpha_{4_h}^2\left\{1+J_0^*+\theta_{4_h}(2\theta_{4_h}+1)J_1-4\theta_{4_h}J_{01}\right\}\\ -2\alpha_{4_h}\left\{1-\theta_{4_h}J_{01}+\frac{\theta_{4_h}(\theta_{4_h}+1)}{2}J_1\right\}\end{array}\right]$$

$$MSE(T^s_{ak_5}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[\begin{array}{l}1+\alpha_{5_h}^2\left\{1+J_0^*+\theta_{5_h}(2\theta_{5_h}+1)J_1^*-4\theta_{5_h}J_{01}^*\right\}\\ -2\alpha_{5_h}\left\{1-\theta_{5_h}J_{01}^*+\frac{\theta_{5_h}(\theta_{5_h}+1)}{2}J_1^*\right\}\end{array}\right]$$

$$MSE(T^s_{ak_6}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[\begin{array}{l}1+\alpha_{6_h}^2\left\{\begin{array}{l}1+J_0^*+\theta_{6_h}(2\theta_{6_h}+1)(J_1^*-J_1)\\ -4\theta_{6_h}(J_{01}^*-J_{01})\end{array}\right\}\\ -2\alpha_{6_h}\left\{1-\theta_{6_h}(J_{01}^*-J_{01})+\frac{\theta_{6_h}(\theta_{6_h}+1)}{2}(J_1^*-J_1)\right\}\end{array}\right]$$

$$MSE(T^s_{ak_7}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[\begin{array}{l}1+\alpha_{7_h}^2\left\{1+J_0^*+3\theta_{7_h}^2J_1-4\theta_{7_h}J_{01}\right\}\\ -2\alpha_{7_h}\left\{1+\theta_{7_h}^2J_1-\theta_{7_h}J_{01}\right\}\end{array}\right]$$

$$MSE(T^s_{ak_8}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[\begin{array}{l}1+\alpha_{8_h}^2\left\{1+J_0^*+3\theta_{8_h}^2J_1^*-4\theta_{8_h}J_{01}^*\right\}\\ -2\alpha_{8_h}\left\{1+\theta_{8_h}^2J_1^*-\theta_{8_h}J_{01}^*\right\}\end{array}\right]$$

$$MSE(T^s_{ak_9}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[\begin{array}{l}1+\alpha_{9_h}^2\left\{\begin{array}{l}1+J_0^*+3\theta_{9_h}^2(J_1^*-J_1)\\ -4\theta_{9_h}(J_{01}^*-J_{01})\end{array}\right\}\\ -2\alpha_{9_h}\left\{1+\theta_{9_h}^2(J_1^*-J_1)-\theta_{9_h}(J_{01}^*-J_{01})\right\}\end{array}\right]$$

The minimum $MSEs$ of the sequent estimators consisting of the proffered IMs is given as

$$minMSE(T^s_{ak_j}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2(1-\alpha_{j_h(opt)});\ j=1,2,3 \tag{3.3}$$

$$minMSE(T^s_{ak_j}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left(1-\frac{A_{j_h}^2}{B_{j_h}}\right);\ j=4,5,6,7,8,9 \tag{3.4}$$

Appendix C contains the derivations of these MSE expressions along with brief annotations.

4. Theoretical conditions

4.1. Combined imputation methods

The following theoretical conditions are obtained by comparing the minimum mean square error of the suggested combined IMs $y^{c}_{.i_{ak_j}}$, $j=1,2,\dots,9$ provided in (3.1) and (3.2) with the minimum mean square error of the conventional combined IMs provided from (A.1) to (A.10).

$$MSE(T_m) > MSE(T^c_{ak_j}) \implies \frac{A_j^2}{B_j} > 1-I_0^*,\ j=1,2,\dots,9$$

$$MSE(T^c_{R_1}) > MSE(T^c_{ak_j}) \implies \frac{A_j^2}{B_j} > 1-I_0^*-I_1+2I_{01}$$

$$MSE(T^c_{R_2}) > MSE(T^c_{ak_j}) \implies \frac{A_j^2}{B_j} > 1-I_0^*-I_1^*+2I_{01}^*$$

$$MSE(T^c_{R_3}) > MSE(T^c_{ak_j}) \implies \frac{A_j^2}{B_j} > 1-I_0^*+I_1-I_1^*+2(I_{01}^*-I_{01})$$

$$MSE(T^c_{DP_1}) > MSE(T^c_{ak_j}) \implies \frac{A_j^2}{B_j} > 1-I_0^*+\frac{I_{01}^2}{I_1}$$

$$MSE(T^c_{DP_2}) > MSE(T^c_{ak_j}) \implies \frac{A_j^2}{B_j} > 1-I_0^*+\frac{I_{01}^{*2}}{I_1^*}$$

$$MSE(T^c_{DP_3}) > MSE(T^c_{ak_j}) \implies \frac{A_j^2}{B_j} > 1-I_0^*+\frac{(I_{01}^*-I_{01})^2}{(I_1^*-I_1)}$$

$$MSE(T^c_{S_1}) > MSE(T^c_{ak_j}) \implies \frac{A_j^2}{B_j} > 1-I_0^*+\frac{I_{01}^2}{I_1}$$

$$MSE(T^c_{S_2}) > MSE(T^c_{ak_j}) \implies \frac{A_j^2}{B_j} > 1-I_0^*+\frac{I_{01}^{*2}}{I_1^*}$$

$$MSE(T^c_{S_3}) > MSE(T^c_{ak_j}) \implies \frac{A_j^2}{B_j} > 1-I_0^*+\frac{(I_{01}^*-I_{01})^2}{(I_1^*-I_1)}$$

The proposed combined IMs will be reasonable provided these conditions hold.

4.2. Separate imputation methods

The following theoretical conditions are obtained by comparing the minimum mean square error of the suggested combined IMs $y^{s}_{.i_{ak_j}}$, $j=1,2,\dots,9$ provided in (3.3) and (3.4) with the minimum mean square error of the conventional combined IMs provided in (A.1) to (B.19).

$$MSE(T_m) > MSE(T^s_{ak_j}) \implies \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left(1-\frac{A_{j_h}^2}{B_{j_h}}\right) < \bar{Y}^2 I_0^*,\ j=1,2,\dots,9$$

$$MSE(T^s_{R_1}) > MSE(T^s_{ak_j}) \implies \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left(1-\frac{A_{j_h}^2}{B_{j_h}}\right) < \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[J_0^*+J_1-2J_{01}\right]$$

$$MSE(T^s_{R_2}) > MSE(T^s_{ak_j}) \implies \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left(1-\frac{A_{j_h}^2}{B_{j_h}}\right) < \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[J_0^*+J_1^*-2J_{01}^*\right]$$

$$MSE(T^s_{R_3}) > MSE(T^s_{ak_j}) \implies \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 \left(1 - \frac{A^2_{j_h}}{B_{j_h}}\right) < \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 \begin{bmatrix} J_0^* + J_1^* - J_1 \\ -2(J_{01}^* - J_{01}) \end{bmatrix}$$

$$MSE(T^s_{DP_1}) > MSE(T^s_{ak_j}) \implies \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 \left(1 - \frac{A^2_{j_h}}{B_{j_h}}\right) < \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 \left[J_0^* - \frac{J_{01}^2}{J_1}\right]$$

$$MSE(T^s_{DP_2}) > MSE(T^s_{ak_j}) \implies \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 \left(1 - \frac{A^2_{j_h}}{B_{j_h}}\right) < \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 \left[J_0^* - \frac{J_{01}^{*2}}{J_1^*}\right]$$

$$MSE(T^s_{DP_3}) > MSE(T^s_{ak_j}) \implies \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 \left(1 - \frac{A^2_{j_h}}{B_{j_h}}\right) < \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 \left[J_0^* - \frac{\left(J_{01}^* - J_{01}\right)^2}{\left(J_1^* - J_1\right)}\right]$$

$$MSE(T^s_{S_1}) > MSE(T^s_{ak_j}) \implies \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 \left(1 - \frac{A^2_{j_h}}{B_{j_h}}\right) < \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 \left[J_0^* - \frac{J_{01}^2}{J_1}\right]$$

$$MSE(T^s_{S_2}) > MSE(T^s_{ak_j}) \implies \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 \left(1 - \frac{A^2_{j_h}}{B_{j_h}}\right) < \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 \left[J_0^* - \frac{J_{01}^{*2}}{J_1^*}\right]$$

$$MSE(T^s_{S_3}) > MSE(T^s_{ak_j}) \implies \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 \left(1 - \frac{A^2_{j_h}}{B_{j_h}}\right) < \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 \left[J_0^* - \frac{\left(J_{01}^* - J_{01}\right)^2}{\left(J_1^* - J_1\right)}\right]$$

The suggested separate IMs can be reasonable provided the above conditions hold.

4.3. Comparison of suggested CSIMs

The following condition is obtained by comparing the minimum MSE of the suggested combined and separate classes of estimators provided in (3.1), (3.2) and (3.3), (3.4).

$$minMSE(T^c_{ak_j}) - minMSE(T^s_{ak_j}) = \sum_{h=1}^{L} \left[(\bar{Y}^2 - W_h^2 \bar{Y}_h^2) - \left(\bar{Y}^2 \frac{A_j^2}{B_j} - W_h^2 \bar{Y}_h^2 \frac{A^2_{j_h}}{B_{j_h}}\right)\right] \tag{4.1}$$

The last term of (4.1) is generally insignificant and it becomes faded in if:

1. the sequent estimators are conclusive;
2. within each stratum, if both the variables (x, y) are linearly related and the regression line goes through the origin.

The suggested combined estimators are to be strongly preferred having very small sample in every stratum, as the separate estimators outperform in every stratum with the exception of R_h becoming stable from stratum to stratum, given the sample in every stratum is to be enough big so that the approximate expression of $MSE(T^s_{ak_j})$, $j = 1, 2, ..., 9$ is valid.

Table 1
PRE of suggested combined estimators.

Correlation coefficient	0.6	0.7	0.8	0.9
T_m	100	100	100	100
$x^* \sim N(2,7)$				
$y^* \sim N(9,11)$				
Strategy I				
$T^c_{R_1}$	55.2264	55.6970	56.8299	58.5490
$T^c_{DP_1}/T^c_{S_i}$, $i = 1,4$	100.6625	100.8165	100.6375	100.5616
$T^c_{ak_j}$, $j = 1,7$	110.7267	110.8560	111.3526	112.1161
$T^c_{ak_4}$	111.4849	111.6909	112.1184	112.8627
Strategy II				
$T^c_{R_2}$	36.8794	37.3240	38.4072	40.0865
$T^c_{DP_2}/T^c_{S_i}$, $i = 1,4$	101.4089	101.7394	101.3554	101.1931
$T^c_{ak_j}$, $j = 2,8$	111.4732	111.7789	112.0705	112.7476
$T^c_{ak_5}$	113.1226	113.6026	113.7344	114.3652
Strategy III				
$T^c_{R_3}$	52.6092	53.0839	54.2288	55.9712
$T^c_{DP_3}/T^c_{S_i}$, $i = 3,6$	100.7366	100.9080	100.7088	100.6244
$T^c_{ak_j}$, $j = 3,9$	110.8009	110.9475	111.4239	112.1789
$T^c_{ak_6}$	111.6458	111.8783	112.2773	113.0106
$x^* \sim Exp(1.0)$				
$y^* \sim Exp(0.5)$				
Strategy I				
$T^c_{R_1}$	52.6742	49.0842	51.2637	55.3784
$T^c_{DP_1}/T^c_{S_i}$, $i = 1,4$	100.3880	100.4866	100.4600	100.5303
$T^c_{ak_j}$, $j = 1,7$	100.7498	100.7982	100.7993	100.9215
$T^c_{ak_4}$	100.7715	100.8209	100.8222	100.9477
Strategy II				
$T^c_{R_2}$	34.5214	31.3491	33.2554	37.0227
$T^c_{DP_2}/T^c_{S_i}$, $i = 1,4$	100.8228	101.0329	100.9761	101.1263
$T^c_{ak_j}$, $j = 2,8$	101.1845	101.3445	101.3153	101.5174
$T^c_{ak_5}$	101.2314	101.3935	101.3649	101.574
Strategy III				
$T^c_{R_3}$	50.0427	46.4561	48.6303	52.7625
$T^c_{DP_3}/T^c_{S_i}$, $i = 3,6$	100.4313	100.541	100.5114	100.5896
$T^c_{ak_j}$, $j = 3,9$	100.7931	100.8526	100.8506	100.9808
$T^c_{ak_6}$	100.8173	100.8778	100.8761	101.0099
$x^* \sim \chi^2(7)$				
$y^* \sim \chi^2(11)$				
Strategy I				
$T^c_{R_1}$	37.7957	36.9291	38.9137	40.5873
$T^c_{DP_1}/T^c_{S_i}$, $i = 1,4$	100.1268	100.3388	100.3466	100.8760
$T^c_{ak_j}$, $j = 1,7$	100.9514	101.1073	101.1787	101.7163
$T^c_{ak_4}$	100.9896	101.1671	101.2414	101.8149
Strategy II				
$T^c_{R_2}$	22.3490	21.7129	23.1803	24.4481
$T^c_{DP_2}/T^c_{S_i}$, $i = 1,4$	100.2680	100.7179	100.7345	101.8676
$T^c_{ak_j}$, $j = 2,8$	101.0927	101.4864	101.5666	102.7079
$T^c_{ak_5}$	101.1743	101.6157	101.7022	102.9248
Strategy III				
$T^c_{R_3}$	35.3523	34.5106	36.4404	38.0739
$T^c_{DP_3}/T^c_{S_i}$, $i = 3,6$	100.1409	100.3765	100.3852	100.9743
$T^c_{ak_j}$, $j = 3,9$	100.9655	101.1451	101.2173	101.8146
$T^c_{ak_6}$	101.008	101.2117	101.2872	101.9246

5. Simulation study

To stimulate the soundness of the theoretical findings, motivated by [31], [32], and [33], a simulation study is presented by utilizing three artificial populations, namely, normal, exponential, and chi-square, each consisting of $N = 300$ units having the variables x and y which can be obtained as

$$y_i = 8.2 + \sqrt{(1 - \rho_{xy}^2)}\, y_i^* + \rho_{xy} \left(\frac{S_y}{S_x}\right) x_i^*$$

Table 2
PRE of suggested separate estimators.

Correlation coefficient	0.6	0.7	0.8	0.9
T_m	100	100	100	100
$x^* \sim N(2,7)$				
$y^* \sim N(4,9)$				
Strategy I				
$T^s_{R_1}$	47.4202	48.2471	48.4044	50.0764
$T^s_{DP_1}/T^s_{S_i}$, $i=1,4$	100.6625	100.8165	100.6375	100.5616
$T^s_{ak_j}$, $j=1,7$	107.5355	107.7830	107.7864	108.2394
$T^s_{ak_4}$	108.1420	108.4567	108.3925	108.8289
Strategy II				
$T^s_{R_2}$	29.9329	30.6324	30.7665	32.2095
$T^s_{DP_2}/T^s_{S_i}$, $i=1,4$	101.4089	101.7394	101.3554	101.1931
$T^s_{ak_j}$, $j=2,8$	108.282	108.7059	108.5043	108.8708
$T^s_{ak_5}$	109.6051	110.1820	109.8252	110.1515
Strategy III				
$T^s_{R_3}$	44.8028	45.6236	45.7799	47.4446
$T^s_{DP_3}/T^s_{S_i}$, $i=3,6$	100.7366	100.908	100.7088	100.6244
$T^s_{ak_j}$, $j=3,9$	107.6097	107.8745	107.8577	108.3022
$T^s_{ak_6}$	108.2858	108.6259	108.5333	108.9590
$x^* \sim Exp(1.0)$				
$y^* \sim Exp(0.5)$				
Strategy I				
$T^s_{R_1}$	49.3701	45.5457	47.9883	52.5620
$T^s_{DP_1}/T^s_{S_i}$, $i=1,4$	100.3880	100.4866	100.4600	100.5303
$T^s_{ak_j}$, $j=1,7$	100.7141	100.7651	100.7663	100.8899
$T^s_{ak_4}$	100.7351	100.7868	100.7884	100.9153
Strategy II				
$T^s_{R_2}$	31.5957	28.3765	30.4126	34.4198
$T^s_{DP_2}/T^s_{S_i}$, $i=1,4$	100.8228	101.0329	100.9761	101.1263
$T^s_{ak_j}$, $j=2,8$	101.1488	101.3113	101.2824	101.4858
$T^s_{ak_5}$	101.1941	101.3585	101.3302	101.5409
Strategy III				
$T^s_{R_3}$	46.7407	42.9472	45.3665	49.9303
$T^s_{DP_3}/T^s_{S_i}$, $i=3,6$	100.4313	100.5410	100.5114	100.5896
$T^s_{ak_j}$, $j=3,9$	100.7574	100.8194	100.8177	100.9492
$T^s_{ak_6}$	100.7808	100.8436	100.8423	100.9775
$x^* \sim \chi^2(7)$				
$y^* \sim \chi^2(11)$				
Strategy I				
$T^s_{R_1}$	34.4668	33.5150	34.8819	36.1917
$T^s_{DP_1}/T^s_{S_i}$, $i=1,4$	100.1268	100.3388	100.3466	100.876
$T^s_{ak_j}$, $j=1,7$	100.8635	101.0234	101.0715	101.6031
$T^s_{ak_4}$	100.9001	101.0808	101.1312	101.6970
Strategy II				
$T^s_{R_2}$	19.9444	19.2757	20.2386	21.1774
$T^s_{DP_2}/T^s_{S_i}$, $i=1,4$	100.2680	100.7179	100.7345	101.8676
$T^s_{ak_j}$, $j=2,8$	101.0047	101.4025	101.4595	102.5947
$T^s_{ak_5}$	101.0832	101.5270	101.5887	102.8017
Strategy III				
$T^s_{R_3}$	32.1275	31.2095	32.5284	33.7957
$T^s_{DP_3}/T^s_{S_i}$, $i=3,6$	100.1409	100.3765	100.3852	100.9743
$T^s_{ak_j}$, $j=3,9$	100.8776	101.0612	101.1102	101.7014
$T^s_{ak_6}$	100.9184	101.1252	101.1767	101.8062

$$x_i = 4.2 + x_i^*$$

where x_i^* and y_i^* are independent proportional distribution variables. Following the division of each population into three equal strata, a random sample of size 10 units is taken from each stratum. The suggested combined and separate estimators' percent relative efficiency (PRE) regarding the traditional mean estimator is derived after 15000 iterations as

$$PRE = \frac{\frac{1}{15{,}000}\sum_{i=1}^{15{,}000}(T_m - \bar{Y})^2}{\frac{1}{15{,}000}\sum_{i=1}^{15{,}000}(T - \bar{Y})^2} \times 100$$

where T denotes the conventional and suggested sequent combined and separate estimators.

The findings of the simulation are unfolded herein from Table 1 to Table 2 in terms of PRE for reasonably chosen amounts of the correlation coefficients 0.6, 0.7, 0.8, 0.9 and non responding unit $r = 5$.
From Table 1 and Table 2 based on the results of each population, we observe that the suggested CSIMs $y^c_{.i_{ak_j}}$ and $y^s_{.i_{ak_j}}$, $j = 1,2,...,9$ are the most efficient than the other existing IMs for different values of correlation coefficient ρ_{xy_h} under each strategy. It is also seen that the suggested CSIMs $y^c_{.i_{ak_j}}$ and $y^s_{.i_{ak_j}}$, $j = 4,5,6$ are superior among the suggested CSIMs, respectively.

5.1. Results and discussion

From the theoretical and simulation results, we have drawn the following observations:

(i). The suggested CSIMs $y^c_{.i_{ak_j}}$ and $y^s_{.i_{ak_j}}$, $j = 1,2,...,9$ dominate combined and separate mean IM $y^c_{.i_m}$ and $y^s_{.i_m}$, classical ratio IM $y^c_{.i_{R_j}}$ and $y^s_{.i_{R_j}}$, $j = 1,2,3$, ratio kind of IMs $y^c_{.i_{S_j}}$ and $y^s_{.i_{S_j}}$, $j = 1,2,...,6$ and regression IMs $y^c_{.i_{DP_j}}$ and $y^s_{.i_{DP_j}}$, $j = 1,2,3$ for different values of correlation coefficient ρ_{xy_h} under strategies I to III.
(ii). The most effective methods in the recommended class of CSIMs are $y^c_{.i_{ak_j}}$ and $y^s_{.i_{ak_j}}$, $j = 4,5,6$ under strategies I to III, respectively.
(iii). The suggested combined IMs $y^c_{.i_{ak_j}}$, $j = 1,2,...,9$ outperform the suggested separate IMs $y^s_{.i_{ak_j}}$, $j = 1,2,...,9$ under each strategy in each population.
(iv). The suggested CSIMs $y^c_{.i_{ak_j}}$ and $y^s_{.i_{ak_j}}$, $j = 1,2,...,9$ perform better under strategy II as compare to strategies I and III.
(v). The PRE of the suggested CSIMs $y^c_{.i_{ak_j}}$ and $y^s_{.i_{ak_j}}$, $j = 1,2,...,9$ increases with the successive incremental change in the value of correlation coefficient from 0.6 to 0.9.

6. Illustration

The illustration of the suggested and traditional IMs is demonstrated utilizing real data from [34], page 228. The production of $N = 80$ factories is taken as the main variable y and the fixed capital of the factories is taken as the auxiliary variable x. These variables are noted from four areas (strata) of the 80 factories. Neyman allocation is implemented to select a total sample of size $n = 45$ from $h = 4$ strata. The descriptives are given in Table 3.
Based on the descriptives given in Table 3, we have computed the PRE of combined and separate estimators T with respect to the usual mean estimator T_m utilizing the following formula:

$$PRE = \frac{MSE(T_m)}{MSE(T)} \times 100$$

Table 4 displays the outcomes produced utilizing real data, demonstrating the supremacy of the suggested estimators over the conventional estimators. Moreover, the combined class of estimators perform better than the separate class of estimators in each strategy.

7. Conclusion

This study considers some novel classes of CSIMs in the case of missing data in SSRS. The characteristics of the sequent combined and separate estimators obtained from the respective IMs are reported. The members of the suggested combined and separate classes of IMs are identified as the combined and separate conventional mean IM, ratio IM, ratio type IM defined on the lines of [27], and regression IM defined on the lines of [4]. The theoretical conditions are derived and sustained with the simulation study accomplished on the hypothetically drawn one symmetrical population such as normal and two asymmetrical populations such as exponential and Chi-square. The simulation

Table 3
Descriptive statistics of real data.

	Total	Symbol for stratum h	1	2	3	4
Population size	$N=80$	N_h	19	32	14	15
Sample size	$n=45$	n_h	11	18	8	8
Responding units	$r=26$	r_h	6	12	4	4
Population mean	$\bar{X}=1126.46$	$\bar{X}_h$	349.68	706.59	1539.57	2620.53
Population mean	$\bar{Y}=5182.64$	$\bar{Y}_h$	2967.95	4657.63	6537.21	7843.67
Correlation coefficient	$\rho_{xy}=0.94$	ρ_{xy_h}	0.93	0.92	0.98	0.96
Standard deviation	$S_x=845.61$	S_{x_h}	109.44	109.22	277.18	370.97
Standard deviation	$S_y=1835.66$	S_{y_h}	757.08	669.12	416.11	645.68

Table 4
PRE of suggested combined estimators w.r.t. usual mean estimator.

Combined estimators	PRE	Separate estimators	PRE
Strategy I			
$T^c_{R_1}$	113.5945	$T^s_{R_1}$	112.9952
$T^c_{DP_1}/T^c_{S_i},\ i=1,4$	142.6297	$T^s_{DP_1}/T^s_{S_i},\ i=1,4$	141.1211
$T^c_{ak_j},\ j=1,7$	143.0664	$T^s_{ak_j},\ j=1,7$	142.7539
$T^c_{ak_4}$	143.2226	$T^s_{ak_4}$	142.7983
Strategy II			
$T^c_{R_2}$	160.2879	$T^s_{R_2}$	159.8929
$T^c_{DP_2}/T^c_{S_i},\ i=2,5$	1648.7680	$T^s_{DP_2}/T^s_{S_i},\ i=2,5$	1647.9714
$T^c_{ak_j},\ j=2,8$	1649.2040	$T^s_{ak_j},\ j=2,8$	1648.8920
$T^c_{ak_5}$	1670.3520	$T^s_{ak_5}$	1654.8650
Strategy III			
$T^c_{R_3}$	134.4895	$T^s_{R_3}$	133.1935
$T^c_{DP_3}/T^c_{S_i},\ i=3,6$	278.1370	$T^s_{DP_3}/T^s_{S_i},\ i=3,6$	277.7919
$T^c_{ak_j},\ j=3,9$	278.5738	$T^s_{ak_j},\ j=3,9$	278.2613
$T^c_{ak_6}$	279.3849	$T^s_{ak_6}$	278.4918

results exhibit the dominance of the suggested CSIMs over the existing CSIMs. The numerical findings based on real data also exhibit the superiority of the proposed IMs over their traditional counterparts. Furthermore, the combined IMs perform superior than the separate IMs. As a result, it is necessary to favor the suggested combined and separate classes of IMs for estimating the population mean in case of missing data.
The suggested CSIMs will be evaluated under SRSS in upcoming studies.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Data availability

The data chosen for the empirical study are provided in the manuscript.

Acknowledgement

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R515), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1445).

Appendix A

The MSE of the subsequent estimators used in currently available combined imputation approaches is displayed below.

$$V(T_m)=\bar{Y}^2 I_0^* \tag{A.1}$$

$$MSE(T^c_{R_1})=\bar{Y}^2\left[I_0^*+I_1-2I_{01}\right] \tag{A.2}$$

$$MSE(T^c_{R_2})=\bar{Y}^2\left[I_0^*+I_1^*-2I_{01}^*\right] \tag{A.3}$$

$$MSE(T^c_{R_3})=\bar{Y}^2\left[I_0^*+I_1^*-I_1-2(I_{01}^*-I_{01})\right] \tag{A.4}$$

$$MSE(T^c_{DP_1})=\bar{Y}^2 I_0^*+b_1^2\bar{X}^2 I_1-2b_1\bar{X}\bar{Y}I_{01}$$

$$MSE(T^c_{DP_2})=\bar{Y}^2 I_0^*+b_2^2\bar{X}^2 I_1^*-2b_2\bar{X}\bar{Y}I_{01}^*$$

$$MSE(T^c_{DP_3})=\bar{Y}^2 I_0^*+b_3^2\bar{X}^2 (I_1^*-I_1)-2b_3\bar{X}\bar{Y}(I_{01}^*-I_{01})$$

$$minMSE(T^c_{DP_1})=\bar{Y}^2\left[I_0^*-\frac{I_{01}^2}{I_1}\right] \tag{A.5}$$

$$minMSE(T^c_{DP_2})=\bar{Y}^2\left[I_0^*-\frac{I_{01}^{*2}}{I_1^*}\right] \tag{A.6}$$

$$minMSE(T^c_{DP_3})=\bar{Y}^2\left[I_0^*-\frac{\left(I_{01}^*-I_{01}\right)^2}{\left(I_1^*-I_1\right)}\right] \tag{A.7}$$

$$MSE(T^c_{s_i})=\bar{Y}^2\left[I_0^*+\beta_i^2 I_1-2\beta_i I_{01}\right],\ i=1,4$$

$$MSE(T^c_{S_i})=\bar{Y}^2\left[I_0^*+\beta_i^2 I_1^*-2\beta_i I_{01}^*\right],\ i=2,5$$

$$MSE(T^c_{S_i})=\bar{Y}^2\left[\begin{array}{l} I_0^*+\beta_i^2\left\{I_1^*-I_1\right\} \\ -2\beta_i\left\{I_{01}^*-I_{01}\right\}\end{array}\right],\ i=3,6$$

$$minMSE(T^c_{S_i})=\bar{Y}^2\left[I_0^*-\frac{I_{01}^2}{I_1}\right];\ \ i=1,4 \tag{A.8}$$

$$minMSE(T^c_{S_i})=\bar{Y}^2\left[I_0^*-\frac{I_{01}^{*2}}{I_1^*}\right];\ \ i=2,5 \tag{A.9}$$

$$minMSE(T^c_{S_i})=\bar{Y}^2\left[I_0^*-\frac{\left(I_{01}^*-I_{01}\right)^2}{\left(I_1^*-I_1\right)}\right];\ \ i=3,6 \tag{A.10}$$

The scalars' optimum values for the combined estimators are shown below.
$b_1=R\frac{I_{01}}{I_1}$, $b_2=R\frac{I_{01}^*}{I_1^*}$, $b_3=R\frac{(I_{01}^*-I_{01})}{(I_1^*-I_1)}$, $\beta_{1(opt)}=\beta_{4(opt)}=\frac{I_{01}}{I_1}$, $\beta_{2(opt)}=\beta_{5(opt)}=\frac{I_{01}^*}{I_1^*}$, and $\beta_{3(opt)}=\beta_{6(opt)}=\frac{\left(I_{01}^*-I_{01}\right)}{(I_1^*-I_1)}$.

Appendix B

The MSE of the subsequent estimators used in currently available separate imputation approaches is displayed below.

$$MSE(T^s_{R_1})=\sum_{h=1}^{L}W_h^2\bar{Y}_h^2\left(J_1+J_0^*-2J_{01}\right) \tag{B.11}$$

$$MSE(T^s_{R_2})=\sum_{h=1}^{L}W_h^2\bar{Y}_h^2\left(J_1^*+J_0^*-2J_{01}^*\right) \tag{B.12}$$

$$MSE(T^s_{R_3})=\sum_{h=1}^{L}W_h^2\bar{Y}_h^2\left[J_0^*+(J_1^*-J_1)-2(J_{01}^*-J_{01})\right] \tag{B.13}$$

$$MSE(T^s_{DP_1}) = \sum_{h=1}^{L} W_h^2[\bar{Y}_h^2 J_0^* + b_{1_h}^2 \bar{X}_h^2 J_1 - 2b_{1_h}\bar{X}_h\bar{Y}_h J_{01}]$$

$$MSE(T^s_{DP_2}) = \sum_{h=1}^{L} W_h^2[\bar{Y}_h^2 J_0^* + b_{2_h}^2 \bar{X}^2 J_1^* - 2b_{2_h}\bar{X}_h\bar{Y}_h J_{01}^*]$$

$$MSE(T^s_{DP_3}) = \sum_{h=1}^{L} W_h^2[\bar{Y}_h^2 J_0^* + b_{3_h}^2 \bar{X}_h^2 (J_1^* - J_1) - 2b_{3_h}\bar{X}_h\bar{Y}_h (J_{01}^* - J_{01})]$$

$$minMSE(T^s_{DP_1}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[J_0^* - \frac{J_{01}^2}{J_1}\right] \tag{B.14}$$

$$minMSE(T^s_{DP_2}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[J_0^* - \frac{J_{01}^{*2}}{J_1^*}\right] \tag{B.15}$$

$$minMSE(T^s_{DP_3}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[J_0^* - \frac{\left(J_{01}^* - J_{01}\right)^2}{\left(J_1^* - J_1\right)}\right] \tag{B.16}$$

$$MSE(T^s_{S_i}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[J_0^* + \beta_{i_h}^2 J_1 - 2\beta_{i_h} J_{01}\right],\ i = 1,4$$

$$MSE(T^s_{S_i}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[J_0^* + \beta_{i_h}^2 J_1^* - 2\beta_{i_h} J_{01}^*\right],\ i = 2,5$$

$$MSE(T^s_{S_i}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[J_0^* + \beta_{i_h}^2\left(J_1^* - J_1\right) - 2\beta_{i_h}\left(J_{01}^* - J_{01}\right)\right],$$

$i = 3,6$

$$minMSE(T^s_{S_i}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[J_0^* - \frac{J_{01}^2}{J_1}\right];\ i = 1,4 \tag{B.17}$$

$$minMSE(T^s_{S_i}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[J_0^* - \frac{J_{01}^{*2}}{J_1^*}\right];\ i = 2,5 \tag{B.18}$$

$$minMSE(T^s_{S_i}) = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2\left[J_0^* - \frac{\left(J_{01}^* - J_{01}\right)^2}{\left(J_1^* - J_1\right)}\right];\ i = 3,6 \tag{B.19}$$

The scalars' optimum values for the combined estimators are shown below.
$b_{1_h} = R_h\frac{J_{01}}{J_1}$, $b_{2_h} = R_h\frac{J_{01}^*}{J_1^*}$, $b_{3_h} = R_h\frac{(J_{01}^* - J_{01})}{(J_1^* - J_1)}$, $\beta_{1_h(opt)} = \beta_{4_h(opt)} = \frac{J_{01}}{J_1}$,
$\beta_{2_h(opt)} = \beta_{5_h(opt)} = \frac{J_{01}^*}{J_1^*}$, $\beta_{3_h(opt)} = \beta_{6_h(opt)} = \frac{\left(J_{01}^* - J_{01}\right)}{J_1^* - J_1}$.

Appendix C

This section contains the proof of MSE and minimum MSE expressions discussed in Section 3.1.1 and Section 3.2.1.
Consider the following estimator as part of Strategy I:

$$T^c_{ak_1} = \alpha_1\bar{y}_{r_{st}} + \theta_1(\bar{x}_{n_{st}} - \bar{X})$$

Employing the notations discussed in subsection 2.1, we get

$$T^c_{ak_1} - \bar{Y} = \alpha_1\bar{Y}\varepsilon_0 + \theta_1\bar{X}\varepsilon_1 + (\alpha_1 - 1)\bar{Y} \tag{C.20}$$

By squaring both sides of (C.20) and taking the expectation, we get

$$MSE(T^c_{ak_1}) = (\alpha_1 - 1)^2\bar{Y}^2 + \alpha_1^2\bar{Y}^2 I_0^* + \theta_1^2 I_1 + 2\alpha_1\theta_1\bar{X}\bar{Y}I_{01} \tag{C.21}$$

By minimizing (C.21) w.r.t α_1 and θ_1, we get the optimum values of α_1 and θ_1 as

$$\alpha_{1(opt)} = \frac{1}{\left(1 + I_0^* - \frac{I_{12}^2}{I_1}\right)} = \alpha_{7(opt)}$$

$$\theta_{1(opt)} = -\frac{\bar{Y}}{\bar{X}}\frac{I_{01}}{I_1}\alpha_{1(opt)}$$

Putting $\alpha_{1(opt)}$ and $\theta_{1(opt)}$ in (C.21), we get the minimum MSE as

$$MSE_{min}(T^c_{ak_1}) = \bar{Y}^2(1 - \alpha_{1(opt)})$$

In a similar manner, the minimum MSE of additional suggested estimators can be determined. The scalars' optimum values for the suggested combined estimators are shown below.

$$\alpha_{2(opt)} = \frac{1}{\left(1 + I_0^* - \frac{I_{01}^{*2}}{I_1^*}\right)} = \alpha_{8(opt)}$$

$$\theta_{2(opt)} = -\frac{\bar{Y}}{\bar{X}}\frac{I_{01}^*}{I_1^*}\alpha_{2(opt)}$$

$$\alpha_{3(opt)} = \frac{1}{\left(1 + I_0^* - \frac{\left(I_{01}^* - I_{01}\right)^2}{(I_1^* - I_1)}\right)}$$

$$\theta_{3(opt)} = -\frac{\bar{Y}}{\bar{X}}\left(\frac{I_{01}^* - I_{01}}{I_1^* - I_1}\right)\alpha_{3(opt)}$$

$$\alpha_{j(opt)} = \frac{A_j}{B_j};\ j = 4,5,6,7,8,9$$

$$\theta_{j(opt)} = \frac{I_{01}}{I_1};\ j = 4,7$$

$$\theta_{j(opt)} = \frac{I_{01}^*}{I_1^*};\ j = 5,8.$$

$$\theta_{j(opt)} = \frac{(I_{01}^* - I_{01})}{(I_1^* - I_1)};\ j = 6,9$$

where $A_4 = \left(1 + \frac{I_{01}}{2} - \frac{I_{12}^2}{2I_1}\right)$, $B_4 = \left(1 + I_0^* + I_{01} - \frac{2I_{12}^2}{I_1}\right)$, $A_5 = \left(1 + \frac{I_{01}^*}{2} - \frac{I_{01}^{*2}}{2I_1^*}\right)$, $B_5 = \left(1 + I_0^* + I_{01}^* - \frac{2I_{01}^{*2}}{I_1^*}\right)$, $A_6 = \left[1 - \frac{1}{2}\frac{\left\{I_{01}^* - I_{01}\right\}^2}{I_1^* - I_1} + \frac{1}{2}(I_{01}^* - I_{01})\right]$,
$B_6 = \left[1 + I_0^* - 2\frac{\left\{I_{01}^* - I_{01}\right\}^2}{I_1^* - I_1} + (I_{01}^* - I_{01})\right]$, $A_7 = 1$, $B_7 = \left(1 + I_0^* - \frac{I_{12}^2}{I_1}\right)$,
$A_9 = 1$ and $B_9 = \left[1 + I_0^* - \frac{(I_{01}^* - I_{01})^2}{(I_1^* - I_1)}\right]$.
Similarly, the outline of the proof of MSE and minimum MSE expressions given in Section 3.2.1 can be obtained.

References

[1] R.B. Rubin, Inference and missing data, Biometrika 63 (3) (1976) 581–592.
[2] D.F. Heitjan, S. Basu, Distinguishing 'missing at random' and 'missing completely at random', Am. Stat. 50 (1996) 207–213.
[3] C. Kadilar, H. Cingi, Estimators for the population mean in the case of missing data, Commun. Stat., Theory Methods 37 (2008) 2226–2236.
[4] G. Diana, P.F. Perri, Improved estimators of the population mean for missing data, Commun. Stat., Theory Methods 39 (2010) 3245–3251.
[5] S. Bhushan, A.P. Pandey, Optimal imputation of missing data for estimation of population mean, J. Stat. Manag. Syst. 19 (6) (2016) 755–769.
[6] C. Mohamed, S.A. Sedory, S. Singh, Imputation using higher order moments of an auxiliary variable, Commun. Stat., Simul. Comput. 46 (2016) 6588–6617.
[7] S. Bhushan, A.P. Pandey, A. Pandey, On optimality of imputation methods for estimation of population mean using higher order moment of an auxiliary variable, Commun. Stat., Simul. Comput. 49 (6) (2020) 1560–1574.
[8] G.N. Singh, M. Usman, Generalized class of estimators using dual of auxiliary variable under non-response, J. Stat. Theory Pract. 15 (2) (2021) 1–35.
[9] M.M. Anas, Z. Huang, U. Shahzad, T. Zaman, S. Shahzadi, Compromised imputation based mean estimators using robust quantile regression, Commun. Stat., Theory Methods (2022) 1–16.
[10] S. Bhushan, A. Kumar, A.P. Pandey, S. Singh, Estimation of population mean in presence of missing data under simple random sampling, Commun. Stat., Simul. Comput. 52 (12) (2023) 6048–6069.
[11] S. Bhushan, A. Kumar, T. Zaman, A. Al Mutairi, Efficient difference and ratio-type imputation methods under ranked set sampling, Axioms 12 (2023) 558.
[12] S. Bhushan, A. Kumar, Imputation of missing data using multi auxiliary information under ranked set sampling, Commun. Stat., Simul. Comput. (2023) 1–23, https://doi.org/10.1080/03610918.2023.2288796.

[13] U. Shahzad, M. Hanif, N. Koyuncu, A new estimator for mean under stratified random sampling, Math. Sci. (2018) 163–169.
[14] T. Zaman, C. Kadilar, On estimating the population mean using auxiliary character in stratified random sampling, J. Stat. Manag. Syst. 23 (8) (2020) 1415–1426.
[15] T. Zaman, An efficient exponential estimator of the mean under stratified random sampling, Math. Popul. Stud. 28 (2) (2021) 104–121.
[16] M. Subzar, S.A. Lone, M. Aslam, A.H. AL-Marshadi, S. Maqbool, Exponential ratio estimator of the median: an alternative to the regression estimator of the median under stratified sampling, J. King Saud Univ., Sci. 35 (3) (2023) 102536.
[17] S. Ahmad, S. Hussain, M. Aamir, U. Yasmeen, J. Shabbir, Z. Ahmad, Dual use of auxiliary information for estimating the finite population mean under the stratified random sampling scheme, J. Math. 2021 (2021) 1–2.
[18] S. Ahmad, S. Hussain, U. Yasmeen, M. Aamir, J. Shabbir, M. El-Morshedy, A. Al-Bossly, Z. Ahmad, A simulation study: using dual ancillary variable to estimate population mean under stratified random sampling, PLoS ONE 17 (11) (2022) e0275875.
[19] T. Zaman, C. Kadilar, Exponential ratio and product type estimators of the mean in stratified two-phase sampling, AIMS Math. 6 (5) (2021) 4265–4279.
[20] Q. Rana, M.N. Qureshi, M. Hanif, Generalized estimator for population mean using auxiliary attribute in stratified two-phase sampling, J. Stat. Theory Appl. 21 (2) (2022) 44–57.
[21] S. Bhushan, A. Kumar, S. Singh, Some efficient classes of estimators under stratified sampling, Commun. Stat., Theory Methods (2021) 1–30, https://doi.org/10.1080/03610926.2021.1939052.
[22] S. Bhushan, A. Kumar, R. Onyango, S. Singh, Some improved classes of estimators in stratified sampling using bivariate auxiliary information, J. Probab. Stat. 2022 (2) (2022) 1–23.
[23] S. Bhushan, A. Kumar, U. Shahzad, A.I. Al-Omari, A.I. Almanjahie, On some improved class of estimators by using stratified ranked set sampling, Mathematics 10 (3283) (2022) 1–32.
[24] S. Bhushan, A. Kumar, S. Shahab, S.A. Lone, M.T. Akhtar, On efficient estimation of population mean under stratified ranked set sampling, J. Math. 2022 (3) (2022) 1–20.
[25] S. Bhushan, A. Kumar, J. Banerjie, Mean estimation using logarithmic estimators in stratified ranked set sampling, Life Cycle Reliab. Saf. Eng. (2022) 1–9.
[26] N.S. Thakur, K. Yadav, S. Pathak, Estimation of mean with imputation of missing data in stratified random sampling, Res. Rev.: J. Stat. 3 (1) (2014) 1–12.
[27] M.S. Ahmed, O. Al-Titi, Z. Al-Rawi, W. Abu-Dayyeh, Estimation of a population mean using different imputation methods, Stat. Transit. 7 (6) (2006) 1247–1264.
[28] R. Pandey, N.S. Thakur, K. Yadav, Separate regression type imputation methods to estimate population mean, Int. J. Eng. Technol. Sci. Res. 2 (2015) 36–45.
[29] S. Bhushan, A.P. Pandey, Optimality of ratio type estimation methods for population mean in the presence of missing data, Commun. Stat., Theory Methods 47 (11) (2018) 2576–2589.
[30] S. Bhushan, A. Kumar, On optimal classes of estimators under ranked set sampling, Commun. Stat., Theory Methods 51 (8) (2022) 2610–2639.
[31] S. Singh, S. Horn, An alternative survey in multi-character survey, Metrika 48 (1998) 99–107.
[32] S. Bhushan, A. Kumar, M.T. Akhtar, S.A. Lone, Logarithmic type predictive estimators under simple random sampling, AIMS Math. 7 (7) (2022) 11992–12010.
[33] S. Bhushan, A. Kumar, Predictive estimation approach using difference and ratio type estimators in ranked set sampling, J. Comput. Appl. Math. 410 (2022) 114214.
[34] M.N. Murthy, Sampling Theory and Methods, Statistical Publishing Society, Calcutta, India, 1967.