Brazilian Journal of Biometrics, **44**, 1–23 (2026)
doi:10.28951/bjb.v44i1.847

BRAZILIAN JOURNAL OF BIOMΣTRICS
ISSN:2764-5290

ARTICLE

Improved classes of logarithmic type estimators of finite population mean in case of missing data

Lovleen Kumar Grover *[1], Anchal Sharma [2]
[1]Department of Mathematics, Guru Nanak Dev University, Amritsar (143005), Punjab, India
[2]P. G. Department of Mathematics, Khalsa College, Amritsar (143005), Punjab, India.
*Corresponding author. Email: lovleen_2@yahoo.co.in.

(Received: February 14, 2025; Revised: November, 27, 2025; Accepted: December 10, 2025; Published: March 2, 2026.)

Abstract
The situation of missing observations is a common issue in surveys and other data collection processes. This issue can seriously affect the statistical inference. The imputation techniques are widely recognized as an effective means of addressing this challenge. Over the years, several estimators or classes of estimators have been proposed to reduce the bias and mean square error of various existing estimators of interest in the presence of missing data, particularly by utilizing auxiliary information. In this paper, we have proposed some new classes of logarithmic type estimators of the population mean when missing values are there in the sample data. We derive the approximate expressions for the bias and mean square errors of these classes. The optimum mean square errors of these proposed estimators are compared with those of the existing estimators. A numerical investigation is used to validate the conclusions so acquired. This investigation illustrate that the proposed estimators outperform their present counterparts across all the situations under consideration, as evidenced by consistently higher percent relative efficiencies.

Keywords: Simple random sampling without replacement; Study variable; Auxiliary variable; Imputation; Bias; Mean square error; Percent relative efficiency.

1. Introduction

Incompleteness or missingness in data is a common problem, faced by practitioners in the sample surveys, like: drug testing experiments, agricultural experiments, socio-economic surveys, etc. The inferences regarding the population parameters can be spoiled if the stochastic nature of incompleteness of the required data is not known. Hansen and Hurwitz (1946) were the first researchers to deal with the problem of incomplete samples in mail surveys. Rubin (1976) introduced the following concepts: Missing at Random (MAR), Observed at Random (OAR), Missing Completely at Random (MCAR) and Parameter Distinctness (PD).

(a) The data are MAR if the probability of the observed missingness pattern, given the observed and unobserved data, does not depend upon the value of the unobserved data.
(b) The data are OAR if for every possible value of the missing data, the probability of the observed absence pattern, given the observed and unobserved data, does not depend on

the values of the observed data.

(c) MCAR means that the missing data is independent of observed and unobserved data that is, when data set is MCAR, it can be considered as a simple random sample of the full data set of interest.

(d) PD holds if there are no priori ties between the parameters of the absence model and those of data model.

Heitzan and Basu (1996) have distinguished MAR and MCAR in a very effective way. Many methods are used by the practitioners to reduce the negative impact of the missingness in making conclusions about population parameters. Imputation is one of the popular methods wherein non response units are imputed suitably using fabricated values by involving the respondent units of the population. Sande (1979) and Kalton *et al.* (1981) suggested imputation methods that make an incomplete data set structurally complete. In past two decades, many pioneering advancements are made to reduce the negative effect if the data is missing. A number of authors such as Singh and Horn (2000), Singh and Deo (2003), Ahmed *et al.* (2006), Kadilar and Cingi (2008), Diana and Perri (2010), Thakur *et al.* (2012) have suggested several imputation techniques with the help of auxiliary variable. Yang and Kim (2016) suggested Fractional Imputation (FI) for handling item nonresponses. In FI, each missing item is "filled in" with several imputed values with their fractional weights, where each fractional weight represents the conditional probability of the imputed value given the observed data. Kumar *et al.* (2017) suggested a class of exponential chain type estimator for population mean under double sampling scheme using imputation techniques. Zahid and Shabbir (2018) suggested an improved class of estimators in the presence of measurement error and non-response under stratified random sampling for estimating the finite population mean. Sohail and Shabbir (2019) gave another imputation method using raw moments. Bhushan *et al.* (2020) described some improved imputation methods by using higher order moment of auxiliary variable. Hussain *et al.* (2020) suggested two new families of estimators using the auxiliary information in terms of the population distribution functions with the case of non-response under simple random sampling. Ahmad *et al.* (2022) addressed the issue of estimating the population mean using rank of the auxiliary variable in simple random sampling scheme when non response is present in the data. Zahid *et al.* (2022) gave a general class of estimators for estimating the finite population mean for sensitive variable, in the presence of measurement error and non-response in simple random sampling. Prasad and Yadav (2023) gave some other product type exponential imputation methods to tackle the problem of incomplete values of study variable in the sample. Pandey *et al.* (2024) suggested two other classes of imputation techniques based on the logarithmic function. Yadav and Prasad (2024) introduced some efficient generalized class of factor-type exponential imputation techniques and their corresponding estimators using auxiliary information. Bhushan and Kumar (2025) utilized the multi auxiliary information available under Ranked Set Sampling (RSS) for the imputation of missing data and gave some other imputation methods.

In the present paper, we have proposed three imputation strategies, by making use of the logarithmic relationship between the study variable y and the auxiliary variable x, for estimating finite population mean in case of missing data. The use of logarithmic function is motivated by its wide-ranging real-world applications, such as measuring the magnitude of earthquakes, assessing noise levels, determining the pH value of chemical substances, and calculating the half-life of radioactive elements. In most of the cases, log transformation is employed to approximate normality which is a fundamental assumption underlying many statistical tests and estimation methods. This transformation not only stabilizes variance but also enhances the reliability and interpretability of statistical inferences. In this paper, three classes of logarithmic-type estimators are developed under different assumptions regarding the availability of the population mean ($\bar{X}$) of the auxiliary variable x. Then, a comparative study with respect to their mean square errors is carried out and theoretical results so obtained are finally verified with the help of some numerical populations and simulation techniques, confirming the practical utility and robustness of the proposed approaches.

This article is organized as follows: Section 2 deals with the methodology and notations used in this research, followed by the description of existing estimators in Table 1. Section 3 introduces the formulation of proposed classes of logarithmic estimators of the population mean, followed by description of their properties (Biases and Mean Square Errors (MSEs)). Section 4 provides a comparative analysis of the efficiencies of the proposed estimators with the existing ones. Numerical illustrations based on real and hypothetical datasets are given in Section 5, followed by a discussion of the obtained results in Section 6. The concluding remarks are presented in Section 7.

2. Methodology and Notations

Let $\Omega = \{1,2,3, \ldots, N\}$ be a finite population consisting of N identifiable units. Let y be the character under study and x be the auxiliary variate which is positively correlated with y. The information about x may be available on the entire population. Let S be the Simple Random Sample Without Replacement (SRSWOR) of size n $(n < N)$ which is drawn from Ω. Let the number of responding units be denoted by r, so that number of non-responding units be equal to (n-r). Let the responding set of sampling units be denoted by R and that of non-responding set by R^c. So, we have $S= R \cup R^c$. For each $i \in R$, the value of y_i is known. However, for the units $i \in R^c$, the y_i values are missing and thus imputed values are derived in this case. The imputation is carried out with the aid of auxiliary variate x, such that, the value of x for unit i (say x_i), is known for each $i \in S$. Consider the following rule:

$$y_{.i} = \begin{cases} y_i & i \in R \\ \hat{y}_i & i \in R^c \end{cases}, i \in S,$$

where $\hat{y}_i$ denotes the imputed value of the variable y for the i^{th} non-responding unit.

The value of $\hat{y}_i$ is different for each imputation technique.

The following notations are used throughout the paper:

$\bar{Y}, \bar{X}$: population means of study variable and auxiliary variable respectively.

$\bar{y}_r, \bar{x}_r$: response means of the respective variables for the responding set R of the given sample.

$\bar{y}_n, \bar{x}_n$: sample means of the variables y and x based on sample S of size n.

(Y_i, X_i) denote the values of (y, x) on the i^{th} population unit.

$S_{yx} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})(X_i - \bar{X}), \quad S_y{}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2, \; S_x{}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$

$\rho_{yx} (or\ \rho) = \frac{S_{yx}}{S_y S_x}$: population correlation coefficient between the variables y and x,

$C_y = \frac{S_y}{\bar{Y}}, \quad C_x = \frac{S_x}{\bar{X}}, \quad R^* = \frac{\bar{Y}}{\bar{X}},$

$\lambda_1 = \frac{1}{r} - \frac{1}{N}, \quad \lambda_2 = \frac{1}{n} - \frac{1}{N}, \quad \lambda_3 = \frac{1}{r} - \frac{1}{n} = \lambda_1 - \lambda_2.$

Also define for the sample S,

$s_{yx} = \frac{1}{n-1}\sum_{i=1}^{n}(y_{.i} - \bar{y}_n)(x_i - \bar{x}_n), \quad s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2, \quad s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_{.i} - \bar{y}_n)^2.$

Correspondingly for the responding set R, we take,

$s_{yx}^* = \frac{1}{r-1}\sum_{i=1}^{r}(y_i - \bar{y}_r)(x_i - \bar{x}_r)$, $\quad s_x^{*2} = \frac{1}{r-1}\sum_{i=1}^{r}(x_i - \bar{x}_r)^2$, $\quad s_y^{*2} = \frac{1}{r-1}\sum_{i=1}^{r}(y_i - \bar{y}_r)^2$

Let $\varepsilon = \frac{\bar{y}_r}{\bar{Y}} - 1$, $\quad \delta = \frac{\bar{x}_r}{\bar{X}} - 1$, $\quad \eta = \frac{\bar{x}_n}{\bar{X}} - 1$.

On using the concept of two-phase sampling, given by Rao and Sitter (1995), and the mechanism of MCAR, for given r and n, we have the following expectations under SRSWOR:
$E(\varepsilon) = 0$, $E(\delta) = 0$, $E(\eta) = 0$.
Up to first order of approximation, further we have the following expectations:
$E(\varepsilon^2) = \lambda_1 C_y{}^2$, $\quad E(\delta^2) = \lambda_1 C_x{}^2$, $\quad E(\epsilon\delta) = \lambda_1 \rho C_y C_x$,
$E(\eta^2) = \lambda_2 C_x{}^2$, $\quad E(\delta\eta) = \lambda_2 C_x{}^2$, $\quad E(\epsilon\eta) = \lambda_2 \rho C_y C_x$.

The population mean $\bar{X}$ of the auxiliary variable x may or may not be known. For the formulation of the proposed imputed estimators, three specific situations have been taken into consideration, as outlined below.

Strategy I: When $\bar{X}$ is known and $\bar{x}_r$ is used.
Strategy II: When $\bar{X}$ is known and $\bar{x}_n$ is used.
Strategy III: When $\bar{X}$ is unknown and both $\bar{x}_n$ and $\bar{x}_r$ are used.
The general point estimator for the unknown population mean $\bar{Y}$ takes the form:

$$\bar{y}_t = \frac{1}{n}\sum_{i\in S} y_{.i} = \frac{1}{n}\left[\sum_{i\in R} y_i + \sum_{i\in R^c} \hat{y}_i\right]$$

In Table 1, we have presented a list of some existing imputation techniques for the estimation of $\bar{Y}$ along with their notations, structures and Variances/ Mean Square Errors (MSEs) for the purpose of comparative analysis in the coming sections.

Table 1. Existing imputation methods, notations, structures and Variances\MSE of the estimators of $\bar{Y}$

Imputation Methods	Notations	Structures	Variances /Mean square errors
1. Simple mean method of imputation $y_{.i} = \begin{cases} y_i & i \in R \\ \bar{y}_r & i \in R^c \end{cases}$	$\bar{y}_{(1)}$	$\bar{y}_r$	$\mathrm{Var}(\bar{y}_{(1)}) = \lambda_1 \bar{Y}^2 C_y{}^2$
2. Ratio Estimators method of imputation (Reformulating by Singh and Horn (2000))			
Under strategy I $y_{(21)} = \begin{cases} y_i & i \in R \\ \frac{1}{n-r}\left[n\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right) - r\bar{y}_r\right] & i \in R^c \end{cases}$	$\bar{y}_{(21)}$	$\bar{y}_r \frac{\bar{X}}{\bar{x}_r}$	$MSE(\bar{y}_{(21)}) = \lambda_1 \bar{Y}^2 (C_x^2 + C_y^2 - 2\rho C_y C_x)$
Under strategy II $y_{(22)} = \begin{cases} y_i & i \in R \\ \frac{1}{n-r}\left[n\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right) - r\bar{y}_r\right] & i \in R^c \end{cases}$	$\bar{y}_{(22)}$	$\bar{y}_r \frac{\bar{X}}{\bar{x}_n}$	$MSE(\bar{y}_{(22)}) = \bar{Y}^2(\lambda_1 C_y^2 + \lambda_2 (C_x^2 - 2\rho C_y C_x))$
Under strategy III $y_{(23)} = \begin{cases} y_i & i \in R \\ \frac{1}{n-r}\left[n\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right) - r\bar{y}_r\right] & i \in R^c \end{cases}$	$\bar{y}_{(23)}$	$\bar{y}_r \frac{\bar{x}_n}{\bar{x}_r}$	$MSE(\bar{y}_{(23)}) = \bar{Y}^2(\lambda_1 C_y^2 + \lambda_3 (C_x^2 - 2\rho C_y C_x))$
3. Regression estimators method of imputation (Diana and Perri (2010))			

Under strategy I			
$y_{(31)}=\begin{cases}\frac{n}{r}y_i-\frac{bnx_i}{r} & i\in R\\ \frac{bn\bar{X}}{n-r} & i\in R^c\end{cases}$	$\bar{y}_{(31)}$	$\bar{y}_r+b(\bar{X}-\bar{x}_r)$	$MSE(\bar{y}_{(31)})=\lambda_1\bar{Y}^2C_y^2(1-\rho^2)$
Under strategy II			
$y_{(32)}=\begin{cases}\frac{n}{r}y_i+b(\bar{X}-x_i) & i\in R\\ b(\bar{X}-x_i) & i\in R^c\end{cases}$	$\bar{y}_{(32)}$	$\bar{y}_r+b(\bar{X}-\bar{x}_n)$	$MSE(\bar{y}_{(32)})=\lambda_3\bar{Y}^2C_y^2+\lambda_2\bar{Y}^2C_y^2(1-\rho^2)$
Under strategy III			
$y_{(33)}=\begin{cases}\frac{n}{r}y_i-\frac{bnx_i}{r} & i\in R\\ \frac{bn\bar{x}_n}{n-r} & i\in R^c\end{cases}$	$\bar{y}_{(33)}$	$\bar{y}_r+b(\bar{x}_n-\bar{x}_r)$	$MSE(\bar{y}_{(33)})=\lambda_2\bar{Y}^2C_y^2+\lambda_3\bar{Y}^2C_y^2(1-\rho^2)$
4. Compromised Method of Imputation			
(Singh and Horn (2000))			
Under strategy I			
$y_{(41)}=\begin{cases}y_i & i\in R\\ \frac{1}{n-r}\left[n\left(\alpha\bar{y}_r+(1-\alpha)\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)\right)-r\bar{y}_r\right] & i\in R^c\end{cases}$	$\bar{y}_{(41)}$	$\alpha\bar{y}_r+(1-\alpha)\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)$	$MSE_{min}(\bar{y}_{(41)})=\lambda_1\bar{Y}^2C_y^2(1-\rho^2)$
Under strategy II			
$y_{(42)}=\begin{cases}y_i & i\in R\\ \frac{1}{n-r}\left[n\left(\alpha\bar{y}_r+(1-\alpha)\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right)\right)-r\bar{y}_r\right] & i\in R^c\end{cases}$	$\bar{y}_{(42)}$	$\alpha\bar{y}_r+(1-\alpha)\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right)$	$MSE_{min}(\bar{y}_{(42)})=\lambda_3\bar{Y}^2C_y^2+\lambda_2\bar{Y}^2C_y^2(1-\rho^2)$
Under strategy III			
$y_{(43)}=\begin{cases}y_i & i\in R\\ \frac{1}{n-r}\left[n\left(\alpha\bar{y}_r+(1-\alpha)\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)\right)-r\bar{y}_r\right] & i\in R^c\end{cases}$	$\bar{y}_{(43)}$	$\alpha\bar{y}_r+(1-\alpha)\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)$	$MSE_{min}(\bar{y}_{(43)})=\lambda_2\bar{Y}^2C_y^2+\lambda_3\bar{Y}^2C_y^2(1-\rho^2)$
where α is suitably chosen constant.			
5. Kadilar and Cingi (2008)'s estimators method of imputation			
Under strategy I			
$y_{(51)}=\begin{cases}\left(\frac{n}{r}y_i-\frac{bnx_i}{r}\right)\left(\frac{\bar{X}}{\bar{x}_r}\right) & i\in R\\ \left(\frac{bn\bar{X}}{n-r}\right)\left(\frac{\bar{X}}{\bar{x}_r}\right) & i\in R^c\end{cases}$	$\bar{y}_{(51)}$	$(\bar{y}_r+\mathrm{b}(\bar{X}-\bar{x}_r))\frac{\bar{X}}{\bar{x}_r}$	$MSE_{min}(\bar{y}_{(51)})=\lambda_1\bar{Y}^2[C_y^2(1-\rho^2)+C_x^2]$
Under strategy II			
$y_{(52)}=\begin{cases}\left(\frac{n}{r}y_i+b(\bar{X}-x_i)\right)\left(\frac{\bar{X}}{\bar{x}_n}\right) & i\in R\\ b(\bar{X}-x_i)\left(\frac{\bar{X}}{\bar{x}_n}\right) & i\in R^c\end{cases}$	$\bar{y}_{(52)}$	$(\bar{y}_r+\mathrm{b}(\bar{X}-\bar{x}_n))\frac{\bar{X}}{\bar{x}_n}$	$MSE_{min}(\bar{y}_{(52)})=\bar{Y}^2[\lambda_1C_y^2+\lambda_2(C_x^2-\rho^2C_y^2)]$
Under strategy III			
$y_{(53)}=\begin{cases}\left(\frac{n}{r}y_i-\frac{bnx_i}{r}\right)\left(\frac{\bar{x}_n}{\bar{x}_r}\right) & i\in R\\ \frac{bn\bar{x}_n}{n-r}\left(\frac{\bar{x}_n}{\bar{x}_r}\right) & i\in R^c\end{cases}$	$\bar{y}_{(53)}$	$(\bar{y}_r+\mathrm{b}(\bar{x}_n-\bar{x}_r))\frac{\bar{x}_n}{\bar{x}_r}$	$MSE_{min}(\bar{y}_{(53)})=\bar{Y}^2[\lambda_1C_y^2+\lambda_3(C_x^2-\rho^2C_y^2)]$
6. Singh *et al.* (2010)'s estimators method of imputation			
Under strategy I			
$y_{(61)}=\begin{cases}y_i & i\in R\\ \frac{1}{(n-r)}\bar{y}_r\left[\frac{n\{(A+C)\bar{X}+fB\bar{x}_r\}}{(A+fB)\bar{X}+C\bar{x}_r}-r\right] & i\in R^c\end{cases}$	$\bar{y}_{(61)}$	$\bar{y}_r\left[\frac{n\{(A+C)\bar{X}+fB\bar{x}_r\}}{(A+fB)\bar{X}+C\bar{x}_r}\right]$	$MSE_{min}(\bar{y}_{(61)})=\lambda_1\bar{Y}^2C_y^2(1-\rho^2)$
Under strategy II			
$y_{(62)}=\begin{cases}y_i & i\in R\\ \frac{1}{(n-r)}\bar{y}_r\left[\frac{n\{(A+C)\bar{X}+fB\bar{x}_n\}}{(A+fB)\bar{X}+C\bar{x}_n}-r\right] & i\in R^c\end{cases}$	$\bar{y}_{(62)}$	$\bar{y}_r\left[\frac{n\{(A+C)\bar{X}+fB\bar{x}_n\}}{(A+fB)\bar{X}+C\bar{x}_n}\right]$	$MSE_{min}(\bar{y}_{(62)})=\bar{Y}^2C_y^2(\lambda_1-\lambda_2\rho^2)$
where $A=(d-1)(d-2)$, $B=(d-1)(d-4)$, $C=(d-2)(d-3)(d-4)$, $f=\frac{n}{N}(0\le f\le 1)$ and d is a non-negative real constant (i.e. $d>0$).			
7. Pandey *et al.* (2015)'s estimator method of imputation			
Under strategy I			

$y_{(71)} = \begin{cases} y_i & i \in R \\ \frac{n}{n-r}\bar{y}_r\left[exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right)-\frac{r}{n}\right] & i \in R^c \end{cases}$	$\bar{y}_{(71)}$	$\bar{y}_r exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right)$	$MSE_{min}(\bar{y}_{(71)}) = \lambda_1\bar{Y}^2\left(C_y^2+\frac{1}{4}C_x^2 - \rho C_y C_x\right)$
Under strategy II $y_{(72)} = \begin{cases} y_i & i \in R \\ \frac{n}{n-r}\bar{y}_r\left[exp\left(\frac{\bar{X}-\bar{x}_n}{\bar{X}+\bar{x}_n}\right)-\frac{r}{n}\right] & i \in R^c \end{cases}$	$\bar{y}_{(72)}$	$\bar{y}_r exp\left(\frac{\bar{X}-\bar{x}_n}{\bar{X}+\bar{x}_n}\right)$	$MSE_{min}(\bar{y}_{(72)}) = \bar{Y}^2\left[\lambda_1 C_y^2 + \lambda_2\left(\frac{1}{4}C_x^2 - \rho C_y C_x\right)\right]$
Under strategy III $y_{(73)} = \begin{cases} y_i & i \in R \\ \frac{n}{n-r}\bar{y}_r\left[exp\left(\frac{\bar{x}_r-\bar{x}_n}{\bar{x}_r-\bar{x}_n}\right)-\frac{r}{n}\right] & i \in R^c \end{cases}$	$\bar{y}_{(73)}$	$\bar{y}_r exp\left(\frac{\bar{x}_r-\bar{x}_n}{\bar{x}_r-\bar{x}_n}\right)$	$MSE_{min}(\bar{y}_{(73)}) = \frac{\bar{Y}^2}{4}\left[4\lambda_1 C_y^2 + \lambda_1 C_x^2 + (\lambda_1-\lambda_2)\rho C_y C_x\right]$
8. Prasad (2018)'s estimator method of imputation			
Under strategy I $y_{(81)} = \begin{cases} y_i exp(\varphi(a^*,b^*,\bar{x}_r)) & i \in R \\ \frac{\bar{y}_r}{\bar{x}_r}\left[x_i - \frac{n}{n-r}(\bar{x}_n-\bar{x}_r)\right]exp(\varphi(a^*,b^*,\bar{x}_r)) & i \in R^c \end{cases}$	$\bar{y}_{(81)}$	$\bar{y}_r exp(\varphi(a^*,b^*,\bar{x}_r))$	$MSE_{min}(\bar{y}_{(81)}) = \lambda_1\bar{Y}^2\left(C_y^2+\frac{1}{4}\theta^2C_x^2 - \theta\rho C_y C_x\right)$
Under strategy II $y_{(82)} = \begin{cases} y_i exp(h(a^*,b^*,\bar{x}_n)) & i \in R \\ \frac{\bar{y}_r}{\bar{x}_r}\left[x_i - \frac{n}{n-r}(\bar{x}_n-\bar{x}_r)\right]exp(h(a^*,b^*,\bar{x}_n)) & i \in R^c \end{cases}$ where $\varphi(a^*,b^*,\bar{x}_r) = \frac{a^*(\bar{X}-\bar{x}_r)}{a^*(\bar{X}-\bar{x}_r)+2b^*}$, $h(a^*,b^*,\bar{x}_n) = \frac{a^*(\bar{X}-\bar{x}_n)}{a^*(\bar{X}-\bar{x}_n)+2b^*}$ with $a^* \neq 0$ and b^* are either some real numbers or the known values of population parameters of the auxiliary variable *x*.	$\bar{y}_{(82)}$	$\bar{y}_r exp(h(a^*,b^*,\bar{x}_n))$	$MSE_{min}(\bar{y}_{(82)}) = \bar{Y}^2\left(\lambda_1 C_y^2 + \lambda_2\left(\frac{1}{4}\theta^2C_x^2 - \theta\rho C_y C_x\right)\right)$ where $\theta = \frac{a^*\bar{X}}{a^*\bar{X}+b^*}$.

In the next section, we have proposed three imputation methods (according to three strategies I-III) to tackle the problem of missing data by replacing the missing values with the fabricated imputed values. Ultimately, we proposed various classes of estimators of $\bar{Y}$ under these strategies.

3. Proposed imputation procedure and corresponding estimators of $\bar{Y}$.

By getting motivation from Bhushan *et al.* (2015), we propose logarithmic type estimators of $\bar{Y}$, using Searl's technique (1964).

Under Strategy I: When $\bar{X}$ is known and $\bar{x}_r$ is used.

The proposed imputation procedure is as follows:

$$y_{.i} = \begin{cases} y_i & i \in R \\ \frac{1}{n-r}\left[n\left(\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)^g + b_1(\bar{X}-\bar{x}_r)\right)\left(1+log\left(\frac{\bar{x}_r^*}{\bar{X}^*}\right)^{\beta}\right) - r\bar{y}_r\right] & i \in R^c \end{cases} \quad (1)$$

$$y_{.i} = \begin{cases} \alpha_1 y_i & i \in R \\ \frac{\alpha_1}{n-r}\left[n\left(\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)^g + b_1(\bar{X}-\bar{x}_r)\right)\left(1+log\left(\frac{\bar{x}_r^*}{\bar{X}^*}\right)^{\beta}\right) - r\bar{y}_r\right] & i \in R^c \end{cases} \quad (2)$$

where $\bar{x}_r^* = p\bar{x}_r + q$, $\bar{X}^* = p\bar{X} + q$, g and β are design parameters which may take values either -1 or 0 or 1; b_1 and α_1 are constants which are to be suitably determined; $p(\neq 0)$ and q are either known real numbers or the known values of some population parameters of auxiliary variable x, such as correlation coefficient (ρ), coefficient of variation (C_x), coefficient of kurtosis $(\beta_2(x))$, standard deviation (S_x), coefficient of skewness $(\beta_1(x))$, etc.

Theorem 3.1 The point estimators of $\bar{Y}$ under the proposed method of imputation are respectively:

$$\bar{y}_{d1} = \left[\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)^g + b_1(\bar{X}-\bar{x}_r)\right]\left[1 + log\left(\frac{\bar{x}_r^*}{\bar{X}^*}\right)^\beta\right]$$

$$\bar{y}_{P1} = \alpha_1\left[\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)^g + b_1(\bar{X}-\bar{x}_r)\right]\left[1 + log\left(\frac{\bar{x}_r^*}{\bar{X}^*}\right)^\beta\right]$$

Proof: The point estimator of $\bar{Y}$ is

$$\bar{y}_t = \frac{1}{n}\left[\sum_{i\in R} y_{.i} + \sum_{i\in R^c} y_{.i}\right] \qquad (3)$$

On putting the values from (1) and (2) in (3), we obtained the proposed classes of estimators of $\bar{Y}$.

Theorem 3.2 The biases of proposed estimators $\bar{y}_{d1}$ and $\bar{y}_{P1}$, up to the terms of $o(n^{-1})$, are respectively:

$$Bias(\bar{y}_{d1}) = \bar{Y}\left(-\frac{1}{2}\beta k^2\lambda_1 C_x^2 - g\beta k\lambda_1 C_x^2 + \frac{1}{2}g(g+1)\lambda_1 C_x^2 + (\beta k - g)\lambda_1\rho C_y C_x\right) - b_1\bar{X}\beta k\lambda_1 C_x^2 .$$

$$Bias(\bar{y}_{P1})=\bar{Y}\left(\alpha_1(1 + (-\frac{1}{2}\beta k^2 - g\beta k + \frac{1}{2}g(g+1))\lambda_1 C_x^2 + (\beta k - g)\lambda_1\rho C_y C_x\right) - \alpha_1 b_1\bar{X}\beta k\lambda_1 C_x^2.$$

Proof: For derivation of above results, refer to Appendix A.

Theorem 3.3 The mean square errors of proposed estimators of $\bar{y}_{d1}$ and $\bar{y}_{P1}$, up to the terms of $o(n^{-1})$, are respectively:

$$\text{MSE}(\bar{y}_{d1}) = \bar{Y}^2\lambda_1 C_y^2 + (\beta^2k^2 + g^2 - 2g\beta k)\bar{Y}^2\lambda_1 C_x^2 + 2(\beta k - g)\bar{Y}^2\lambda_1\rho C_y C_x + 2b_1(g - \beta k)\bar{X}\bar{Y}\lambda_1 C_x^2 + b_1^2\bar{X}^2\lambda_1 C_x^2 - 2b_1\bar{X}\bar{Y}\lambda_1\rho C_y C_x$$

$$\text{MSE}(\bar{y}_{P1})=\alpha_1^2[\bar{Y}^2 + \bar{Y}^2\lambda_1 C_y^2 + \left(\beta^2k^2 + g^2 - \beta k^2 - 4g\beta k + g(g+1)\right)\bar{Y}^2\lambda_1 C_x^2 + 4(\beta k - g)\bar{Y}^2\lambda_1\rho C_y C_x + b_1^2\bar{X}^2\lambda_1 C_x^2 + 2b_1(g - 2\beta k)\bar{X}\bar{Y}\lambda_1 C_x^2 - 2b_1\bar{X}\bar{Y}\lambda_1\rho C_y C_x] + \alpha_1[-2\bar{Y}^2 + \left(\beta k^2 + 2g\beta k - g(g+1)\right)\bar{Y}^2\lambda_1 C_x^2 - 2(\beta k - g)\bar{Y}^2\lambda_1\rho C_y C_x + 2b_1\beta k\bar{X}\bar{Y}\lambda_1 C_x^2] + \bar{Y}^2 .$$

Here $k = \frac{p\bar{X}}{p\bar{X}+q}$.

Proof: For derivation of above results, refer to Appendix A.

Theorem 3.4 The minimum mean square errors of proposed estimators $\bar{y}_{d1}$ and $\bar{y}_{P1}$ w.r.t. b_1 and α_1, up to the terms of $o(n^{-1})$, are respectively:

$$MSE_{min}(\bar{y}_{d1}) = \lambda_1\bar{Y}^2(1-\rho^2)\, C_y^2 .$$

$$MSE_{min}(\bar{y}_{P1}) = \frac{\bar{Y}^2 MSE_{\min}(\bar{y}_{d1}) - \bar{Y}^2(Bias(\bar{y}_{d1}))^2}{\bar{Y}^2 + MSE_{\min}(\bar{y}_{d1}) + 2\bar{Y}Bias(\bar{y}_{d1})}$$

Proof: For derivation of above results, refer to Appendix A

Under Strategy II: When $\bar{X}$ is known and $\bar{x}_n$ is used.

The proposed imputation procedure is as follows:

$$y_{.i} = \begin{cases} y_i & i \in R \\ \frac{1}{n-r}\left[n\left(\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right)^g + b_2(\bar{X}-\bar{x}_n)\right)\left(1 + log\left(\frac{\bar{x}_n^*}{\bar{X}^*}\right)^\beta\right) - r\bar{y}_r\right] & i \in R^c \end{cases} \qquad (4)$$

$$y_{.i} = \begin{cases} \alpha_2 y_i & i \in R \\ \frac{\alpha_2}{n-r}\left[n\left(\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right)^g + b_2(\bar{X}-\bar{x}_n)\right)\left(1+log\left(\frac{\bar{x}_n^*}{\bar{X}^*}\right)^\beta\right) - r\bar{y}_r\right] & i \in R^c \end{cases} \quad (5)$$

where $\bar{x}_n{}^* = p\bar{x}_n + q, \bar{X}^* = p\bar{X} + q,$ g and β are design parameters which may take values either -1 or 0 or 1; b_2 and α_2 are constants which are to be suitably determined; $p(\neq 0)$ and q are either known real numbers or the known values of some population parameters of auxiliary variable as mentioned earlier.

Theorem 3.5 The point estimators of $\bar{Y}$ under the proposed method of imputation are respectively:

$$\bar{y}_{d2} = \left[\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right)^g + b_2(\bar{X}-\bar{x}_n)\right]\left[1+log\left(\frac{\bar{x}_n^*}{\bar{X}^*}\right)^\beta\right]$$

$$\bar{y}_{P2} = \alpha_2\left[\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right)^g + b_2(\bar{X}-\bar{x}_n)\right]\left[1+log\left(\frac{\bar{x}_n^*}{\bar{X}^*}\right)^\beta\right]$$

Proof: The point estimator of $\bar{Y}$ is

$$\bar{y}_t = \frac{1}{n}\left[\sum_{i\in R} y_{.i} + \sum_{i\in R^c} y_{.i}\right] \quad (6)$$

On putting the values from (4) and (5) in (6), we obtained the proposed classes of estimators of $\bar{Y}$.

Theorem 3.6 The biases of proposed estimators $\bar{y}_{d2}$ and $\bar{y}_{P2}$, up to the terms of $o(n^{-1})$, are respectively:

$$Bias(\bar{y}_{d2}) = \bar{Y}\left(-\frac{1}{2}\beta k^2\lambda_2 C_x^2 - g\beta k\lambda_2 C_x^2 + \frac{1}{2}g(g+1)\lambda_2 C_x^2 + (\beta k - g)\lambda_2\rho C_y C_x\right) - b_2\bar{X}\beta k\lambda_2 C_x^2$$

$$Bias(\bar{y}_{P2}) = \bar{Y}\left(\alpha_2(1 + (-\frac{1}{2}\beta k^2 - g\beta k + \frac{1}{2}g(g+1))\lambda_2 C_x^2 + (\beta k - g)\lambda_2\rho C_y C_x\right) - \alpha_2 b_2\bar{X}\beta k\lambda_2 C_x^2$$

Proof: This is similar to proof of Theorem 3.2.

Theorem 3.7 The mean square errors of proposed estimators $\bar{y}_{d2}$ and $\bar{y}_{P2}$, up to the terms of $o(n^{-1})$, are respectively:

$$MSE(\bar{y}_{d2}) = \bar{Y}^2\lambda_1 C_y^2 + (\beta^2k^2 + g^2 - 2g\beta k)\bar{Y}^2\lambda_2 C_x^2 + 2(\beta k - g)\bar{Y}^2\lambda_2\rho C_y C_x + 2b_2(g - \beta k)\bar{X}\bar{Y}\lambda_2 C_x^2 + b_2^2\bar{X}^2\lambda_2 C_x^2 - 2b_2\bar{X}\bar{Y}\lambda_2\rho C_y C_x \quad .$$

$$\text{MSE}(\bar{y}_{P2}) = \alpha_2^2[\bar{Y}^2 + \bar{Y}^2\lambda_1 C_y^2 + (\beta^2k^2 + g^2 - \beta k^2 - 4g\beta k + g(g + +1))\bar{Y}^2\lambda_2 C_x^2 + 4(\beta k - g)\bar{Y}^2\lambda_2\rho C_y C_x + b_2^2\bar{X}^2\lambda_2 C_x^2 + 2b_2(g - 2\beta k)\bar{X}\bar{Y}\lambda_2 C_x^2 - 2b_2\bar{X}\bar{Y}\lambda_2\rho C_y C_x] + \alpha_2[-2\bar{Y}^2 + \left(\beta k^2 + 2g\beta k - g(g+1)\right)\bar{Y}^2\lambda_2 C_x^2 - 2(\beta k - g)\bar{Y}^2\lambda_2\rho C_y C_x + 2b_2\beta k\bar{X}\bar{Y}\lambda_2 C_x^2] + \bar{Y}^2.$$

Here $k = \frac{p\bar{X}}{p\bar{X}+q}$.

Proof: This is similar to proof of Theorem 3.3.

Theorem 3.8 The minimum mean square errors of proposed estimators $\bar{y}_{d2}$ and $\bar{y}_{P2}$ w.r.t. b_2 and α_2, up to the terms of $o(n^{-1})$, are respectively:

$$MSE_{min}(\bar{y}_{d2}) = \bar{Y}^2[\lambda_2(1-\rho^2)\ C_y^2\ + \lambda_3\ \ C_y^2\ ].$$

$$\mathrm{MS}E_{\min}(\bar{y}_{P2}) = \frac{\bar{Y}^2 MSE_{\min}(\bar{y}_{d2}) - \bar{Y}^2\left(Bias(\bar{y}_{d2})\right)^2}{\bar{Y}^2 + MSE_{\min}(\bar{y}_{d2}) + 2\bar{Y}Bias(\bar{y}_{d2})}.$$

Proof: This is similar to proof of Theorem 3.4.

Under Strategy III: When $\bar{X}$ is unknown and both $\bar{x}_n$ and $\bar{x}_r$ are used.

The proposed imputation procedure is as follows:

$$y_{.i} = \begin{cases} y_i & i \in R \\ \frac{1}{n-r}\left[n\left(\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^g + b_3(\bar{x}_n - \bar{x}_r)\right)\left(1 + log\left(\frac{\bar{x}_r^*}{\bar{x}_n^*}\right)^\beta\right) - r\bar{y}_r\right] & i \in R^c \end{cases} \quad (7)$$

$$y_{.i} = \begin{cases} \alpha_3 y_i & i \in R \\ \frac{\alpha_3}{n-r}\left[n\left(\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^g + b_3(\bar{x}_n - \bar{x}_r)\right)\left(1 + log\left(\frac{\bar{x}_r^*}{\bar{x}_n^*}\right)^\beta\right) - r\bar{y}_r\right] & i \in R^c \end{cases} \quad (8)$$

where b_3 and α_3 are constants which are to be suitably determined.

Theorem 3.9 The point estimators of $\bar{Y}$ under the proposed method of imputation are respectively:

$$\bar{y}_{d3} = \ [\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^g + b_3(\bar{x}_n - \bar{x}_r)][1 + log\left(\frac{\bar{x}_r{}^*}{\bar{x}_n{}^*}\right)^\beta]$$

$$\bar{y}_{P3} = \ \alpha_3[\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^g + b_3(\bar{x}_n - \bar{x}_r)][1 + log\left(\frac{\bar{x}_r{}^*}{\bar{x}_n{}^*}\right)^\beta]$$

where $\bar{x}_n{}^* = p\bar{x}_n + q, \bar{x}_r{}^* = p\bar{x}_r + q,\ ,\ g$ and β are design parameters which may take values either -1 or 0 or 1; b_3 and α_3 are constants which are to be suitably determined; $p(\neq 0)$ and q are either known real numbers or the known values of some population parameters of auxiliary variable x, as mentioned earlier.

Proof: The point estimator of $\bar{Y}$ is

$$\bar{y}_t = \frac{1}{n}[\textstyle\sum_{i\in R} y_{.i} + \sum_{i\in R^c} y_{.i}]$$

(9)

Putting the values from (7) and (8) in (9), we obtained the proposed classes of estimators of $\bar{Y}$.

Theorem 3.10 The biases of proposed estimators $\bar{y}_{d3}$ and $\bar{y}_{P3}$, up to the terms of $o(n^{-1})$, are respectively:

$$Bias(\bar{y}_{d3}) = \ \bar{Y}\left(-\frac{1}{2}\beta k^2\lambda_3 C_x^2 - g\beta k\lambda_3 C_x^2 + \frac{1}{2}g(g+1)\lambda_3 C_x^2 + (\beta k - g)\lambda_3\rho C_y C_x\right) - b_3\bar{X}\beta k\lambda_3 C_x^2\ .$$

$$Bias(\bar{y}_{P3}) = \bar{Y}\left(\alpha_3(1 + (-\frac{1}{2}\beta k^2 - g\beta k + \frac{1}{2}g(g+1))\lambda_3 C_x^2 + (\beta k - g)\lambda_3\rho C_y C_x\right) - \alpha_3 b_3\bar{X}\beta k\lambda_3 C_x^2\ .$$

Proof: This is similar to proof of Theorem 3.2.

Theorem 3.11 The mean square errors of proposed estimators $\bar{y}_{d3}$ and $\bar{y}_{P3}$, up to the terms of $o(n^{-1})$, are respectively:

$$MSE(\bar{y}_{d3})= \bar{Y}^2\lambda_1 C_y^2 + (\beta^2k^2 + g^2 - 2g\beta k)\bar{Y}^2\lambda_3 C_x^2 + 2(\beta k - g)\bar{Y}^2\lambda_3\rho C_y C_x + 2b_3(g-\beta k)\bar{X}\bar{Y}\lambda_3 C_x^2 + b_3^2\bar{X}^2\lambda_3 C_x^2 - 2b_3\bar{X}\bar{Y}\lambda_3\rho C_y C_x \ .$$

$$\text{MSE}(\bar{y}_{P3})=\alpha_3^2[\bar{Y}^2 + \bar{Y}^2\lambda_1 C_y^2 + (\beta^2k^2 + g^2 - \beta k^2 - 4g\beta k + g(g+1))\bar{Y}^2\lambda_3 C_x^2 + 4(\beta k - g)\bar{Y}^2\lambda_3\rho C_y C_x + b_3^2\bar{X}^2\lambda_3 C_x^2 + 2b_3(g-2\beta k)\bar{X}\bar{Y}\lambda_3 C_x^2 - 2b_3\bar{X}\bar{Y}\lambda_3\rho C_y C_x] + \alpha_3[-2\bar{Y}^2 + \left(\beta k^2 + 2g\beta k - g(g+1)\right)\bar{Y}^2\lambda_3 C_x^2 - 2(\beta k - g)\bar{Y}^2\lambda_3\rho C_y C_x + 2b_3\beta k\bar{X}\bar{Y}\lambda_3 C_x^2] + \bar{Y}^2.$$

Proof: This is similar to proof of Theorem 3.3.

Theorem 3.12 The minimum mean square errors of proposed estimators $\bar{y}_{d3}$ and $\bar{y}_{P3}$ w.r.t. b_3 and α_3, up to the terms of $o(n^{-1})$, are respectively:

$$MSE_{min}(\bar{y}_{d3}) = \bar{Y}^2[\lambda_3(1-\rho^2)\ C_y^2 \ + \lambda_2 C_y^2]$$

$$MSE_{\min}(\bar{y}_{P3}) = \frac{\bar{Y}^2 MSE_{\min}(\bar{y}_{d3}) - \bar{Y}^2(Bias(\bar{y}_{d3}))^2}{\bar{Y}^2 + MSE_{\min}(\bar{y}_{d3}) + 2\bar{Y}Bias(\bar{y}_{d3})} \ .$$

Proof: This is similar to proof of Theorem 3.4.

Remark 3.1: If the population parameters are not known, then we can replace them by their consistent estimators in the optimum values of constants b_i $(i = 1,2,3)$ and α_i $(i = 1,2,3)$, based on the same sample observations. The mean square errors of the estimators so obtained will remain the same up to the terms of order n^{-1} [see Srivastava and Jhajj (1983)].

Remark 3.2: As shown in the following table, for particular values of g, β, α, b_i $(i = 1,2,3)$ and k, we can see that the proposed classes of estimators $\bar{y}_{Pi}$ (i=1,2,3) reduces to some of the existing imputed estimators.

Table 2. Various members of the proposed classes of imputed estimators

Values of g	Values of β	Values of α	Values of b_i $(i = 1,2,3)$	Values of k	Existing Estimators
0	0	1	0	1	Mean per unit estimator
1	0	1	0	1	Ratio estimator
-1	0	1	0	1	Product estimator
0	0	1	b	1	Diana & Perri (2010)'s estimator

Remark 3.3: As we know that bias has negligible effect on the accuracy of an estimator when the bias is less than one tenth of the Standard Deviation (SD) of an estimator so we can be confident that the ratio Bias/SD will not exceed 0.1 if the sample size is large enough. Keeping in view this fact, we shall consider only mean square errors of various estimators and not their biases in the comparative study to be taken up

in the next section (see pp. 14–15 of Cochran (1977)).

4. Relative Performances of Proposed Estimators

In this section, we present a comparative analysis of the MSEs of the proposed classes of estimators of $\bar{Y}$ with that of the several existing estimators. The detailed results of this analysis are presented below.

4.1 Under Strategy I: Comparison of minimum MSEs of proposed classes $\bar{y}_{d1}$ and $\bar{y}_{P1}$ with that of existing estimators

Here, we have compared mean square errors of $\bar{y}_{d1}$ and $\bar{y}_{P1}$ with that of existing estimators namely $\bar{y}_{(1)}, \bar{y}_{(21)}, \bar{y}_{(31)}, \bar{y}_{(41)}, \bar{y}_{(51)}, \bar{y}_{(61)}, \bar{y}_{(71)}, \bar{y}_{(81)}$. Taking the following:

$\mathrm{Var}(\bar{y}_{(1)}) - MSE_{min}(\bar{y}_{d1}) = \lambda_1 \rho^2 \bar{Y}^2 C_y^2 > 0$, always.

$\mathrm{MSE}(\bar{y}_{(21)}) - MSE_{min}(\bar{y}_{d1}) = \lambda_1 \bar{Y}^2 (\rho C_y - C_x)^2 > 0$, always.

$MSE_{min}(\bar{y}_{d1}) = MSE(\bar{y}_{(31)}) = MSE_{min}(\bar{y}_{(41)}) = MSE_{min}(\bar{y}_{(61)})$

$\mathrm{MSE}(\bar{y}_{(51)}) - MSE_{min}(\bar{y}_{d1}) = \lambda_1 \bar{Y}^2 C_x^2 > 0$, always.

$\mathrm{MSE}(\bar{y}_{(71)}) - MSE_{min}(\bar{y}_{d1}) = \lambda_1 \bar{Y}^2 \left(\frac{C_x}{2} - \rho C_y\right)^2 > 0$, always.

$\mathrm{MSE}(\bar{y}_{(81)}) - MSE_{min}(\bar{y}_{d1}) = \lambda_1 \bar{Y}^2 \left(\rho C_y - \frac{\theta_j C_x}{2}\right)^2 > 0$, always.

Further note that,

$$MSE_{min}(\bar{y}_{d1}) - \mathrm{MSE}_{\min}(\bar{y}_{P1}) = MSE_{min}(\bar{y}_{d1}) - \frac{\bar{Y}^2 MSE_{\min}(\bar{y}_{d1}) - \bar{Y}^2 \left(Bias(\bar{y}_{d1})\right)^2}{\bar{Y}^2 + MSE_{\min}(\bar{y}_{d1}) + 2\bar{Y} Bias(\bar{y}_{d1})}$$
$$= \left(\frac{(MSE_{min}(\bar{y}_{d1}) + \bar{Y} Bias(\bar{y}_{d1}))^2}{\bar{Y}^2 + MSE_{\min}(\bar{y}_{d1}) + 2\bar{Y} Bias(\bar{y}_{d1})}\right) > 0, \text{ provided that } Bias(\bar{y}_{d1}) > 0. \quad (10)$$

Thus when $Bias(\bar{y}_{d1}) > 0$ then using result (10), we have

$\mathrm{Var}(\bar{y}_{(1)}) - \mathrm{MSE}_{\min}(\bar{y}_{P1}) > 0$, always.

$\mathrm{MSE}(\bar{y}_{(l1)}) - \mathrm{MSE}_{\min}(\bar{y}_{P1}) > 0$, always, where $l = 2,3,4,5,6,7,8$.

4.2 Under strategy II: Comparison of minimum MSEs of proposed classes $\bar{y}_{d2}$ and $\bar{y}_{P2}$ with that of existing estimators

We have compared the MSEs of the estimators $\bar{y}_{d2}$ and $\bar{y}_{P2}$ with that of the existing estimators namely $\bar{y}_{(1)}, \bar{y}_{(22)}, \bar{y}_{(32)}, \bar{y}_{(42)}, \bar{y}_{(52)}, \bar{y}_{(62)}, \bar{y}_{(72)}, \bar{y}_{(82)}$. Taking the following:

$\mathrm{Var}(\bar{y}_{(1)}) - MSE_{min}(\bar{y}_{d2}) = \lambda_2 \rho^2 \bar{Y}^2 C_y^2 > 0$, always.

$\mathrm{MSE}(\bar{y}_{(22)}) - MSE_{min}(\bar{y}_{d2}) = \lambda_2 \bar{Y}^2 (\rho C_y - C_x)^2 > 0$, always.

$MSE_{min}(\bar{y}_{d2}) = MSE(\bar{y}_{(32)}) = MSE_{min}(\bar{y}_{(42)}) = MSE_{min}(\bar{y}_{(62)})$.

$\mathrm{MSE}(\bar{y}_{(52)}) - MSE_{min}(\bar{y}_{d2}) = \lambda_2 \bar{Y}^2 C_x^2 > 0$, always.

$\mathrm{MSE}(\bar{y}_{(72)}) - MSE_{min}(\bar{y}_{d2}) = \lambda_2 \bar{Y}^2 \left(\frac{C_x}{2} - \rho C_y\right)^2 > 0$, always.

$\mathrm{MSE}(\bar{y}_{(82)}) - MSE_{min}(\bar{y}_{d2}) = \lambda_2 \bar{Y}^2 \left(\rho C_y - \frac{\theta_j C_x}{2}\right)^2 > 0$, always.

Further note that,

$$MSE_{min}(\bar{y}_{d2}) - \mathrm{MSE}_{\min}(\bar{y}_{P2}) = MSE_{min}(\bar{y}_{d2}) - \frac{\bar{Y}^2 MSE_{\min}(\bar{y}_{d2}) - \bar{Y}^2 \left(Bias(\bar{y}_{d2})\right)^2}{\bar{Y}^2 + MSE_{\min}(\bar{y}_{d2}) + 2\bar{Y} Bias(\bar{y}_{d2})}$$
$$= \left(\frac{(MSE_{min}(\bar{y}_{d2}) + \bar{Y} Bias(\bar{y}_{d2}))^2}{\bar{Y}^2 + MSE_{\min}(\bar{y}_{d2}) + 2\bar{Y} Bias(\bar{y}_{d2})}\right) > 0, \text{ provided that } Bias(\bar{y}_{d2}) > 0. \quad (11)$$

Thus when $Bias(\bar{y}_{d2}) > 0$ then using result (11), we have

$\mathrm{Var}(\bar{y}_{(1)}) - \mathrm{MSE}_{\min}(\bar{y}_{P2}) > 0$, always

$\text{MSE}(\bar{y}_{(l2)}) - MSE_{\min}(\bar{y}_{P2}) > 0$, always, where $l = 2,3,4,5,6,7,8$.

4.3 Under strategy III: Comparison of minimum MSEs of proposed classes $\bar{y}_{d2}$ and $\bar{y}_{P2}$ with that of existing estimators

Here, we have compared MSEs of $\bar{y}_{d3}$ and $\bar{y}_{P3}$ with that of existing estimators namely $\bar{y}_{(1)}$, $\bar{y}_{(23)}, \bar{y}_{(33)}, \bar{y}_{(43)}, \bar{y}_{(53)}, \bar{y}_{(73)}$. Taking the following:

$\text{Var}(\bar{y}_{(1)}) - MSE_{min}(\bar{y}_{d3}) = \lambda_3 \rho^2 \bar{Y}^2 C_y^2 > 0$, always.

$\text{MSE}(\bar{y}_{(23)}) - MSE_{min}(\bar{y}_{d3}) = \lambda_3 \bar{Y}^2 (\rho C_y - C_x)^2 > 0$, always.

$MSE_{min}(\bar{y}_{d3}) = MSE(\bar{y}_{(33)}) = MSE_{min}(\bar{y}_{(43)})$.

$\text{MSE}(\bar{y}_{(53)}) - MSE_{min}(\bar{y}_{d3}) = \lambda_3 \bar{Y}^2 C_x^2 > 0$, always.

$\text{MSE}(\bar{y}_{(73)}) - MSE_{min}(\bar{y}_{d3}) = \bar{Y}^2 [\lambda_3 (\rho^2 C_y^2 + \rho C_y C_x) + \lambda_1 C_x^2] > 0$, always.

Further note that,

$$MSE_{min}(\bar{y}_{d3}) - MSE_{\min}(\bar{y}_{P3}) = MSE_{min}(\bar{y}_{d3}) - \frac{\bar{Y}^2 MSE_{\min}(\bar{y}_{d3}) + (Bias(\bar{y}_{d3}))^2}{\bar{Y}^2 + MSE_{\min}(\bar{y}_{d3}) + 2\bar{Y} Bias(\bar{y}_{d3})}$$

$$= \left(\frac{(MSE_{min}(\bar{y}_{d3}) + \bar{Y} Bias(\bar{y}_{d3}))^2}{\bar{Y}^2 + MSE_{\min}(\bar{y}_{d3}) + 2\bar{Y} Bias(\bar{y}_{d3})} \right) > 0, \text{ provided that } Bias(\bar{y}_{d3}) > 0. \qquad (12)$$

Thus when $Bias(\bar{y}_{d3}) > 0$ then using result (12), we have

$\text{Var}(\bar{y}_{(1)}) - MSE_{\min}(\bar{y}_{P3}) > 0$, always

$\text{MSE}(\bar{y}_{(l3)}) - MSE_{\min}(\bar{y}_{P3}) > 0$, always, where $l = 2,3,4,5,7$.

From the above theoretical results, we can see that the proposed classes of estimators $\bar{y}_{Pi}$ (i=1, 2, 3) have mostly lesser mean square error than that of the existing estimators, considered in this paper.

5 Relative Performances of Proposed Estimators

To support the theoretical results obtained in the previous section, we have considered four real life data sets and three hypothetical data sets.

5.1 Empirical Study

The four real life data sets have been taken into consideration for the practical applications of proposed classes of estimators. The descriptions of these four populations are given in Table 3. The performances of existing and proposed imputation techniques are compared on the basis of their Percent Relative Efficiencies (PRE). The PRE of an estimator $\bar{y}_m$ w.r.t. $\bar{y}_1$ is defined as:

$$PRE(\bar{y}_m) = \frac{Var(\bar{y}_1)}{MSE(\bar{y}_m)} \times 100 \text{ or } PRE(\bar{y}_m) = \frac{Var(\bar{y}_1)}{MSE_{min}(\bar{y}_m)} \times 100$$

To investigate the sensitivity and robustness of the proposed estimators, we have taken multiple combinations of the parameters p and q. Just to get an idea, we have considered the following values of (p, q): $(1, \beta_2(x)), (1, \rho), (1, C_x), (C_x, \rho), (C_x, \beta_2(x))$ and $(\beta_2(x), \rho)$. Obviously, we have got different values of $k = \frac{p\bar{X}}{p\bar{X}+q}$. This comprehensive evaluation enables a better understanding of how variations in p and q jointly influence the efficiency behavior of the estimators.

The tables 4 - 6 give the values of MSEs and PREs of various estimators under the three considered imputation Strategies I, II and III respectively. Just to have a bird eye view, we have made a bar chart in

Figure 1, which represents *PRE*s of various considered estimators in population 2.

Table 3. Data description of real population data sets

Population Parameters	**Population 1 {Source: Kadilar and Cingi (2006)}**	**Population 2 {Source: Cochran (1977)}**	**Population 3 {Source: Murthy (1967)}**	**Population 4 {Source: Kadilar and Cingi (2006)}**
N	106	10	80	104
n	20	5	25	20
r	18	4	20	17
$\bar{Y}$	15.37	56.9	5182.638	6.254
$\bar{X}$	243.76	54.2961	285.125	13931.683
$\beta_2(x)$	25.71	2.5932	3.5360	17.516
C_y	4.1802	0.1840	0.3542	1.866
C_x	2.0179	0.1621	0.9485	1.653
ρ	0.82	0.9237	0.9140	0.865

Table 4. PREs of the existing and proposed imputation techniques under Strategy I

Estimators	**Population 1**		**Population 2**		**Population 3**		**Population 4**	
	MSE	**PRE**	**MSE**	**PRE**	**MSE**	**PRE**	**MSE**	**PRE**
$\bar{y}_{(1)}$	190.3908	100	16.4418	100	126365.9688	100	6.7015	100
$\bar{y}_{(21)}$	84.1723	226.5766	2.4433	672.9318	41395.9136	305.2619	1.6901	396.4952
$\bar{y}_{(31)}$	62.4781	304.7320	2.4133	681.2995	20800.3439	607.5186	1.6872	397.1800
$\bar{y}_{(41)}$	62.4781	304.7320	2.4133	681.2995	20800.3439	607.5186	1.6872	397.1800
$\bar{y}_{(51)}$	106.9194	178.3723	15.1742	108.3539	92696.1877	136.3227	5.9462	112.7022
$\bar{y}_{(61)}$	62.4781	304.7320	2.4133	681.2995	20800.3439	607.5186	1.6872	397.1800
$\bar{y}_{(71)}$	126.3332	150.9617	6.2523	262.9703	43617.9802	289.7107	2.8811	232.6006
$\bar{y}_{(81)}$	131.5165	145.0121	6.5780	249.9506	41890.5522	301.6574	2.8842	232.3467
$\boldsymbol{\bar{y}_{P1(1)}}$ $(\boldsymbol{p=1, q=\beta_2(x)})$	**58.2856**	**326.6515**	**2.3369**	**703.5619**	**10561.1244**	**1196.5200**	**0.2375**	**2821.4719**
$\boldsymbol{\bar{y}_{P1(2)}}$ $(\boldsymbol{p=1, q=\rho})$	**54.3625**	**350.2245**	**2.0456**	**803.7641**	**10534.4589**	**1199.5487**	**0.1982**	**3381.1806**
$\boldsymbol{\bar{y}_{P1(3)}}$ $(\boldsymbol{p=1, q=C_x})$	**59.6325**	**319.2735**	**2.1963**	**748.6135**	**10600.9632**	**1192.0234**	**0.3256**	**2058.2002**
$\boldsymbol{\bar{y}_{P1(4)}}$ $(\boldsymbol{p=C_x, q=\rho})$	**55.1556**	**345.1885**	**2.1023**	**782.0862**	**10563.2365**	**1196.2807**	**0.3452**	**1941.3383**
$\boldsymbol{\bar{y}_{P1(5)}}$ $(\boldsymbol{p=C_x, q=\beta_2(x)})$	**50.2635**	**378.7854**	**1.9563**	**840.4539**	**10494.7456**	**1204.0874**	**0.1352**	**4956.7307**
$\boldsymbol{\bar{y}_{P1(6)}}$ $(\boldsymbol{p=\beta_2(x), q=\rho})$	**53.2658**	**357.4353**	**2.0004**	**821.9256**	**10507.9563**	**1202.5741**	**0.1854**	**3614.6170**

Table 5. PREs of the existing and proposed imputation techniques under Strategy II

Estimators	Population 1		Population 2		Population 3		Population 4	
	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
$\bar{y}_{(1)}$	190.3908	100	16.4418	100	126365.9688	100	6.7015	100
$\bar{y}_{(22)}$	96.8411	196.6011	7.1095	231.266	53726.0616	235.2042	2.5887	258.8686
$\bar{y}_{(32)}$	70.2795	270.9052	6.285	261.6024	43404.4188	291.1361	2.2838	293.4327
$\bar{y}_{(42)}$	70.2795	270.9052	6.285	261.6024	43404.4188	291.1361	2.2838	293.4327
$\bar{y}_{(52)}$	116.8145	162.9856	15.5967	105.4184	71347.7960	177.1126	5.9023	113.5404
$\bar{y}_{(62)}$	70.2795	270.9052	6.285	261.6024	43404.4188	291.1361	2.2838	293.4327
$\bar{y}_{(72)}$	133.8605	142.2307	9.6488	170.4020	65684.1105	192.3844	3.5661	187.9196
$\bar{y}_{(82)}$	138.4121	137.5536	9.8659	166.6520	64417.3299	196.1676	3.5687	187.7835
$\bar{y}_{P2(1)}$ $(p=1, q=\beta_2(x))$	**70.1269**	**271.4944**	**6.2747**	**262.0307**	**34057.1998**	**371.0403**	**1.7057**	**392.8879**
$\bar{y}_{P2(2)}$ $(p=1, q=\rho)$	**68.3214**	**278.6693**	**6.1523**	**267.2463**	**32056.3262**	**394.1997**	**1.5632**	**428.7039**
$\bar{y}_{P2(3)}$ $(p=1, q=C_x)$	**69.3201**	**274.0907**	**5.9215**	**277.6627**	**36123.7852**	**349.8137**	**2.0145**	**332.6631**
$\bar{y}_{P2(4)}$ $(p=C_x, q=\rho)$	**65.2145**	**291.9455**	**5.9652**	**275.6286**	**31256.7896**	**404.2832**	**1.4562**	**460.2046**
$\bar{y}_{P2(5)}$ $(p=C_x, q=\beta_2(x))$	**60.1456**	**316.9458**	**5.4269**	**302.9685**	**30102.4556**	**419.7862**	**1.0025**	**668.4788**
$\bar{y}_{P2(6)}$ $(p=\beta_2(x), q=\rho)$	**63.5639**	**299.5266**	**5.7896**	**283.9885**	**31052.8796**	**406.9380**	**1.2651**	**529.7209**

Table 6. PREs of the existing and proposed imputation techniques under Strategy III

Estimators	Population 1		Population 2		Population 3		Population 4	
	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
$\bar{y}_{(1)}$	190.3908	100	16.4418	100	126365.9688	100	6.7015	100
$\bar{y}_{(23)}$	177.5791	107.2146	11.7756	139.6255	105652.8208	119.6049	5.8029	115.4849
$\bar{y}_{(33)}$	174.9704	108.8131	11.7656	139.7442	98215.1355	128.6624	5.8024	115.4953
$\bar{y}_{(43)}$	174.9704	108.8131	11.7656	139.7442	98215.1355	128.6624	5.8024	115.4953
$\bar{y}_{(53)}$	180.3145	105.5882	16.0193	102.6377	99859.3605	126.5439	6.1256	109.4015
$\bar{y}_{(73)}$	182.6490	104.2386	13.0453	126.0360	104299.8385	121.1564	6.0165	111.3859
$\bar{y}_{P3(1)}$ $(p=1, q=\beta_2(x))$	**136.7851**	**139.1897**	**11.2656**	**145.9469**	**90852.4323**	**139.0892**	**5.7929**	**115.6851**
$\bar{y}_{P3(2)}$ $(p=1, q=\rho)$	**132.5695**	**143.6158**	**10.2647**	**160.1780**	**88256.4789**	**143.1803**	**5.3215**	**125.9325**
$\bar{y}_{P3(3)}$ $(p=1, q=C_x)$	**139.5412**	**136.4405**	**11.6489**	**141.1446**	**92546.8523**	**136.5426**	**5.5125**	**121.5691**
$\bar{y}_{P3(4)}$ $(p=C_x, q=\rho)$	**125.0321**	**152.2735**	**10.0025**	**164.3769**	**87659.4781**	**144.1555**	**5.0785**	**131.9582**
$\bar{y}_{P3(5)}$ $(p=C_x, q=\beta_2(x))$	**121.6547**	**156.5000**	**9.1256**	**180.1722**	**85741.6632**	**147.3798**	**4.9236**	**136.9097**
$\bar{y}_{P3(6)}$ $(p=\beta_2(x), q=\rho)$	**130.2566**	**146.1659**	**10.7856**	**152.4421**	**90045.5548**	**140.3355**	**5.2145**	**128.5166**

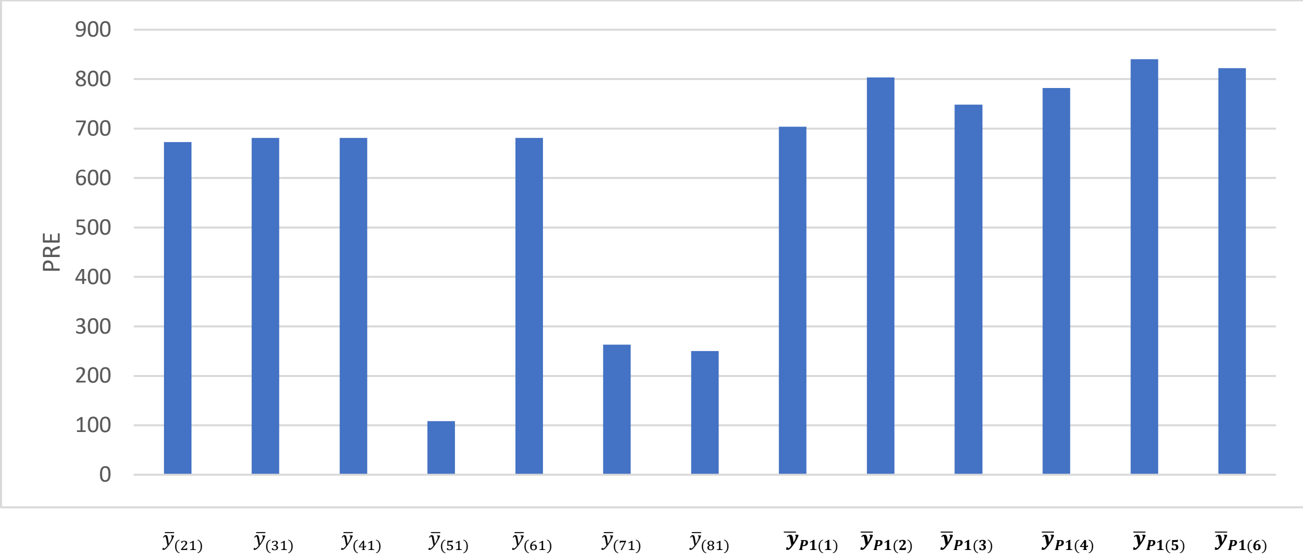

Figure 1. Population 2 (Real Data Set (Strategy I)).

5.2. Simulation Study

In order to obtain the actual picture of the performances of the proposed classes of estimators, we have generated artificial data sets, using R language. Through this software, we have generated three artificial data sets by using bivariate normal population. The study variable y is generated as $y = \rho^2 x + e$, where $e \sim N(0,1)$.

The parameters for these three bivariate normal populations are:

For Population 5:

$\bar{Y} = 756.2626$, $\bar{X} = 1181.6$, $S_y = 1117.4536$, $S_x = 1746.0503$, $\rho = 0.8$

For Population 6:

$\bar{Y} = 425.4131$, $\bar{X} = 1181.6$, $S_y = 628.5903$, $S_x = 1746.0503$, $\rho = 0.6$

For Population 7:

$\bar{Y} = 1202.943$, $\bar{X} = 1485$, $S_y = 1147.891$, $S_x = 1417.161$, $\rho = 0.9$

The algorithm for our used simulation process is as follows:

(i) A Simple Random Sample Without Replacement (SRSWOR) of size $n = 80$ is drawn from a population of size $N = 600$.

(ii) From this selected sample S, $n - r$ sample units (with $r = 35$) are deleted for the observation of the study variable y only.

(iii) For these deleted units, the missing values of the variable y are replaced by the corresponding imputed values according to the specified imputation strategies.

(iv) The above procedure is repeated 20000 times and obtained the value of estimator $\hat{t}$ of $\bar{Y}$ as $\hat{t}_1, \hat{t}_2, \dots, \hat{t}_{20000}$

(v) The mean square of error of $\hat{t}$ is obtained by using the following formula:

$$MSE(\hat{t}) = \frac{1}{20000}\sum_{i=1}^{20000}(\hat{t}_i - \bar{Y})^2$$

(vi) Finally, to assess the comparative performance of different estimators, we have calculated the PREs of the various estimators with respect to simple mean estimator $\bar{y}_{(1)}$ by using the following expression:

$$PRE(\hat{t}) = \frac{Var(\bar{y}_{(1)})}{MSE(\hat{t})} \times 100$$

For all artificial populations, the MSEs and PREs of various estimators under the three considered strategies are given in Tables 7 - 9. Further, to provide a clearer and more comprehensive comparison of the performance of all the proposed strategies, the PREs of various estimators for Population 7 have been illustrated diagrammatically (bar chart). This visual representation helps in identifying the trends in relative efficiencies of different estimators more effectively (see Figure 2).

Table 7. PREs of the existing and proposed imputation techniques under Strategy I

Estimators	Population 5		Population 6		Population 7	
	MSE	PRE	MSE	PRE	MSE	PRE
$\bar{y}_{(1)}$	33600.2178	100	10631.1942	100	35451.1603	100
$\bar{y}_{(21)}$	13440.8911	249.9850	8505.8217	124.9872	7090.3609	499.9909
$\bar{y}_{(31)}$	12096.0784	277.7777	6803.9642	156.25	6735.7204	526.3157
$\bar{y}_{(41)}$	12096.0784	277.7777	6803.9642	156.25	6735.7204	526.3157
$\bar{y}_{(51)}$	15700.3158	214.0098	7437.3242	142.9438	7088.1693	500.1455
$\bar{y}_{(61)}$	12096.0784	277.7777	6803.9642	156.25	6735.7204	526.3157
$\bar{y}_{(71)}$	15119.4951	222.2311	6910.1679	153.8485	12407.6484	285.7202
$\bar{y}_{(81)}$	15132.0987	222.0459	6911.4995	153.8189	12416.7578	285.5106
$\bar{y}_{P1(1)}$ $(p = 1, q = \beta_2(x))$	**10104.0682**	**332.5414**	**6660.8090**	**159.6081**	**5381.6642**	**658.7397**
$\bar{y}_{P1(2)}$ $(p = 1, q = \rho)$	**10000.2145**	**335.9949**	**6561.0123**	**162.0345**	**5263.7485**	**673.4964**
$\bar{y}_{P1(3)}$ $(p = 1, q = C_x)$	**11041.6587**	**304.3040**	**6789.9563**	**156.5723**	**5564.9856**	**637.0359**
$\bar{y}_{P1(4)}$ $(p = C_x, q = \rho)$	**9985.5563**	**336.4881**	**6452.7456**	**164.7545**	**5142.6854**	**689.3511**
$\bar{y}_{P1(5)}$ $(p = C_x, q = \beta_2(x))$	**9532.2236**	**352.4908**	**6152.4789**	**172.7952**	**5005.2415**	**708.2807**
$\bar{y}_{P1(6)}$ $(p = \beta_2(x), q = \rho)$	**10523.2856**	**319.2939**	**6785.0756**	**156.6849**	**5287.4521**	**670.4711**

Table 8. PREs of the existing and proposed imputation techniques under Strategy II

Estimators	Population 5		Population 6		Population 7	
	MSE	PRE	MSE	PRE	MSE	PRE
$\bar{y}_{(1)}$	33600.2178	100	10631.1942	100	35451.1603	100
$\bar{y}_{(22)}$	25482.9668	131.8536	9775.4026	108.7545	24031.5464	147.5192
$\bar{y}_{(32)}$	24941.4714	134.7162	9090.1414	116.9530	23888.7487	148.4010
$\bar{y}_{(42)}$	24941.4714	134.7162	9090.1414	116.9530	23888.7487	148.4010
$\bar{y}_{(52)}$	25072.3812	134.0128	9371.7156	113.4391	25163.8498	140.8813
$\bar{y}_{(62)}$	24941.4714	134.7162	90901414	116.9530	23888.7487	148.4010
$\bar{y}_{(72)}$	26158.8649	128.4467	9132.9048	116.4053	26172.5781	135.4515
$\bar{y}_{(82)}$	26163.9398	128.4218	9133.4410	116.3985	26176.2460	135.4325
$\bar{y}_{P2(1)}$ $(p = 1, q = \beta_2(x))$	**24843.9297**	**135.2451**	**8983.1508**	**118.3459**	**23877.9486**	**148.4681**
$\bar{y}_{P2(2)}$ $(p = 1, q = \rho)$	**23659.6654**	**142.0147**	**8856.4456**	**120.0390**	**22856.4823**	**155.1033**
$\bar{y}_{P2(3)}$ $(p = 1, q = C_x)$	**23489.3145**	**143.0446**	**9074.4789**	**117.5148**	**22456.3302**	**157.8671**
$\bar{y}_{P2(4)}$ $(p = C_x, q = \rho)$	**22458.1256**	**149.6127**	**8693.1402**	**122.2940**	**22854.2245**	**155.1186**
$\bar{y}_{P2(5)}$ $(p = C_x, q = \beta_2(x))$	**20456.0021**	**164.2560**	**8564.6301**	**124.1290**	**20652.4412**	**171.6560**
$\bar{y}_{P2(6)}$ $(p = \beta_2(x), q = \rho)$	**23785.9456**	**141.2608**	**8765.6513**	**121.2824**	**23674.3326**	**149.7451**

Table 9. PREs of the existing and proposed imputation techniques under Strategy III

Estimators	Population 5		Population 6		Population 7	
	MSE	PRE	MSE	PRE	MSE	PRE
$\bar{y}_{(1)}$	33600.2178	100	10631.1942	100	35451.1603	100
$\bar{y}_{(23)}$	21558.1421	155.8585	9361.6133	113.5651	18509.9748	191.5246
$\bar{y}_{(33)}$	20754.8248	161.8911	8345.0170	127.3957	18298.1321	193.7419
$\bar{y}_{(43)}$	20754.8248	161.8911	8345.0170	127.3957	18298.1321	193.7419
$\bar{y}_{(53)}$	23828.1525	141.0105	8496.8028	125.1199	19475.4799	182.0297
$\bar{y}_{(73)}$	22560.8481	148.9315	8408.4573	126.4345	21686.2306	163.4731
$\bar{y}_{P3(1)}$ $(p = 1, q = \beta_2(x))$	**20703.5650**	**162.2919**	**8334.1242**	**127.5622**	**18207.1791**	**194.7097**
$\bar{y}_{P3(2)}$ $(p = 1, q = \rho)$	**19456.3012**	**172.6958**	**8254.9920**	**128.7850**	**17526.6214**	**202.2073**
$\bar{y}_{P3(3)}$ $(p = 1, q = C_x)$	**20654.8410**	**162.6747**	**8174.0301**	**130.0606**	**17745.9712**	**199.7701**
$\bar{y}_{P3(4)}$ $(p = C_x, q = \rho)$	**20001.3201**	**167.9900**	**8024.5412**	**132.4835**	**17563.1596**	**201.8495**
$\bar{y}_{P3(5)}$ $(p = C_x, q = \beta_2(x))$	**18742.9510**	**179.2685**	**7865.5203**	**135.1619**	**16632.7530**	**213.1406**
$\bar{y}_{P3(6)}$ $(p = \beta_2(x), q = \rho)$	**19803.1402**	**169.6711**	**8162.4123**	**130.2457**	**17453.8523**	**203.1136**

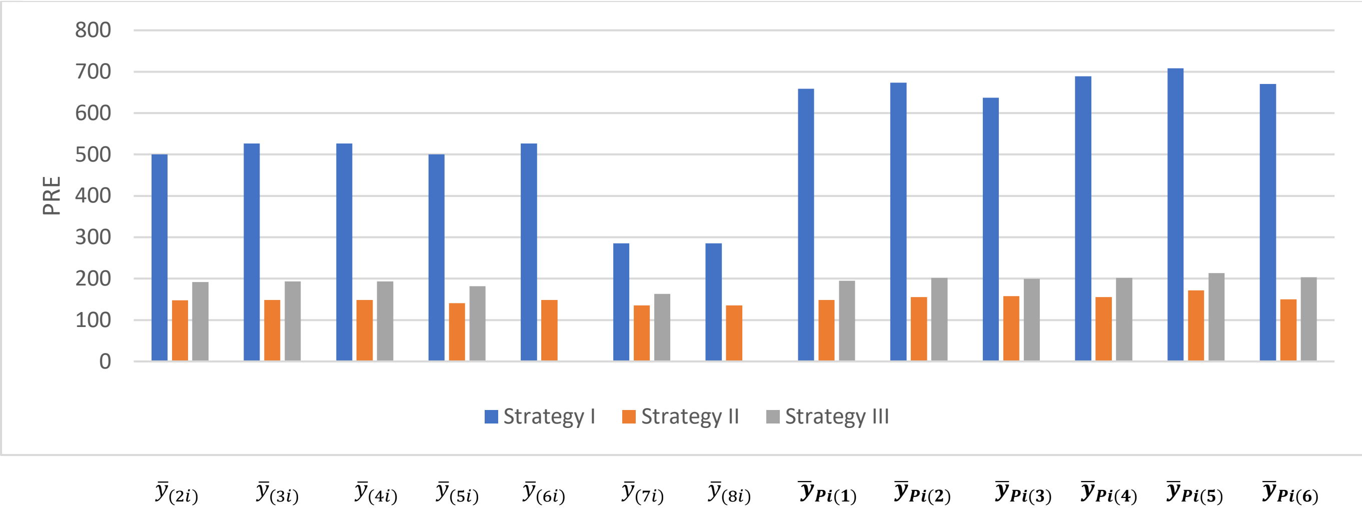

Figure 2. Population 7 (All strategies).

6. Results and Discussions

From the previous section, we observe the following facts.

(I) Tables 4 and 7 reveal that the estimators $\bar{y}_{(31)}, \bar{y}_{(41)}$and $\bar{y}_{(61)}$ exhibit identical efficiency levels, indicating their comparable performance. Furthermore, these estimators are found to be considerably more efficient than $\bar{y}_{(21)}, \bar{y}_{(51)}, \bar{y}_{(71)}$, and $\bar{y}_{(81)}$ under the same strategy. Among all the considered estimators, the various members of proposed class $\bar{y}_{P1(i)}$ demonstrates the highest PRE. This fact confirms the superior performance of strategy I over all the existing estimators.

(II) The findings from Tables 5 and 8, corresponding to Strategy II, show that the estimators $\bar{y}_{(32)}, \bar{y}_{(42)}$ and $\bar{y}_{(62)}$ are equally efficient and these estimators outperform $\bar{y}_{(22)}, \bar{y}_{(52)}, \bar{y}_{(72)}$ and $\bar{y}_{(82)}$. All the members of proposed class $\bar{y}_{P2(i)}$ attains the highest PRE among all the considered estimators when the Imputation strategy II is applied, highlighting the effectiveness of the proposed approach in this setting.

(III) Under Strategy III, as shown in Tables 6 and 9, the estimators $\bar{y}_{(33)}$ and $\bar{y}_{(43)}$ are observed to have identical efficiencies. These both surpass than $\bar{y}_{(23)}, \bar{y}_{(53)}$ and $\bar{y}_{(73)}$ in terms of PRE. It is noteworthy that when the population mean of the auxiliary variable is not known, the various members of the proposed class $\bar{y}_{P3(i)}$ achieves the maximum PRE among all the existing considered estimators under imputation strategy III. This fact confirms its practical utility in these situations.

(IV) A comparative analysis across all three proposed classes reveals that the proposed class $\bar{y}_{P1}$ consistently yields the highest PRE, indicating its robustness and superior performance across the different strategies.

(V) From all the tables 4 - 9, it is consistently observed that the PREs of the proposed estimators attain their maximum values when the parameters p and q are taken as $p = C_x, q = \beta_2(x)$. This demonstrates that the chosen combination of parameters (p, q) plays a crucial role in enhancing estimator's performance and achieving optimal efficiency.

7. Conclusion

The present study introduces and evaluates several new classes of estimators of $\bar{Y}$ designed to address the problem of missing data through improved imputation strategies. The numerical results, supported by the Percent Relative Efficiencies (PREs) presented in Tables 4 - 9, clearly demonstrate that the proposed estimators outperform the existing ones under all the three strategies. Among them, the proposed class $\bar{y}_{P1}$ emerges as the most efficient across varying conditions, as compared to proposed classes $\bar{y}_{P2}$ and $\bar{y}_{P3}$. A key finding of this study is that the efficiency of the proposed estimators reaches its peak when the value of parameter (p, q) is $(C_x, \beta_2(x))$. It indicates that the suitable choices of known auxiliary information as values of (p, q) can significantly enhance estimation accuracy. From a practical perspective, these estimators can be effectively employed in survey sampling and other real-life data collection processes where non-response or missing values are inevitable. By leveraging available auxiliary information, the proposed methods offer a reliable, computationally simple, and more efficient alternative to traditional imputation and estimation techniques. Hence, the proposed estimators hold strong potential for application in national surveys, socioeconomic studies, and agricultural or industrial research where data incompleteness often affects the reliability of population estimates.

Acknowledgments

The Authors would like to thank all the learned reviewers and the editor/co-editor for their valuable suggestions that leads to improving the quality of the research paper in the present version

Conflicts of Interest

The authors declared no conflict of interest.

Author Contributions

Conceptualization: GROVER, L.K.; Data curation: GROVER, L.K., SHARMA, A.; Formal analysis: SHARMA, A.; Investigation: GROVER, L.K., SHARMA, A.; Methodology: GROVER, L.K., SHARMA, A.; Software: GROVER, L.K., SHARMA, A.; Supervision: GROVER, L.K., SHARMA, A.; Validation: GROVER, L.K., SHARMA, A.; Visualization: GROVER, L.K., SHARMA, A.; Writing - original draft: SHARMA, A. Writing - review and editing: GROVER, L.K.,

References

1. Ahmad, S., Hussain, S., Aamir, M., Khan, F., Alshahrani, M. N., & Alqawba, M. Estimation of finite population mean using dual auxiliary variable for non-response using simple random sampling. *Aims Mathematics* **7**(3), 4592-4613 (2022). https://doi.org/10.1016/j.aej.2024.04.058
2. Ahmed, M.S., Al-Titi, O., Al-Rawi, Z., Abu-Dayyeh, W. Estimation of population mean using different imputation methods. *Statistics in Transition* **7**(6), 1247–1264 (2006).
3. Bhushan, S., Gupta, R., Pandey, S. K. Some log-type classes of estimators using auxiliary information. *International Journal of Agricultural and Statistical Sciences* **11**(2), 487 - 491 (2015).
4. Bhushan, S., Pandey, A.P., Pandey, A. On optimality of imputation methods for estimation of population mean using higher order moment of an auxiliary variable. *Communications in Statistics – Simulation and Computation* **49** (6), 1560 – 1574 (2020). https://doi.org/10.1080/03610918.2018.1500595

5. Bhushan, S., & Kumar, A. Imputation of missing data using multi auxiliary information under ranked set sampling. *Communications in Statistics - Simulation and Computation* **54**(5), 1500-1521 (2025). https://doi.org/10.1080/03610918.2023.2288796
6. Cochran. W. G. Sampling Techniques, 3rd edition, Wiley and Sons. (1977).
7. Diana, G., Perri, P.F. Improved estimators of the population mean for Missing Data. *Communications in Statistics-Theory and Methods* **39**, 3245-3251 (2010).
8. Hansen, M. H., Hurwitz, W.N. The Problem of non-response in sample surveys. *Journal of the American Statistical Association* **41**, 517-529 (1946).
9. Heitjan, D.F., Basu S. Distinguishing "Missing at Random" and "Missing Completely at Random." *The American Statistician* **50**, 207-213 (1996). https://doi.org/10.2307/2684656
10. Hussain, S., Zichuan, M., Hussain, S., Iftikhar, A., Asif, M., Akhtar, S., & Ahmad, S. On estimation of distribution function using dual auxiliary information under non-response using simple random sampling. *Journal of Probability and Statistics* (1), 1693612 (2020).
11. Kadilar, C., Cingi, H. Ratio estimators for population variance in simple and stratifed sampling. *Applied Mathematics and Computation* **173**, 1047-1058 (2006). https://doi.org/10.1016/j.amc.2005.04.032
12. Kadilar, C., Cingi, H. Improvement in variance estimation using auxiliary information. *Hacettepe Journal of Mathematics and Statistics* **35** (1), 111-115 (2006).
13. Kadilar, C., Cingi, H. Estimators for the population mean in the case of missing data. *Communications in Statistics - Theory and Methods* **37** (14), 2226–2236 (2008). https://doi.org/10.1080/03610920701855020
14. Kalton, G., Kasprzyk, D., Santos, R. Issues of nonresponse and imputation in the Survey of Income and Program Participation. *Current Topics in Survey Sampling*. (D. Krewski, R. Platek and J.N.K. Rao, eds.) 455-480. Academic Press, New York (1981).
15. Kumar, A., Singh, A.K., Singh, P., Singh, V.K. A class of exponential chain type estimator for population mean with imputation of missing data under double sampling scheme. *Journal of Statistics Applications and Probability* **6**(3), 479-485 (2017). http://dx.doi.org/10.18576/jsap/060304
16. Murthy, M. N. Sampling theory and methods. *Statistical Publishing Society*, Calcutta, India, (1967).
17. Pandey R, Singh, N.S., Yadav, K. Estimation of population mean using exponential ratio type imputation method under survey non-response. *Journal of the Indian Statistical Association* **53**(1),89–107 (2015). http://dx.doi.org/10.22237/jmasm/1462076280
18. Pandey, A. K., Singh, G. N., Bhattacharyya, D., & Singh, P. K. Efficient and alternative approaches for imputing missing data to estimate population mean. *Quality & Quantity* **58**(6), 5883-5897 (2024). http://dx.doi.org/10.1007/s11135-024-01914-w
19. Prasad, S. A study on new methods of ratio exponential type imputation in sample surveys. *Hacettepe Journal of Mathematics and Statistics* **47**(5), 1281-1301 (2018). https://doi.org/10.15672/HJMS.2016.392
20. Prasad, S., & Yadav, V. K. Imputation of missing data through product type exponential methods in sampling theory. *Revista Colombiana de Estadística* **46**(1), 111-127 (2023).
21. Rao, J.N.K. & Sitter, R.R. Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika* **82**(2), 453-460 (1995). https://doi.org/10.1093/biomet/82.2.453
22. Rubin, D.B. Inference and missing data. *Biometrika* **63**, 581-593 (1976).

23. Sande, I.G. A personal view of hot deck imputation procedure. *Survey Methodology* **5**, 238-258 (1979).
24. Searls, D.T. The utilization of a known coefficient of variation in the estimation procedure. *Journal of the American Statistical Association* **59**(308), 1225-1226 (1964).
25. Singh, S.& Horn, S. Compromised imputation in survey sampling. *Metrika* **51**, 266-276 (2000). https://doi.org/10.1007/s001840000054
26. Singh, S.& Deo, B. Imputation by Power Transformation. *Statistical Papers* **44**(4), 555-579 (2003).
27. Singh, G.N., Priyanka, K., Kim, J.M., Singh, S. Estimation of population mean using imputation techniques in sample surveys. *Journal of the Korean Statistical Society* **39**(1), 67–74 (2010).
28. Sohail, M.U., Shabbir, J. Imputation of missing values by using raw moments. *Statistics in Transition* **20**(1), 21-40 (2019).
29. Srivastava, S. K., Jhajj, H. S. A class of estimators of the population mean using multi-auxiliary information. *Calcutta Statistical. Association. Bulletin* **32**, 47-56 (1983). https://doi.org/10.1177/0008068319830104
30. Thakur, N.S., Yadav, K., Pathak, S. Mean estimation with imputation in two-phase sampling. *International Journal of Modern Engineering Research (IJMER)* **2**(5), 3561-3571 (2012).
31. Yadav, V. K., & Prasad, S. Generalized class of factor type exponential imputation techniques for population mean using simulation approach. *Journal of Statistical Computation and Simulation* **94**(9), 1997-2039 (2024). https://doi.org/10.1080/00949655.2024.2310699
32. Yang, S. & Kim, J.K. Fractional imputation in survey sampling: A Comparative Review. *Statistical Science* **31**(3), 415-432 (2016). http://dx.doi.org/10.1214/16-STS569
33. Zahid, E., & Shabbir, J. Estimation of population mean in the presence of measurement error and non-response under stratified random sampling. *PloS one* **13**(2), e0191572 (2018). https://doi.org/10.1371/journal.pone.0191572
34. Zahid, E., Shabbir, J., Gupta, S., Onyango, R., & Saeed, S. A generalized class of estimators for sensitive variable in the presence of measurement error and non-response. *Plos one*, **17**(1), e0261561 (2022). https://doi.org/10.1371/journal.pone.0261561

Appendix A

Proofs of Theorems 3.2-3.4.

Taking

$$\bar{y}_{d1} = \left[\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)^g + b_1(\bar{X} - \bar{x}_r)\right]\left[1 + log\left(\frac{\bar{x}_r^*}{\bar{X}^*}\right)^\beta\right].$$

By using the notations as given in Section 2, $\bar{y}_{d1}$ can be rewritten in terms of ϵ and δ, as follows:

$$\bar{y}_{d1} = [(1+\epsilon)(1+\delta)^{-g}\bar{Y} + b_1(\bar{X} - (1+\delta)\bar{X})][1 + \beta\log(1 + k\delta)]$$

$$\Rightarrow \bar{y}_{d1} = \bar{Y}\left[1 + \beta k\delta - \frac{\beta k^2\delta^2}{2} - g\beta k\delta^2 + \frac{g(g+1)}{2}\delta^2 - g\delta + \varepsilon + \beta k\varepsilon\delta - g\varepsilon\delta\right] - b_1\bar{X}(\delta + \beta k\delta^2) \quad (13)$$

Subtracting $\bar{Y}$ from both sides of equation (13), we get

$$\bar{y}_{d1} - \bar{Y} = \bar{Y}\left[\beta k\delta - \frac{\beta k^2\delta^2}{2} - g\beta k\delta^2 + \frac{g(g+1)}{2}\delta^2 - g\delta + \varepsilon + \beta k\varepsilon\delta - g\varepsilon\delta\right] - b_1\bar{X}(\delta + \beta k\delta^2) \quad (14)$$

By taking expectation on both sides of (14), the bias of $\bar{y}_{d1}$, up to first order of approximation, is given by:

$$Bias(\bar{y}_{d1}) = \bar{Y}\left(-\frac{1}{2}\beta k^2\lambda_1 C_x^2 - g\beta k\lambda_1 C_x^2 + \frac{1}{2}g(g+1)\lambda_1 C_x^2 + (\beta k - g)\lambda_1\rho C_y C_x\right) - b_1\bar{X}\beta k\lambda_1 C_x^2 .$$

On squaring both sides of equation (14), and retaining terms up to second degree of ϵ's and δ's we get:

$$(\bar{y}_{d1} - \bar{Y})^2 = \bar{Y}^2(\beta^2k^2\delta^2 + g^2\delta^2 + \varepsilon^2 + 2\beta k\varepsilon\delta - 2\beta kg\delta^2 - 2g\varepsilon\delta) + {b_1}^2\bar{X}^2\delta^2 - 2b_1\bar{X}\bar{Y}(\beta k\delta^2 + \varepsilon\delta - g\delta^2) \quad (15)$$

By taking expectations on both sides of equation (15) and using the expected values of $\epsilon's$ and $\delta's$ as given in Section 2, we get *MSE* of $\bar{y}_{d1}$ up to first order of approximation, as follows:

$$MSE(\bar{y}_{d1}) = \bar{Y}^2\lambda_1 C_y^2 + (\beta^2k^2 + g^2 - 2g\beta k)\bar{Y}^2\lambda_1 C_x^2 + 2(\beta k - g)\bar{Y}^2\lambda_1\rho C_y C_x + 2b_1(g - \beta k)\bar{X}\bar{Y}\lambda_1 C_x^2 + b_1^2\bar{X}^2\lambda_1 C_x^2 - 2b_1\bar{X}\bar{Y}\lambda_1\rho C_y C_x. \quad (16)$$

The MSE($\bar{y}_{d1}$) at equation (16) is minimized for

$$b_{1opt} = R^*\left[\frac{(\beta k - g)C_x^2 + \rho C_y C_x}{C_x^2}\right].$$

The resulting Bias($\bar{y}_{d1}$) and Minimum MSE($\bar{y}_{d1}$) are given as

$$Bias(\bar{y}_{d1}) = \left(-\frac{\beta k^2}{2} + \frac{g(g+1)}{2} - \beta^2k^2\right)\lambda_1 C_x^2\bar{Y} - g\rho C_y C_x\lambda_1\bar{Y},$$

$$MSE_{min}(\bar{y}_{d1}) = \lambda_1\bar{Y}^2(1-\rho^2)\, C_y^2 .$$

Similarly, taking

$$\bar{y}_{P1} = \alpha_1\left[\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)^g + b_1(\bar{X} - \bar{x}_r)\right]\left[1 + log\left(\frac{\bar{x}_r^*}{\bar{X}^*}\right)^\beta\right].$$

By using the notations as given in Section 2, $\bar{y}_{P1}$ can be rewritten in terms of $\epsilon's$ and $\delta's$, as follows:

$$\bar{y}_{P1} = \alpha_1[(1+\epsilon)(1+\delta)^{-g}\bar{Y} + b_1(\bar{X} - (1+\delta)\bar{X})][1 + \beta\log(1+k\delta)] \quad (17)$$

Subtracting $\bar{Y}$ from both sides of equation (17), we get

$$\bar{y}_{P1} - \bar{Y} = \alpha_1[(1+\epsilon)(1+\delta)^{-g}\bar{Y} + b_1(\bar{X} - (1+\delta)\bar{X})][1 + \beta\log(1+k\delta)] - \bar{Y} \quad (18)$$

By taking expectation on both sides of equation (18), the bias of $\bar{y}_{P1}$, up to first order of approximation, is given by:

$$Bias(\bar{y}_{P1}) = \bar{Y}\left(\alpha_1(1 + (-\tfrac{1}{2}\beta k^2 - g\beta k + \tfrac{1}{2}g(g+1))\lambda_1 C_x^2 + (\beta k - g)\lambda_1\rho C_y C_x\right) - \alpha_1 b_1\bar{X}\beta k\lambda_1 C_x^2.$$

On squaring both sides of equation (18), and retaining terms up to second degree of $\epsilon's$ and δ's. Then, taking expectations on both sides and using expected values of $\epsilon's$ and $\delta's$ as given in Section 2, we get *MSE* of $\bar{y}_{P1}$, up to first order of approximation, as follows:

$$MSE(\bar{y}_{P1}) = \alpha_1^2[\bar{Y}^2 + \bar{Y}^2\lambda_1 C_y^2 + (\beta^2k^2 + g^2 - \beta k^2 - 4g\beta k + g(g+1))\bar{Y}^2\lambda_1 C_x^2 + 4(\beta k - g)\bar{Y}^2\lambda_1\rho C_y C_x + b_1^2\bar{X}^2\lambda_1 C_x^2 + 2b_1(g - 2\beta k)\bar{X}\bar{Y}\lambda_1 C_x^2 - 2b_1\bar{X}\bar{Y}\lambda_1\rho C_y C_x] + \alpha_1[-2\bar{Y}^2 + \left(\beta k^2 + 2g\beta k - g(g+1)\right)\bar{Y}^2\lambda_1 C_x^2 - 2(\beta k - g)\bar{Y}^2\lambda_1\rho C_y C_x + 2b_1\beta k\bar{X}\bar{Y}\lambda_1 C_x^2] + \bar{Y}^2. \quad (19)$$

The $MSE(\bar{y}_{P1})$ at equation (19) after putting $b_1 = b_{1opt}$, is minimized for

$$\alpha_{1opt} = \frac{2\bar{Y}^2 + (-2\beta^2k^2 - \beta k^2 + g(g+1))\lambda_1\bar{Y}^2 C_x^2 - 2g\lambda_1\bar{Y}^2\rho C_y C_x}{2(\bar{Y}^2 + (1-\rho^2)\lambda_1\bar{Y}^2 C_y^2 + (-2\beta^2k^2 - \beta k^2 + g(g+1))\lambda_1\bar{Y}^2 C_x^2 - 2g\lambda_1\bar{Y}^2\rho C_y C_x}.$$

The resulting minimum MSE($\bar{y}_{P1}$) is given by $MSE_{min}(\bar{y}_{P1}) = \frac{\bar{Y}^2 MSE_{\min}(\bar{y}_{d1}) - \bar{Y}^2(Bias(\bar{y}_{d1}))^2}{\bar{Y}^2 + MSE_{\min}(\bar{y}_{d1}) + 2\bar{Y}Bias(\bar{y}_{d1})}$.

On applying similar steps for estimators $\bar{y}_{di}$ and $\bar{y}_{Pi}$ (i = 2, 3), we can obtain their biases and the minimum MSEs.