



Some General Classes of Efficient Estimators in Case of Missing Data

G. N. Singh¹ · M. Usman¹ · B. Khatoun²

Accepted: 25 February 2022 / Published online: 30 March 2022
© Grace Scientific Publishing 2022

Abstract

To handle the missing data problem in sample surveys, the imputation technique for missing values may suits well in reducing the negative impact of non-response in estimating the population mean. The socio-economic data yield fruitful results while imputation method employed to missing observations. Keeping this in mind, we have proposed three new general classes of difference-cum-ratio type imputation methods and the corresponding estimators in three different sampling strategies using the dual (rank) of an auxiliary variable in the presence of non-response. The biases and mean square errors of the proposed estimators are obtained up to the first-order approximation. The theoretical comparisons of the proposed estimators with usual mean imputation and the works Lee et al. [12], Kadilar and Cingi [11], Gira [9], Diana and Perri [8], Bhusan and Pandey [4], and Bhusan and Pandey [5] have been made which are also the special cases of the proposed estimators apart from being less efficient. The results are computed under an empirical study where the proposed work shows the efficacious performance over the above discussed works.

Keywords Population mean · Study variable · Dual of auxiliary variable · Missing data · Imputation

Mathematics Subject Classification 62D05

✉ M. Usman
mahmoodu33@gmail.com

G. N. Singh
gnsingh@iitism.ac.in

B. Khatoun
bushra.stats@gmail.com

¹ Department of Mathematics & Computing, Indian Institute of Technology (Indian School of Mines), Dhanbad 826007, India

² Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh, India

1 Introduction

Missing data are an inherent phenomena in sample surveys which may need an appropriate methodology to handle the data sets. For example, in an experiment of three kinds of new drinks in the market for 30 days, some of the data may accidentally be missing for some days; in an experiment of animals, some of the animals may die in a laboratory study, or the technician may accidentally omit some of the results of the experiments; in a medical investigation patients may not turn up or may not co-operate or die before completing the actual periods. Similarly, in agricultural experiments, the plants may be eaten away by animals or washed away by floods, *etc.* In all such situations, the experiments result in incomplete data may mislead the inferences. In case of missing data, several statisticians have proved that the inferences or predictions regarding the population parameters may be highly distorted, especially when the respondents and non-respondents are differ. Therefore, the knowledge of the appropriate pattern of the incomplete mechanism is needed at the estimation stage to overcome the missing data problems. Rubin [17] introduced three fundamental concepts of the missing patterns of the data: missing at random (MAR), observed at random (OAR) and parameter distribution (PD). The combination of MAR and OAR termed as the notion of missing completely at random (MCAR). Heitzan and Basu (1996) have differentiated the meaning of MAR and MCAR very systematically. Following these authors, we have assumed the MCAR mechanism in the present study to deal with the problem of missing data.

Imputation is one of the effective techniques in surveys to compensate for the missing data. In various fields like the energy storage system, which provides a peak reduction service to local electricity network, the food composition databases, the clinical trials, the industrial databases, *etc.*, the imputation technique for missing observations play a contributory role regarding the estimation of population parameters under non-response. A number of imputation methods using the MCAR mechanism have been discussed by the authors including Lee et al. [12], Singh and Horn [23], Singh and Deo [22], Kadilar and Cingi [11], Singh [21], Diana and Perri [8], Al-Omari et al. [1], Gira [9] and recently Bhusan and Pandey [4], Prasad [15], Bhusan and Pandey [5], Singh and Suman [24], Singh et al. [25], Bhusan et al. [6] and Singh and Usman [26]. These authors have made the utilization of information available on each unit of an auxiliary variable which is often used in surveys to increase the precision of estimate of the population mean. Some other related references are Prasad [16], Bouza et al. [3] and Bouza-Herrera and Viada [2].

The aim of the present study is to develop the imputation methods and subsequent estimators with enhanced precision by incorporating the double use of an auxiliary information to estimate the population mean over some relevant estimators which are based on the prime/single use of auxiliary information in case of missing data under the MCAR mechanism. For this, we have considered the rank(dual) of an auxiliary variable which may behaves like an additional auxiliary variable. To our knowledge, no one has tried this type of work for imputation to handle the missing data in estimating the population mean so far. The rest part of the paper is organized as follows: In Sect. 2, the methodology and notations have been discussed and some conventional imputation methods have been reviewed in Sect. 3. In Sect. 4, we have suggested three

general class of estimators using imputation techniques and studied their properties. Section 5 talks about the theoretical comparisons of the estimators and the empirical comparisons based on real data sets are presented in Sect. 6. Finally, some concluding remarks are made in Sect. 7.

2 Methodology and Notations

Consider an identifiable population $U = \{U_1, U_2, U_3, \dots, U_N\}$ of size N where our goal is to estimate the population mean \bar{Y} of study variable y which possess a proper correlation with an auxiliary variable x . Let (y_i, x_i) be the i^{th} observations of y and x . Suppose that the information is readily available at each unit of auxiliary variable x in the population. Let $R_x = \{r_{x,1}, r_{x,2}, \dots, r_{x,N}\}$ denote the values of corresponding ranks of $X = \{x_1, x_2, \dots, x_N\}$ in U . Remember that the rank R_x can also hold an adequate amount of correlation with the study variable y . Let a sample of size n be drawn using simple random sampling without replacement (SRSWOR) technique from the population and surveyed. Unfortunately, response is observed only on $r (< n)$ units for study variable y . For the remaining $(n - r)$ non-responding units, we propose some new imputation methods using the rank of an auxiliary variable, given in section 4. Let A and \bar{A} be the sets of responding units and non-responding units, respectively, in the sample. For the sampled units $i \in A$, the values y_i are observed while for the units $i \in \bar{A}$ some imputation techniques are used.

We define some useful notations as follows:

- $\bar{Y} = \sum_{i=1}^N y_i/N, \bar{y}_r = \sum_{i=1}^r y_i/r$: The population mean and the response mean of study variable y ,
- $\bar{X} = \sum_{i=1}^N x_i/N, \bar{x}_r = \sum_{i=1}^r x_i/r, \bar{x}_n = \sum_{i=1}^n x_i/n$: The population mean, sample mean and the response mean of auxiliary variable x ,
- $\bar{R}_x = \sum_{i=1}^N r_{x,i}/N, \bar{r}_{x(n)} = \sum_{i=1}^n r_{x,i}/n, \bar{r}_{x(r)} = \sum_{i=1}^r r_{x,i}/r$: The population mean, sample mean and the response mean of R_x ,
- $S_y^2 = \sum_{i=1}^N (y_i - \bar{Y})^2/(N - 1)$: The population variance of y ,
- $S_x^2 = \sum_{i=1}^N (x_i - \bar{X})^2/(N - 1)$: The population variance of x ,
- $S_{R_x}^2 = \sum_{i=1}^N (r_{x,i} - \bar{R}_x)^2/(N - 1)$: The population variance of R_x ,
- $C_y = S_y/\bar{Y}$: The coefficient of variation of y ,
- $C_x = S_x/\bar{X}$: The coefficient of variation of x ,
- $C_{R_x} = S_{R_x}/\bar{R}_x$: The coefficient of variation of R_x ,
- $\rho_{yx} = S_{yx}/S_y S_x$: The correlation coefficient between y and x .
- $\rho_{yR_x} = S_{yR_x}/S_y S_{R_x}$: The correlation coefficient between y and R_x .
- $\rho_{xR_x} = S_{xR_x}/S_x S_{R_x}$: The correlation coefficient between x and R_x .

To obtain the biases and mean square errors (MSEs) of the proposed estimators, we define the following error transformation, as:

$$\frac{\bar{y}_r}{\bar{Y}} = (1 + \xi_0), \quad \frac{\bar{x}_r}{\bar{X}} = (1 + \xi'_1), \quad \frac{\bar{x}_n}{\bar{X}} = (1 + \xi_1), \quad \frac{\bar{r}_{x(r)}}{\bar{R}_x} = (1 + \xi'_2), \quad \frac{\bar{r}_{x(n)}}{\bar{R}_x} = (1 + \xi_2),$$

such that

$$E(\xi_0) = E(\xi'_1) = E(\xi_1) = E(\xi'_2) = E(\xi_2) = 0$$

and

$$\begin{aligned} E(\xi_0^2) &= f_2 C_y^2, & E(\xi_1'^2) &= f_2 C_x^2, & E(\xi_1^2) &= f_1 C_x^2, & E(\xi_2'^2) &= f_2 C_{r_x}^2, \\ E(\xi_2^2) &= f_1 C_{r_x}^2, \\ E(\xi_0 \xi_1') &= f_2 \rho_{yx} C_y C_x, & E(\xi_0 \xi_1) &= f_1 \rho_{yx} C_y C_x, & E(\xi_0 \xi_2') &= f_2 \rho_{y r_x} C_y C_{r_x}, \\ E(\xi_0 \xi_2) &= f_1 \rho_{y r_x} C_y C_{r_x}, & E(\xi_2' \xi_2) &= f_1 C_{r_x}^2, & E(\xi_1' \xi_1) &= f_1 C_x^2, \\ E(\xi_1' \xi_2') &= f_2 \rho_{x r_x} C_x C_x, & E(\xi_1' \xi_2) &= f_1 \rho_{x r_x} C_x C_{r_x} \end{aligned}$$

where $f_1 = (\frac{1}{n} - \frac{1}{N})$ and $f_2 = (\frac{1}{r} - \frac{1}{N})$. We also write $f_3 = (\frac{1}{r} - \frac{1}{n}) = f_2 - f_1$.

3 Some Conventional Imputation Methods

In this section, we discuss some customary imputation methods and corresponding estimators under three different sampling strategies which are discussed as follows:

Strategy I: When \bar{X} and \bar{x}_n are used.

Strategy II: When \bar{X} and \bar{x}_r are used.

Strategy III: When \bar{x}_n and \bar{x}_r are used.

3.1 Mean Imputation Method

The usual mean method of imputation which is free from auxiliary information, is given by

$$y_{i,m} = \begin{cases} y_i, & \text{if } i \in A \\ \bar{y}_r, & \text{if } i \in \bar{A}. \end{cases} \tag{3.1}$$

The estimator of population mean \bar{Y} under mean method of imputation is, say $t_m = \bar{y}_r$ whose variance is given by

$$V(t_m) = V(\bar{y}_r) = \bar{Y}^2 f_2 C_y^2 \tag{3.2}$$

3.2 Lee et al. [12] Imputation Methods

When there is an auxiliary information, then the ratio method of imputation for the data due to Lee et al. [12] can be considered in three strategies as:

$$y_{i,R_1} = \begin{cases} y_i, & \text{if } i \in A \\ \frac{1}{n-r} \left[n \bar{y}_r \left(\frac{\bar{X}}{\bar{x}_n} \right) - r \bar{y}_r \right], & \text{if } i \in \bar{A}. \end{cases} \tag{3.3}$$

$$y_{i,R_2} = \begin{cases} y_i, & \text{if } i \in A \\ \frac{1}{n-r} \left[n\bar{y}_r \left(\frac{\bar{X}}{\bar{x}_r} \right) - r\bar{y}_r \right], & \text{if } i \in \bar{A}. \end{cases} \tag{3.4}$$

$$y_{i,R_3} = \begin{cases} y_i, & \text{if } i \in A \\ \frac{1}{n-r} \left[n\bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r} \right) - r\bar{y}_r \right], & \text{if } i \in \bar{A}. \end{cases} \tag{3.5}$$

The corresponding ratio type estimators are defined as:

$$t_{R_1} = \bar{y}_r \left(\frac{\bar{X}}{\bar{x}_n} \right) \tag{3.6}$$

$$t_{R_2} = \bar{y}_r \left(\frac{\bar{X}}{\bar{x}_r} \right) \tag{3.7}$$

$$t_{R_3} = \bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r} \right) \tag{3.8}$$

The MSEs of t_{R_i} ($i = 1, 2, 3$) to the first-order approximation, are given by

$$MSE(t_{R_i}) = V(\bar{y}_r) + \bar{Y}^2 f_i C_x (C_x - 2\rho C_y) \tag{3.9}$$

The ratio estimators t_{R_i} ($i = 1, 2, 3$) are better than mean estimator \bar{y}_r if $(\rho C_y/C_x) > 1/2$.

3.3 Kadilar and Cingi [11] Imputation Methods

The imputation methods under three strategies are given by

$$y_{i,KC_1} = \begin{cases} \left[\frac{ny_i}{r} + b(\bar{X} - x_i) \right] \frac{\bar{X}}{\bar{x}_n}, & \text{if } i \in A \\ \left[b(\bar{X} - x_i) \right] \frac{\bar{X}}{\bar{x}_n}, & \text{if } i \in \bar{A} \end{cases} \tag{3.10}$$

$$y_{i,KC_2} = \begin{cases} \left[\frac{ny_i}{r} - b\frac{nx_i}{r} \right] \frac{\bar{X}}{\bar{x}_r}, & \text{if } i \in A \\ b\frac{n\bar{X}}{n-r} \frac{\bar{X}}{\bar{x}_r}, & \text{if } i \in \bar{A} \end{cases} \tag{3.11}$$

$$y_{i,KC_3} = \begin{cases} \left[\frac{ny_i}{r} - b\frac{nx_i}{r} \right] \frac{\bar{x}_n}{\bar{x}_r}, & \text{if } i \in A \\ b\frac{n\bar{x}_n}{n-r} \frac{\bar{x}_n}{\bar{x}_r}, & \text{if } i \in \bar{A} \end{cases} \tag{3.12}$$

The respective estimators are given as:

$$t_{KC_1} = \frac{\bar{y}_r + b(\bar{X} - \bar{x}_n) \bar{X}}{\bar{x}_n} \tag{3.13}$$

$$t_{KC_2} = \frac{\bar{y}_r + b(\bar{X} - \bar{x}_r) \bar{X}}{\bar{x}_r} \tag{3.14}$$

$$t_{KC_3} = \frac{\bar{y}_r + b(\bar{x}_n - \bar{x}_r) \bar{x}_n}{\bar{x}_r} \tag{3.15}$$

where $b = s_{yx}/s_x^2$ is the least squares estimated regression coefficient of y on x . Here $s_{yx} = \sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)/(n - 1)$ and $s_x^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2/(n - 1)$.

The MSEs of the estimators t_{KC_i} ($i = 1, 2, 3$) to the first-order approximation, are respectively, given as:

$$MSE(t_{KC_i}) = V(\bar{y}_r) + f_i \bar{Y}^2 (C_x^2 - \rho_{yx}^2 C_y^2) \tag{3.16}$$

The estimator t_{KC_i} ($i = 1, 2, 3$) are better than mean estimator \bar{y}_r if $(\rho C_y/C_x) > 1$.

3.4 Gira [9] Imputation Methods

On the lines of Gira [9], we have considered three ratio type imputation methods to deal with missing data, given by

$$y_{i,G_1} = \begin{cases} y_i, & \text{if } i \in A \\ \bar{y}_r \left[n \left(\frac{v_1 - \bar{x}_n}{v_1 - \bar{X}} \right) - r \right] \frac{x_i}{\sum_{i \in \bar{A}} x_i}, & \text{if } i \in \bar{A} \end{cases} \tag{3.17}$$

$$y_{i,G_2} = \begin{cases} y_i, & \text{if } i \in A \\ \bar{y}_r \left[n \left(\frac{v_2 - \bar{x}_r}{v_2 - \bar{X}} \right) - r \right] \frac{x_i}{\sum_{i \in \bar{A}} x_i}, & \text{if } i \in \bar{A} \end{cases} \tag{3.18}$$

$$y_{i,G_3} = \begin{cases} y_i, & \text{if } i \in A \\ \bar{y}_r \left[n \left(\frac{v_3 - \bar{x}_r}{v_3 - \bar{x}_n} \right) - r \right] \frac{x_i}{\sum_{i \in \bar{A}} x_i}, & \text{if } i \in \bar{A} \end{cases} \tag{3.19}$$

where v_i ($i = 1, 2, 3$) are the suitably chosen constants.

The point estimators under (3.17), (3.18) and (3.19) are, respectively, given by

$$t_{G_1} = \bar{y}_r \left(\frac{v_1 - \bar{x}_n}{v_1 - \bar{X}} \right) \tag{3.20}$$

$$t_{G_2} = \bar{y}_r \left(\frac{v_2 - \bar{x}_r}{v_2 - \bar{X}} \right) \tag{3.21}$$

$$t_{G_3} = \bar{y}_r \left(\frac{v_3 - \bar{x}_r}{v_3 - \bar{x}_n} \right) \tag{3.22}$$

The minimum MSEs of the estimators t_{G_i} ($i = 1, 2, 3$) are given by

$$\min.MSE(t_{G_i}) = \bar{Y}^2 C_y^2 (f_2 - f_i \rho_{yx}^2) \tag{3.23}$$

The optimum values are given as: $v_{i(opt)} = \bar{X} \left[1 + \frac{1}{\rho_{yx} \frac{C_y}{C_x}} \right]$.

From (3.2), (3.9) (3.16) and (3.23), it is clear that t_{G_i} ($i = 1, 2, 3$) is always better than $\bar{y}_r, t_{R_i}, t_{KC_i}$ in the respective strategies. Note that Gira [9] showed both theoretically and empirically that his method is equally efficient to the methods propounded by Singh and Horn [23], Singh and Deo [22] and Singh [21].

3.5 Diana and Perri [8] Estimators

Diana and Perri [8] established three regression-type imputation methods, under which the resultant data take the form as:

$$y_{i,DP_1} = \begin{cases} \frac{ny_i}{r} + b(\bar{X} - x_i), & \text{if } i \in A \\ b(\bar{X} - x_i), & \text{if } i \in \bar{A} \end{cases} \tag{3.24}$$

$$y_{i,DP_2} = \begin{cases} \frac{ny_i}{r} - b\frac{nx_i}{r}, & \text{if } i \in A \\ b\frac{n\bar{X}}{n-r}, & \text{if } i \in \bar{A} \end{cases} \tag{3.25}$$

$$y_{i,DP_3} = \begin{cases} \frac{ny_i}{r} - b\frac{nx_i}{r}, & \text{if } i \in A \\ b\frac{n\bar{x}_n}{n-r}, & \text{if } i \in \bar{A} \end{cases} \tag{3.26}$$

The subsequent estimators are, respectively, given as:

$$t_{DP_1} = \bar{y}_r + b(\bar{X} - \bar{x}_n) \tag{3.27}$$

$$t_{DP_2} = \bar{y}_r + b(\bar{X} - \bar{x}_r) \tag{3.28}$$

$$t_{DP_3} = \bar{y}_r + b(\bar{x}_n - \bar{x}_r) \tag{3.29}$$

The MSEs of the estimators t_{DP_i} ($i = 1, 2, 3$) are given as:

$$MSE(t_{DP_i}) = \min.MSE(t_{G_i}) = \bar{Y}^2 C_y^2 \left[f_2 - f_i \rho_{yx}^2 \right] \tag{3.30}$$

3.6 Bhusan and Pandey [4] Imputation Methods

Bhusan and Pandey [4] proposed three different types of imputation methods which are paralleling the improvement over the Diana and Perri [8], are given as:

$$y_{i,BP_1} = \begin{cases} \mu_1 y_i, & \text{if } i \in A \\ \mu_1 \bar{y}_r + \frac{n\lambda_1}{n-r}(\bar{X} - \bar{x}_n), & \text{if } i \in \bar{A} \end{cases} \tag{3.31}$$

$$y_{i,BP_2} = \begin{cases} \mu_2 y_i, & \text{if } i \in A \\ \mu_2 \bar{y}_r + \frac{n\lambda_2}{n-r}(\bar{X} - \bar{x}_r), & \text{if } i \in \bar{A} \end{cases} \tag{3.32}$$

$$y_{i,BP_3} = \begin{cases} \mu_3 y_i, & \text{if } i \in A \\ \mu_3 \bar{y}_r + \lambda_3(x_i - \bar{x}_r), & \text{if } i \in \bar{A} \end{cases} \tag{3.33}$$

where (μ_i, λ_i) ($i = 1, 2, 3$) are the arbitrary chosen constants.

The corresponding estimators are defined as:

$$t_{BP_1} = \mu_1 \bar{y}_r + \lambda_1(\bar{X} - \bar{x}_n) \tag{3.34}$$

$$t_{BP_2} = \mu_2 \bar{y}_r + \lambda_2(\bar{X} - \bar{x}_r) \tag{3.35}$$

$$t_{BP_3} = \mu_3 \bar{y}_r + \lambda_3(\bar{x}_n - \bar{x}_r) \tag{3.36}$$

The minimum MSE of the estimators t_{BP_i} ($i = 1, 2, 3$) are, respectively, given as:

$$\min.MSE(t_{BP_i}) = \frac{\bar{Y}^2 MSE(t_{DP_i})}{\bar{Y}^2 + MSE(t_{DP_i})} \tag{3.37}$$

for the optimum values

$$\mu_{1(opt)} = \frac{1}{1 + [f_3 + f_1(1 - \rho_{yx}^2)]C_y^2}, \quad \text{and} \quad \beta_{1(opt)} = \left(\rho_{yx} \frac{S_y}{S_x}\right) \mu_{1(opt)} \tag{3.38}$$

$$\mu_{2(opt)} = \frac{1}{1 + [f_2(1 - \rho_{yx}^2)]C_y^2}, \quad \text{and} \quad \beta_{2(opt)} = \left(\rho_{yx} \frac{S_y}{S_x}\right) \mu_{2(opt)} \tag{3.39}$$

$$\mu_{3(opt)} = \frac{1}{1 + [f_1 + f_3(1 - \rho_{yx}^2)]C_y^2}, \quad \text{and} \quad \beta_{3(opt)} = \left(\rho_{yx} \frac{S_y}{S_x}\right) \mu_{3(opt)} \tag{3.40}$$

Bhusan and Pandey [5] have also given the improvement over the usual ratio type imputation methods due to Lee. et al. [12] under which the data becomes

$$y_{i, BP_1^*} = \begin{cases} \omega_1 y_i, & \text{if } i \in A \\ \frac{\omega_1}{n-r} \left[n\bar{y}_r \left(\frac{\bar{X}}{\bar{x}_n} \right)^{\eta_1} - r\bar{y}_r \right], & \text{if } i \in \bar{A} \end{cases} \tag{3.41}$$

$$y_{i, BP_2^*} = \begin{cases} \omega_2 y_i, & \text{if } i \in A \\ \frac{\omega_2}{n-r} \left[n\bar{y}_r \left(\frac{\bar{X}}{\bar{x}_r} \right)^{\eta_2} - r\bar{y}_r \right], & \text{if } i \in \bar{A} \end{cases} \tag{3.42}$$

$$y_{i, BP_3^*} = \begin{cases} \omega_3 y_i, & \text{if } i \in A \\ \frac{\omega_3}{n-r} \left[n\bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r} \right)^{\eta_3} - r\bar{y}_r \right], & \text{if } i \in \bar{A} \end{cases} \tag{3.43}$$

The respective estimators are given by

$$t_{BP_1}^* = \omega_1 \bar{y}_r \left(\frac{\bar{X}}{\bar{x}_n} \right)^{\eta_1} \tag{3.44}$$

$$t_{BP_2}^* = \omega_2 \bar{y}_r \left(\frac{\bar{X}}{\bar{x}_r} \right)^{\eta_2} \tag{3.45}$$

$$t_{BP_3}^* = \omega_3 \bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r} \right)^{\eta_3} \tag{3.46}$$

where (ω_i, η_i) ($i = 1, 2, 3$) are the arbitrary chosen constants.

The minimum MSEs of the estimators $t_{BP_i}^*$ ($i = 1, 2, 3$) are, respectively, given as:

$$\min.MSE(t_{BP_i}^*) = \bar{Y}^2 \left(1 - \frac{H_i^2}{G_i} \right) \tag{3.47}$$

where

$$G_i = \{1 + f_2 C_y^2 + 2\eta_{i(opt)}^2 f_i C_x^2 + \eta_{i(opt)} f_i C_x (C_x - 4\rho_{yx} C_y)\}$$

$$H_i = \{1 + \frac{\eta_{i(opt)}^2}{2} f_i C_x^2 + \frac{\eta_{i(opt)}}{2} f_i C_x (C_x - 2\rho_{yx} C_y)\}$$

The optimum values are given as: $\omega_{i(opt)} = \frac{H_i}{G_i}$ ($i = 1, 2, 3$) and $\eta_{i(opt)} = \rho_{yx} \frac{C_y}{C_x}$.

The above existing imputation methods and resultant estimators are based on single use of an auxiliary variable . We propose some imputation techniques based on dual use of an auxiliary variable given in next section.

4 Suggested Imputation Methods

In this section, we consider the double (rank) use of an auxiliary variable to impute the missing data in three strategies *ie*, Strategy I, Strategy II and Strategy III which are defined as follows:

Strategy I: When \bar{X} , \bar{x}_n and $\bar{r}_{x(n)}$ are used.

Strategy II: When \bar{X} , \bar{x}_r and $\bar{r}_{x(r)}$ are used.

Strategy III: When (\bar{x}_n, \bar{x}_r) and $(\bar{r}_{x(n)}, \bar{r}_{x(r)})$ are used.

We suggest three generalized class of difference-cum-ratio type imputation methods in three strategies given above under which the data, respectively, take forms as:

$$y_{i,P_1} = \begin{cases} y_i, & \text{if } i \in A \\ \left(1 - \frac{n}{n-r}\right) \bar{y}_r + \left[\frac{n}{n-r} \alpha_1 \bar{y}_r + \beta_1 (x_1^* - x_i) + \gamma_1 (r_1^* - r_i)\right] \left[\frac{u\bar{X}+v}{u\bar{x}_n+v}\right], & \text{if } i \in \bar{A} \end{cases} \tag{4.1}$$

$$y_{i,P_2} = \begin{cases} y_i, & \text{if } i \in A \\ \left(1 - \frac{n}{n-r}\right) \bar{y}_r + \left[\frac{n}{n-r} \alpha_2 \bar{y}_r + \beta_2 (x_2^* + \frac{n}{r} x_i) + \gamma_2 (r_2^* + \frac{n}{r} r_i)\right] \left[\frac{u\bar{X}+v}{u\bar{x}_r+v}\right], & \text{if } i \in \bar{A} \end{cases} \tag{4.2}$$

$$y_{i,P_3} = \begin{cases} y_i, & \text{if } i \in A \\ \left(1 - \frac{n}{n-r}\right) \bar{y}_r + \left[\frac{n}{n-r} \alpha_3 \bar{y}_r + \beta_3 (x_3^* + \frac{n}{r} x_i) + \gamma_3 (r_3^* + \frac{n}{r} r_i)\right] \left[\frac{u\bar{x}_n+v}{u\bar{x}_r+v}\right], & \text{if } i \in \bar{A} \end{cases} \tag{4.3}$$

where

$$x_1^* = \frac{n\bar{X} - r\bar{x}_r}{n - r}, \quad r_1^* = \frac{n\bar{R}_x - r\bar{r}_{x(r)}}{n - r},$$

$$x_2^* = \frac{n}{n - r} \left[\bar{X} - \frac{n}{r} \bar{x}_n\right], \quad r_2^* = \frac{n}{n - r} \left[\bar{R}_x - \frac{n}{r} \bar{r}_{x(n)}\right],$$

$$x_3^* = \frac{n}{n - r} \left(1 - \frac{n}{r}\right) \bar{x}_n, \quad r_3^* = \frac{n}{n - r} \left(1 - \frac{n}{r}\right) \bar{r}_{x(n)}.$$

The corresponding estimators of population mean \bar{Y} of study variable y in case of missing observations are defined as:

$$t_{P_1} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_n) + \gamma_1 (\bar{R}_x - \bar{r}_{x(n)})] \left[\frac{u \bar{X} + v}{u \bar{x}_n + v} \right] \tag{4.4}$$

$$t_{P_2} = [\alpha_2 \bar{y}_r + \beta_2 (\bar{X} - \bar{x}_r) + \gamma_2 (\bar{R}_x - \bar{r}_{x(r)})] \left[\frac{u \bar{X} + v}{u \bar{x}_r + v} \right] \tag{4.5}$$

$$t_{P_3} = [\alpha_3 \bar{y}_r + \beta_3 (\bar{x}_n - \bar{x}_r) + \gamma_3 (\bar{r}_{x(n)} - \bar{r}_{x(r)})] \left[\frac{u \bar{x}_n + v}{u \bar{x}_r + v} \right] \tag{4.6}$$

Here, $(\alpha_i, \beta_i, \gamma_i) (i = 1, 2, 3)$ are the arbitrary chosen constants, $u (\neq 0), v$ are the real numbers or the functions of known parameters of the auxiliary variable which may be readily known or guessed from past surveys such as coefficient of variation C_x , correlation coefficient ρ_{yx} , etc.

4.1 Some Special Cases of Proposed Estimators

(i) If we set $\alpha_i = 1; (i = 1, 2, 3)$ and $(u, v) = (0, 1)$ in (4.1), (4.2) and (4.3), we can get the conventional difference type imputation methods based on dual of an auxiliary variable, given as:

$$y_{i,d_1} = \begin{cases} y_i, & \text{if } i \in A \\ \bar{y}_r + [\phi_1(x_1^* - x_i) + \phi_1(r_1^* - r_i)], & \text{if } i \in \bar{A} \end{cases} \tag{4.7}$$

$$y_{i,d_2} = \begin{cases} y_i, & \text{if } i \in A \\ \bar{y}_r + [\phi_2(x_2^* + \frac{n}{r} x_i) + \phi_2(r_2^* + \frac{n}{r} r_i)], & \text{if } i \in \bar{A} \end{cases} \tag{4.8}$$

$$y_{i,d_3} = \begin{cases} y_i, & \text{if } i \in A \\ \bar{y}_r + [\phi_3(x_3^* + \frac{n}{r} x_i) + \phi_3(r_3^* + \frac{n}{r} r_i)], & \text{if } i \in \bar{A} \end{cases} \tag{4.9}$$

The respective estimators are given by

$$t_{d_1} = [\bar{y}_r + \phi_1 (\bar{X} - \bar{x}_n) + \phi_1 (\bar{R}_x - \bar{r}_{x(n)})] \tag{4.10}$$

$$t_{d_2} = [\bar{y}_r + \phi_2 (\bar{X} - \bar{x}_r) + \phi_2 (\bar{R}_x - \bar{r}_{x(r)})] \tag{4.11}$$

$$t_{d_3} = [\bar{y}_r + \phi_3 (\bar{x}_n - \bar{x}_r) + \phi_3 (\bar{r}_{x(n)} - \bar{r}_{x(r)})] \tag{4.12}$$

(ii) If we set $(u, v) = (0, 1)$ in (4.1), (4.2) and (4.3), we get the improved difference type imputation methods given by

$$y_{i,D_1} = \begin{cases} y_i, & \text{if } i \in A \\ \left(1 - \frac{n}{n-r}\right) \bar{y}_r + \left[\frac{n}{n-r} \alpha_1^* \bar{y}_r + \beta_1^* (x_1^* - x_i) + \gamma_1^* (r_1^* - r_i)\right], & \text{if } i \in \bar{A} \end{cases} \tag{4.13}$$

$$y_{i,D_2} = \begin{cases} y_i, & \text{if } i \in A \\ \left(1 - \frac{n}{n-r}\right) \bar{y}_r + \left[\frac{n}{n-r} \alpha_2^* \bar{y}_r + \beta_2^* (x_2^* + \frac{n}{r} x_i) + \gamma_2^* (r_2^* + \frac{n}{r} r_i)\right], & \text{if } i \in \bar{A} \end{cases} \tag{4.14}$$

$$y_{i,D_3} = \begin{cases} y_i, & \text{if } i \in A \\ \left(1 - \frac{n}{n-r}\right) \bar{y}_r + \left[\frac{n}{n-r} \alpha_3^* \bar{y}_r + \beta_3^* (x_3^* + \frac{n}{r} x_i) + \gamma_3^* (r_3^* + \frac{n}{r} r_i)\right], & \text{if } i \in \bar{A} \end{cases} \tag{4.15}$$

The respective estimators are given by

$$t_{D_1} = [\alpha_1^* \bar{y}_r + \beta_1^* (\bar{X} - \bar{x}_n) + \gamma_1^* (\bar{R}_x - \bar{r}_{x(n)})] \tag{4.16}$$

$$t_{D_2} = [\alpha_2^* \bar{y}_r + \beta_2^* (\bar{X} - \bar{x}_r) + \gamma_2^* (\bar{R}_x - \bar{r}_{x(r)})] \tag{4.17}$$

$$t_{D_3} = [\alpha_3^* \bar{y}_r + \beta_3^* (\bar{x}_n - \bar{x}_r) + \gamma_3^* (\bar{r}_{x(n)} - \bar{r}_{x(r)})] \tag{4.18}$$

which are paralleling improvement over the estimators $t_{d_i} (i = 1, 2, 3)$.

4.2 Properties of Proposed Estimators

Theorem 4.1 *The biases and MSEs of the estimators $t_{P_i} (i = 1, 2, 3)$ to the first-order approximations are given by*

$$B(t_{P_i}) = \bar{Y} [\alpha_i F_i + R_1 \beta_i G_i + R_2 \gamma_i H_i - 1] \tag{4.19}$$

and

$$MSE(t_{P_i}) = \bar{Y}^2 [\alpha_i^2 A_i' + \beta_i^2 B_i' + \gamma_i^2 C_i' + 2\alpha_i \beta_i D_i' + 2\alpha_i \gamma_i E_i' + 2\beta_i \gamma_i F_i' - 2\alpha_i G_i' - 2\beta_i H_i' - 2\gamma_i I_i' + 1] \tag{4.20}$$

The minimum MSEs of the estimators $t_{P_i} (i = 1, 2, 3)$ are given by

$$\min.MSE(t_{P_i}) = \bar{Y}^2 \left[1 - \frac{Q_i^2}{P_i} \right] \tag{4.21}$$

at the optimum values of $\alpha_i (i = 1, 2, 3)$, β_i and γ_i , are given by

$$\alpha_{i(opt)} = \frac{E_i - \beta_{i(opt)} L_i - \gamma_{i(opt)} M_i}{P_i}, \tag{4.22}$$

$$\beta_{i(opt)} = \frac{C_i^* E_i^* - B_i^* D_i^*}{C_i^* D_i^* - A_i^* E_i^*} \gamma_{i(opt)} \tag{4.23}$$

$$\gamma_{i(opt)} = \frac{C_i^* D_i^* - A_i^* E_i^*}{D_i} \tag{4.24}$$

where

$$\begin{aligned}
 A'_i &= [1 + f_2 C_y^2 + f_i \{\psi C_x (3\psi C_x - 4\rho_{yx} C_y)\}], \\
 B'_i &= R_1^2 f_i C_x^2, \quad R_1 = \frac{\bar{X}}{\bar{Y}}, \\
 C'_i &= R_2^2 f_i C_r^2, \quad R_2 = \frac{\bar{R}_x}{\bar{Y}}, \\
 D'_i &= R_1 f_i C_x (2\psi C_x - \rho_{yx} C_y), \quad E'_i = R_2 f_i C_r (2\psi \rho_{xr} C_x - \rho_{yr} C_y), \\
 F'_i &= R_1 R_2 f_i \rho_{xr} C_x C_r, \quad G'_i = [1 + \psi f_i C_x (\psi C_x - \rho_{yx} C_y)], \\
 H'_i &= R_1 \psi f_i C_x^2, \quad I'_i = R_2 \psi f_i \rho_{xr} C_x C_r, \\
 F_i &= 1 - \psi f_i C_x (\rho_{yx} C_y - \psi C_x), \\
 G_i &= \psi f_i C_x^2, \quad H_i = \psi f_i \rho_{xr} C_x C_r, \\
 Q_i^2 &= E_i^2 - (A_i^* E_i^{*2} + B_i^* D_i^{*2} - 2C_i^* D_i^* E_i^*) / D_i, \\
 P_i &= 1 + f_2 C_y^2 + f_i [\psi C_x (3\psi C_x - 4\rho_{yx} C_y)], \\
 D_i &= C_i^{*2} - A_i^* B_i^*, \\
 A_i^* &= R_1^2 f_i C_x^2 [(1 + f_2 C_y^2) - f_i (\rho_{yx}^2) - \psi^2 C_x^2], \\
 B_i^* &= R_2^2 f_i C_r^2 \left[(1 + f_2 C_y^2) - f_i \{ \rho_{yr}^2 - \psi^2 C_x^2 (3 - 4\rho_{xr}^2) - 4\psi C_y C_x (\rho_{yx} - \rho_{yr} \rho_{xr}) \} \right], \\
 C_i^* &= R_2 f_i \rho_{xr} C_x C_r [R_1 (1 + f_2 C_y^2 + 3f_i C_x^2 - 4f_i \rho_{yx} C_y C_x) \\
 &\quad - R_2 \psi f_i C_r (2\psi \rho_{xr} C_x - \rho_{yr} C_y)], \\
 D_i^* &= R_1 f_i C_x [(\psi C_x - \rho_{yx} C_y) (\psi^2 f_i C_x^2 - 1) + \psi f_i C_y^2 C_x (1 - \rho_{yx}^2)], \\
 E_i^* &= R_2 f_i C_r [\rho_{yr} C_y - \psi \rho_{xr} C_x (1 - \psi^2 f_i C_x^2) \\
 &\quad + \psi f_i C_y C_x \{ C_y (\rho_{xr} - \rho_{yx} \rho_{yr}) + \psi C_x (\rho_{yr} - 2\rho_{yx} \rho_{xr}) \}], \\
 E_i &= 1 + \psi f_i C_x (\psi C_x - \rho_{yx} C_y), \quad L_i = R_1 f_i C_x (2\psi C_x - \rho_{yx} C_y), \\
 \text{and } M_i &= R_2 f_i C_r (2\psi \rho_{xr} C_x - \rho_{yr} C_y).
 \end{aligned}$$

Proof Expressing (4.4), (4.5) and (4.6) in terms of errors, we get

$$t_{P_1} = [\bar{Y}\alpha_1(1 + \xi_0) - \bar{X}\beta_1\xi_1 - \bar{R}_x\gamma_1\xi_2] (1 + \psi\xi_1)^{-1} \tag{4.25}$$

$$t_{P_2} = [\bar{Y}\alpha_2(1 + \xi_0) - \bar{X}\beta_2\xi'_1 - \bar{R}_x\gamma_2\xi'_2] (1 + \psi\xi'_1)^{-1} \tag{4.26}$$

$$\begin{aligned}
 t_{P_3} &= [\bar{Y}\alpha_3(1 + \xi_0) - \bar{X}\beta_3(\xi_1 - \xi'_1) - \bar{R}_x\gamma_3(\xi_2 - \xi'_2)] \\
 &\quad (1 + \psi\xi_1)(1 + \psi\xi'_1)^{-1} \tag{4.27}
 \end{aligned}$$

where $\psi = \frac{u\bar{X}}{u\bar{X}+v}$. We assume that $|\xi_1| < 1$ and $|\xi'_1| < 1$, so that $(1 + \xi_1)^{-1}$ and $(1 + \xi'_1)^{-1}$ are expandable. Now, expanding $(1 + \xi_1)^{-1}$ and $(1 + \xi'_1)^{-1}$ binomially, multiplying over right-hand sides (r.h.s.) of (4.25), (4.26) and (4.27) and neglecting the terms of errors having power greater than two, we can get, respectively, as:

$$t_{P_1} = \bar{Y} \left[\alpha_1 \left(1 + \xi_0 - \psi\xi_1 + \psi\xi_2^2 - \psi\xi_0\xi_1 \right) \right]$$

$$-\beta_1 R_1(\xi_1 - \psi \xi_1^2) - \gamma_1 R_2(\xi_2 - \psi \xi_1 \xi_2) \Big] \tag{4.28}$$

$$t_{P_2} = \bar{Y} \left[\alpha_2 \left(1 + \xi_0 - \psi \xi_1' + \psi \xi_2'^2 - \psi \xi_0 \xi_1' \right) - \beta_2 R_1(\xi_1' - \psi \xi_1'^2) - \gamma_2 R_2(\xi_2' - \psi \xi_1' \xi_2') \right] \tag{4.29}$$

$$t_{P_3} = \bar{Y} \left[\alpha_3 \left(1 + \xi_0 - \psi \xi_1' + \psi \xi_2'^2 - \psi \xi_0 \xi_1' \right) - \beta_3 R_1(\xi_1' - \psi \xi_1'^2) - \gamma_3 R_2(\xi_2' - \psi \xi_1' \xi_2') \right] \tag{4.30}$$

The biases of the estimators $t_{P_i} (i = 1, 2, 3)$ can be derived as:

$$B(t_{P_1}) = E [t_{P_1} - \bar{Y}] = E \left[\bar{Y} \left[\alpha_1 \left(1 + \xi_0 - \psi \xi_1 + \psi \xi_2^2 - \psi \xi_0 \xi_1 \right) - \beta_1 R_1(\xi_1 - \psi \xi_1^2) - \gamma_1 R_2(\xi_2 - \psi \xi_1 \xi_2) \right] - \bar{Y} \right] \tag{4.31}$$

$$B(t_{P_2}) = E [t_{P_2} - \bar{Y}] = E \left[\bar{Y} \left[\alpha_2 \left(1 + \xi_0 - \psi \xi_1' + \psi \xi_2'^2 - \psi \xi_0 \xi_1' \right) - \beta_2 R_1(\xi_1' - \psi \xi_1'^2) - \gamma_2 R_2(\xi_2' - \psi \xi_1' \xi_2') \right] - \bar{Y} \right] \tag{4.32}$$

$$B(t_{P_3}) = E [t_{P_3} - \bar{Y}] = E \left[\bar{Y} \left[\alpha_3 \left(1 + \xi_0 - \psi \xi_1' + \psi \xi_2'^2 - \psi \xi_0 \xi_1' \right) - \beta_3 R_1(\xi_1' - \psi \xi_1'^2) - \gamma_3 R_2(\xi_2' - \psi \xi_1' \xi_2') \right] - \bar{Y} \right] \tag{4.33}$$

Taking expectations of both sides of (4.31)–(4.33), we get the expressions for biases given in (4.19).

The MSEs of the estimators $t_{P_i} (i = 1, 2, 3)$ to the first-order approximation can be derived as:

$$M(t_{P_1}) = E [t_{P_1} - \bar{Y}]^2 = E \left[\bar{Y}^2 \left[\alpha_1^2 (1 + 2\xi_0 - 2\psi \xi_1 + \xi_0^2 + 3\psi^2 \xi_1^2 - 4\psi \xi_0 \xi_1) + \beta_1^2 R_1^2 \xi_1^2 + \gamma_1^2 R_2^2 \xi_2^2 + 2\alpha_1 \beta_1 R_1 (2\psi \xi_1^2 - \xi_0 \xi_1) + 2\alpha_1 \gamma_1 R_2 (2\psi \xi_1 \xi_2 - \xi_0 \xi_2) + 2\alpha_1 \beta_1 R_1 R_2 \xi_1 \xi_2 - 2\alpha_1 (1 + \psi^2 \xi_1^2 - \psi \xi_0 \xi_1) - 2\beta_1 R_1 \psi \xi_1^2 - 2\gamma_3 R_2 \psi \xi_1 \xi_2 + 1 \right] \right] \tag{4.34}$$

$$M(t_{P_2}) = E [t_{P_2} - \bar{Y}]^2 = E \left[\bar{Y}^2 \left[\alpha_2^2 (1 + 2\xi_0 - 2\psi \xi_1' + \xi_0^2 + 3\psi^2 \xi_1'^2 - 4\psi \xi_0 \xi_1') + \beta_2^2 R_1^2 \xi_1'^2 + \gamma_2^2 R_2^2 \xi_2'^2 + 2\alpha_2 \beta_2 R_1 (2\psi \xi_1'^2 - \xi_0 \xi_1') + 2\alpha_2 \gamma_2 R_2 (2\psi \xi_1' \xi_2' - \xi_0 \xi_2') + 2\alpha_2 \beta_2 R_1 R_2 \xi_1' \xi_2' - 2\alpha_2 (1 + \psi^2 \xi_1'^2 - \psi \xi_0 \xi_1') - 2\beta_2 R_1 \psi \xi_1'^2 - 2\gamma_2 R_2 \psi \xi_1' \xi_2' + 1 \right] \right] \tag{4.35}$$

$$\begin{aligned}
 M(t_{P_3}) &= E [t_{P_3} - \bar{Y}]^2 \\
 &= E [\bar{Y}^2 [\alpha_3^2(1 + 2\xi_0 - 2\psi\xi'_1 + 2\psi\xi_1 + \xi_0^2 + 3\psi^2\xi_1'^2 + 4\psi\xi_0\xi_1 \\
 &\quad - 4\psi\xi_0\xi'_1 - 4\psi^2\xi_1\xi'_1) \\
 &\quad + \beta_3^2 R_1^2(\xi_1^2 + \xi_1'^2 - 2\xi_1\xi'_1) + \gamma_3^2 R_2^2(\xi_2^2 + \xi_2'^2 - 2\xi_2\xi_2') \\
 &\quad + 2\alpha_3\beta_3 R_1(2\psi\xi_1'^2 + 2\psi\xi_1'^2 + \xi_0\xi_1 - \xi_0\xi'_1 - 4\psi\xi_1\xi'_1) \\
 &\quad + 2\alpha_3\gamma_3 R_2(\xi_0\xi_2 - \xi_0\xi_2' + 2\psi\xi_1\xi_2 - 4\psi\xi_1\xi_2' + 2\psi\xi_1'\xi_2') \\
 &\quad + 2\beta_3\gamma_3 R_1 R_2(\xi_1\xi_2 - \xi_1\xi_2' - \xi_1'\xi_2 - \xi_1'\xi_2') \\
 &\quad - 2\alpha_3(1 + \psi^2\xi_1'^2 + \psi\xi_0\xi'_1 - \psi\xi_0\xi_1' - \psi^2\xi_1\xi_1') \\
 &\quad - 2\beta_3 R_1(\psi\xi_1^2 + \psi\xi_1'^2 - 2\psi\xi_1\xi_1') \\
 &\quad - 2\gamma_3 R_2(\psi\xi_1\xi_2 - \psi\xi_1\xi_2' - \psi\xi_1'\xi_2 + \psi\xi_1'\xi_2') + 1]]
 \end{aligned} \tag{4.36}$$

Now, taking the expectations of both sides of (4.34)–(4.36), we get the expressions for MSEs given in (4.20).

To obtain the optimum choices of α_i , β_i and γ_i , we differentiate the expressions of MSEs given in (3.20) partially with respect to α_i , β_i and γ_i and equate them to zero, we get

$$\frac{\partial}{\partial \alpha_i} MSE(t_{P_i}) = \alpha_i A'_i + \beta_i D'_i + \gamma_i E'_i - G'_i = 0 \tag{4.37}$$

$$\frac{\partial}{\partial \beta_i} MSE(t_{P_i}) = \alpha_i D'_i + \beta_i B'_i + \gamma_i F'_i - H'_i = 0 \tag{4.38}$$

$$\frac{\partial}{\partial \gamma_i} MSE(t_{P_i}) = \alpha_i E'_i + \beta_i F'_i + \gamma_i C'_i - I'_i = 0 \tag{4.39}$$

Solving Eqs. (4.37)–(4.39) for α_i , β_i and γ_i , we get the optimum values given in (4.22)–(4.24). By putting these optimum values in (4.20), we get the expressions for minimum MSEs of t_{P_i} given in (4.21). \square

Corollary 4.1 *The MSEs of the unbiased estimators t_{d_i} ($i = 1, 2, 3$) to the first-order approximations are given by*

$$\begin{aligned}
 MSE(t_{d_i}) &= \bar{Y}^2 \left[f_2 C_y^2 + \beta_i^2 R_1^2 f_i C_x^2 + \gamma_i^2 R_2^2 f_i C_r^2 + 2\beta_i \gamma_i R_1 R_2 f_i \rho_{xr} C_x C_r \right. \\
 &\quad \left. - 2\beta_i R_1 f_i \rho_{yx} C_y C_x + 2\gamma_i R_2 f_i \rho_{yr} C_y C_r \right]
 \end{aligned} \tag{4.40}$$

The minimum MSEs of the estimators t_{d_i} ($i = 1, 2, 3$) to the first-order approximations are given by

$$\min. MSE(t_{d_i}) = \bar{Y}^2 C_y^2 (f_2 - f_i R_{y.xr_x}^2) \tag{4.41}$$

for the optimum values

$$\phi_{i(opt)} = \frac{C_y}{C_x} \frac{(\rho_{yx} - \rho_{yr_x} \rho_{xr_x})}{1 - \rho_{xr_x}^2} \tag{4.42}$$

$$\phi_{i(opt)} = \frac{C_y}{C_r} \frac{(\rho_{yr_x} - \rho_{yx}\rho_{xr_x})}{1 - \rho_{xr_x}^2} \tag{4.43}$$

where $R_{y.xr_x}^2 = \left(\frac{\rho_{yx}^2 + \rho_{yr_x}^2 - 2\rho_{yx}\rho_{yr_x}\rho_{xr_x}}{1 - \rho_{xr_x}^2} \right)$.

Proof By putting $\alpha_i = 1$ and $\psi = 0$ in (4.20), we get the expressions for MSEs of the estimators t_{d_i} given in (4.40).

To obtain minimum MSEs of t_{d_i} , we differentiate the expressions of MSEs given in (4.40) partially with respect to ϕ_i and φ_i and equate them to zero, we get

$$\begin{aligned} \frac{\partial}{\partial \phi_i} MSE(t_{d_i}) &= \phi_i R_1^2 f_i C_x^2 \\ &+ \varphi_i R_1 R_2 f_i \rho_{xr} C_x C_r - R_1 f_i \rho_{yx} C_y C_x = 0 \end{aligned} \tag{4.44}$$

$$\begin{aligned} \frac{\partial}{\partial \varphi_i} MSE(t_{d_i}) &= \phi_i R_1 R_2 f_i \rho_{xr} C_x C_r \\ &+ \varphi_i R_2^2 f_i C_r^2 - R_1 f_i \rho_{yx} C_y C_x = 0 \end{aligned} \tag{4.45}$$

Solving Eqs. (4.44) and (4.45) for ϕ_i and φ_i , we get the optimum values given in (4.42) and (4.43). By putting these optimum values in (4.40), we get the minimum MSE of t_{d_i} given in (4.41). \square

Corollary 4.2 *The biases and MSEs of the estimators t_{D_i} ($i = 1, 2, 3$) to the first-order approximations are given by*

$$B(t_{D_i}) = \bar{Y}(\alpha_i^* - 1) \tag{4.46}$$

and

$$\begin{aligned} MSE(t_{D_i}) &= \bar{Y}^2 \left[\alpha_i^{*2} (1 + f_2 C_y^2) + \beta_i^{*2} R_1^2 f_i C_x^2 + \gamma_i^{*2} R_2^2 f_i C_r^2 - 2\alpha_i^* \beta_i^* R_1 f_i \rho_{yx} C_y C_x \right. \\ &\quad \left. + 2\alpha_i^* \gamma_i^* R_2 f_i \rho_{yr} C_y C_r + 2\beta_i^* \gamma_i^* R_1 R_2 f_i \rho_{xr} C_x C_r - 2\alpha_i^* \right] \end{aligned} \tag{4.47}$$

The minimum MSEs of the estimators t_{D_i} are given by

$$min.MSE(t_{D_i}) = \frac{\bar{Y}^2 min.MSE(t_{d_i})}{\bar{Y}^2 + min.MSE(t_{d_i})} \tag{4.48}$$

for the optimum values

$$\alpha_{i(opt)}^* = \frac{1}{1 + C_y^2 (f_2 - f_i R_{y.xr_x}^2)} \tag{4.49}$$

$$\beta_{i(opt)}^* = \alpha_{i(opt)}^* \frac{C_y (\rho_{yx} - \rho_{yr_x} \rho_{xr_x})}{C_x (1 - \rho_{xr_x}^2)} \tag{4.50}$$

$$\gamma_{i(opt)}^* = \alpha_{i(opt)}^* \frac{C_y (\rho_{yr_x} - \rho_{yx}\rho_{xr_x})}{C_r (1 - \rho_{xr_x}^2)} \tag{4.51}$$

Proof By putting $\psi = 0$ in (4.19) and (4.20), we get the expressions for biases and MSEs of the estimators t_{D_i} given in (4.46) and (4.47).

To obtain minimum MSEs of t_{D_i} , we differentiate the expressions of MSEs given in (4.47) partially with respect to α_i^* , β_i^* and γ_i^* and equate them to zero, we get

$$\frac{\partial}{\partial \alpha_i^*} MSE(t_{D_i}) = \alpha_i^* (1 + f_2 C_y^2) - \beta_i^* R_1 f_i \rho_{yx} C_y C_x + \gamma_i^* R_2 f_i \rho_{yr} C_y C_r - 1 = 0 \tag{4.52}$$

$$\frac{\partial}{\partial \beta_i^*} MSE(t_{D_i}) = \alpha_i^* R_1 f_i \rho_{yx} C_y C_x - \beta_i^* R_1^2 f_i C_x^2 - \gamma_i^* R_1 R_2 f_i \rho_{xr} C_x C_r = 0 \tag{4.53}$$

$$\frac{\partial}{\partial \gamma_i^*} MSE(t_{D_i}) = \alpha_i^* R_2 f_i \rho_{yr} C_y C_r + \beta_i^* R_1 R_2 f_i \rho_{xr} C_x C_r + \gamma_i^* R_2^2 f_i C_r^2 = 0 \tag{4.54}$$

Solving Equations (4.52)–(4.54) for α_i^* , β_i^* and γ_i^* , we get the optimum values given in (4.49)–(4.51). By putting these optimum values in (4.47), we get the minimum MSE of t_{D_i} given in (4.48).

4.3 Practicability of the suggested estimators

The suggested imputation methods t_{P_i} ($i = 1, 2, 3$) are designed using the scalars $(\alpha_i, \beta_i, \gamma_i)$. Therefore, we have to choose the appropriate value for these scalars in order to estimate the population mean. We have seen in (4.22), (4.23) and (4.24) that the optimum values of α_i , β_i , and γ_i depend on the parameters \bar{Y} , \bar{X} , ρ_{yx} , C_y , C_x , etc., which may not be available every time. In such situations, they can be estimated using a pilot survey or guessed from a past survey and subsequently employed for estimation of the population mean. Similarly, the optimum values of scalars used in estimators t_{d_i} and t_{D_i} can be obtained at the estimation stage. □

4.4 Some other members of proposed classes of estimators t_{P_1} , t_{P_2} and t_{P_3}

Many estimators can be generated using t_{P_1} , t_{P_2} and t_{P_3} families of estimators by choosing various values of u and v in (4.4), (4.5) and (4.6). Some of them in strategy I, strategy II and strategy III, respectively, are given in Table 1.

The respective imputations for the data can be formed just by putting the suitable values of u and v in (4.1), (4.2) and (4.3). The biases and minimum MSEs of the estimators $t_{P_i}^{(j)}$ ($i = 1, 2, 3$) and ($j = 1, 2, \dots, 10$) can be easily obtained by putting suitable values in (4.19) and (4.20).

Table 1 Some members of proposed family of estimators t_{P_1} , t_{P_2} and t_{P_3}

Estimators	u	v
Strategy I		
$t_{P_1}^{(1)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_n) + \gamma_1 (\bar{R}_x - \bar{r}_{x(n)})] \left[\frac{\bar{X}}{\bar{x}_n} \right] = t_{P_1}^* \text{ (say)}$	1	0
$t_{P_1}^{(2)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_n) + \gamma_1 (\bar{R}_x - \bar{r}_{x(n)})] \left[\frac{\bar{X} + C_x}{\bar{x}_n + C_x} \right]$	1	C_x
$t_{P_1}^{(3)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_n) + \gamma_1 (\bar{R}_x - \bar{r}_{x(n)})] \left[\frac{\bar{X} + \rho_{yx}}{\bar{x}_n + \rho_{yx}} \right]$	1	ρ_{yx}
$t_{P_1}^{(4)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_n) + \gamma_1 (\bar{R}_x - \bar{r}_{x(n)})] \left[\frac{\rho_{yx} \bar{X} + C_x}{\rho_{yx} \bar{x}_n + C_x} \right]$	ρ_{yx}	C_x
$t_{P_1}^{(5)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_n) + \gamma_1 (\bar{R}_x - \bar{r}_{x(n)})] \left[\frac{C_x \bar{X} + \rho_{yx}}{C_x \bar{x}_n + \rho_{yx}} \right]$	C_x	ρ_{yx}
$t_{P_1}^{(6)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_n) + \gamma_1 (\bar{R}_x - \bar{r}_{x(n)})] \left[\frac{\bar{X} + \beta_{2(x)}}{\bar{x}_n + \beta_{2(x)}} \right]$	1	$\beta_{2(x)}$
$t_{P_1}^{(7)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_n) + \gamma_1 (\bar{R}_x - \bar{r}_{x(n)})] \left[\frac{\beta_{2(x)} \bar{X} + C_x}{\beta_{2(x)} \bar{x}_n + C_x} \right]$	$\beta_{2(x)}$	C_x
$t_{P_1}^{(8)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_n) + \gamma_1 (\bar{R}_x - \bar{r}_{x(n)})] \left[\frac{\beta_{2(x)} \bar{X} + \rho_{yx}}{\beta_{2(x)} \bar{x}_n + \rho_{yx}} \right]$	$\beta_{2(x)}$	ρ_{yx}
$t_{P_1}^{(9)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_n) + \gamma_1 (\bar{R}_x - \bar{r}_{x(n)})] \left[\frac{\rho_{yx} \bar{X} + \beta_{2(x)}}{\rho_{yx} \bar{x}_n + \beta_{2(x)}} \right]$	ρ_{yx}	$\beta_{2(x)}$
$t_{P_1}^{(10)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_n) + \gamma_1 (\bar{R}_x - \bar{r}_{x(n)})] \left[\frac{C_x \bar{X} + \beta_{2(x)}}{C_x \bar{x}_n + \beta_{2(x)}} \right]$	C_x	$\beta_{2(x)}$
Strategy II		
$t_{P_2}^{(1)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_r) + \gamma_1 (\bar{R}_x - \bar{r}_{x(r)})] \left[\frac{\bar{X}}{\bar{x}_r} \right] = t_{P_2}^* \text{ (say)}$	1	0
$t_{P_2}^{(2)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_r) + \gamma_1 (\bar{R}_x - \bar{r}_{x(r)})] \left[\frac{\bar{X} + C_x}{\bar{x}_r + C_x} \right]$	1	C_x
$t_{P_2}^{(3)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_r) + \gamma_1 (\bar{R}_x - \bar{r}_{x(r)})] \left[\frac{\bar{X} + \rho_{yx}}{\bar{x}_r + \rho_{yx}} \right]$	1	ρ_{yx}
$t_{P_2}^{(4)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_r) + \gamma_1 (\bar{R}_x - \bar{r}_{x(r)})] \left[\frac{\rho_{yx} \bar{X} + C_x}{\rho_{yx} \bar{x}_r + C_x} \right]$	ρ_{yx}	C_x
$t_{P_2}^{(5)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_r) + \gamma_1 (\bar{R}_x - \bar{r}_{x(r)})] \left[\frac{C_x \bar{X} + \rho_{yx}}{C_x \bar{x}_r + \rho_{yx}} \right]$	C_x	ρ_{yx}
$t_{P_2}^{(6)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_r) + \gamma_1 (\bar{R}_x - \bar{r}_{x(r)})] \left[\frac{\bar{X} + \beta_{2(x)}}{\bar{x}_r + \beta_{2(x)}} \right]$	1	$\beta_{2(x)}$
$t_{P_2}^{(7)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_r) + \gamma_1 (\bar{R}_x - \bar{r}_{x(r)})] \left[\frac{\beta_{2(x)} \bar{X} + C_x}{\beta_{2(x)} \bar{x}_r + C_x} \right]$	$\beta_{2(x)}$	C_x
$t_{P_2}^{(8)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_r) + \gamma_1 (\bar{R}_x - \bar{r}_{x(r)})] \left[\frac{\beta_{2(x)} \bar{X} + \rho_{yx}}{\beta_{2(x)} \bar{x}_r + \rho_{yx}} \right]$	$\beta_{2(x)}$	ρ_{yx}
$t_{P_2}^{(9)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_r) + \gamma_1 (\bar{R}_x - \bar{r}_{x(r)})] \left[\frac{\rho_{yx} \bar{X} + \beta_{2(x)}}{\rho_{yx} \bar{x}_r + \beta_{2(x)}} \right]$	ρ_{yx}	$\beta_{2(x)}$
$t_{P_2}^{(10)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{X} - \bar{x}_r) + \gamma_1 (\bar{R}_x - \bar{r}_{x(r)})] \left[\frac{C_x \bar{X} + \beta_{2(x)}}{C_x \bar{x}_r + \beta_{2(x)}} \right]$	C_x	$\beta_{2(x)}$

Table 1 continued

Estimators	u	v
Strategy III		
$t_{P_3}^{(1)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{x}_n - \bar{x}_r) + \gamma_1 (\bar{r}_{x(n)} - \bar{r}_{x(r)})] \left[\frac{\bar{x}_n}{\bar{x}_r} \right] = t_{P_3}^* (say)$	1	0
$t_{P_3}^{(2)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{x}_n - \bar{x}_r) + \gamma_1 (\bar{r}_{x(n)} - \bar{r}_{x(r)})] \left[\frac{\bar{x}_n + C_x}{\bar{x}_r + C_x} \right]$	1	C_x
$t_{P_3}^{(3)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{x}_n - \bar{x}_r) + \gamma_1 (\bar{r}_{x(n)} - \bar{r}_{x(r)})] \left[\frac{\bar{x}_n + \rho_{yx}}{\bar{x}_r + \rho_{yx}} \right]$	1	ρ_{yx}
$t_{P_3}^{(4)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{x}_n - \bar{x}_r) + \gamma_1 (\bar{r}_{x(n)} - \bar{r}_{x(r)})] \left[\frac{\rho_{yx} \bar{x}_n + C_x}{\rho_{yx} \bar{x}_r + C_x} \right]$	ρ_{yx}	C_x
$t_{P_3}^{(5)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{x}_n - \bar{x}_r) + \gamma_1 (\bar{r}_{x(n)} - \bar{r}_{x(r)})] \left[\frac{C_x \bar{x}_n + \rho_{yx}}{C_x \bar{x}_r + \rho_{yx}} \right]$	C_x	ρ_{yx}
$t_{P_3}^{(6)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{x}_n - \bar{x}_r) + \gamma_1 (\bar{r}_{x(n)} - \bar{r}_{x(r)})] \left[\frac{\bar{x}_n + \beta_{2(x)}}{\bar{x}_r + \beta_{2(x)}} \right]$	1	$\beta_{2(x)}$
$t_{P_3}^{(7)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{x}_n - \bar{x}_r) + \gamma_1 (\bar{r}_{x(n)} - \bar{r}_{x(r)})] \left[\frac{\beta_{2(x)} \bar{x}_n + C_x}{\beta_{2(x)} \bar{x}_r + C_x} \right]$	$\beta_{2(x)}$	C_x
$t_{P_3}^{(8)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{x}_n - \bar{x}_r) + \gamma_1 (\bar{r}_{x(n)} - \bar{r}_{x(r)})] \left[\frac{\beta_{2(x)} \bar{x}_n + \rho_{yx}}{\beta_{2(x)} \bar{x}_r + \rho_{yx}} \right]$	$\beta_{2(x)}$	ρ_{yx}
$t_{P_3}^{(9)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{x}_n - \bar{x}_r) + \gamma_1 (\bar{r}_{x(n)} - \bar{r}_{x(r)})] \left[\frac{\rho_{yx} \bar{x}_n + \beta_{2(x)}}{\rho_{yx} \bar{x}_r + \beta_{2(x)}} \right]$	ρ_{yx}	$\beta_{2(x)}$
$t_{P_3}^{(10)} = [\alpha_1 \bar{y}_r + \beta_1 (\bar{x}_n - \bar{x}_r) + \gamma_1 (\bar{r}_{x(n)} - \bar{r}_{x(r)})] \left[\frac{C_x \bar{x}_n + \beta_{2(x)}}{C_x \bar{x}_r + \beta_{2(x)}} \right]$	C_x	$\beta_{2(x)}$

5 Theoretical Comparisons

In this section, we compare the proposed estimators with above existing estimators based on their theoretical results.

5.1 Parallel Comparisons

5.1.1 Comparisons of $t_{P_i} (i = 1, 2, 3)$ with Other Existing Estimators

From (3.2), (3.9), (3.16), (3.23), (3.30), (3.37), (3.47) and (4.20), we get

(i) $min.MSE(t_{P_i}) < MSE(\bar{y}_r)$, if

$$\frac{Q_i^2}{P_i} + f_2 C_y^2 > 1 \tag{5.1}$$

(ii) $min.MSE(t_{P_i}) < MSE(t_{R_i})$, if

$$\frac{Q_i^2}{P_i} + [f_2 C_y^2 + f_i C_x (C_x - 2\rho_{yx} C_y)] > 1 \tag{5.2}$$

(iii) $min.MSE(t_{P_i}) < MSE(t_{KC_i})$, if

$$\frac{Q_i^2}{P_i} + [f_2 C_y^2 + f_i (C_x^2 - \rho_{yx}^2 C_y^2)] > 1 \tag{5.3}$$

(iv) $\min.MSE(t_{P_i}) < \min.MSE(t_{G_i})$ or $MSE(t_{DP_i})$, if

$$\frac{Q_i^2}{P_i} + C_y^2(f_2 - f_i \rho_{yx}^2) > 1 \tag{5.4}$$

(v) $\min.MSE(t_{P_i}) < \min.MSE(t_{BP_i})$, if

$$\frac{Q_i^2}{P_i} + \frac{1}{1 + \frac{1}{C_y^2(f_2 - f_i \rho_{yx}^2)}} > 1 \tag{5.5}$$

(vi) $\min.MSE(t_{P_i}) < \min.MSE(t_{BP_i}^*)$, if

$$\frac{Q_i^2}{P_i} + \left(1 - \frac{H_i^2}{G_i}\right) > 1 \tag{5.6}$$

5.1.2 Comparisons of t_{d_i} ($i = 1, 2, 3$) with Other Existing Estimators

From (3.2), (3.9), (3.16), (3.23), (3.30), (3.37), (3.47) and (4.41), we get

(i) $\min.MSE(t_{d_i}) < MSE(\bar{y}_r)$, if

$$R_{y.xr_x}^2 > 0 \tag{5.7}$$

(ii) $\min.MSE(t_{d_i}) < MSE(t_{R_i})$, if

$$R_{y.xr_x}^2 + C_x(C_x - 2\rho_{yx}C_y) > 0 \tag{5.8}$$

(iii) $\min.MSE(t_{d_i}) < MSE(t_{KC_i})$, if

$$R_{y.xr_x}^2 + (C_x^2 - \rho_{yx}^2 C_y^2) > 0 \tag{5.9}$$

(iv) $\min.MSE(t_{d_i}) < \min.MSE(t_{G_i})$ or $MSE(t_{DP_i})$, if

$$(\rho_{y_{r_x}} - \rho_{yx}\rho_{x_{r_x}})^2 > 0 \tag{5.10}$$

(v) $\min.MSE(t_{d_i}) < \min.MSE(t_{BP_i})$, if

$$\left(\frac{1}{f_2 - f_i R_{y.xr_x}^2} - \frac{1}{f_2 - f_i \rho_{yx}^2}\right) - C_y^2 > 0 \tag{5.11}$$

(vi) $\min.MSE(t_{d_i}) < \min.MSE(t_{BP_i}^*)$, if

$$\left(1 - \frac{H_i^2}{G_i}\right) - C_y^2(f_2 - f_i R_{y.xr_x}^2) > 0 \tag{5.12}$$

From (5.10), it is clear that the estimator t_{d_i} ($i = 1, 2, 3$) are always better than t_{DP_i} which contradicts the statement of Diana and Perri [8]. Diana and Perri [8] stated “Using the same amount of auxiliary information, no further improvement upon the regression estimator is possible, at least if the first order approximation is considered” which appears to be false over here.

5.1.3 Comparisons of t_{D_i} ($i = 1, 2, 3$) with Other Existing Estimators

From (3.2), (3.9), (3.16), (3.23), (3.30), (3.37), (3.47) and (4.48), we get

(i) $\min.MSE(t_{D_i}) < MSE(\bar{y}_r)$, if

$$f_2 C_y^2 \left(1 - \frac{f_2}{f_i R_{y.xr_x}^2} \right) > 1 \tag{5.13}$$

(ii) $\min.MSE(t_{D_i}) < MSE(t_{R_i})$, if

$$\left[1 + \frac{1}{C_y^2(f_2 - f_i R_{y.xr_x}^2)} \right] [f_2 C_y^2 + f_i C_x(C_x - 2\rho_{yx} C_y)] > 1 \tag{5.14}$$

(iii) $\min.MSE(t_{D_i}) < MSE(t_{KC_i})$, if

$$\left[1 + \frac{1}{C_y^2(f_2 - f_i R_{y.xr_x}^2)} \right] [f_2 C_y^2 + f_i(C_x^2 - \rho_{yx}^2 C_y^2)] > 1 \tag{5.15}$$

(iv) $\min.MSE(t_{D_i}) < \min.MSE(t_{G_i})$ or $MSE(t_{DP_i})$, if

$$\left[C_y^2 + \frac{1}{(f_2 - f_i R_{y.xr_x}^2)} \right] (f_2 - f_i \rho_{yx}^2) > 1 \tag{5.16}$$

(v) $\min.MSE(t_{D_i}) < \min.MSE(t_{BP_i})$, if

$$\frac{R_{y.xr_x}^2}{\rho_{yx}^2} > 1 \tag{5.17}$$

(vi) $\min.MSE(t_{D_i}) < \min.MSE(t_{BP_i}^*)$, if

$$\left[1 + \frac{1}{C_y^2(f_2 - f_i R_{y.xr_x}^2)} \right] \left(1 - \frac{H_i^2}{G_i} \right) > 1 \tag{5.18}$$

Thus, the estimators t_{P_i} ($i = 1, 2, 3$), t_{d_i} and t_{D_i} are better than the traditional estimators t_{R_i} , t_{G_i} , t_{KC_i} , t_{DP_i} , t_{BP_i} and $t_{BP_i}^*$ in parallel if the conditions (5.1)–(5.6), (5.7)–(5.12) and (5.13)–(5.18) are, respectively, satisfied.

5.2 Mutual Comparisons of the Proposed Estimators

Here, we discuss the mutual comparisons of the estimators t_{P_i} ($i = 1, 2, 3$), t_{d_i} and t_{D_i} .

(i) $min.MSE(t_{\bullet_1}) < min.MSE(t_{\bullet_3})$, if

$$r > \left(\frac{nN}{2N - n} \right) \tag{5.19}$$

(ii) $min.MSE(t_{\bullet_2}) < min.MSE(t_{\bullet_1})$, if

$$r < n \tag{5.20}$$

(iii) $min.MSE(t_{\bullet_2}) < min.MSE(t_{\bullet_3})$, if

$$n < N \tag{5.21}$$

This means that in the respective sets of estimators, the second estimators t_{P_2} , t_{d_2} and t_{D_2} are always better than first estimators t_{P_1} , t_{d_1} and t_{D_1} and third estimators t_{P_3} , t_{d_3} and t_{D_3} , whereas the first estimators t_{P_1} , t_{d_1} and t_{D_1} are better than the estimators t_{P_3} , t_{d_3} and t_{D_3} , respectively, if the condition (5.19) holds. The similar conditions also holds for other existing estimators discussed above in the respective strategies.

6 Empirical Comparisons and Computations

To judge the merits of the proposed class of estimators t_{P_1} , t_{P_2} and t_{P_3} over the other considered estimators in the respective strategies, we have chosen 10 real data sets whose parametric details are given as follows:

Data set-1: [13,p-428]: The data are on capital expenditures (y) and approximations (x) for the years 1953-1967 on a quarterly basis. These data are from the National Industrial Conference Board. The description of the parameters for this data is: $N = 60, n = 20, r = 16, \bar{Y} = 3092.417, \bar{X} = 3319.483, C_y = 0.3725059, C_x = 0.4159578, C_r = 0.5725904, \rho_{yx} = 0.8832073, \rho_{yr_x} = 0.7818964, \rho_{xr_x} = 0.9592037$.

Data set-2: [13,p-108]: The data present experience and salary structure of University of Michigan economists in 1983-1984. Let y be the salary (thousands of dollars) and x be the years of experience (defined as years since receiving Ph.D.). The description of the parameters for these data is: $N = 32, n = 12, r = 8, \bar{Y} = 47.37812, \bar{X} = 18.375, C_y = 0.1819515, C_x = 0.4548528, C_r = 0.5677532, \rho_{yx} = 0.4245114, \rho_{yr_x} = 0.3368753, \rho_{xr_x} = 0.9447145$.

Data set-3: [13,p-41]: The data are on the weekly cash inflows (x) and outflows (y) of a business firm for 30 weeks. The description of the parameters for these data is: $N = 30, n = 12, r = 8, \bar{Y} = 51.73333, \bar{X} = 62.93333, C_y = 0.4261637, C_x = 0.3361672, C_r = 0.5676459, \rho_{yx} = -0.009132783, \rho_{yr_x} = -0.02862014, \rho_{xr_x} = 0.9927513$.

Data set-4: [19,p-108]: A list of 70 villages in a Tehsil of India along with their population in 1981 and cultivated area (in acres) in the same year is taken into consideration. Let y be the cultivated area(in acres) and x be the population of village. The description of the parameters for these data is: $N = 70$, $n = 20$, $r = 15$, $\bar{Y} = 981.2857$, $\bar{X} = 1755.529$, $C_y = 0.625359$, $C_x = 0.8009741$, $C_r = 0.57327$, $\rho_{yx} = 0.7779$, $\rho_{yr_x} = 0.7588$, $\rho_{xr_x} = 0.8497$.

Data set-5: [18]: Let y be the number of successful students and x be the number of teachers considered in a survey data of 923 districts of Turkey in 2007. The description of the parameters for these data is: $N = 261$, $n = 90$, $r = 70$, $\bar{Y} = 222.5824$, $\bar{X} = 306.4483$, $C_y = 1.8654$, $C_x = 1.7595$, $C_r = 0.57623$, $\rho_{yx} = 0.9705$, $\rho_{yr_x} = 0.6371$, $\rho_{xr_x} = 0.6265$.

Data set-6: [20,p-1111]: Let y be the amount (in \$000) of real estate farm loans and x be the amount (in \$000) of non-real estate farm loans in different states of USA during 1997. The details of the parameters for this data set are: $N = 50$, $n = 20$, $r = 8$, $\bar{Y} = 878.1626$, $\bar{X} = 555.4345$, $C_y = 1.235167$, $C_x = 1.052916$, $C_r = 0.571662$, $\rho_{yx} = 0.8038$, $\rho_{yr_x} = 0.7461$, $\rho_{xr_x} = 0.9236$.

Data set-7: [7,p-182]: Let y be the number of paralytic polio cases in the placebo group and x be the number of placebo children. The details of the parameters for this data set are: $N = 34$, $n = 12$, $r = 8$, $\bar{Y} = 2.588235$, $\bar{X} = 4.923529$, $C_y = 1.233278$, $C_x = 1.023331$, $C_r = 0.5687383$, $\rho_{yx} = 0.7328235$, $\rho_{yr_x} = 0.6571887$, $\rho_{xr_x} = 0.8165117$.

Data set-8: [7,p-152]: The data show the number of inhabitants in LARGE UNITED STATES CITIES (in 1000's). Let y be the number of inhabitants in 1930 and x be the number of inhabitants in 1920. The details of the parameters for this data set are: $N = 49$, $n = 15$, $r = 12$, $\bar{Y} = 127.7959$, $\bar{X} = 103.1429$, $C_y = 0.9634205$, $C_x = 1.012237$, $C_r = 0.5714601$, $\rho_{yx} = 0.981742$, $\rho_{yr_x} = 0.7207159$, $\rho_{xr_x} = 0.7915108$.

Data set-9: [7,p-34]: We investigate food cost of family for y and the family size for x . The values of the population parameters are: $N = 33$, $n = 10$, $r = 8$, $\bar{Y} = 27.49091$, $\bar{X} = 3.727273$, $C_y = 0.3685139$, $C_x = 0.4094911$, $C_r = 0.555573$, $\rho_{yx} = 0.432738$, $\rho_{yr_x} = 0.4495658$, $\rho_{xr_x} = 0.9820251$.

Data set-10: [14,p-399]: Consider y as area under wheat in 1964 and x as the area under wheat in 1963. The statistical summary of the population is: $N = 34$, $n = 11$, $r = 8$, $\bar{Y} = 199.4412$, $\bar{X} = 208.8824$, $C_y = 0.7531797$, $C_x = 0.7205298$, $C_r = 0.5689992$, $\rho_{yx} = 0.9800867$, $\rho_{yr_x} = 0.9152007$, $\rho_{xr_x} = 0.9416689$.

We have computed the MSEs of the estimators \bar{y}_r , t_{R_i} ($i = 1, 2, 3$), t_{G_i} , t_{KC_i} , t_{DP_i} , t_{BP_i} , $t_{BP_i}^*$ and the different members of t_{P_i} at their optimum situations based on their theoretical results, are given in Tables 2 and 3. The relative performance of all the above estimators is computed in terms of percentage relative efficiency (PRE) with respect to mean estimator \bar{y}_r . To calculate the PREs of the estimators \bar{y}_r , t_{R_i} , t_{G_i} , t_{KC_i} , t_{DP_i} , t_{BP_i} , $t_{BP_i}^*$, t_{d_i} , t_{D_i} and t_{P_i} with respect to \bar{y}_r , we have used the formula, given by

$$PRE(t., \bar{y}_r) = \frac{V(\bar{y}_r)}{MSE(t.)} \times 100 \quad (6.1)$$

The results are shown in Table 4.

Table 2 MSEs of the different estimators

Estimators	Populations									
	1	2	3	4	5	6	7	8	9	10
\bar{y}_r	60819.64	6.966	44.556	19725.28	1802.178	123535.20	0.973	953.872	9.719	2156.906
Strategy I										
t_{R_1}	28726.30	22.940	60.029	14987.171	621.159	100812.30	0.684	281.351	11.672	824.724
t_{G_1}	26315.88	6.269	44.554	11586.028	620.225	100728.89	0.679	278.019	8.379	823.962
$t_{K C_1}$	81469.41	30.457	59.676	33649.299	1736.665	126377.20	1.057	1052.107	17.212	2093.921
$t_{D P_1}$	26315.88	6.269	44.554	11586.028	620.225	100728.89	0.679	278.019	8.379	823.962
$t_{B P_1}$	26243.66	6.252	43.824	11448.280	612.557	89091.86	0.616	273.366	8.288	807.240
$t_{B^* P_1}^*$	26216.56	6.241	43.827	11379.738	612.957	88950.60	0.613	272.553	8.263	807.868
t_{d_1}	23957.61	6.121	43.911	11123.468	618.481	100725.66	0.673	272.060	8.258	823.232
t_{D_1}	23897.74	6.105	43.202	10996.439	610.855	89089.34	0.612	267.602	8.169	806.540
$t_{P_1}^{(1)*} (sav)$	23897.40	6.104	43.198	10993.495	610.682	88736.66	0.608	267.384	8.168	806.001
Strategy II										
t_{R_2}	16691.30	35.718	72.922	12776.055	106.129	44005.01	0.460	39.046	12.373	86.232
t_{G_2}	13376.97	5.711	44.552	7787.712	104.787	43713.06	0.451	34.514	7.899	85.047
$t_{K C_2}$	89213.07	49.249	72.277	40147.177	1708.096	133482.14	1.121	1087.501	19.900	2059.006
$t_{D P_2}$	13376.97	5.711	44.552	7787.712	104.787	43713.06	0.451	34.514	7.899	85.047
$t_{B P_2}$	13358.28	5.697	43.823	7725.233	104.566	41368.14	0.422	34.441	7.817	84.866
$t_{B^* P_2}^*$	13336.72	5.677	43.828	7649.090	104.655	41117.40	0.418	34.242	7.786	84.956
t_{d_2}	10134.35	5.445	43.373	7109.291	102.282	43701.77	0.441	26.406	7.734	83.913
t_{D_2}	10123.62	5.431	42.681	7057.187	102.072	41358.04	0.414	26.364	7.655	83.737
$t_{P_2}^{(1)*} (sav)$	10123.54	5.430	42.674	7055.389	102.065	41067.88	0.411	26.361	7.654	83.727

Table 2 continued

Estimators	Populations									
	1	2	3	4	5	6	7	8	9	10
Strategy III										
t_{R_3}	48784.64	19.745	57.450	17514.160	1287.147	66727.93	0.750	711.567	10.420	1418.414
t_{G_3}	47880.73	6.409	44.554	15926.960	1286.740	66519.39	0.746	710.366	9.239	1417.991
t_{KC_3}	68563.30	25.759	57.156	26223.154	1773.609	130640.17	1.038	989.266	12.407	2121.991
t_{DP_3}	47880.73	6.409	44.554	15926.960	1286.740	66519.39	0.746	710.366	9.239	1417.991
t_{BP_3}	47642.19	6.391	43.825	15667.811	1254.167	61237.20	0.671	680.756	9.127	1369.182
$t_{BP_3}^*$	47624.94	6.382	43.827	15627.072	1254.523	60982.63	0.668	680.070	9.593	1369.767
t_{d_3}	46996.38	6.290	44.018	15711.099	1285.979	66511.33	0.742	708.219	9.195	1417.587
t_{D_3}	46766.55	6.273	43.306	15458.871	1253.444	61230.37	0.668	678.784	9.084	1368.805
$t_{P_3}^{(1)*} (sav)$	46766.06	6.270	43.303	15456.189	1253.129	60792.63	0.665	678.294	9.081	1367.957

Table 3 *MSEs* of the different estimators belong to proposed family of estimators t_{P_1} , t_{P_2} and t_{P_2}

Estimators	Populations									
	1	2	3	4	5	6	7	8	9	10
Strategy I										
$t_{P_1}^{(1)*} (say)$	23897.40	6.104	43.198	10993.495	610.682	88736.66	0.608	267.384	8.168	806.001
$t_{P_1}^{(2)}$	23897.40	6.104	43.198	10993.495	610.682	88736.69	0.608	267.384	8.163	806.001
$t_{P_1}^{(3)}$	23897.40	6.104	43.198	10993.495	610.682	88736.70	0.605	267.395	8.162	806.001
$t_{P_1}^{(4)}$	23897.40	6.104	43.198	10993.495	610.682	88736.60	0.606	267.394	8.163	806.001
$t_{P_1}^{(5)}$	23897.40	6.104	43.198	10993.495	610.682	88736.62	0.607	267.395	8.163	806.001
$t_{P_1}^{(6)}$	23897.40	6.104	43.198	10993.495	610.682	88736.63	0.608	267.365	8.164	806.001
$t_{P_1}^{(7)}$	23897.40	6.104	43.198	10993.495	610.682	88736.69	0.609	267.360	8.165	806.001
$t_{P_1}^{(8)}$	23897.40	6.104	43.198	10993.495	610.682	88736.72	0.608	267.361	8.166	806.001
$t_{P_1}^{(9)}$	23897.40	6.104	43.198	10993.495	610.682	88736.75	0.609	267.369	8.167	806.001
$t_{P_1}^{(10)}$	23897.40	6.104	43.198	10993.495	610.682	88736.73	0.608	267.366	8.166	806.001

Table 3 continued

Estimators	Populations									
	1	2	3	4	5	6	7	8	9	10
Strategy II										
$t_{P_2}^{(1)} = t_{P_2}^*(s\alpha\gamma)$	10123.54	5.431	42.674	7055.389	102.065	41067.88	0.411	26.361	7.654	83.727
$t_{P_2}^{(2)}$	10123.54	5.431	42.674	7055.389	102.065	41067.82	0.412	26.331	7.652	83.727
$t_{P_2}^{(3)}$	10123.54	5.431	42.674	7055.389	102.065	41067.85	0.413	26.331	7.642	83.727
$t_{P_2}^{(4)}$	10123.54	5.431	42.674	7055.389	102.065	41067.85	0.412	26.332	7.643	83.727
$t_{P_2}^{(5)}$	10123.54	5.431	42.674	7055.389	102.065	41067.84	0.413	26.333	7.642	83.727
$t_{P_2}^{(6)}$	10123.54	5.431	42.674	7055.389	102.065	41067.83	0.412	26.324	7.643	83.727
$t_{P_2}^{(7)}$	10123.54	5.431	42.674	7055.389	102.065	41067.96	0.412	26.325	7.645	83.727
$t_{P_2}^{(8)}$	10123.54	5.431	42.674	7055.389	102.065	41067.86	0.413	26.331	7.647	83.727
$t_{P_2}^{(9)}$	10123.54	5.431	42.674	7055.389	102.065	41067.82	0.414	26.332	7.645	83.727
$t_{P_2}^{(10)}$	10123.54	5.431	42.674	7055.389	102.065	41067.75	0.415	26.334	7.646	83.727

Table 3 continued

Estimators	Populations									
	1	2	3	4	5	6	7	8	9	10
Strategy III										
$t_{P_3}^{(1)} = t_{P_3}^*(s\alpha\gamma)$	46766.06	6.273	43.303	15456.189	1253.129	60792.63	0.665	678.294	9.084	1367.957
$t_{P_3}^{(2)}$	46766.06	6.273	43.303	15456.189	1253.129	60792.67	0.655	678.284	9.085	1367.957
$t_{P_3}^{(3)}$	46766.06	6.273	43.303	15456.189	1253.129	60792.66	0.654	678.265	9.086	1367.957
$t_{P_3}^{(4)}$	46766.06	6.273	43.303	15456.189	1253.129	60792.66	0.655	678.266	9.085	1367.957
$t_{P_3}^{(5)}$	46766.06	6.273	43.303	15456.189	1253.129	60792.67	0.654	678.265	9.086	1367.957
$t_{P_3}^{(6)}$	46766.06	6.273	43.303	15456.189	1253.129	60792.66	0.653	678.269	9.085	1367.957
$t_{P_3}^{(7)}$	46766.06	6.273	43.303	15456.189	1253.129	60792.89	0.654	678.275	9.086	1367.957
$t_{P_3}^{(8)}$	46766.06	6.273	43.303	15456.189	1253.129	60792.70	0.655	678.281	9.088	1367.957
$t_{P_3}^{(9)}$	46766.06	6.273	43.303	15456.189	1253.129	60792.73	0.654	678.288	9.089	1367.957
$t_{P_3}^{(10)}$	46766.06	6.273	43.303	15456.189	1253.129	60792.71	0.653	678.285	9.087	1367.957

Table 4 P REs of the different estimators with respect to mean estimator \bar{y}_r

Estimators	Populations									
	1	2	3	4	5	6	7	8	9	10
\bar{y}_r	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Strategy I										
t_{R1}	211.72	30.37	74.22	131.61	290.13	122.54	142.38	339.03	83.27	261.53
t_{G1}	231.11	111.13	100.00	170.25	290.57	122.64	143.46	343.10	115.99	261.77
t_{KC1}	74.65	22.87	74.66	58.62	103.77	97.75	92.13	90.66	56.47	103.01
t_{DP1}	231.11	111.13	100.00	170.25	290.57	122.64	143.46	343.10	115.99	261.77
t_{BP1}	231.75	111.44	101.67	172.30	294.21	138.66	158.00	348.94	117.27	267.20
t_{BP1}^*	231.99	111.63	101.66	173.34	294.01	138.88	158.85	349.98	117.61	266.99
t_{d1}	253.86	113.82	101.47	177.33	291.39	122.65	144.68	350.61	117.69	262.00
t_{D1}	254.50	114.13	103.13	179.38	295.03	138.66	159.21	356.45	118.98	267.43
t_{P1}^* or t_{P1}	254.51	114.15	103.14	179.43	295.11	139.22	160.08	356.74	118.99	267.61
Strategy II										
t_{R2}	364.38	19.51	61.10	154.39	1698.11	280.73	211.70	2442.96	78.55	2501.29
t_{G2}	454.66	121.98	100.01	253.29	1719.84	282.60	216.00	2763.75	123.046	2536.13
t_{KC2}	68.17	14.15	61.65	49.13	105.51	92.55	86.84	87.71	48.84	104.75
t_{DP2}	454.66	121.98	100.01	253.29	1719.84	282.60	216.00	2763.75	123.04	2536.13
t_{BP2}	455.30	122.29	101.67	255.34	1723.48	298.62	230.54	2769.59	124.33	2541.55
t_{BP2}^*	456.03	122.72	101.66	257.88	1722.01	300.45	232.86	2785.66	124.83	2538.86
t_{d2}	600.13	127.96	102.73	277.46	1761.96	282.68	220.95	3612.27	125.67	2570.40
t_{D2}	600.74	128.22	104.39	279.51	1765.60	298.70	235.49	3618.11	126.95	2575.83
t_{P2}^* or t_{P2}	600.77	128.28	104.41	279.58	1765.72	300.81	237.11	3618.51	126.98	2576.11

Table 4 continued

Estimators	Populations									
	1	2	3	4	5	6	7	8	9	10
Strategy III										
t_{R_3}	124.67	35.28	77.56	112.62	140.01	185.13	129.87	134.05	93.28	152.06
t_{G_3}	127.02	108.71	100.00	123.85	140.06	185.71	130.56	134.28	105.20	152.11
t_{KC_3}	88.71	27.05	77.95	75.22	101.61	94.56	93.81	96.42	78.34	101.65
t_{DP_3}	127.02	108.71	100.00	123.85	140.06	185.71	130.56	134.28	105.20	152.11
t_{BP_3}	127.66	109.02	101.67	125.90	143.70	201.73	145.10	140.12	106.49	157.53
$t_{BP_3}^*$	127.71	109.16	101.66	126.23	143.65	202.57	145.70	140.26	101.31	157.47
t_{d_3}	129.41	110.76	101.22	125.55	140.14	185.74	131.34	134.69	105.70	152.15
t_{D_3}	130.05	111.07	102.87	127.60	143.78	201.75	145.88	140.53	106.99	157.58
$t_{P_3}^*$ or t_{P_3}	130.06	111.09	102.89	127.62	143.81	203.21	146.54	140.63	106.99	157.67

Interpretation of the results:

In Table 2, we see that the lowest amount of MSE values occurred for the members $t_{P_i}^*$ ($i = 1, 2, 3$) of suggested class of estimators t_{P_i} , respectively, in parallel comparison with the estimators t_{R_i} , t_{G_i} , t_{KC_i} , t_{DP_i} , t_{BP_i} , $t_{BP_i}^*$, t_{d_i} and t_{D_i} as well as \bar{y}_r . Further, we see that the special members t_{D_i} ($i = 1, 2, 3$) of proposed estimators t_{P_i} , respectively, have the second lowest amount of MSE values in the parallel comparisons of other existing estimators. In Table 3, we observe that the MSE values of all the discussed members of the proposed estimators in respective strategies are same in the Populations 1, 2, 3, 4, 5 and 10 but a tiny bit change in the Populations 6, 7, 8 and 9 which can also be admitted as negligible difference. Thus, it can be argued that the MSEs of all the members of proposed class of estimators (in Table 3) are equal and smaller than all other discussed estimators in the respective strategies. Subsequently, their PREs with respect to \bar{y}_r will also be same. Therefore, we have considered the notation t_{P_i} ($i = 1, 2, 3$) only, for all the members $t_{P_i}^{(j)}$ ($i = 1, 2, 3; j = 1, 2, \dots, 10$) in Table 4 for our convenience which present the PREs of all the members of proposed classes of estimators in the respective strategies.

From Table 4, we report that

- (i) The performance of ratio estimators t_{R_i} ($i = 1, 2, 3$) is good in Populations 1, 4, 5, 6, 7, 8 and 10, while in Populations 2, 3 and 9 it is poor. Note that in the populations 2, 3 and 9, the values of $\rho_{yx} \frac{C_y}{C_x}$ are 0.16981, -0.01157 and 0.38943, respectively, which are not satisfying the condition $(\rho C_y / C_x) > 1/2$ for t_{R_i} to overcome the mean estimator \bar{y}_r .
- (ii) Theoretically it has been stated above that the performances of the estimators t_{G_i} ($i = 1, 2, 3$) are always better than \bar{y}_r , t_{R_i} and t_{KC_i} which confirmed by this empirical study.
- (iii) The estimators t_{KC_i} ($i = 1, 2, 3$) perform good only in Populations 5 and 10 where the numerical values of $\rho_{yx} \frac{C_y}{C_x}$ are 1.0289 and 1.02449, respectively. Note that the condition $(\rho C_y / C_x) > 1$ is satisfied for the said populations while for the remaining populations it does not holds.
- (iv) We see that t_{DP_i} ($i = 1, 2, 3$) and t_{G_i} are equally efficient and t_{BP_i} are paralleling the improvement over t_{DP_i} in all the populations.
- (v) The performance of the estimators $t_{BP_i}^*$ ($i = 1, 2, 3$) in parallel is very near to t_{BP_i} .
- (vi) We see that the proposed estimator t_{d_i} ($i = 1, 2, 3$) are always better than t_{DP_i} and t_{BP_i} , respectively, in all the populations. Therefore, it can be argued that the regression type estimator based on dual use of auxiliary information always outperform both the conventional regression and difference type estimators which are based on only the prime information of an auxiliary variable. Hence, a better overcome on missing data problem can be attained just by using dual of an auxiliary variable.
- (vii) The proposed estimators t_{D_i} ($i = 1, 2, 3$) are paralleling the improvement over t_{d_i} .
- (viii) The performance of the proposed class of estimators t_{P_i} ($i = 1, 2, 3$) is:
 - (a) Good (better than \bar{y}_r) in all the Populations 1-10.

- (b) Paralleling more efficient than the estimators $t_{R_i}, t_{G_i}, t_{KC_i}, t_{DP_i}, t_{BP_i}, t_{B^*P_i}$ and their special members t_{d_i} ($i = 1, 2, 3$) and t_{D_i} . Thus, t_{P_i} paralleling accomplish the maximum gain in efficiency among all the other estimators in all the populations considered in this empirical study.
- (c) We observe that the second one t_{P_2} is always better than first one t_{P_1} and third one t_{P_3} .
- (c) We also observe that the PREs of first proposed estimator t_{P_1} are greater than the third proposed estimator t_{P_3} in all the populations where the condition (5.19) holding.

Thus, the suggested class of estimators t_{P_i} ($i = 1, 2, 3$) in parallel outperform all other estimators $t_{R_i}, t_{G_i}, t_{KC_i}, t_{DP_i}, t_{BP_i}, t_{B^*P_i}, t_{d_i}$ and t_{D_i} considered in this study. We see that the second proposed estimator t_{P_2} is always better than the first and third proposed estimators t_{P_1} and t_{P_3} . We also see that the proposed estimator t_{P_2} is the most efficient among all the other estimators discussed in this study.

On the basis of this empirical study, we conclude that the proposed estimators formulated under double(rank) use of an auxiliary variable are capable to enhance their precision over relevant estimators based on single/prime use of an auxiliary variable.

7 Conclusions

To exercise the problem of missing data efficiently, there are several notable researchers, but no one has discussed the imputation technique using dual of auxiliary information in literature so far. In the present study, we have suggested three imputation techniques and corresponding estimators using an auxiliary information as well as its rank in three different sampling strategies (based on different amounts of auxiliary information) under non-response. In the empirical study consisting 10 real data sets, it has been found that the performance of the suggested set of estimators paralleling are more efficient than usual mean estimator and the works Lee et al. [12], Kadilar and Cingi [11], Gira [9], Diana and Perri [8], Bhusan and Pandey [4] and Bhusan et al. (2018). The present study is important in survey sampling to estimate the population mean because it overcomes to the missing data more effectively than the works (based on prime/single use of an auxiliary information) discussed above, just by using the dual(rank) of an auxiliary information. Since, for the optimal solution, both from a theoretical and practical perspectives, the proposed estimators are very simple to apply. Hence, the suggested classes of estimators are appreciable and recommended to use for sampling practitioners when the non-response cannot be ignored.

Acknowledgements Authors are highly thankful to the Indian Institute of Technology (Indian School of Mines), Dhanbad, for providing financial assistance and necessary infrastructure to accomplish the present research work.

Declarations

Conflict of interest There is no conflict of interest associated with the present article.

References

1. Al-Omari AI, Bouza CN, Herrera C (2013) Imputation methods of missing data for estimating the population mean using simple random sampling with known correlation coefficient. *Qual Quant* 47(1):353–365
2. Bouza-Herrera CN, Viada CE (2021) Imputation of individual values of a variable using product predictors. *Revista Investig Oper* 41(7):979–989
3. Bouza CN, Viada C, Vishwakarma GK (2020) Studying the total under missingness by guessing the value of a superpopulation model for imputation. *Revista Investig Oper* 42(3):321–325
4. Bhushan S, Pandey AP (2016) Optimal imputation of missing data for estimation of population mean. *J Stat Manage Syst* 19(6):755–769
5. Bhushan S, Pandey AP (2018) Optimality of ratio type estimation methods for population mean in the presence of missing data. *Commun Stat Theory Methods* 47(11):2576–2589
6. Bhushan S, Pratap, Pandey A, Pandey A (2020) On optimality of imputation methods for estimation of population mean using higher order moment of an auxiliary variable. *Commun Stat Simul Comput* 49(6):1560–1574
7. Cochran WG (1977) *Sampling techniques*, 3rd edn. Wiley, New York
8. Diana G, Peri PF (2010) Improved estimators of the population mean for missing data. *Commun Stat Theory Methods* 39:3245–3251
9. Gira AA (2015) Estimation of population mean with a new imputation method. *Appl Math Sci* 9(34):1663–1672
10. Heitzin FD, Basu S (1996) Distinguishing missing at random and missing completely at random. *Am Stat* 50(3):207–213
11. Kadilar C, Cingi H (2008) Estimators for the population mean in the case of missing data. *Commun Stat Theory Methods* 37(14):2226–2236
12. Lee H, Rancourt E, Sarndal CE (1994) Experiments with variance estimation from survey data with imputed values. *J Official Stat* 10(3):231–243
13. Maddala GS (1992) *Introduction to econometrics*, 2nd edn. Macmillan, New York
14. Murthy MN (1967) *Sampling theory and methods*. Statistical Publishing Society Calcutta (India). Vol. I and, 2, 1-1220
15. Prasad S (2017) Ratio exponential type estimators with imputation for missing data in sample surveys. *Model Assist Stat Appl* 12(2):95–106
16. Prasad S (2017) An exponential imputation in the case of missing data. *J Stat Manage Syst* 20(6):1127–1140. <https://doi.org/10.1080/09720510.2017.1407515>
17. Rubin DB (1976) Inference and missing data. *Biometrika* 63:581–593
18. Satıcı E, KAdilar, C. (2011) Ratio estimator for the population mean at the current occasion in the presence of non-response in successive sampling. *Haceteppe J Math Stat* 40(1):115–124
19. Singh D, Chaudhary FS (1986) *Theory and analysis of sample survey designs*. Wiley, New York
20. Singh S (2003) *Advanced sampling theory with applications*, vol 2. Springer, Berlin
21. Singh S (2009) A new method of imputation in survey sampling. *Statistics* 43(5):499–511
22. Singh S, Deo B (2003) Imputation by power transformation. *Stat Papers* 44(4):555–579
23. Singh S, Horn S (2000) Compromised imputation in survey sampling. *Metrika* 51(3):267–276
24. Singh GN, Suman S (2019) Estimation of population mean using imputation methods for missing data under two-phase sampling design. *J Stat Theory Pract.* <https://doi.org/10.1007/s42519-018-0016-5>
25. Singh GN, Pandey AK, Sharma AK (2020) Some improved and alternative imputation methods for finite population mean in presence of missing information. *Commun Stat Theory Methods.* <https://doi.org/10.1080/03610926.2020.1713375>
26. Singh GN, Usman M (2021) Some efficient estimators in case of missing data. *Commun Stat Simul Comput* 1-27

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.