



OPEN

# A computational strategy for estimation of mean using optimal imputation in presence of missing observation

Subhash Kumar Yadav<sup>1</sup>, Gajendra K. Vishwakarma<sup>2✉</sup> & Dinesh K. Sharma<sup>3</sup>

In this study, we suggest an optimal imputation strategy for the elevated estimation of the population mean of the primary variable utilizing the known auxiliary parameters for the missing observations. Under this strategy, we suggest a new modified Searls type estimator, and we study its sampling properties, mainly bias and mean squared error (MSE), for an approximation of order one. The introduced estimator is compared theoretically with the estimators of population mean in competition under the imputation method. The efficiency conditions for the introduced estimator to be more efficient than the estimators in the competition are derived. To be sure about the efficiencies, these efficiency conditions are verified through the three natural populations. We have also conducted a simulation study and generated an artificial population with the same parameters as a natural population. The estimator with minimum MSE and the highest Percentage Relative Efficiency (PRE) is recommended for practical use in different areas of applications.

**Keywords** Estimation, Imputation, Missing data

The arithmetic mean is the most accurate way to measure central tendency when the population being studied is homogeneous for the characteristic being studied. Because of time and financial restrictions, estimating the arithmetic mean is crucial for large populations for determining various policy decisions and other uses. These days, missing data or non-response is a prevalent and unavoidable problem and is a prevalent problem with data received from sampling surveys. These missing numbers make data analysis, processing, and handling more complex. Missing data is a concern in clinical or life-saving drug testing trials since some experimental units have to be eliminated during the experiment. Similar to this, during agricultural experiments, crops are destroyed by disease or other natural disasters. For a variety of reasons, responses from every unit in the sample are typically not available in demographic and socioeconomic surveys. This kind of incompleteness is referred to as non-response, and judgments about the population parameters may be tainted if the necessary information regarding the nature of non-response is not present.

For a long time, sample survey professionals have known that failing to account for the random character of incompleteness or non-response might degrade the data quality. In surveys, there are two forms of non-responses: unit non-response and item non-response. When an eligible sample unit is entirely absent from the survey, unit non-response occurs. Missing values can be replaced by imputations and then treated as any other auxiliary variable. However, this depends on the way the imputations are derived. It is well known that in sampling designs the use of auxiliary information improves the precision of an estimator substantially for instance see Vishwakarma and Kumar<sup>1</sup> and Kumar and Vishwakarma<sup>2</sup> and Vishwakarma et al.<sup>3</sup>. In contrast, item non-response occurs when a sampled unit is present in the survey but fails to provide information about a component of a unit in the sample survey. Missing data is an issue in such cases. Numerous studies adopted the imputation method to address this issue, which involves substituting values for missing data. It is a highly suggested method for resolving non-response issues in sample surveys, Singh et al.<sup>4</sup>.

Rubin<sup>5</sup> suggested three methods for missing observations in survey sampling, namely, Missing At Random (MAR), Observed At Random (OAR), and Parameter Distinctness (PD). The difference between MAR and Missing Completely at Random (MCAR) was discussed by Heitjan and Basu<sup>6</sup>. Various authors, including, Singh and

<sup>1</sup>Department of Statistics, Babasaheb Bhimrao Ambedkar University, Lucknow 226025, India. <sup>2</sup>Department of Mathematics and Computing, Indian Institute of Technology (ISM) Dhanbad, Dhanbad 826004, India. <sup>3</sup>Department of Business, Management and Accounting, University of Maryland Eastern Shore, Princess Anne, MD 21853, USA. ✉email: vishwagk@rediffmail.com; vishwagk@iitism.ac.in

Vishwakarma<sup>7</sup>, Kadilar and Cingi<sup>8</sup>, Diana and Perri<sup>9</sup>, Gira<sup>10</sup>, Bhushan and Pandey<sup>11</sup>, Prasad<sup>12</sup>, Audu et al.<sup>13</sup>, Audu et al.<sup>14</sup>, Singh et al.<sup>4</sup> and others worked on the MCAR technique and suggested different imputation methods for efficient estimation of population mean of the main variable under the situation of missing observations. The theory of sample surveys was expanded by Shahzad et al.<sup>15</sup> to include imputation-based mean estimators in the event of missing data, using variance–covariance matrices and robust regression. Shahzad et al.<sup>16</sup> proposed estimators of the quantile regression-ratio type for estimating means under both partial and complete auxiliary data. In order to achieve enhanced population mean estimate, Alomair and Shahzad<sup>17</sup> worked on compromised imputation and EWMA based memory-type mean estimators employing quantile regression. Lawson<sup>18</sup> presented a novel approach to imputation for estimating population mean in survey sampling when there are missing data. Robust quantile regression was used by Anas et al.<sup>19</sup> to develop compromised imputation-based mean estimators for better population mean estimation.

Let  $U = (U_1, U_2, \dots, U_N)$  be the finite population under consideration consisting of  $N$  distinct and identifiable units. Further, let  $Y$  be the main characteristic under investigation with a population mean as  $\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$  and

population variance as  $\sigma^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2$ . To estimate  $\bar{Y}$ , we draw a random sample  $s$  of size  $n$  using Simple Random Sampling Without Replacement (SRSWOR). Let  $r$  be the responding units belongs to  $R$ , the set of all responding units and  $n - r$  be the non-responding units belongs to  $R^c$ , the set of all non-responding or missing units out of  $n$  sampled units from the above population of size  $N$ . For every  $i$ th unit belonging to  $R$  that is  $i \in R$ , the corresponding value of  $y_i$  is observed, while for  $i \in R^c$ , the value of  $y_i$  is missing and is estimated through different imputation methods. For elevated estimation of  $\bar{Y}$  under imputation methods, many authors utilized the known population mean ( $\bar{X}$ ) of the auxiliary variable  $x$ . Let  $x_i$  be the observation for the  $i$ th unit of  $x$  and is positive for all  $i \in s$ . Now let  $y_{.i}$  be the observation on  $Y$  such that:

$$y_{.i} = \begin{cases} y_i & \text{if } i \in R \\ \tilde{y}_i & \text{if } i \in R^c \end{cases}$$

where,  $\tilde{y}_i$  is the imputed value for the  $i$ th non-responding unit and by the utilization of above data, the point estimator of  $\bar{Y}$  under the imputation method is given by:

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

In the present study, we also suggested a new imputation method for elevated estimation of  $\bar{Y}$  using the known auxiliary parameters under MCAR mechanism. We study the large sampling properties of the suggested estimator for the first order of approximation. The conditions of efficiency for the proposed estimator over the estimators in competition are derived and are verified using the three natural populations along with one simulated population. The most efficient estimator is recommended for practical use in different areas of applications.

## Review of imputation estimators

The most appropriate imputation estimator for estimating  $\bar{Y}$  is obtained through the following imputation method as:

$$y_{.i} = \begin{cases} y_i & \text{if } i \in R \\ \tilde{y}_i & \text{if } i \in R^c \end{cases}$$

and the resultant point estimator of  $\bar{Y}$  is,

$$t_0 = \frac{1}{n} \left[ \sum_{i \in R} y_i + \sum_{i \in R^c} \tilde{y}_i \right] = \frac{1}{r} \sum_{i=1}^r y_i = \bar{y}_r.$$

It is unbiased for  $\bar{Y}$  and its variance for an approximation of order one is,

$$V(t_0) = \theta_{r,N} \bar{Y}^2 C_y^2 \quad (1)$$

where,  $\theta_{r,N} = \left( \frac{1}{r} - \frac{1}{N} \right) S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})^2$ ,  $C_y = \frac{S_y}{\bar{Y}}$ .

The estimator, making the use of auxiliary parameters is the ratio estimator and under the missing observation technique is given by the following imputation method as,

$$y_{.i} = \begin{cases} y_i & \text{if } i \in R \\ \hat{\beta} \tilde{y}_i & \text{if } i \in R^c. \end{cases}$$

where,  $\hat{\beta} = \frac{\sum_{i=1}^r y_i}{\sum_{i=1}^r x_i}$ .

The resulting estimator of  $\bar{Y}$  is presented by,

$$t_r = \frac{\bar{y}_r \bar{x}_n}{\bar{x}_r}$$

where,  $\bar{x}_r = \frac{1}{r} \sum_{i \in R} x_i$  and  $\bar{x}_n = \frac{1}{r} \sum_{i \in S} x_i$ .

The bias and MSE of  $t_r$  for an approximation of order one respectively are,

$$\begin{aligned} B(t_r) &= \theta_{r,n} \bar{Y} (C_x^2 - C_{yx}), \\ MSE(t_r) &= \theta_{r,n} \bar{Y}^2 [C_y^2 + C_x^2 - 2C_{yx}] \end{aligned} \tag{2}$$

where,  $C_x = \frac{S_x}{\bar{X}}$ ,  $S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2$ ,  $S_{yx} = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})(x_i - \bar{X})$ ,  $\rho = \frac{S_{yx}}{S_y S_x}$ ,  $C_{yx} = \rho C_y C_x$ ,  $\theta_{r,n} = (\frac{1}{r} - \frac{1}{n})$ ,  $\bar{X} = \sum_{i=1}^N x_i$ .

Singh and Horn<sup>20</sup> defined the following study variable after utilizing the imputed observation for the responding and non-responding units under consideration as,

$$y_i = \begin{cases} \lambda \frac{n}{r} y_i + (1 - \lambda) \hat{\beta} x_i & \text{if } i \in R \\ (1 - \lambda) \hat{\beta} x_i & \text{if } i \in R^c. \end{cases}$$

Under the imputation method, the resultant estimator of  $\bar{Y}$  is as follows,

$$t_1 = \bar{y}_r \left( \lambda + (1 - \lambda) \frac{\bar{x}_n}{\bar{x}_r} \right)$$

where,  $\lambda$  is a characterizing constant to be obtained so that  $MSE(t_1)$  is least.

The bias and MSE of  $t_1$  for an approximation of degree one are respectively given by,

$$\begin{aligned} B(t_1) &= (1 - \lambda) \theta_{r,n} \bar{Y} (C_x^2 - C_{yx}), \\ MSE(t_1) &= \theta_{r,n} \bar{Y}^2 C_y^2 + \theta_{r,n} \bar{Y}^2 [(1 - \lambda)^2 C_x^2 - 2(1 - \lambda) C_{yx}] \end{aligned}$$

The optimum value of  $\lambda$  is given by,

$$\lambda_{opt} = 1 - \frac{C_{yx}}{C_x^2}.$$

The minimum MSE of  $t_1$  for  $\lambda_{opt}$  is,

$$MSE_{min}(t_1) = \bar{Y}^2 C_y^2 [\theta_{r,n} - \theta_{r,n} \rho^2]. \tag{3}$$

Singh and Deo<sup>21</sup> utilized the transformed auxiliary information and defined the following variable after using the imputed observation for missing value as,

$$y_i = \begin{cases} y_i & \text{if } i \in R \\ \bar{y}_r \left[ n \left( \frac{\bar{x}_n}{\bar{x}_r} \right)^\beta - r \right] \frac{x_i}{\sum_{i \in R^c} x_i} & \text{if } i \in R^c. \end{cases}$$

The resultant estimator of  $\bar{Y}$  is given by,

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

where,  $\beta$  is a characterizing constant and is obtained so that  $MSE(t_1)$  is least.

The bias and the minimum MSE of  $t_2$  for an approximation of degree one respectively are,

$$B(t_2) = \theta_{r,n} \bar{Y} \left[ \frac{\beta(\beta - 1)}{2} C_x^2 - \beta C_{yx} \right].$$

The optimal value of  $\beta$  for which  $MSE(t_2)$  is least is given by,

$$\beta_{opt} = \rho \frac{C_y}{C_x}.$$

The least value of  $MSE(t_2)$  for  $\beta_{opt}$  is,

$$MSE_{\min}(t_2) = MSE_{\min}(t_1) - \theta_{r,n} S_x^2 \left( \frac{S_{yx}}{S_x^2} - \frac{\bar{Y}}{\bar{X}} \right)^2. \tag{4}$$

Kadilar and Cingi<sup>8</sup> suggested the following regression type estimators of  $\bar{Y}$  under the case of missing observations as,

$$t_3 = \left[ \bar{y}_r + \hat{\beta}(\bar{X} - \bar{x}_r) \right] \frac{\bar{X}}{\bar{x}_r}, \quad t_4 = \left[ \bar{y}_r + \hat{\beta}(\bar{X} - \bar{x}_n) \right] \frac{\bar{X}}{\bar{x}_n}, \quad t_5 = \left[ \bar{y}_r + \hat{\beta}(\bar{X} - \bar{x}_n) \right] \frac{\bar{x}_n}{\bar{x}_r}.$$

The biases and MSEs of  $t_i (i = 3, 4, 5)$  for an approximation of order are respectively given by,  $B(t_3) \cong \theta_{n,N} \bar{Y} C_x^2$ ,  $B(t_4) \cong \theta_{r,N} \bar{Y} C_x^2$ ,  $B(t_5) \cong \theta_{r,n} \bar{Y} C_{yx}$  and

$$MSE_{\min}(t_3) = MSE_{\min}(t_1) - \theta_{r,n} S_x^2 (R^2 - \beta^2), \tag{5}$$

$$MSE_{\min}(t_4) = MSE_{\min}(t_1) - \theta_{n,N} S_x^2 (R^2 - \beta^2), \tag{6}$$

$$MSE_{\min}(t_5) = MSE_{\min}(t_1) - \theta_{r,n} [(R + \beta)^2 S_x^2 - 2(R + \beta) S_{yx}], \tag{7}$$

where,  $R = \frac{\bar{Y}}{\bar{X}}$  and  $\beta = \frac{S_{yx}}{S_x^2}$ .

Singh<sup>22</sup> suggested a new modified imputation method for estimation of  $\bar{Y}$  and suggested the following variable as,

$$y_i = \begin{cases} y_i & \text{if } i \in R \\ \bar{y}_r \left[ \frac{(n-r)\bar{x}_n + r(\bar{x}_n - \bar{x}_r)}{\alpha \bar{x}_n - (1-\alpha)\bar{x}_r} \right] \frac{x_i}{\sum_{i \in R^c} x_i} & \text{if } i \in R^c. \end{cases}$$

The resultant estimator of  $\bar{Y}$  is given by,

$$t_6 = \frac{\bar{y}_r \bar{x}_n}{\alpha \bar{x}_r + (1-\alpha)\bar{x}_n}$$

where,  $\alpha$  is a scalar to be obtained so that  $MSE(t_6)$  is least.

The bias and MSE of  $t_6$  for an approximation of degree one respectively are,

$$B(t_6) = \bar{Y} [\theta_{n,N} C_{yx} + \alpha^2 \theta_{r,n} C_x^2 + (1-\alpha)^2 \theta_{n,N} C_x^2 - \alpha(\theta_{r,n} C_{yx} + \theta_{n,N} C_x^2) + 2\alpha(1-\alpha)\theta_{n,N} C_x^2 - (1-\alpha)\theta_{n,N} (C_{yx} + C_x^2)]$$

The optimal value of  $\alpha$  for which  $MSE(t_6)$  is least is given by  $\alpha_{opt} = \rho C_y / C_x$ . The least value of  $MSE(t_6)$  for  $\alpha_{opt}$  is,

$$MSE_{\min}(t_6) = MSE_{\min}(t_1) - \theta_{r,n} S_x^2 \left( \frac{S_{yx}}{S_x^2} - \frac{\bar{Y}}{\bar{X}} \right)^2. \tag{8}$$

Singh et al.<sup>23</sup> suggested a new imputation method using exponential function as,

$$y_i = \begin{cases} \alpha \frac{n}{r} y_i + (1-\alpha)\bar{y}_r \exp\left(\frac{\bar{X} - \bar{x}_r}{\bar{X} + \bar{x}_r}\right) & \text{if } i \in R \\ (1-\alpha)\bar{y}_r \exp\left(\frac{\bar{X} - \bar{x}_r}{\bar{X} + \bar{x}_r}\right) & \text{if } i \in R^c. \end{cases}$$

The resultant estimator of  $\bar{Y}$  under imputation method is given by,

$$t_7 = \alpha \bar{y}_r + (1-\alpha)\bar{y}_r \exp\left(\frac{\bar{X} - \bar{x}_r}{\bar{X} + \bar{x}_r}\right).$$

where,  $\alpha$  is a constant to be obtained so that  $MSE(t_7)$  is least.

The bias and MSE of  $t_7$  for an approximation of degree one respectively are,

$$B(t_7) = (1-\alpha)\theta_{r,N} \bar{Y} \left( \frac{3}{8} C_x^2 - \frac{1}{2} C_{yx} \right),$$

$$MSE(t_7) = \theta_{r,N} \bar{Y}^2 \left[ C_y^2 + \frac{(1-\alpha)^2}{4} C_x^2 - (1-\alpha) C_{yx} \right].$$

The optimal value of  $\alpha$  is given by,

$$\alpha_{opt} = 1 - 2 \frac{C_{yx}}{C_x^2}.$$

The least MSE of  $t_7$  for  $\alpha_{opt}$  is given by,

$$MSE_{min}(t_7) = MSE_{min}(t_1) - \theta_{r,n} S_x^2 \left( \frac{S_{yx}}{S_x^2} - \frac{\bar{Y}}{\bar{X}} \right)^2. \tag{9}$$

Gira<sup>10</sup> suggested a novel imputation method for the estimation of  $\bar{Y}$  as,

$$y_i = \begin{cases} y_i & \text{if } i \in R \\ \bar{y}_r \left[ n \frac{(\delta - \bar{x}_r)}{(\delta - \bar{x}_n)} - r \right] \frac{x_i}{\sum_{i \in R^c} x_i} & \text{if } i \in R^c. \end{cases}$$

The resultant estimator of  $\bar{Y}$  is given by,

$$t_8 = \bar{y}_r \left[ \frac{(\delta - \bar{x}_r)}{(\delta - \bar{x}_n)} \right].$$

where,  $\delta$  is the scalar to be obtained so that  $MSE(t_8)$  is least.

The bias of  $t_8$  is given by,

$$B(t_8) = -\theta_{r,n} \frac{\bar{X} \bar{Y}}{\delta - \bar{X}} C_{yx}.$$

The optimal value of  $\delta$  for which  $MSE(t_8)$  is least, is given by,

$$\delta_{opt} = \bar{X} \left( \frac{C_x}{\rho C_y} - 1 \right).$$

The least value of  $MSE(t_8)$  for  $\delta_{opt}$  is,

$$MSE_{min}(t_8) = V(t_0) - \theta_{r,n} \bar{Y}^2 \rho^2 C_y^2. \tag{10}$$

Singh et al.<sup>4</sup> worked on a new imputation method as,

$$y_i = \begin{cases} y_i & \text{if } i \in R \\ \bar{y}_r \left[ \frac{\{m(n+r) - r\} \bar{x}_n + \{(1-m)n - mr\} \bar{x}_r}{m \bar{x}_n - (1-m) \bar{x}_r} \right] \frac{x_i}{\sum_{i \in R^c} x_i} & \text{if } i \in R^c. \end{cases}$$

The resultant point estimator of  $\bar{Y}$  is given by,

$$t_9 = \bar{y}_r \left[ \frac{m \bar{x}_r + (1-m) \bar{x}_n}{m \bar{x}_n + (1-m) \bar{x}_r} \right].$$

The bias and MSE of  $t_9$  for an approximation of degree one respectively are,

$$B(t_9) = \theta_{r,n} \bar{Y} [(2m^2 - 3m + 1) C_x^2 + (2m - 1) C_{yx}],$$

$$MSE(t_9) = \bar{Y}^2 [\theta_{r,n} C_y^2 + (1 - 2m)^2 \theta_{r,n} C_x^2 - 2(1 - 2m) \theta_{r,n} C_{yx}].$$

The optimum value of  $m$  is given by,

$$m_{opt} = \frac{1}{2} \left( 1 - \rho \frac{C_y}{C_x} \right).$$

The least value of  $MSE(t_9)$  for  $m_{opt}$  is,

$$MSE_{min}(t_9) = V(t_0) - \theta_{r,n} \bar{Y}^2 \rho^2 C_y^2. \tag{11}$$

Aliyu et al.<sup>24</sup> suggested a new imputation method for the estimation of  $\bar{Y}$  as,

$$y_i = \begin{cases} y_i & \text{if } i \in R \\ \frac{1}{n-r} \left[ n \left\{ \alpha \bar{y}_r + (1-\alpha) \bar{y}_r \frac{\bar{x}_r}{\bar{X}} \right\} \exp \left( \frac{\bar{X} - \bar{x}_r}{\bar{X} + \bar{x}_r} \right) - r \bar{y}_r \right] & \text{if } i \in R^c. \end{cases}$$

The resultant estimator of  $\bar{Y}$  is,

$$t_{10} = \left\{ \alpha \bar{y}_r + (1 - \alpha) \bar{y}_r \frac{\bar{x}_r}{\bar{X}} \right\} \exp \left( \frac{\bar{X} - \bar{x}_r}{\bar{X} + \bar{x}_r} \right).$$

The bias and MSE of  $t_{10}$  for an approximation of degree one respectively are,

$$B(t_{10}) = \theta_{r,N} \bar{Y} \left[ - \left( \frac{1}{8} - \alpha \right) C_x^2 + \left( \frac{1}{2} + \alpha \right) C_{yx} \right].$$

$$MSE(t_{10}) = \theta_{r,N} \bar{Y}^2 \left[ C_y^2 + \left( \frac{1}{2} - \alpha \right)^2 C_x^2 + 2 \left( \frac{1}{2} - \alpha \right) C_{yx} \right].$$

The optimum value of  $\alpha$  is given by,

$$\alpha_{opt} = \frac{1}{2} + \rho \frac{C_y}{C_x}$$

The least value of  $MSE(t_{10})$  for  $\alpha_{opt}$  is,

$$MSE_{min}(t_{10}) = \theta_{r,N} \bar{Y}^2 C_y^2 (1 - \rho^2). \tag{12}$$

### Suggested computational strategy

Searls<sup>25</sup> suggested and proved that an estimator which is a constant multiple of sample mean estimator, is more efficient than the sample mean estimator. This has also been proven by various authors for ratio and product estimators. Therefore, motivated by Searls<sup>25</sup> and Aliyu et al.<sup>24</sup>, we suggest the following imputation strategy for the estimation of  $\bar{Y}$  as,

$$y_i = \begin{cases} y_i & \text{if } i \in R \\ \frac{1}{n-r} \left[ n \left\{ \kappa_1 \bar{y}_r + \kappa_2 \bar{y}_r \frac{\bar{x}_r}{\bar{X}} \right\} \exp \left( \frac{\bar{X} - \bar{x}_r}{\bar{X} + \bar{x}_r} \right) - r \bar{y}_r \right] & \text{if } i \in R^c. \end{cases}$$

The resultant point estimator of  $\bar{Y}$  is,

$$t_p = \left\{ \kappa_1 \bar{y}_r + \kappa_2 \bar{y}_r \frac{\bar{x}_r}{\bar{X}} \right\} \exp \left( \frac{\bar{X} - \bar{x}_r}{\bar{X} + \bar{x}_r} \right)$$

where,  $\kappa_1$  and  $\kappa_2$  are the Searls constants to be obtained so that the MSE of  $t_p$  is least and  $\kappa_1 + \kappa_2 \neq 1$ . If  $\kappa_1 + \kappa_2 = 1$ , then the suggested estimator reduced to Aliyu et al.<sup>24</sup> estimator.

The properties of the sampling distribution of  $t_p$ , the standard approximations are given as,

$$\bar{y}_r = \bar{Y}(1 + e_0), \bar{x}_r = \bar{X}(1 + e_1).$$

$$\text{such that } E(e_0) = E(e_1) = 0 \text{ and } E(e_0^2) = \theta_{r,N} C_y^2, E(e_1^2) = \theta_{r,N} C_x^2, E(e_0 e_1) = \theta_{r,N} C_{yx}.$$

Expressing  $t_p$  in terms of  $e_i$ ,  $s(i = 0, 1)$ , expanding and retaining the terms for an approximation of order one, we have,

$$t_p = \left[ \kappa_1 \bar{Y}(1 + e_0) + \kappa_2 \bar{Y}(1 + e_0) \frac{\bar{X}(1 + e_1)}{\bar{X}} \right] \exp \left[ \frac{\bar{X} - \bar{X}(1 + e_1)}{\bar{X} + \bar{X}(1 + e_1)} \right],$$

$$= [\kappa_1 \bar{Y}(1 + e_0) + \kappa_2 \bar{Y}(1 + e_0)(1 + e_1)] \exp \left[ \frac{-e_1}{2 + e_1} \right],$$

$$= \bar{Y} [\kappa_1(1 + e_0) + \kappa_2(1 + e_0)(1 + e_1)] \exp \left[ -\frac{e_1}{2} + \frac{e_1^2}{4} \right],$$

$$= \bar{Y} [\kappa_1(1 + e_0) + \kappa_2(1 + e_0 + e_1 + e_0 e_1)] \left[ 1 - \frac{e_1}{2} + \frac{3e_1^2}{8} \right],$$

$$= \left[ \kappa_1 \left( 1 + e_0 - \frac{e_1}{2} - \frac{e_0 e_1}{2} + \frac{3e_1^2}{8} \right) + \kappa_2 \left( 1 + e_0 + \frac{e_1}{2} + \frac{e_0 e_1}{2} - \frac{e_1^2}{8} \right) \right].$$

Subtracting  $\bar{Y}$  on both sides of above equation, we get,

$$t_p - \bar{Y} = \bar{Y} \left[ \kappa_1 \left( 1 + e_0 - \frac{e_1}{2} - \frac{e_0 e_1}{2} + \frac{3e_1^2}{8} \right) + \kappa_2 \left( 1 + e_0 + \frac{e_1}{2} + \frac{e_0 e_1}{2} - \frac{e_1^2}{8} \right) - 1 \right]. \tag{13}$$

Taking expectation on both sides and putting values of different expectations, we have bias of  $t_p$  as,

$$B(t_p) = \bar{Y} \left[ \kappa_1 \left( 1 - \frac{1}{2} \theta_{r,N} C_{yx} + \frac{3}{8} \theta_{r,N} C_x^2 \right) + \kappa_2 \left( 1 + \frac{1}{2} \theta_{r,N} C_{yx} - \frac{1}{8} \theta_{r,N} C_x^2 \right) - 1 \right].$$

Squaring on both sides of (13), simplifying and putting the terms for an approximation of degree one, we get the MSE of  $t_p$  as,

$$\begin{aligned} MSE(t_p) &= \bar{Y}^2 E \left[ \kappa_1 \left( 1 + e_0 - \frac{e_1}{2} - \frac{e_0 e_1}{2} + \frac{3e_1^2}{8} \right) + \kappa_2 \left( 1 + e_0 + \frac{e_1}{2} + \frac{e_0 e_1}{2} - \frac{e_1^2}{8} \right) - 1 \right]^2, \\ &= \bar{Y}^2 E \left[ 1 + \kappa_1^2 (1 + e_0^2 + e_1^2 - 2e_0 e_1) + \kappa_2^2 (1 + e_0^2 + 2e_0 e_1) + 2\kappa_1 \kappa_2 (1 + e_0^2) \right. \\ &\quad \left. - 2\kappa_1 \left( 1 - \frac{1}{2} e_0 e_1 + \frac{3}{8} e_1^2 \right) - 2\kappa_2 \left( 1 - \frac{1}{2} e_0 e_1 - \frac{1}{8} e_1^2 \right) \right]. \end{aligned}$$

Putting the values of different expectations, we get,

$$\begin{aligned} MSE(t_p) &= \bar{Y}^2 \left[ 1 + \kappa_1^2 \left\{ 1 + \theta_{r,N} (C_y^2 + C_x^2 - 2C_{yx}) \right\} + \kappa_2^2 \left\{ 1 + \theta_{r,N} (C_y^2 - 2C_{yx}) \right\} + 2\kappa_1 \kappa_2 (1 + \theta_{r,N} C_y^2) \right. \\ &\quad \left. - 2\kappa_1 \left( 1 - \frac{1}{2} \theta_{r,N} C_{yx} + \frac{3}{8} \theta_{r,N} C_x^2 \right) - 2\kappa_2 \left( 1 - \frac{1}{2} \theta_{r,N} C_{yx} - \frac{1}{8} \theta_{r,N} C_x^2 \right) \right], \\ MSE(t_p) &= \bar{Y}^2 [1 + \kappa_1^2 A + \kappa_2^2 B + 2\kappa_1 \kappa_2 C - 2\kappa_1 D - 2\kappa_2 F] \end{aligned} \tag{14}$$

where  $A = \{1 + \theta_{r,N} (C_y^2 + C_x^2 - 2C_{yx})\}$ ,  $B = \{1 + \theta_{r,N} (C_y^2 - 2C_{yx})\}$ ,  $C = (1 + \theta_{r,N} C_y^2)$ ,  $D = (1 - \frac{1}{2} \theta_{r,N} C_{yx} + \frac{3}{8} \theta_{r,N} C_x^2)$ ,  $F = (1 - \frac{1}{2} \theta_{r,N} C_{yx} - \frac{1}{8} \theta_{r,N} C_x^2)$ .

The optimal values of  $\kappa_1$  and  $\kappa_2$ , which reduces the MSE of  $t_p$  respectively are,

$$\kappa_{1(opt)} = \frac{BD - CF}{AB - C^2} \text{ and } \kappa_{2(opt)} = \frac{AF - DC}{AB - C^2}.$$

The minimum value of  $MSE(t_p)$  for the optimum values of  $\kappa_1$  and  $\kappa_2$  is,

$$\begin{aligned} MSE_{min}(t_p) &= \bar{Y}^2 \left[ 1 - \frac{\left\{ \begin{aligned} &2(AF - DC)F + 2(BD - CF)D - 2(AF - DC)(BD - CF)C \\ &-(BD - CF)^2 A - (AF - DC)^2 B \end{aligned} \right\}}{(AB - C^2)^2} \right], \\ MSE_{min}(t_p) &= \bar{Y}^2 \left[ 1 - \frac{L}{M^2} \right] \end{aligned} \tag{15}$$

where,

$$\begin{aligned} L &= \left\{ \begin{aligned} &2(AF - DC)F + 2(BD - CF)D - 2(AF - DC)(BD - CF)C \\ &-(BD - CF)^2 A - (AF - DC)^2 B \end{aligned} \right\}, \\ M &= (AB - C^2). \end{aligned}$$

### Theoretical efficiency comparison

In this section, the introduced estimator has been compared with the estimators of  $\bar{Y}$  in competition under imputation methods in terms of their efficiencies. The efficiency conditions for which the introduced estimator is more efficient than the completing estimators are obtained.

The introduced estimator  $t_p$  is better than  $t_0$  under imputation method if,

$$\begin{aligned} V(t_0) - MSE_{min}(t_p) &> 0, \text{ or} \\ \theta_{r,N} C_y^2 - \left[ 1 - \frac{L}{M^2} \right] &> 0. \end{aligned} \tag{16}$$

The introduced estimator  $t_p$  has lesser MSE than the usual ratio estimator  $t_r$  for the condition if,

$$\begin{aligned} MSE(t_r) - MSE_{min}(t_p) &> 0, \text{ or} \\ \theta_{r,n} [C_y^2 + C_x^2 - 2C_{yx}] + \frac{L}{M^2} &> 1. \end{aligned} \tag{17}$$

The suggested estimator  $t_p$  performs better than the Singh and Horn<sup>20</sup> estimator  $t_1$  for the condition if,

$$\begin{aligned} MSE_{min}(t_1) - MSE_{min}(t_p) &> 0, \text{ or} \\ C_y^2 [\theta_{r,N} - \theta_{r,n} \rho^2] + \frac{L}{M^2} &> 1. \end{aligned} \tag{18}$$

The proposed estimator  $t_p$  is more efficient than the Singh and Deo<sup>21</sup> estimator  $t_2$  if,

$$MSE_{\min}(t_2) - MSE_{\min}(t_p) > 0, \text{ or}$$

$$MSE_{\min}(t_1) - \theta_{r,n} S_x^2 \left( \frac{S_{yx}}{S_x^2} - \frac{\bar{Y}}{\bar{X}} \right)^2 + \frac{L}{M^2} > 1. \tag{19}$$

The introduced estimator  $t_p$  has lesser MSE than the Kadilar and Cingi<sup>8</sup> estimators  $t_i$  ; ( $i = 3, 4, 5$ ) under the conditions if,

$$MSE_{\min}(t_i) - MSE_{\min}(t_p) > 0; \ i = 3, 4, 5; \text{ or}$$

$$MSE_{\min}(t_1) - \theta_{r,N} S_x^2 (R^2 - \beta^2) + \frac{L}{M^2} > 1, \tag{20}$$

$$MSE_{\min}(t_1) - \theta_{n,N} S_x^2 (R^2 - \beta^2) + \frac{L}{M^2} > 1, \tag{21}$$

$$MSE_{\min}(t_1) - \theta_{r,n} [(R + \beta)^2 S_x^2 - 2(R + \beta) S_{yx}] + \frac{L}{M^2} > 1. \tag{22}$$

The suggested estimator  $t_p$  performs better than the Singh<sup>22</sup> estimator  $t_6$  for the condition if,

$$MSE_{\min}(t_6) - MSE_{\min}(t_p) > 0, \text{ or}$$

$$MSE_{\min}(t_1) - \theta_{r,n} S_x^2 \left( \frac{S_{yx}}{S_x^2} - \frac{\bar{Y}}{\bar{X}} \right)^2 + \frac{L}{M^2} > 1. \tag{23}$$

The proposed estimator  $t_p$  is more efficient than the Singh et al.<sup>23</sup> estimator  $t_7$  for the condition if,

$$MSE_{\min}(t_7) - MSE_{\min}(t_p) > 0, \text{ or}$$

$$MSE_{\min}(t_1) - \theta_{r,n} S_x^2 \left( \frac{S_{yx}}{S_x^2} - \frac{\bar{Y}}{\bar{X}} \right)^2 + \frac{L}{M^2} > 1. \tag{24}$$

The introduced estimator  $t_p$  performs better than the Gira<sup>10</sup> estimator  $t_8$  for the condition if,

$$MSE_{\min}(t_8) - MSE_{\min}(t_p) > 0, \text{ or}$$

$$V(t_0) - \theta_{r,n} \bar{Y}^2 \rho^2 C_y^2 + \frac{L}{M^2} > 1. \tag{25}$$

The suggested estimator  $t_p$  has lesser MSE than the Singh et al.<sup>4</sup> estimator  $t_9$  if the following condition is satisfied.

$$MSE_{\min}(t_9) - MSE_{\min}(t_p) > 0, \text{ or}$$

$$V(t_0) - \theta_{r,n} \bar{Y}^2 \rho^2 C_y^2 + \frac{L}{M^2} > 1. \tag{26}$$

The introduced estimator  $t_p$  is more efficient than the Aliyu et al.<sup>24</sup> estimator  $t_{10}$  under the condition if,

$$MSE_{\min}(t_{10}) - MSE_{\min}(t_p) > 0, \text{ or}$$

$$\theta_{r,N} \bar{Y}^2 C_y^2 (1 - \rho^2) + \frac{L}{M^2} > 1. \tag{27}$$

### Empirical study

To verify the efficiency conditions of the introduced estimator over the estimators in competition, we have considered the following three natural populations. The data of Population-1 has been taken from Murthy<sup>26</sup>, while Population-2 has been taken into account from Cochran<sup>27</sup> and the data set of Population-3 has been obtained from Sarndal et al.<sup>28</sup> for both study and the auxiliary variables. These populations are of different natures as per

Population-1: Murthy <sup>26</sup>	Population-2: Cochran <sup>27</sup>	Population-3: Sarndal et al. <sup>28</sup>
$N = 80, n = 25, r = 20$	$N = 10, n = 5, r = 4$	$N = 284, n = 35, r = 25$
$\bar{Y} = 5182.638, \bar{X} = 285.125$	$\bar{Y} = 56.900, \bar{X} = 54.296$	$\bar{Y} = 29.360, \bar{X} = 245.088$
$C_y = 0.354, C_x = 0.949$	$C_y = 0.184, C_x = 0.162$	$C_y = 1.760, C_x = 2.430$
$\beta_1(x) = 0.949$	$\beta_1(x) = 0.496$	$\beta_1(x) = 8.770$
$\beta_2(x) = 3.536, \rho_{yx} = 0.914$	$\beta_2(x) = 2.593, \rho_{yx} = 0.924$	$\beta_2(x) = 88.880, \rho_{yx} = 0.961$

**Table 1.** Parameters of the three natural populations.



their populations sizes and, we have taken samples of different sizes to see the performances of the competing and introduced estimators. The parameters of the considered three populations along with their sources are presented in Table 1.

Table 2 represents the MSE of different estimators in competition and the introduced estimator along with the Percentage Relative Efficiency (PRE) of various estimators with respect to  $t_0$  for all the three natural populations.

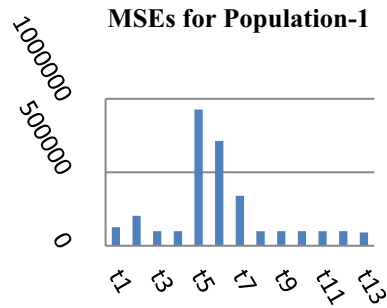


Figure 1. MSEs of Different estimators.

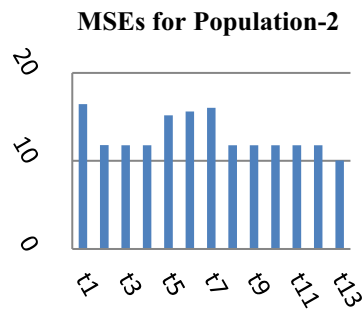


Figure 2. MSEs of Different estimators.

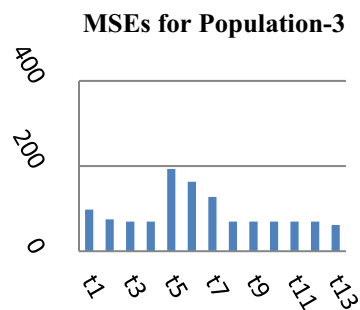


Figure 3. MSEs of Different estimators.

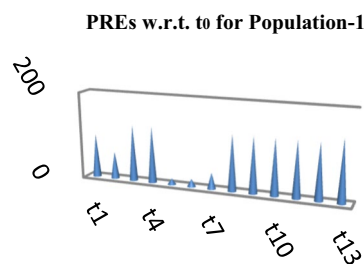
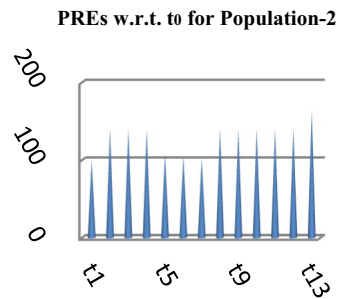
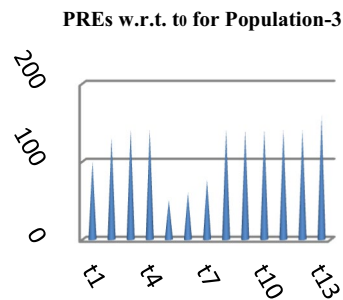


Figure 4. PREs of Different estimators.



**Figure 5.** PREs of Different estimators.



**Figure 6.** PREs of Different estimators.

The following Figs. 1, 2 and 3 represent the MSEs of the estimators in competition and the introduced estimators for the three real natural populations respectively.

The following Figs. 4, 5 and 6 represent the PREs of the competing and the suggested estimators with respect to  $t_0$  for the three natural populations respectively.

### Simulation study

In this section, an artificial data has been generated for the comparison of competing and introduced imputation methods for the large simulated population to see the nature of different estimators under comparison. We have generated a population using the same parameters of the real population-3. We have taken a sizable sample to examine the nature of the suggested estimator and other estimators for large samples, rather than taking into account the changing sample size in the simulation study because it is well-established in the literature that the estimate becomes closer to the true parameter as the sample size increases. We created the population using a normal distribution because it is widely known that all theoretical and sampling distributions approach a normal distribution for big sample sizes. This means that the findings of more complicated real-world scenarios with various sample sizes would remain unchanged. A bivariate normal distribution with mean vectors and a variance-covariance matrix is used to construct the population as:

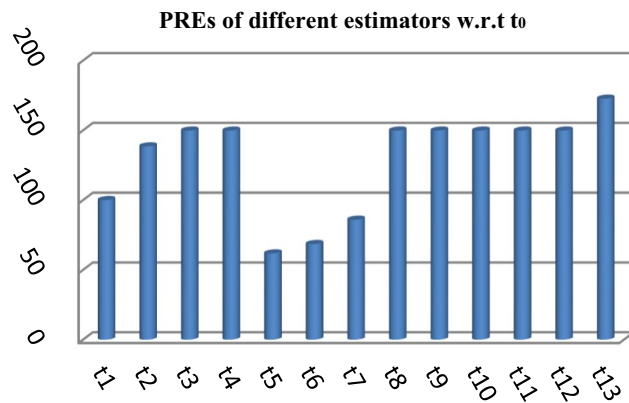
Means of  $[Y, X]$  as  $\mu = [29.360, 245.088]$ .

Variances and covariance of  $[Y, X]$  as  $\sigma^2 = \begin{bmatrix} 2670.161 & 29574.704 \\ 29574.704 & 354696.288 \end{bmatrix}$ .

Correlation  $\rho_{yx} = 0.961$ .

The following steps have been used for the simulation of the required population:

- A bivariate normal distribution of  $X$  and  $Y$  of size  $N = 5000$  have been generated through these parameters using R Program.
- The parameters have been computed for this simulated population of size  $N = 5000$ .
- A sample of size  $n = 200$  has been selected from this simulated population with response rate  $r = 160$ .
- Sample statistics that is sample mean, sample variance and the values of the introduced and competing estimators  $t_i, i = 0, 1, \dots, 10, p$  of  $\bar{Y}$  are calculated for this sample under imputation technique.
- Steps (c) and (d) are repeated  $m = 50,000$  times.
- The MSE of every estimator  $t_i$  is calculated through the formula,  $MSE(t_i) = \frac{1}{m} \sum_{j=1}^m (t_{ij} - \bar{Y})^2$ .
- The PRE of each of the estimator  $t_i$  with respect to  $t_0$  has been calculated using the formula:
- $PRE(t_i) = \frac{MSE(t_0)}{MSE(t_i)} \times 100, i = 1, 2, \dots, 10, p$



**Figure 7.** PREs of different estimators for the simulated population.

Table 3 represents the PRE of various estimators of  $\bar{Y}$  under imputation methods with respect to  $t_0$  for the simulated population.

These results are also presented in the form of graph in Fig. 7 given below as,

### Results and discussion

From Table 2, it may be observed that the MSEs of the competing estimator of  $\bar{Y}$  under imputation methods lie in the intervals [98215.14, 926,966.20], [11.77, 16.44] and [69.22, 97.40] while for the suggested estimator, these are 90,056.57, 10.04 and 61.75 for Population-1, Population-2 and Population-3 respectively. On the other hand the PREs of various estimators with respect to  $t_0$  lie in the intervals [13.63, 128.66], [102.62, 139.68] and [50.43, 140.41] while the PREs of the suggested estimator are 140.32, 163.75 and 158.41 for Population-1, Population-2 and Population-3 respectively. The same results may also be verified from the figures from Figs. 1, 2, 3, 4, 5, and 6 for the three real populations under consideration. It may also be observed from Table 3 that the PREs of various estimators in competition with respect to  $t_0$  lie in the interval [61.76, 149.81] and for the introduced estimator is 172.68, which may also be verified from Fig. 7 for the simulated data.

Estimator	Population-1		Population-2		Population-3	
	MSE	PRE	MSE	PRE	MSE	PRE
$t_0$	126,366.00	100.00	16.44	100.00	97.40	100.00
$t_r$	203,055.80	62.23	11.78	139.56	74.60	130.56
$t_1$	98,215.14	128.66	11.77	139.68	69.22	140.71
$t_2$	98,215.14	128.66	11.77	139.68	69.22	140.71
$t_3$	926,966.20	13.63	15.17	108.37	193.13	50.43
$t_4$	713,472.80	17.71	15.60	105.38	163.14	59.71
$t_5$	339,859.40	37.18	16.02	102.62	127.39	76.46
$t_6$	98,215.14	128.66	11.77	139.68	69.22	140.71
$t_7$	98,215.14	128.66	11.77	139.68	69.22	140.71
$t_8$	98,215.14	128.66	11.77	139.68	69.22	140.71
$t_9$	98,215.14	128.66	11.77	139.68	69.22	140.71
$t_{10}$	98,215.14	128.66	11.77	139.68	69.22	140.71
$t_p$	90,056.57	140.32	10.04	163.75	61.49	158.41

**Table 2.** MSE and PRE of different estimators with respect to  $t_0$ .

Estimator	$t_0$	$t_r$	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$	$t_7$	$t_8$	$t_9$	$t_{10}$	$t_p$
PRE	100.00	138.47	149.81	149.81	61.76	68.52	85.94	149.81	149.81	149.81	149.81	149.81	172.68

**Table 3.** PRE of different estimators with respect to  $t_0$  for the simulated population.

## Conclusion

In this manuscript, we have introduced a new class of estimators of  $\bar{Y}$  under imputation method. The bias and the MSE of the introduced family have been studied for an approximation of degree one. Through theoretical comparison with competing estimators using imputation techniques, efficiency requirements over competing estimators are produced for the suggested estimator. Along with a simulated population, three actual natural populations are used to confirm these efficiency criteria of the presented estimators. It has been found that the suggested estimators is having least MSE in all three real populations and has highest PRE for all real and simulated populations. Thus it is evident that the introduced estimator is the most efficient among the class of all estimators of  $\bar{Y}$  in competition under the imputation methods. As the proposed estimator is most efficient, therefore it is recommended for use in different areas of applications including Agricultural Sciences, Biological Sciences, Commerce, Engineering, Economics, Fisheries, Medical Science, Social Science and other areas of applications. For instance, as certain experimental units must be discarded during the experiment, in therapeutic or life-saving drug testing studies. In a manner similar to this, diseases, livestock grazing, or other natural calamities kill crops during agricultural experiments.

## Data availability

All relevant data is within the manuscript.

Received: 1 October 2023; Accepted: 15 March 2024

Published online: 18 March 2024

## References

- Vishwakarma, G. K. & Kumar, M. An efficient class of estimators for the mean of a finite population in two-phase sampling using multi-auxiliary variates. *Commun. Math. Stat.* **3**(4), 477–489 (2015).
- Kumar, M. & Vishwakarma, G. K. Estimation of mean in double sampling using exponential technique on multi-auxiliary variates. *Commun. Math. Stat.* **5**(4), 429–445 (2017).
- Vishwakarma, G. K., Singh, N. & Kumar, N. A computational framework for estimation of mean in presence of observational error. *Concurr. Comput. Pract. Exp.* **34**(11), e6842 (2022).
- Singh, G. N., Jaiswal, A. K., Singh, C. & Usman, M. An improved alternative method of imputation for missing data in survey sampling. *J. Stat. Appl. Probab.* **11**(2), 535–543 (2022).
- Rubin, D. B. Inference and missing data. *Biometrika* **63**, 581–592 (1976).
- Heitjan, D. F. & Basu, S. Distinguishing, “missing at random” and “missing completely at random”. *Am. Stat.* **50**, 207–213 (1996).
- Singh, H. P. & Vishwakarma, G. K. Modified exponential ratio and product estimators for finite population mean in double sampling. *Austrian J. Stat.* **36**(3), 217–225 (2007).
- Kadilar, C. & Cingi, H. Estimators for the population mean in the case of missing data. *Commun. Stat. Theory Methods* **37**, 2226–2236 (2008).
- Diana, G. & Perri, P. F. Improved estimators of the population mean for missing data. *Commun. Stat. Theory Methods* **39**, 3245–3251 (2010).
- Gira, A. A. Estimation of population mean with a new imputation methods. *Appl. Math. Sci.* **9**(34), 1663–1672 (2015).
- Bhushan, S. & Pandey, A. P. Optimal imputation of missing data for estimation of population mean. *J. Stat. Manag. Syst.* **19**(6), 755–769 (2016).
- Prasad, S. A study on new methods of ratio exponential type imputation in sample surveys. *Hacet. J. Math. Stat.* **47**(2), 1–11 (2017).
- Audu, A. *et al.* On the class of exponential-type imputation estimators of population mean with known population mean of auxiliary variable. *NIPES J. Sci. Technol. Res.* **2**(4), 1–11 (2020).
- Audu, A. *et al.* Regression-type imputation class of estimators using auxiliary attribute. *Asian Res. J. Math.* **17**(5), 1–13 (2021).
- Shahzad, U., Al-Noor, N. H., Hanif, M., Sajjad, I. & Anas, M. M. Imputation based mean estimators in case of missing data utilizing robust regression and variance-covariance matrices. *Commun. Stat. Simul. Comput.* **51**(8), 4276–4295 (2022).
- Shahzad, U., Hanif, M., Sajjad, I. & Anas, M. M. Quantile regression-ratio-type estimators for mean estimation under complete and partial auxiliary information. *Sci. Iran.* **29**(3), 1705–1715 (2022).
- Alomair, M. A. & Shahzad, U. Compromised-imputation and EWMA-based memory-type mean estimators using quantile regression. *Symmetry* **15**(10), 1888. <https://doi.org/10.3390/sym15101888> (2023).
- Lawson, N. New imputation method for estimating population mean in the presence of missing data. *Lobachevskii J. Math.* **44**, 3740–3748. <https://doi.org/10.1134/S1995080223090202> (2023).
- Anas, M. M., Huang, Z., Shahzad, U., Zaman, T. & Shahzadi, S. Compromised imputation based mean estimators using robust quantile regression. *Commun. Stat. Theory Methods* **53**(5), 1700–1715. <https://doi.org/10.1080/03610926.2022.2108057> (2024).
- Singh, S. & Horn, S. Compromised imputation in survey sampling. *Metrika* **51**, 267–276 (2000).
- Singh, S. & Deo, B. Imputation by power transformation. *Stat. Pap.* **44**, 555–579 (2003).
- Singh, S. A new method of imputation in survey sampling. *Statistics* **43**, 499–511 (2009).
- Singh, A. K., Singh, P. & Singh, V. K. Exponential-type compromised imputation in survey sampling. *J. Stat. Appl. Probab.* **3**(2), 211–217 (2014).
- Aliyu, Y. H. *et al.* Modified compromised type method of imputation for estimating population mean. *J. Sci. Res.* **66**(1), 404–410 (2022).
- Searls, D. T. The utilization of a known coefficient of variation in the estimation procedure. *J. Am. Stat. Assoc.* **59**(308), 1225–1226 (1964).
- Murthy, M. N. *Sampling Theory and Methods* (Statistical Publishing Society, 1967).
- Cochran, W. G. *Sampling Techniques* 3rd edn. (Wiley and Sons, 1977).
- Sarndal, C. E., Swensson, B. & Wretman, J. H. *Model Assisted Survey Sampling* (Springer-Verlag, 1992).

## Acknowledgements

The authors express their heartfelt gratitude to the editor and the learned referees for their critical reviews, which improved the earlier draft.

## Author contributions

All authors contribute equally.

## Funding

The authors received no specific funding for this work.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to G.K.V.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024