Original Article
ISSN (Online): 2454-1907

International Journal of Engineering Technologies and Management Research
September 2022 9(9), 1–16

AN EFFICIENT COMPROMISED IMPUTATION METHOD FOR ESTIMATING POPULATION MEAN

Sandeep Mishra [1]

[1] Association of Indian Universities, New Delhi, India

Received 16 July 2022
Accepted 19 August 2022
Published 05 September 2022

Corresponding Author

Sandeep Mishra,
smisra1983@yahoo.co.in

DOI 10.29121/ijetmr.v9.i9.2022.1216

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

This paper suggests a modified new ratio-product-exponential imputation procedure to deal with missing data in order to estimate a finite population mean in a simple random sample without replacement. The bias and mean squared error of our proposed estimator are obtained to the first degree of approximation. We derive conditions for the parameters under which the proposed estimator has smaller mean squared error than the sample mean, ratio, and product estimators. We carry out an empirical study which shows that the proposed estimator outperforms the traditional estimators using real data.

Keywords: Missing Data, Mean Square Error, Imputation, Bias, Ratio Estimator

1. INTRODUCTION

Imputation means replacing a missing value with another value based on a reasonable estimate. Information on the related auxiliary variable is generally used to recreate the missing values for completing datasets. Incomplete data is usually categorized into three different response mechanisms: Missing Completely at Random (MCAR); Missing at Random (MAR); and Missing Not at Random (MNAR or NMAR) Little and Rubin (2002). Missing completely at random (MCAR): Missing data are randomly distributed across the variable and unrelated to other variables. Missing at random (MAR): Missing data are not randomly distributed but they are accounted for by other observed variables. Missing not at random (MNAR): Missing data systematically differ from the observed values. From the above-mentioned classifications of missing data, we, in the present study, have assumed MCAR.

Auxiliary information is important for survey practitioner as it is utilized to improve the performance of the methods. It may be utilized at the design stage or the estimation stage of the survey to get the more efficient estimator. At estimation stage ratio, product and regression methods are traditionally used. Bhal and Tuteja (1991) introduced exponential ratio and product estimator for estimation of population mean. Many modifications have been proposed using these methods till date. For handling missing data on the study variable several extensions and developments were proposed in the literature. Singh (2003) suggested product estimation for imputation. Shakti Prasad (2018) adapts exponential product type estimator given by Bahal and Tuteja (1991) and proposed exponential estimators for imputation. Kadilar and Cingi (2008) investigated some ratio-type imputation methods and proposed three new estimators to overcome the problem of the missing data. Diana and Perri (2010) proposed three regression type estimators which were more efficient than the Kadilar and Cingi (2008). The present article suggests a general ratio product exponential type method of imputation and accordingly proposed three estimators using the different amount of available auxiliary information as utilized by Ahmad et al. (2006), Kadilar and Cingi (2008), and Diana and Perri (2010). The proposed methods are than compared by traditional procedure of imputation. The proposed estimators come out to be more efficient than the usual ratio, product, regression, and exponential method for handling missing observations to estimate the population mean.

Given a finite population $\Omega = \{1,2,\cdots,N\}$, the objective is to estimate the population mean $\bar{Y} = N^{-1}\sum_{i=1}^{N} y_i$. A simple random sample wor, s, of size n, is drawn from the population Ω. Let the responding units be r from the n sampled units. Let us denote R as the set of responding units and R^c the set of non-responding units, i.e., y_i is observed for $i \in R$ but for units in R^c the values are not available and hence imputed values are derived by some method. In this paper we shall use the following notations:

N: Population Size; n: Sample size; r: Number of responding units; $\bar{Y}, \bar{X}$: Population means of study variate y and auxiliary variate x respectively; S_y, S_x: Standard Deviation of study variate y and auxiliary variate x respectively; C_y, C_x: Coefficient of variation of study variate y and auxiliary variate x respectively; ρ: Correlation coefficient between y and x; $f_r = \left(\frac{1}{r} - \frac{1}{N}\right)$, $f_n = \left(\frac{1}{n} - \frac{1}{N}\right)$, $f_{rn} = \left(\frac{1}{r} - \frac{1}{n}\right)$.

2. SOME EXISTING METHODS OF IMPUTATION

1) The *mean method of imputation* suggests replacing the missing observations with the mean of the observations available on response units i.e.

$$y_{\cdot i} = \begin{cases} y_i & if \quad i \in R \\ \bar{y}_r & if \quad i \in R^c \end{cases}$$

Then the estimator of the population mean $\bar{Y}$ is given by

$\bar{y}_m = \frac{1}{r}\sum_{i \in R} y_i = \bar{y}_r$ and its MSE is given by

$$MSE(\bar{y}_m) = f_r\, \bar{Y}^2 C_y^2 \tag{1.1}$$

2) The ratio method of imputation uses information on one auxiliary variable x and calculates the missing values by

$$y_{\cdot i} = \begin{cases} y_i & if \;\; i \in R \\ \hat{b}x_i & if \;\; i \in R^c \end{cases}$$

Where $\hat{b} = \frac{\sum_{i=1}^{r} y_i}{\sum_{i=1}^{r} x_i}$

This gives the resulting estimator by

$$\bar{y}_{RAT} = \bar{y}_r \cdot \bar{x}_n / \bar{x}_r$$

The MSE of $\bar{y}_{RAT}$ is given by

$$MSE(\bar{y}_{RAT}) = \bar{Y}^2 [f_n C_y^2 + f_{rn}(C_y^2 + C_x^2 - 2\rho_{yx} C_y\, C_x)] \tag{1.2}$$

It is noted that, in the presence of missing data, the availability of information on auxiliary variable x in the data set supports suggesting efficient estimators.

3) Diana and Perri (2010) proposed three estimators as by using different regression-type method of imputation such that the imputed data is given by

$$y_{\cdot DP_1} = \begin{cases} \frac{ny_i}{r} + b(\bar{X} - x_i) & if \; i \in R \\ b(\bar{X} - x_i) & if \; i \in R^c \end{cases}$$

$$y_{\cdot DP_2} = \begin{cases} \frac{ny_i}{r} - b\frac{nx_i}{r} & if \; i \in R \\ b\frac{n\bar{X}}{n-r} & if \; i \in R^c \end{cases}$$

$$y_{\cdot DP_1} = \begin{cases} \frac{ny_i}{r} - b\frac{nx_i}{r} & if \; i \in R \\ b\frac{n\bar{x}_n}{n-r} & if \; i \in R^c \end{cases}$$

For these methods the resulting estimators are

$$T_{DP_1} = \bar{y}_r + b(\bar{X} - \bar{x}_n)$$
$$T_{DP_2} = \bar{y}_r + b(\bar{X} - \bar{x}_r)$$
$$T_{DP_3} = \bar{y}_r + b(\bar{x}_n - \bar{x}_r)$$

$$MSE\left(T_{DP_1}\right) = S_y^2\left[\left(\frac{1}{r}-\frac{1}{n}\right)+\left(\frac{1}{n}-\frac{1}{N}\right)(1-\rho^2)\right] \tag{1.2}$$

$$MSE\left(T_{DP_2}\right) = \left(\frac{1}{r}-\frac{1}{N}\right)S_y^2(1-\rho^2) \tag{1.3}$$

$$MSE\left(T_{DP_3}\right) = S_y^2\left(\left(\frac{1}{n}-\frac{1}{N}\right)+\left(\frac{1}{r}-\frac{1}{n}\right)(1-\rho^2)\right) \tag{1.4}$$

They proved that the suggested estimators are more efficient than the Kadilar and Cingi (2008) estimators. T_{DP_2} is always more efficient than both T_{DP_1} and T_{DP_3}, whereas T_{DP_3} perform better than T_{DP_1} if the condition

$$r < \frac{nN}{2N-n}$$

3. THE PROPOSED ESTIMATOR

The estimator suggested here is inspired by the Sahai (1985) estimator of population mean in case of simple random sampling, and is defined as

$$y_{.i} = \begin{cases} y_i & if \quad i \in R \\ \dfrac{1}{n-r}\left[n\bar{y}_r\left\{W_1\left(\dfrac{(1-a)\bar{x}_r + a\,\bar{x}_n}{(1-a)\bar{x}_n + a\bar{x}_r}\right) + W_2\exp\left(\dfrac{(1-2a)(\bar{x}_r-\bar{x}_n)}{\bar{x}_n+\bar{x}_r}\right)\right\} - r\bar{y}_r\right] & if \quad i \in R^c \end{cases}$$

With the above imputation method, the resulting estimator of the population mean $\bar{Y}$ is obtained as

$$T = \frac{1}{n}\sum_{i=1}^{n} y_{\cdot i} = \frac{1}{n}\left[\sum_{i=1}^{r} y_I + \frac{n}{n-r}\bar{y}_r\left\{\sum_{i=r+1}^{n} W_1\left\{\frac{(1-a)\bar{x}_r+a\,\bar{x}_n}{(1-a)\bar{x}_n+a\bar{x}_r}\right\} + W_2\exp\left\{\frac{(1-2a)(\bar{x}_r-\bar{x}_n)}{\bar{x}_n+\bar{x}_r}\right\} - r\right\}\right] \tag{2.1}$$

W_1 and W_2 are constant chosen suitably so that their choice minimizes the mean square error of the resultant estimator and a is a real constant. Our goal in this paper is to discuss the suggested estimators for different values of a and have a comparative study of the suggested estimator for these values of a in order to get the minimum MSE.

4. FIRST DEGREE APPROXIMATION TO THE BIAS

To derive the Bias and MSE expressions of the proposed estimator T upto $O(1/n)$, we define

$$e_0 = \frac{\bar{y}_r - \bar{Y}}{\bar{Y}}, \qquad e_1 = \frac{\bar{x}_r - \bar{Y}}{\bar{Y}}, \qquad e_2 = \frac{\bar{x}_n - \bar{Y}}{\bar{Y}}$$

Thus, we have $\bar{y}_r = \bar{Y}(1+e_0)$, $\quad \bar{x}_r = \bar{X}(1+e_1)$, $\quad \bar{x}_n = \bar{X}(1+e_2)$

The expectation of these $e_i's$ are $E(e_i) = 0 \quad i = 0,1,2$

And under simple random sampling without replacement,

$$E({e_0}^2) = f_r C_y^2, \qquad E(e_1^2) = f_r C_x^2, \qquad E(e_2^2) = f_n C_y^2$$

$$E(e_0 e_1) = f_r \rho C_y C_x, \qquad E(e_0 e_2) = f_n \rho C_y C_x, \quad E(e_2 e_1) = f_n C_x^2$$

where $\rho = cor(x,y)$, $C_y^2 = S_y^2/\bar{Y}^2$, $C_x^2 = S_x^2/\bar{X}^2$. $f_r = \left(\frac{1}{r} - \frac{1}{N}\right), f_n = \left(\frac{1}{n} - \frac{1}{N}\right),$ $and\ f_{rn} = \left(\frac{1}{r} - \frac{1}{n}\right)$

Now representing (2.1) in terms of $e_i's$, we have

$$T = \bar{Y}(1+e_0)\left[W_1\left\{\frac{(1-a)\bar{X}(1+e_1) + a\,\bar{X}(1+e_2)}{(1-a)\bar{X}(1+e_2) + a\bar{X}(1+e_1)}\right\} + W_2 \exp\left\{\frac{(1-2a)(\bar{X}(1+e_1) - \bar{X}(1+e_2))}{\bar{X}(1+e_2) + \bar{X}(1+e_1)}\right\}\right]$$

$$= \bar{Y}(1+e_0)\left[W_1\left\{\frac{(1+ae_2+(1-a)e_1)}{(1+ae_1+(1-a)e_2)}\right\} + W_2 \exp\left\{\frac{(1-2a)\bar{X}(1+e_1-1-e_2)}{\bar{X}(1+e_2+1+e_1)}\right\}\right]$$

$$= \bar{Y}(1+e_0)\left[W_1\{(1+ae_2+(1-a)e_1)\,(1+ae_1+(1-a)e_2)^{-1}\} + W_2 \exp\left\{\frac{(1-2a)(e_1-e_2)}{(2+e_2+e_1)}\right\}\right]$$

We assume that the sample is large enough to make $|e_1|$ and $|e_2|$ so small that contributions from powers of degree higher than two are negligible. By retaining powers up to e_1^2 and e_2^2, we get

$$T = \bar{Y}(1+e_0)\left[W_1\{(1+ae_2+(1-a)e_1)\,(1-\{ae_1+(1-a)e_2\} + \{ae_1+(1-a)e_2\}^2)\} + W_2 \exp\left\{\frac{(1-2a)}{2}(e_1-e_2)\left(1+\frac{(e_2+e_1)}{2}\right)^{-1}\right\}\right]$$

$$= \bar{Y}(1+e_0)\left[W_1\{(1+ae_2+(1-a)e_1)\,(1-\{ae_1+(1-a)e_2\} + \{ae_1+(1-a)e_2\}^2)\} + W_2 \exp\left\{\frac{(1-2a)}{2}(e_1-e_2)\left(1+\frac{(e_2+e_1)}{2}\right)^{-1}\right\}\right]$$

$$= \bar{Y}(1+e_0)\Bigg[W_1\{1-\{ae_2+(1-a)e_1\}+\{ae_1+(1-a)e_2\}^2 - \{ae_1+(1-a)e_2\}^2)\} + W_2 \exp\left\{\frac{(1-2a)}{2}(e_1-e_2)\left\{1-\frac{(e_2+e_1)}{2}+\frac{(e_2+e_1)^2}{4}\right\}\right\}\Bigg]$$

$$\cong \bar{Y}(1+e_0)\Bigg[W_1\{1-\{ae_1+(1-a)e_2\}+\{ae_2+(1-a)e_1\}+\{ae_1+(1-a)e_2\}^2 - \{(ae_2+(1-a)e_1)(ae_1+(1-a)e_2)\})\} + W_2 \exp\left\{\frac{(1-2a)}{2}\left\{(e_1-e_2)-\frac{(e_2-e_1)^2}{2}\right\}\right\}\Bigg]$$

$$\cong \bar{Y}(1+e_0)\Bigg[W_1\{1-\{a(e_1-e_2)+e_2\}+\{a(e_2-e_1)+e_1\}+\{a(e_1-e_2)+e_2\}^2 - \{(a^2e_1e_2+a(1-a)e_2^2+a(1-a)e_1^2+(1-a)^2e_1e_2)\})\} + W_2\left\{1+\frac{(1-2a)}{2}\left\{(e_1-e_2)-\frac{e_1^2-e_2^2}{2}\right\} + \frac{(1-2a)^2}{4}\left\{(e_1-e_2)-\frac{e_1^2-e_2^2}{2}\right\}^2\right\}\Bigg]$$

$$= \bar{Y}(1+e_0)\Bigg[W_1\Big\{1+(e_1-e_2)(1-2a)+e_1^2\{-a(1-2a)\}+e_2^2\{(1-2a)(1-a)\} - e_1e_2\{(1-2a)(1-2a)\}\Big\} + W_2\left\{1+\left\{\frac{(1-2a)}{2}(e_1-e_2)-\frac{(1-2a)}{4}(e_1^2-e_2^2)\right\} + \frac{(1-2a)^2}{8}(e_1^2+e_2^2-2e_1e_2)\right\}\Bigg]$$

$$= \bar{Y}\Big[W_1\{1+(e_1-e_2)(1-2a)+e_1^2\{-a(1-2a)\}+e_2^2\{(1-2a)(1-a)\} - e_1e_2\{(1-2a)(1-2a)\}+e_0+(e_1-e_2)e_0(1-2a)\} + W_2\Big\{1+\frac{(1-2a)}{2}(e_1-e_2) + e_1^2\left\{-\frac{(1-2a)}{4}\left(1-\frac{(1-2a)}{2}\right)\right\}+e_2^2\left\{\frac{(1-2a)}{4}\left(1+\frac{(1-2a)}{2}\right)\right\}+e_1e_2\left(-\frac{(1-2a)^2}{4}\right)+e_0+ \frac{(1-2a)}{2}(e_1-e_2)e_0\Big\}\Big] \quad (2.2)$$

Theorem 2.1. *The conditional bias up to the first order of approximation of the estimator T is given by the estimator is given as*

$$Bias(T) = \bar{Y}\,\{[W_1Q_4 + W_2Q_5] - 1\}$$

Where $Q_4 = 1 + (1-2a)\, f_{rn}(-a\, C_x^2 + \rho C_y C_x)$ and $Q_5 = 1 + \frac{(1-2a)}{2} f_{rn}\left\{-\left(\frac{(1+2a)}{4}\right)C_x^2 + \rho C_y C_x\right\}$

Proof: From (2.2) we have

$$T-\bar{Y}=\bar{Y}\Big[W_1\{1+(e_1-e_2)(1-2a)+e_1^2\{-a(1-2a)\}+e_2^2\{(1-2a)(1-a)\}-e_1e_2\{(1-2a)(1-2a)\}+e_0+(e_1-e_2)e_0(1-2a)\}\ +W_2\Big\{1+\tfrac{(1-2a)}{2}(e_1-e_2)+e_1^2\Big\{-\tfrac{(1-2a)}{4}\Big(1-\tfrac{(1-2a)}{2}\Big)\Big\}+e_2^2\Big\{\tfrac{(1-2a)}{4}\Big(1+\tfrac{(1-2a)}{2}\Big)\Big\}+e_1e_2\Big(-\tfrac{(1-2a)^2}{4}\Big)+e_0+\tfrac{(1-2a)}{2}(e_1-e_2)e_0\Big\}-\bar{Y}\Big] \quad (2.3)$$

Taking expectation on both side we obtain the bias of T to order $O(n^{-1})$ as

$$\begin{aligned}Bias\ (T)&=E(T-\bar{Y})\\&=\bar{Y}\,E\Big\{\Big[W_1\{1+f_rC_x^2\{-a(1-2a)\}+f_nC_x^2\{(1-2a)(1-a)\}\\&\quad-f_nC_x^2\{(1-2a)(1-2a)\}+(f_r\rho C_yC_x-f_n\rho C_y\,C_x)(1-2a)\}\\&\quad+W_2\Big\{1+f_rC_x^2\Big\{-\frac{(1-2a)}{4}\Big(1-\frac{(1-2a)}{2}\Big)\Big\}\\&\quad+f_nC_x^2\Big\{\frac{(1-2a)}{4}\Big(1+\frac{(1-2a)}{2}\Big)\Big\}+f_nC_x^2\Big(-\frac{(1-2a)^2}{4}\Big)\\&\quad+\frac{(1-2a)}{2}(f_r\rho C_yC_x-f_n\rho C_y\,C_x)\Big\}\Big]-1\Big\}\end{aligned}$$

$$=\bar{Y}\ \Big\{\Big[W_1\{1+(1-2a)\,f_{rn}(\,C_x^2\{-a\}+\rho C_yC_x)\}+W_2\Big\{1+\frac{(1-2a)}{2}f_{rn}\Big\{-\Big(\frac{(1+2a)}{4}\Big)C_x^2+\rho C_yC_x\Big\}\Big\}\Big]-1\Big\} \quad (2.4)$$

Letting $Q_4=1+(1-2a)\,f_{rn}(-a\,C_x^2+\rho C_yC_x)$, $Q_5=1+\frac{(1-2a)}{2}f_{rn}\Big\{-\Big(\frac{(1+2a)}{4}\Big)C_x^2+\rho C_yC_x\Big\}$ in eq. (A.1)

$$Bias(T)=\bar{Y}\,\{W_1Q_4+W_2Q_5-1\}$$

5. MEAN SQUARED ERROR OF T

We calculate the mean squared error of T up to order $O(n^{-1})$ by squaring (2.2) and retaining terms up to squares in e_0, e_1 and e_2, and then taking the expectation. This yields the first-degree approximation of the MSE

Theorem 2.2. *The minimum mean square error of the proposed estimator* $\bar{y}_{N1}$ *is given by*

$$MSE(T)=\bar{Y}^2[1+W_1^2Q_1+W_2^2Q_2+2W_1W_2Q_3-2\,W_1Q_4-2W_2Q_5]$$

The optimum values of W_1 and W_2 are:

$$W_{1opt} = \frac{Q_2Q_4 - Q_3Q_5}{Q_1Q_2 - Q_3^2}, \qquad W_{2opt} = \frac{Q_1Q_5 - Q_3Q_4}{Q_1Q_2 - Q_3^2}$$

Where

$$Q_1 = (1 + f_r C_y^2 + f_{rn}(1-2a)\{4\rho_{yx}C_xC_y + (1-4a)C_x^2)$$

$$Q_2 = \left(1 + f_r C_y^2 + f_{rn}(1-2a)(2\rho C_xC_y - aC_x^2)\right)$$

, $$Q_3 = 1 + 3\, f_{rn}(1-2a)\left\{\rho\, C_yC_x + \frac{(1-6a)}{8}\; C_x^2\right\},$$

$$Q_4 = 1 + f_{rn}(1-2a)\;(\rho C_yC_x - a\, C_x^2)$$

$$Q_5 = 1 + f_{rn}\frac{(1-2a)}{2}\left\{\rho C_yC_x - \left(\frac{1+2a}{4}\right)C_x^2\right\}$$

And the minimum MSE of the proposed estimator is given by

$$\min_{W_1W_2} MSE(T) = \bar{Y}^2\left[1 - \frac{Q_2Q_4^2 - 2\,Q_3Q_4Q_5 + Q_1Q_5^2}{(Q_2Q_2 - Q_3^2)}\right]$$

Proof:

$$MSE(T) = E(T-\bar{Y})^2 = \bar{Y}^2E\left\{\left[W_1\{1 + (e_1-e_2)(1-2a) + e_1^2\{-a(1-2a)\} + e_2^2\{(1-2a)(1-a)\} - e_1e_2\{(1-2a)(1-2a)\} + e_0 + (e_1-e_2)e_0(1-2a)\} + W_2\left\{1 + \tfrac{(1-2a)}{2}(e_1-e_2) + e_1^2\left\{-\tfrac{(1-2a)}{4}\left(1 - \tfrac{(1-2a)}{2}\right)\right\} + e_2^2\left\{\tfrac{(1-2a)}{4}\left(1 + \tfrac{(1-2a)}{2}\right)\right\} + e_1e_2\left(-\tfrac{(1-2a)^2}{4}\right) + e_0 + \tfrac{(1-2a)}{2}(e_1-e_2)e_0\right\}\right] - 1\right\}^2 \quad (A.2)$$

Let coefficient of W_1 and W_2 in eq. (A.2) are A and B respectively then $eq.(A.2)$ is

$$MSE(T) = E(T-\bar{Y})^2 = \bar{Y}^2E\{W_1A + W_2B - 1\}^2$$

$$= \bar{Y}^2E[W_1^2A^2 + W_2^2B^2 + 1 + 2W_1W_2A\,B - 2\,W_1A - 2W_2B]$$

$$= \bar{Y}^2[W_1^2E(A^2) + W_2^2E(B^2) + 1 + 2W_1W_2E(A\,B) - 2\,W_1E(A) - 2W_2E(B)] \ldots\ldots.\ldots\ldots \quad (2.5)$$

Now, let

$$Q_1 = E(A^2) = E\left(1 + (e_1-e_2)(1-2a) + e_1^2\{-a(1-2a)\} + e_2^2\{(1-2a)(1-a)\} - e_1e_2\{(1-2a)(1-2a)\} + e_0 + (e_1-e_2)e_0(1-2a)\right)^2$$

$$= E(1 + (e_1-e_2)^2(1-2a)^2 + 2(e_1-e_2)(1-2a) + e_0^2 + 2e_1^2\{-a(1-2a)\} + 2e_2^2\{(1-2a)(1-a)\} - 2e_1e_2\{(1-2a)(1-2a)\} + 2(e_1-e_2)e_0(1-2a))$$

$$= E(1 + 2(e_1-e_2)(1-2a) + e_1^2\{(1-2a)^2 - 2a(1-2a)\} + e_2^2\{(1-2a)^2 + 2(1-2a)(1-a)\} + e_0^2 + 2e_1e_2\{-(1-2a)^2 - (1-2a)^2\} + 2e_0 + 4(e_1-e_2)e_0(1-2a))$$

$$= (1 + (1-2a)f_rC_x^2\{(1-2a)^2 - 2a\} + f_nC_x^2\{(1-2a)^2 + 2(1-2a)(1-a)\} + f_rC_y^2 + 2f_nC_x^2\{-(1-2a)^2 - (1-2a)^2\} + 4\rho_{yx}C_xC_y(1-2a))$$

$$= (1 + f_r C_y^2 + f_r C_x^2\{(1-2a)^2 - 2a(1-2a)\} + f_n C_x^2\{(1-2a)^2 + 2(1-2a)(1 - a) - 4(1-2a)^2\} + 4\rho_{yx} C_x C_y (1-2a))$$

$$= (1 + f_r C_y^2 + f_r C_x^2 (1-2a)\{1-4a\} + f_n C_x^2 (1-2a)\{-(1-4a)\} + 4\rho_{yx} C_x C_y (1-2a))$$

$$Q_1 = (1 + f_r C_y^2 + f_{rn} C_x^2 (1-2a)\{1-4a\} + 4\rho_{yx} C_x C_y (1-2a))$$

$$Q_1 = (1 + f_r C_y^2 + f_{rn} C_x^2 (1-2a)\{1-4a\} + 4\rho_{yx} C_x C_y (1-2a))$$

$$Q_2 = E(B^2) = E\left(1 + \frac{(1-2a)}{2}(e_1 - e_2) + e_1^2\left\{-\frac{(1-2a)}{4}\left(1 - \frac{(1-2a)}{2}\right)\right\} + e_2^2\left\{\frac{(1-2a)}{4}\left(1 + \frac{(1-2a)}{2}\right)\right\} + e_1 e_2\left(-\frac{(1-2a)^2}{4}\right) + e_0 + \frac{(1-2a)}{2}(e_1 - e_2)e_0\right)^2$$

$$= E\left(1 + \frac{(1-2a)^2}{4}(e_1 - e_2)^2 + e_0^2 + 2\frac{(1-2a)}{2}(e_1 - e_2)e_0 + 2\frac{(1-2a)}{2}(e_1 - e_2) + 2e_1^2\left\{-\frac{(1-2a)}{4}\left(1 - \frac{(1-2a)}{2}\right)\right\} + 2e_2^2\left\{\frac{(1-2a)}{4}\left(1 + \frac{(1-2a)}{2}\right)\right\} + 2e_1 e_2\left(-\frac{(1-2a)^2}{4}\right) + 2\, e_0 + 2\frac{(1-2a)}{2}(e_1 - e_2)e_0\right)$$

$$= E\left(1 + e_0^2 + e_1^2\left\{\frac{(1-2a)^2}{4} - 2\frac{(1-2a)}{4}\frac{(1+2a)}{2}\right\} + e_2^2\left\{-\frac{(1-2a)^2}{4} + \frac{(1-2a)(3-2a)}{4}\right\} + 2e_1 e_2\left(-\frac{(1-2a)^2}{4} - \frac{(1-2a)^2}{4}\right) + (e_1 - e_2)e_0\left\{2\frac{(1-2a)}{2} + 2\frac{(1-2a)}{2}\right\} + 2\frac{(1-2a)}{2}(e_1 - e_2) + 2\, e_0\right)$$

$$= \left(1 + f_r C_y^2 + \frac{(1-2a)}{4} f_r C_x^2\{-4a\} + \frac{(1-2a)}{4} f_n C_x^2\{-4a\} + 2 f_{rn} \rho C_x C_y \left\{\frac{(1-2a)}{2} + \frac{(1-2a)}{2}\right\}\right)$$

$$= \left(1 + f_r C_y^2 + (1-2a) f_{rn}\{-a C_x^2 + +2 f_{rn} \rho C_x C_y\right)$$

$$Q_3 = E(AB) = E\Bigg(\Big(1 + (e_1 - e_2)(1-2a) + e_1^2\{-a(1-2a)\} + e_2^2\{(1-2a)(1-a)\} - e_1e_2\{(1-2a)(1-2a)\} + e_0 + (e_1 - e_2)e_0(1-2a)\Big)\ \Big(1 + \frac{(1-2a)}{2}(e_1 - e_2) + e_1^2\left\{-\frac{(1-2a)}{4}\left(1 - \frac{(1-2a)}{2}\right)\right\} + e_2^2\left\{\frac{(1-2a)}{4}\left(1 + \frac{(1-2a)}{2}\right)\right\} + e_1e_2\left(-\frac{(1-2a)^2}{4}\right) + e_0 + \frac{(1-2a)}{2}(e_1 - e_2)e_0\Big)\Bigg)$$

$$= 1 + 3\, f_{rn}(1-2a)\left\{\rho\, C_yC_x + \frac{(1-6a)}{8}\ C_x^2\right\}$$

From previous theorem

$$Q_4 = E(A) = 1 + (1-2a)\, f_{rn}(\rho C_yC_x - a\, C_x^2)$$

,

$$Q_5 = E(B) = 1 + \frac{(1-2a)}{2} f_{rn}\left\{\rho C_yC_x - \left(\frac{1+2a}{4}\right)C_x^2\right\}$$

Placing these values of Q_i^s in (A.2) we get

$$MSE(T) = E(T - \bar{Y})^2 =$$

$$= \bar{Y}^2[1 + W_1^2Q_1 + W_2^2Q_2 + 2W_1W_2Q_3 - 2\,W_1Q_4 - 2W_2Q_5]$$

Differentiating MSE with respect to W_1 and W_2, and equating to zero, we get

$$\frac{\partial}{\partial W_1}MSE(T) = 0 \Rightarrow 2\,W_1Q_1 + 2W_2Q_3 - 2Q_4 = 0$$

$$\frac{\partial}{\partial W_2}MSE(T) = 0 \Rightarrow 2\,W_2Q_2 + 2W_1Q_3 - 2Q_5 = 0$$

On solving these equations, we get

$$W_{1opt} = \frac{Q_2Q_4 - Q_3Q_5}{Q_1Q_2 - Q_3^2}, \qquad W_{2opt} = \frac{Q_1Q_5 - Q_3Q_4}{Q_1Q_2 - Q_3^2} \ldots\ldots\ldots..()$$

And after placing the values of W_1 and W_2 from () to the expression of MSE (?) the minimum MSE expression is

$$\min_{W_1W_2} MSE(T) = \bar{Y}^2\left[1 - \frac{Q_2Q_4^2 - 2\,Q_3Q_4Q_5 + Q_1Q_5^2}{(Q_2Q_2 - Q_3^2)}\right]$$

6. EXPRESSIONS OF MSE FOR DIFFERENT CHOICES OF a

Here we consider the different forms of the proposed estimator for various values of a.

Case 1. $a = \frac{1}{2}$

$$T = \bar{y}_r [W_1 \left\{ \frac{\frac{1}{2}\bar{x}_r + \frac{1}{2}\bar{x}_n}{\frac{1}{2}\bar{x}_n + \frac{1}{2}\bar{x}_r} \right\} + W_2 \exp[0]$$

$$= \bar{y}_r [W_1 + W_2]$$

$$= v\,\bar{y}_r$$

$$bias\,(T)|_{a=1/2} = \bar{Y}\,\{v - 1\}$$

$$MSE(T) = E(T - \bar{Y})^2 =$$

$$= \bar{Y}^2 [1 + W_1^2 Q_1 + W_2^2 Q_2 + 2W_1 W_2 Q_3 - 2\,W_1 Q_4 - 2W_2 Q_5]$$

$$Q_i = 1 + f_r C_y^2, \quad i = 1,2$$

, $$Q_j = 1, \qquad j = 3,4,5$$

$$MSE(T) = \bar{Y}^2 [1 + W_1^2 (1 + f_r C_y^2) + W_2^2 (1 + f_r C_y^2) + 2W_1 W_2 - 2\,W_1 - 2W_2]$$

$$= \bar{Y}^2 [1 + (W_1^2 + W_2^2)(1 + f_r C_y^2) + 2W_1 W_2 - 2\,W_1 - 2W_2]$$

$$W_{1opt} = \frac{1}{2 + f_r C_y^2}, \qquad W_{2opt} = \frac{1}{2 + f_r C_y^2}$$

$$MSE(T) = \bar{Y}^2 \left[1 + 2\left(\frac{1}{2 + f_r C_y^2}\right)^2 (1 + f_r C_y^2) + 2\left(\frac{1}{2 + f_r C_y^2}\right)^2 - 2\,\frac{1}{2 + f_r C_y^2} - 2\,\frac{1}{2 + f_r C_y^2} \right]$$

$$MSE(T) = \bar{Y}^2 \left[\frac{f_r C_y^2}{2 + f_r C_y^2} \right] = f_r C_y^2 \bar{Y}^2 \left[\frac{1}{2 + f_r C_y^2} \right] = f_r S_y^2 \left[\frac{\bar{Y}^2}{2\,\bar{Y}^2 + f_r S_y^2} \right]$$

Which is better than the mean estimator $\bar{y}_r$ in terms of efficiency as

$$MSE(T) - MSE(\bar{y}) = f_r S_y^2 - f_r S_y^2 \left[\frac{\bar{Y}^2}{2\,\bar{Y}^2 + f_r S_y^2} \right] > 0$$

And if $W_1 + W_2 = 1$, The proposed estimator at $a = \frac{1}{2}$ is equivalent to the mean estimator$\bar{y}_r$

Case 2. $a = 1$

$$T = \bar{y}_r \left[W_1 \left\{ \frac{\bar{x}_n}{\bar{x}_r} \right\} + W_2 \exp\left\{ \frac{-1(\bar{x}_r - \bar{x}_n)}{\bar{x}_n + \bar{x}_r} \right\} \right]$$

$$T = \bar{y}_r\left[W_1\left\{\frac{\bar{x}_n}{\bar{x}_r}\right\} + W_2\exp\left\{\frac{\bar{x}_n - \bar{x}_r}{\bar{x}_n + \bar{x}_r}\right\}\right]$$

$$Bias(T) = \bar{Y}\left\{\left[W_1\{1 - f_{rn}(-C_x^2 + \rho C_y C_x)\} + W_2\left\{1 - \frac{1}{2}f_{rn}\left\{-\left(\frac{3}{4}\right)C_x^2 + \rho C_y C_x\right\}\right\}\right] - 1\right\}$$

$$Q_1 = (1 + f_r C_y^2 - f_{rn}\{4\rho_{yx}C_xC_y - 3C_x^2)$$

$$Q_2 = \left(1 + f_r C_y^2 - f_{rn}(2\rho C_x C_y - C_x^2\right)$$

, $$Q_3 = 1 - 3\,f_{rn}\left\{\rho\,C_yC_x - \tfrac{5}{8}\,C_x^2\right\},$$

$$Q_4 = 1 - f_{rn}(\rho C_y C_x - C_x^2)$$

$$Q_5 = 1 - f_{rn}\frac{1}{2}\left\{\rho C_y C_x - \frac{3}{4}C_x^2\right\}$$

$$MSE(T) = \bar{Y}^2[1 + W_1^2Q_1 + W_2^2Q_2 + 2W_1W_2Q_3 - 2\,W_1Q_4 - 2W_2Q_5]$$

$$W_{1opt} = \frac{Q_2Q_4 - Q_3Q_5}{Q_1Q_2 - Q_3^2}, \qquad W_{2opt} = \frac{Q_1Q_5 - Q_3Q_4}{Q_1Q_2 - Q_3^2}$$

For $W_1 = 1, W_2 = 0$

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

$$Bias(T) = \bar{Y}\left\{f_{rn}(C_x^2 - \rho C_y C_x)\right\}$$

$$MSE(T) = E(T - \bar{Y})^2 =$$

$$= \bar{Y}^2[1 + Q_1 - 2Q_4] = \bar{Y}^2\left[f_rC_y^2 + f_{rn}\{C_x^2 - 2\rho_{yx}C_xC_y)\right]$$

For $W_1 = 0, W_2 = 1$

$$T = \exp\left\{\frac{\bar{x}_n - \bar{x}_r}{\bar{x}_n + \bar{x}_r}\right\}$$

$$Bias(T) = \bar{Y}\left\{\frac{1}{2}f_{rn}\left\{\left(\frac{3}{4}\right)C_x^2 - \rho C_y C_x\right\}\right\}$$

$$MSE(T) = \bar{Y}^2[1 + Q_2 - 2Q_5] = \bar{Y}^2\left[f_rC_y^2 + f_{rn}\left(-\rho C_xC_y + \frac{C_x^2}{4}\right)\right]$$

Case 3. $a = 0$

$$T = \bar{y}_r\left[W_1\left\{\frac{\bar{x}_r}{\bar{x}_n}\right\} + W_2\exp\left\{\frac{(\bar{x}_r - \bar{x}_n)}{(\bar{x}_n + \bar{x}_r)}\right\}\right]$$

A linear combination of product estimator and product type exponential estimator,

$$Q_4 = 1 + f_{rn}\,\rho C_y C_x,\ Q_5 = 1 + \tfrac{1}{2}f_{rn}\left\{-\left(\tfrac{1}{4}\right)C_x^2 + \rho C_y C_x\right\}$$

$$Bias(T) = \bar{Y}\{W_1Q_4 + W_2Q_5 - 1\}$$

$$MSE(T) = \bar{Y}^2[1 + W_1^2Q_1 + W_2^2Q_2 + 2W_1W_2Q_3 - 2\,W_1Q_4 - 2W_2Q_5]$$

$$Q_1 = (1 + f_rC_y^2 + f_{rn}\{4\rho_{yx}C_xC_y + C_x^2)$$

,
$$Q_2 = \left(1 + f_r C_y^2 + 2 f_{rn} \rho C_x C_y\right)$$
$$Q_3 = 1 + 3\, f_{rn} \left\{\rho\, C_y C_x + \frac{1}{8}\, C_x^2\right\},$$
$$Q_4 = 1 + f_{rn}\, \rho C_y C_x$$
$$Q_5 = 1 + f_{rn} \frac{1}{2}\left\{\rho C_y C_x - \left(\frac{1}{4}\right) C_x^2\right\}$$

and further

For $W_1 = 1, W_2 = 0$ it reduces to product estimator

$$T = \bar{y}_r \frac{\bar{x}_r}{\bar{x}_n}$$
$$Bias(T) = \bar{Y}\left\{f_{rn}\, \rho C_y C_x\right\}$$
$$MSE(T) = \bar{Y}^2 [f_r C_y^2 + f_{rn}\{C_x^2 + 2 f_{rn}\, \rho C_y C_x\,]$$

and for $W_1 = 0, W_2 = 1$ it reduces to product type exponential estimator

$$T = \bar{y}_r \exp\left\{\frac{(\bar{x}_r - \bar{x}_n)}{(\bar{x}_n + \bar{x}_r)}\right\}$$
$$Bias(T) = \bar{Y}\left\{f_{rn} \frac{1}{2}\left\{\rho C_y C_x - \left(\frac{1}{4}\right) C_x^2\right\}\right\}$$
$$MSE(T) = \bar{Y}^2 [f_r C_y^2 + 2 f_{rn} \rho C_x C_y - f_{rn}\left\{\rho C_y C_x - \left(\frac{1}{4}\right) C_x^2\right\}]$$
$$MSE(\bar{y}_{AE}) = f_r \bar{Y}^2 \left(C_y^2 + \frac{1}{4}\, C_x^2 + \rho C_y C_x\right)$$

7. EMPIRICAL STUDY

For the purpose of comparison of the proposed estimator we conducted the empirical study and computed the Percentage relative efficiency (PRE) of the estimators $\bar{y}_m, \bar{y}_{RAT}, \bar{y}_{COMP}, \bar{y}_{INP}$ for (i) real data and (ii) artificially generated data.

The empirical study has been carried out to illustrate and compare the performance of the proposed imputation methods with the existing conventional imputation methods and the method proposed by Diana and Perri (2010) and Bhushan and Pandey (2010) utilizing Searl () constant with the Diana and Perri (2010) estimator for (i) real data described in Horvitz and Thompson (1952), Singh (2003), described in Table 1, Table 2.

Table 1

Table 1 Description of Populations considered for empirical study		
Parameters	**Population 1**	**Population 2**
Description	Horvitz and Thompson (1952), y: no. of houses on i th block. x: eye estimate of no. of houses on i th block.	Singh (2003, pp 1114), the season-average price for commercial crop: season average price in $ per pound, by states,1994-96; y: seasons average price per pound during 1996; x: season average price per pound during 1995
N	20	36
n	7	18
r	5	8

$\bar{Y}$	21.15	0.2033
$\bar{X}$	19.7	0.1856
S_y^2	94.028947	0.006458
S_x^2	61.8	0.005654
ρ_{yx}	0.8526154	0.8775

The mean square errors (MSEs) and percent relative efficiencies (PREs) are calculated for the methods considered with respect to the mean method of imputation ($\bar{y}_r$). The methods

1) Mean Method of imputation, $\bar{y}_r = T_0$
2) Ratio Method of imputation, $\bar{y}_{RAT} = T_1$
3) Diana and Perri method of imputation $T_{DP_1} = T_2$
4) Diana and Perri method of imputation $T_{DP_2} = T_3$
5) Diana and Perri method of imputation $T_{DP_3} = T_4$
6) Proposed method of imputation, $T = T_5$

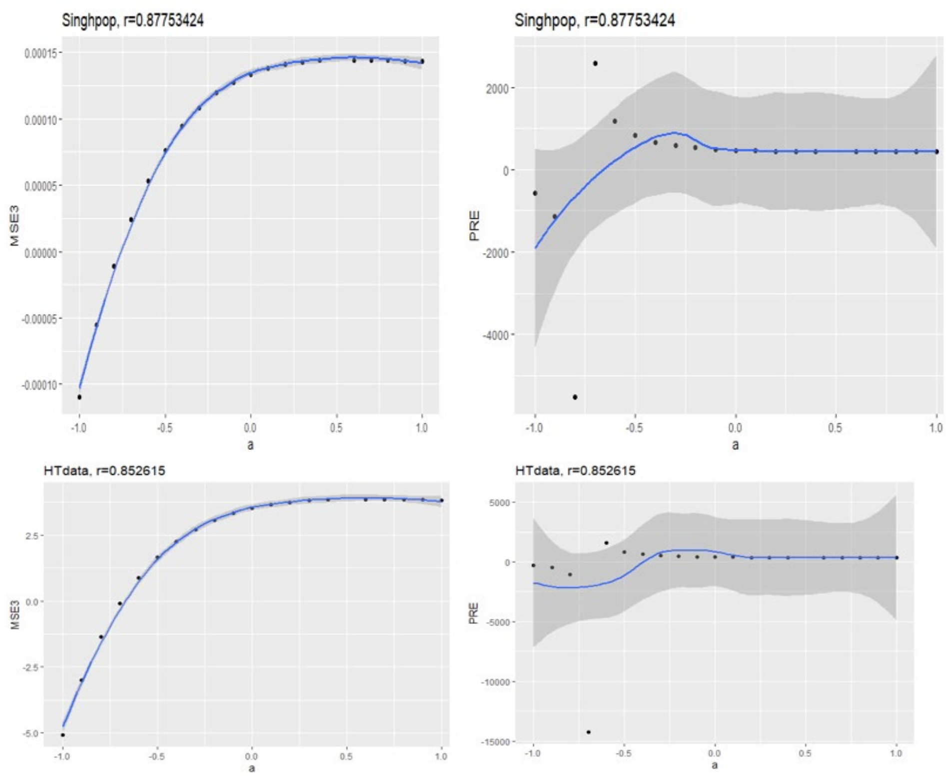
$$PRE = \frac{MSE(T_0)}{MSE(T_i)} \times 100, \quad i = 1, 2, 3,4,5,$$

Table 2

Table 2 Mean square error and percent relative efficiency based on Population I(Singhpop)

Estimator		MSE	PRE	MSE	PRE
$T_0 = m1$		0.000627874	100	14.1043421	100
$T_1 = \bar{Y}_{RAT}$		0.00029219	214.8858518	10.2000587	138.2771
$T_2 = DP_1$		0.00048973	128.2082427	7.75712694	181.8243
$T_3 = DP_2$		0.000144369	434.9080453	3.85114837	366.2373
$T_4 = DP_3$		0.000282514	222.2456233	10.1983635	138.3001
T_5	a	MSE and PRE of Proposed Estimator for different choices of a			
	-1	*-0.000109667	*-572.5264505	*-5.07540636	*-277.8958
	-0.9	*-5.54806E-05	*-1131.701446	*-2.98949277	*-471.7972
	-0.8	*-1.13765E-05	*-5519.055982	*-1.36603234	*-1032.5043
	-0.7	2.43566E-05	2577.844737	*-0.09897433	*-14250.5046
	-0.6	5.31303E-05	1181.763804	0.89001809	1584.7253
	-0.5	7.61179E-05	824.8708681	1.66003196	849.6428
	-0.4	9.4301E-05	665.8194857	2.25644120	625.0702
	-0.3	0.000108506	578.656658	2.71463222	519.5673
	-0.2	0.00011943	525.7257808	3.06257328	460.5389
	-0.1	0.000127667	491.8054912	3.32262502	424.4939
	0	0.000133722	469.5370246	3.51284069	401.5082
	0.1	0.000138025	454.8999563	3.64791560	386.6411
	0.2	0.000140943	445.4819583	3.73989064	377.1325

0.3	0.000142789	439.7214493	3.79868012	371.2959
0.4	0.00014383	436.5391782	3.83247219	368.0220
0.5	NaN	#VALUE!	NaN	-
0.6	0.000144359	434.9406963	3.85095252	366.2559
0.7	0.00014419	435.4483623	3.84580584	366.7461
0.8	0.000143911	436.2925635	3.83631355	367.6535
0.9	0.00014362	437.1774822	3.82543836	368.6987
1	0.000143389	437.8833031	3.81546698	369.6623



8. INTERPRETATIONS OF THE COMPUTATIONAL RESULTS

The following interpretations may be read out from above Tables:

1) For the real populations, HT data and Singh Population where the correlation between y and x is 0.852615 and 0.877534, the results are shown in Table 1. It is clear that the proposed imputation method T is superior to all the imputation methods i.e., $\bar{y}$, $\bar{y}_{RAT}$ and the imputation methods suggested by Diana and Perri DP_1, DP_2, DP_3 for a wide choice of the constant a for the proposed estimator.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Heitjan, D.F. and Basu, S. (1996). Distinguishing 'Missing at Random' and 'Missing Completely at Random'. The American Statistician. 50(3), 207-213. https://doi.org/10.1080/00031305.1996.10474381.

Horvitz, D. G. and Thompson, D.J. (1952). A Generalization of Sampling Without Replacement From a Finite Universe. Journal of the American Statistical Association, 47(260), 663-685. https://doi.org/10.1080/01621459.1952.10483446.

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

Rubin, R. B. (1978). Multiple Imputation for Nonresponse in Surveys. New York : John Wiley.

Singh, S. (2003). Advanced Sampling Theory with Applications. How Michael Selected Amy. Kluwer, Dordrecht, 1(2). https://doi.org/10.1007/978-94-007-0789-4.

Singh, S. and Horn, S. (2000). Compromised Imputation in Survey Sampling. Metrika, 51, 267-276. https://doi.org/10.1007/s001840000054.