Revista Colombiana de Estadística - Theorical Statistics
January 2023, volume 46, issue 1, pp. 111–127
http://doi.org/10.15446/rce.v46n1.102308

Imputation of Missing Data Through Product Type Exponential Methods in Sampling Theory

Imputación de datos faltantes a través de métodos exponenciales de tipo de producto en la teoría del muestreo

Shakti Prasad[a], Vinay Kumar Yadav[b]

Department of Basic and Applied Science, National Institute of Technology Arunachal Pradesh, Jote, India

Abstract

Some efficient product type exponential imputation methods are proposed in this article to tackle the problem of incomplete values in sampling theory. To investigate the effectiveness of proposed exponential methods, the behaviours of the considered estimators are compared in two scenarios: with and without nonresponse. The simulation studies show that the proposed resultant estimators outperform other existing estimators in this literature.

Key words: Auxiliary variable; Product type estimator; Imputation.

Resumen

En este artículo se proponen algunos métodos eficientes de imputación exponencial de tipo de producto para abordar el problema de los valores incompletos en la teoría del muestreo. Para investigar la efectividad de los métodos exponenciales propuestos, se comparan los comportamientos de los estimadores considerados en dos escenarios: con y sin falta de respuesta. Los estudios de simulación muestran que los estimadores resultantes propuestos superan a otros estimadores existentes en esta literatura.

Palabras clave: Variable auxiliar; Estimador de tipo de producto; Imputación.

[a]Assistant Professor. E-mail: shakti.pd@gmail.com
[b]Research Scholar. E-mail: vkyadavbhu@gmail.com

1. Introduction

In sampling theory, imputation is an appealing strategy for analysing missing data. In order to successfully handle the missing values in survey data, Sande (1979) selected imputation strategies that aggregate missing data sets consistently and make simple and direct analysis. In a very nice way Heitjan & Basu (1996) distinguished, the meanings of MAR (missing at random) and MCAR (missing completely at random) were delineated. Hyunshik Lee & Särndal (1994), Lee et al. (1995) for purpose of the imputation, employed the auxiliary information. Using MCAR response mechanism, Ahmed et al. (2006), Singh & Horn (2000), Toutenburg et al. (2008), Singh (2009), Singh et al. (2010), Singh et al. (2016), Gira (2015), Kadilar & Cingi (2008), Diana & Francesco Perri (2010), Prasad (2017), Prasad (2018*a*), Prasad (2018*b*), Prasad (2019), Prasad (2021) have proposed numerous new imputation techniques in sampling theory.

If the study and auxiliary variables have a negative coefficient of correlation, then the product methods of imputation have been used for missing data. Singh & Deo (2003), consider the product approach of imputation to be extremely effective. There are Multiple social science or medical variables that decreases as people grow older. For examples, when peoples get older, then following factors exhibit a negative correlations with the age: (a) visual acuity, (b) head hair density, (c) hearing capabilities, and so on. In sampling theory, the product imputation technique is useful if in-formations on any of the studied variable is unavailable but the ages of the subjects is accessible.

Motivated by the above work of Singh & Deo (2003), to deal with the problem of incomplete values in sampling theory, and four efficient product type exponential imputation approaches have been proposed. The proposed estimator's behaviour is examined in two conditions (with and without nonresponse), and conclusions are drawn.

The remaining sections of the manuscript are organised as follows: We provided the various notations used in this manuscript in Section 2. In Section 3, we looked at a review of existing methods in the literature. In Section 4, we proposed four product-type exponential estimators and derived expressions for their bias and MSE. In Sections 5 and 6, we conducted a numerical illustration and simulation study. Section 7 contains an analysis of the numerical illustration and simulation study. We arrived at conclusions in Section 8.

2. Notations

Suppose that the $U = (U_1, U_2, \ldots, U_N)$ be the finite population of size N and the variables under study and auxiliary are denoted by y and x respectively. Let $(\bar{Y}, \bar{X})$ be the population means of (y, x) respectively. We define some parameters that are used in this manuscript.

N: Population size.
n: Sample size.

r: Responding unit
Y: Study Variable.
X: Auxiliary Variable.
$\bar{Y}$: The Population mean of Study Variable.
$\bar{X}_N$: The Population mean of Auxiliary Variable.
$\bar{x}_r$: The Sample mean of responding unit of r.
A: Number of responding unit of r.
A^c: Number of nonresponding unit of $(n-r)$.
C_y, C_x: Coefficient of variation of variable y and x respectively.
$\beta_1(x)$: Coefficient of skewness.
$\beta_2(x)$: Coefficient of kurtosis.

3. Some Existing Methods

Some imputation methods that are commonly used are as follows.

3.1. Mean Imputation Approach

After imputation, data in this technique have the following structure:

$$y_{.i} = \begin{cases} y_i, & i \epsilon A \\ \bar{y}_r, & i \epsilon A^c \end{cases} \tag{1}$$

Under this approach, the resultant point estimator of $\bar{Y}$ takes the form

$$\bar{y}_r = \frac{1}{r} \sum_{i=1}^{r} y_i \tag{2}$$

And variance of response sample mean $\bar{y}_r$, is calculated using the following:

$$Var(\bar{y}_r) = \left(\frac{1}{r} - \frac{1}{N} \right) \bar{Y}^2 C_y^2 \tag{3}$$

where $\bar{y}_r$ and C_y are the response sample mean and correspondingly, the coefficients of variation having the studied variable y.

3.2. Singh and Deo (2003) Approach

In sampling theory, Singh & Deo (2003) proposed the product imputation approach. After imputation, data in this approach take the following structure:

$$y_{.i} = \begin{cases} y_i, & i \epsilon A \\ \bar{y}_r \left[\frac{n\bar{x}_r - r\bar{x}_n}{\bar{x}_n} \right] \frac{x_i}{\sum_{i \epsilon R^c} x_i}, & i \epsilon A^c \end{cases} \tag{4}$$

In this method, the resultant point estimator of $\bar{Y}$ becomes

$$\bar{y}_{SD} = \frac{\bar{y}_r}{\bar{x}_n}\bar{x}_r \tag{5}$$

where $\bar{x}_r$ and $\bar{x}_n$ having response and sample mean of variables x respectively.

The MSE of this estimator, is as follows

$$MSE(\bar{y}_{SD}) = Var(\bar{y}_r) + \left(\frac{1}{r} - \frac{1}{n}\right)\left(C_x^2 + 2\rho C_y C_x\right)\bar{Y}^2 \tag{6}$$

where C_x, C_y and ρ are the coefficient of variation of variable x, coefficient of variation of variable y and correlation coefficient between the variables y and x simultaneously.

4. Proposed Imputation Methods

With auxiliary variables X and study variables Y, a finite population of size N was considered. In this article, let the values of the auxiliary variables be negatively correlated with the values of the study variates. We have population coefficient of skewness $\beta_1(x)$ and coefficient of kurtosis $\beta_2(x)$ of an auxiliary variables are supposed to be known. Now, random sample of size n is drawn according to the SRSWOR schemes process to estimate the population mean of the study variables $\bar{Y}$.

Assume that the nonresponse is random and that the numbers of responding unit out of a sample of n unit is represented by r. Let A represent the number of responding unit and A^c denote the number of nonresponding unit.

Whenever the nonresponse observation are discarded, it is customary to estimates the population mean $\bar{Y}$, is given in equation (2).

When the nonresponse observation are not discarded and We have some imputation Approach are followed,then the completes data set is as:

$$y_{.i} = \begin{cases} y_i, & i\epsilon A \\ \tilde{y_i}, & i\epsilon A^c \end{cases} \tag{7}$$

The point estimators of the population mean takes the form:

$$\tau = \frac{1}{n}\sum_{i=1}^{n} y_{.i} = \frac{1}{n}\left[\sum_{i\epsilon R} y_i + \sum_{i\epsilon R^c}\tilde{y_i}\right] \tag{8}$$

Where, $\tilde{y_i}$ represent the imputed values of study variable corresponding to the i^{th} nonresponding unit.

If product type exponential imputation methods are imputed for missing values of study variate, there are four simple choices of $\tilde{y_i}$ in equation (8), we have

$$\tilde{y_i} = \frac{\bar{y}_r}{n-r}\left[n\lambda_1 \exp\left(\theta_1 \frac{\left(\bar{x}_r/\bar{X}_N - 1\right)}{1 + \left(\left(\beta_1(x)\bar{x}_r + 1\right)/\left(\beta_1(x)\bar{X}_N + 1\right)\right)}\right) - r\right] \tag{9}$$

$$\tilde{y_i} = \frac{\bar{y}_r}{n-r}\left[n\lambda_2 \exp\left(\theta_2 \frac{\left(\bar{x}_r/\bar{X}_N - 1\right)}{1+\left((\beta_2(x)\bar{x}_r+1)/(\beta_2(x)\bar{X}_N+1)\right)}\right) - r\right] \tag{10}$$

$$\tilde{y_i} = \frac{\bar{y}_r}{n-r}\left[n\lambda_3 \exp\left(\theta_3 \frac{\left(\bar{x}_r/\bar{X}_N - 1\right)}{1+\left((\beta_1(x)\bar{x}_r+\beta_2(x))/(\beta_1(x)\bar{X}_N+\beta_2(x))\right)}\right) - r\right] \tag{11}$$

$$\tilde{y_i} = \frac{\bar{y}_r}{n-r}\left[n\lambda_4 \exp\left(\theta_4 \frac{\left(\bar{x}_r/\bar{X}_N - 1\right)}{1+\left((\beta_2(x)\bar{x}_r+\beta_1(x))/(\beta_2(x)\bar{X}_N+\beta_1(x))\right)}\right) - r\right] \tag{12}$$

where $\theta_1 = \frac{\beta_1(x)\bar{X}_N}{\beta_1(x)\bar{X}_N+1}$, $\theta_2 = \frac{\beta_2(x)\bar{X}_N}{\beta_2(x)\bar{X}_N+1}$, $\theta_3 = \frac{\beta_1(x)\bar{X}_N}{\beta_1(x)\bar{X}_N+\beta_2(x)}$, $\theta_4 = \frac{\beta_2(x)\bar{X}_N}{\beta_2(x)\bar{X}_N+\beta_1(x)}$.

Utilizing (9-12) in (8), we obtain the following four suggested proposed estimators for estimating the population mean $\bar{Y}$:

$$\tau_1 = \lambda_1 \bar{y_r} \exp\left(\theta_1 \frac{\left(\bar{x}_r/\bar{X}_N - 1\right)}{1+\left((\beta_1(x)\bar{x}_r+1)/(\beta_1(x)\bar{X}_N+1)\right)}\right) \tag{13}$$

$$\tau_2 = \lambda_2 \bar{y_r} \exp\left(\theta_2 \frac{\left(\bar{x}_r/\bar{X}_N - 1\right)}{1+\left((\beta_2(x)\bar{x}_r+1)/(\beta_2(x)\bar{X}_N+1)\right)}\right) \tag{14}$$

$$\tau_3 = \lambda_3 \bar{y_r} \exp\left(\theta_3 \frac{\left(\bar{x}_r/\bar{X}_N - 1\right)}{1+\left((\beta_1(x)\bar{x}_r+\beta_2(x))/(\beta_1(x)\bar{X}_N+\beta_2(x))\right)}\right) \tag{15}$$

$$\tau_4 = \lambda_4 \bar{y_r} \exp\left(\theta_4 \frac{\left(\bar{x}_r/\bar{X}_N - 1\right)}{1+\left((\beta_2(x)\bar{x}_r+\beta_1(x))/(\beta_2(x)\bar{X}_N+\beta_1(x))\right)}\right) \tag{16}$$

where λ_1, λ_2, λ_3 and λ_4 are an unknown real constants to be determines by minimization of the Mean Square Error (MSE)of the suggested estimators.

To acquire Bias and MSE of the suggested estimators, let's defines $(\frac{\bar{y}_r}{\bar{Y}} - 1) = \varepsilon_0$ and $(\frac{\bar{x}_r}{\bar{X}} - 1) = \varepsilon_1$ such that $E(\varepsilon_m) = 0$, $|\varepsilon_m| < 1 \forall\ m = 0, 1$.

Under the above transformation, the Bias and MSE of the suggested estimators τ_i (where $i = 1, 2, 3, 4$) are derived up-to the first order of large sample approximations are follows: the suggested estimators τ_i (where $i = 1, 2, 3, 4$) take the following form:

$$\tau_i = \lambda_i \bar{Y}\left(1+\varepsilon_0\right) exp\left[\frac{1}{2}\theta_i\varepsilon_1\left(1+\frac{1}{2}\theta_i\varepsilon_1\right)^{-1}\right] \tag{17}$$

Neglecting the higher power terms of $\varepsilon' s$, the equation(17) can be written as

$$\tau_i - \bar{Y} \cong \bar{Y}\left[(\lambda_i - 1) + \lambda_i\left(\varepsilon_0 + \frac{1}{2}\theta_i\varepsilon_1 + \frac{1}{2}\theta_i\varepsilon_0\varepsilon_1 - \frac{1}{8}\theta_i^2\varepsilon_1^2\right)\right] \tag{18}$$

Taking expectation on both sides of (18), we obtained the bias of the suggested estimators, are given as

$$\begin{aligned} Bias(\tau_i) &= E\left(\tau_i - \bar{Y}\right) \\ &= \bar{Y}\left[(\lambda_i - 1) - \frac{1}{8}\theta_i C_x \lambda_i \left(\frac{1}{r} - \frac{1}{N}\right)(\theta_i C_x - 4\rho_{yx} C_y)\right] \end{aligned} \tag{19}$$

Now, after squaring of (18) and neglecting the higher power terms of ε's, we have

$$\left(\tau_i - \bar{Y}\right)^2 \cong \bar{Y}^2\left[(\lambda_i - 1) + \lambda_i\left(\varepsilon_0 + \frac{1}{2}\theta_i\varepsilon_1 + \frac{1}{2}\theta_i\varepsilon_0\varepsilon_1 - \frac{1}{8}\theta_i^2\varepsilon_1^2\right)\right]^2 \tag{20}$$

Taking expectation on both sides of (20), we get the MSEs of the suggested estimators $\tau_i (i = 1, 2, 3, 4)$ as

$$MSE(\tau_i) = E\left(\tau_i - \bar{Y}\right)^2 = \bar{Y}^2\left[(\lambda_i - 1)^2 + \lambda_i^2 A_1 + 2(\lambda_i^2 - \lambda_i)B_1\right] \tag{21}$$

where

$$A_1 = \left(\frac{1}{r} - \frac{1}{N}\right)\left(C_y^2 + \frac{1}{4}\theta_i^2 C_x^2 + \theta_i\rho_{yx}C_y C_x\right),$$

$$B_1 = -\frac{1}{8}\theta_i C_x\left(\frac{1}{r} - \frac{1}{N}\right)(\theta_i C_x - 4\rho_{yx}C_y).$$

Taking partial derivatives of equation (21) with respect to λ_i and its equating to zero, we can get the optimum values of λ_i are as follows:

$$\lambda_{iopt} = \frac{1 - \frac{1}{8}\theta_i C_x\left(\frac{1}{r} - \frac{1}{N}\right)(\theta_i C_x - 4\rho_{yx}C_y)}{1 + \left(\frac{1}{r} - \frac{1}{N}\right)C_y\left(C_y + 2\theta_i\rho_{yx}C_x\right)} \tag{22}$$

After the optimum values have been substituted of λ_i i.e., λ_{iopt} in equations (21), we obtained the minimum MSEs of the proposed estimators $\tau_i (i = 1, 2, 3, 4)$ are as follows:

$$MSE(\tau_i)_{opt} = \left[1 - \frac{\left(1 - \frac{1}{8}\theta_i C_x\left(\frac{1}{r} - \frac{1}{N}\right)(\theta_i C_x - 4\rho_{yx}C_y)\right)^2}{1 + \left(\frac{1}{r} - \frac{1}{N}\right)C_y\left(C_y + 2\theta_i\rho_{yx}C_x\right)}\right]\bar{Y}^2 \tag{23}$$

5. Numerical Illustration

In order to determine the numerical efficiencies, we have considered the three real data sets (given in Table 1) for taking the sample population of size (n) between 33% to 40% and response rate (r), are between 60% to 92% and performs some numerical study to evaluates performance of the considered estimators.

Now, the percent relative losses in efficiency of the suggested estimators τ_i $(i = 1, 2, 3, 4)$ in the respect to the another suggested estimators ξ_i $(i = 1, 2, 3, 4)$ for the similar circumstances but under the complete response have been obtained to study the effects of nonresponse on the precision of estimates under sampling theory. Another suggested estimators ξ_i $(i = 1, 2, 3, 4)$ is defined under the same circumstance as the estimators τ_i $(i = 1, 2, 3, 4)$, but in the absences of nonresponse and shown by

$$\xi_1 = \phi_1 \bar{y}_n \exp\left(\theta_1 \frac{\left(\bar{x}_n/\bar{X}_N - 1\right)}{1 + \left((\beta_1(x)\bar{x}_n + 1)/(\beta_1(x)\bar{X}_N + 1)\right)}\right) \tag{24}$$

$$\xi_2 = \phi_2 \bar{y}_n \exp\left(\theta_2 \frac{\left(\bar{x}_n/\bar{X}_N - 1\right)}{1 + \left((\beta_2(x)\bar{x}_n + 1)/(\beta_2(x)\bar{X}_N + 1)\right)}\right) \tag{25}$$

$$\xi_3 = \phi_3 \bar{y}_n \exp\left(\theta_3 \frac{\left(\bar{x}_n/\bar{X}_N - 1\right)}{1 + \left((\beta_1(x)\bar{x}_n + \beta_2(x))/(\beta_1(x)\bar{X}_N + \beta_2(x))\right)}\right) \tag{26}$$

$$\xi_4 = \phi_4 \bar{y}_n \exp\left(\theta_4 \frac{\left(\bar{x}_n/\bar{X}_N - 1\right)}{1 + \left((\beta_2(x)\bar{x}_n + \beta_1(x))/(\beta_2(x)\bar{X}_N + \beta_1(x))\right)}\right) \tag{27}$$

where ϕ_1, ϕ_2, ϕ_3 and ϕ_4 are an unknown real constants to be determined after the minimization of the mean squared error of the estimators ξ_i $(i = 1, 2, 3, 4)$. Following the methods discussed in Section 4, the bias and MSE of the another considered estimators ξ_i $(i = 1, 2, 3, 4)$ are obtained as

$$\begin{aligned} Bias(\xi_i) &= E\left(\xi_i - \bar{Y}\right) \\ &= \bar{Y}\left[(\phi_i - 1) - \frac{1}{8}\theta_i C_x \phi_i \left(\frac{1}{n} - \frac{1}{N}\right)(\theta_i C_x - 4\rho_{yx} C_y)\right] \end{aligned} \tag{28}$$

$$MSE(\xi_i) = E\left(\xi_i - \bar{Y}\right)^2 = \bar{Y}^2\left[(\phi_i - 1)^2 + \phi_i^2 A_2 + 2(\phi_i^2 - \phi_i) B_2\right] \tag{29}$$

where

$$A_2 = \left(\frac{1}{n} - \frac{1}{N}\right)\left(C_y^2 + \frac{1}{4}\theta_i^2 C_x^2 + \theta_i \rho_{yx} C_y C_x\right),$$

$$B_2 = -\frac{1}{8}\theta_i C_x \left(\frac{1}{n} - \frac{1}{N}\right)(\theta_i C_x - 4\rho_{yx} C_y).$$

Taking partial derivatives of equation (29) with respect to ϕ_i and equating to zero, we get the optimum values of ϕ_i are given by

$$\phi_{i_{opt}} = \frac{1 - \frac{1}{8}\theta_i C_x \left(\frac{1}{n} - \frac{1}{N}\right)(\theta_i C_x - 4\rho_{yx} C_y)}{1 + \left(\frac{1}{n} - \frac{1}{N}\right) C_y \left(C_y + 2\theta_i \rho_{yx} C_x\right)} \tag{30}$$

Using the optimum values of ϕ_i i.e., $\phi_{i_{opt}}$ in equations (29), we obtains the minimum MSE of the suggested estimators ξ_i $(i = 1, 2, 3, 4)$ as given by

$$MSE(\xi_i)_{opt} = \left[1 - \frac{\left(1 - \frac{1}{8}\theta_i C_x \left(\frac{1}{n} - \frac{1}{N}\right)(\theta_i C_x - 4\rho_{yx} C_y)\right)^2}{1 + \left(\frac{1}{n} - \frac{1}{N}\right) C_y \left(C_y + 2\theta_i \rho_{yx} C_x\right)}\right]\bar{Y}^2 \tag{31}$$

For taking different choices of sample population and response rate, the percent relative losses (PRL_i $(i = 1, 2, 3, 4)$) in precision of the suggested estimators

τ_i $(i = 1, 2, 3, 4)$ are computed in the respect to the another suggested estimators ξ_i $(i = 1, 2, 3, 4)$ respectively and shown in Table 2, where $PRL_1 = \frac{MSE(\tau_1) - MSE(\xi_1)}{MSE(\tau_1)} \times 100$, $PRL_2 = \frac{MSE(\tau_2) - MSE(\xi_2)}{MSE(\tau_2)} \times 100$, $PRL_3 = \frac{MSE(\tau_3) - MSE(\xi_3)}{MSE(\tau_3)} \times 100$, and $PRL_4 = \frac{MSE(\tau_4) - MSE(\xi_4)}{MSE(\tau_4)} \times 100$. Now, PREs of the suggested estimators $\tau_i (i = 1, 2, 3, 4)$ and ξ_i $(i = 1, 2, 3, 4)$ in the respect to the mean imputation approach and Singh & Deo (2003) estimators are computed as $PRE_i = PRE(\tau_i, \bar{y}_r) = \frac{Var(\bar{y}_r)}{MSE(\tau_i)} \times 100$, $(i = 1, 2, 3, 4)$ $PRE_i = PRE(\tau_i, \bar{y}_{SD}) = \frac{MSE(\bar{y}_{SD})}{MSE(\tau_i)} \times 100$ $(i = 1, 2, 3, 4)$.

6. Simulation Study

We demonstrate the performance of all estimators by generating random number from bi-variate normal distribution by using R-Software Team et al. (2021). The auxiliary information on variable X has been generated by artificial data set's having population size $N = 5000$ generated from bi-variate normal distribution for (X, Y) having negative correlation between Study variables and auxiliary variable. This type of population is very relevant in most socio-economic situation with our interest.

The model under which the populations are generated is given below

$$X \leftarrow rnorm(N, m1, s1)$$

$$Y \leftarrow s2\ p(X - m1)/s1 + m2 + s2\ rnorm(N, 0, sqrt(1 - p^2))$$

where $rnorm()$ in R Team et al. (2021) is a built-in function that generates a vector of normally distributed random numbers, $sqrt()$ function in the R programming language is used to determine the square-root of a value that is passed to it as an argument, N is the population size, $m1$ and $m2$ are the means of variables X and Y respectively, $s1$ and $s2$ are the standard deviations of variables X and Y respectively, and p is the correlation coefficient between the variables X and Y. This setting is used to generate bivariate normal numbers between (X, Y).

By using the generated random variables, the PRL and PRE of our considered estimators τ_i $(i = 1, 2, 3, 4)$ in the respect to another suggested estimators ξ_i $(i = 1, 2, 3, 4)$ are calculated and shown in Tables 5-7. Based on the simulation studies, We have observed that the our considered estimators are more effective than the compared estimators in this literature.

7. Analysis of Numerical Illustration and Simulation Study

From the Tables 1-7, the following interpretation can be found:

I. We present descriptions of three real-world data sets in the Table 1 to demonstrate the applications of our research. We are taking different values of N, n, and r.

II. From the Table 2

(a) For Data set-A, the PRL_i in the precision of the considered estimators $\tau_i(i = 1, 2, 3, 4)$ with respect to the other suggested estimators $\xi_i(i = 1, 2, 3, 4)$ remains between 19.14 percent to 49.99 percent for the sample sizes of 35% and 40% and response rates between 62% to 87%.

(b) For Data set B, with sample sizes ranging from 33% to 40% and response rates ranging from 60% to 91.66%, the PRL_i in the precision of the considered estimators $\tau_i(i = 1, 2, 3, 4)$ with respect to the other suggested estimators $\xi_i(i = 1, 2, 3, 4)$ remain between 13.15% to 49.99%.

(c) For Data set C, the PRL_i in the precision of the considered estimators $\tau_i(i = 1, 2, 3, 4)$ with respect to the other suggested estimators $\xi_i(i = 1, 2, 3, 4)$ remains between 13.15% to 49.99%, with sample sizes ranging from 33% to 40% and a response rate ranging from 60% to 91.66%.

III. From the Table 3

(a) For Data set A, the PRE_i over the estimator $\bar{y_r}$ remains between 252.61% to 347.02% for sample sizes of 35% and 40% and response rates ranging from 62% to 87%.

(b) For Data set B, the sample sizes varied from 33% to 40%, and the response rates ranged from 60% to 91.66%; the PRE_i over the estimator $\bar{y_r}$ remains between 219.33% to 226.58%.

(c) For Data set C, the PRE_i over the estimator $\bar{y_r}$ remains between 192.87% to 208.48% with sample sizes ranging from 33% to 40% and response rates ranging from 60% to 91.66%.

IV. From the Table 4

(a) For Data set A, the PRE_i over the estimator $y_{\bar{SD}}$ remains within the range of 150.31% to 289.07% for sample sizes of 35% and 40% with response rates ranging from 62% to 87%.

(b) For Data Set-B, the PRE_i over the estimator $y_{\bar{SD}}$ remains between 133.85% to 204.79% for sample sizes of 33% to 40% and response rates of 60% to 91.66%.

(c) For Data set C: The sample sizes range from 33% to 40%, and the response rate ranges from 60% to 91.66%. The PRE_i over the estimator $y_{\bar{SD}}$ remains within the range of 121.24% to 188.11%.

V. From Table 5 based on the simulation study, it can be seen that sample sizes range from 34% to 38%, response rates range from 61.76% to 86.84%, and the PRL_i in the precision of the considered estimators $\tau_i(i = 1, 2, 3, 4)$ with respect to the other suggested estimators $\xi_i(i = 1, 2, 3, 4)$ remain between 11.71% to 28.62%.

VI. According to Table 6 based on a simulation study, the PRE_i over the estimator $\bar{y_r}$ remains between 184.15% to 255.91% for sample sizes of 34% to 38% with response rates ranging from 61.76% to 86.84%.

VII. According to Table 7, which is based on a simulation study, the PRE_i over the estimator $y_{\bar{S}D}$ remains between 184.12% to 255.82% for sample sizes of 34% to 38% with response rates ranging from 61.76% to 86.84%.

8. Conclusions

In the present article, four efficient product type exponential estimators with imputation τ_i $(i = 1, 2, 3, 4)$ are considered in sampling theory to estimate the population mean of the study variable. According to the Tables 2 and 5, for the fixed values of sample population, the values of PRL_i $(i = 1, 2, 3, 4)$ now decrease with the increasing values of response rate. This behaviour indicates that the higher the correlation coefficient between the auxiliary variables and study variables, the fewer fresh samples are required in sampling theory, and the amount of loss in precision also decreases. The behaviour of the proposed estimators are well supported by numerical illustrations and simulation studies presented in Tables 2-7 for taking different values of sample population and response rate. It is shows that the proposed estimators are more effective than the mean imputation approach and Singh & Deo (2003) estimator in this literature.

Based on analysis of numerical and simulation studies, one may conclude that using the population coefficient of skewness $\beta_1(x)$ and coefficient of kurtosis $\beta_2(x)$ of as auxiliary variables in developing methods of imputation and, as a result, its application in proposed estimators are extremely valuable in terms of estimate precision and survey cost reduction.

TABLE 1: Description of data sets

Parameters	Data set A Pandey & Dubey (1988)	Data set B Singh (2003), p. 1113	Data set C Singh & Mangat (2013), p. 187
N	20	30	30
n	7, 8	10, 11, 12	10, 11, 12
r	(5, 6), (5, 6, 7)	(6,7,8,9), (7, 8, 9, 10), (8, 9, 10, 11)	(6,7,8,9), (7, 8, 9, 10), (8, 9, 10, 11)
$\bar{Y}$	19.55	384.2	6.3766
$\bar{X}$	18.8	67.2667	66.9333
C_y	0.3552	0.1558	0.0266
C_x	0.3943	0.1373	0.0206
$\beta_1(x)$	0.5473	0.3449	0.2833
$\beta_2(x)$	3.0613	2.2389	2.1600
ρ_{yx}	-0.9199	-0.8552	-0.8668

TABLE 2: Percent relative losses PRL_i in the precision of the considered estimators $\tau_i(i = 1, 2, 3, 4)$ in comparison to the other suggested estimators $\xi_i(i = 1, 2, 3, 4)$, respectively.

Data set	N	n	r	PRL_1	PRL_2	PRL_3	PRL_4
A	20	7	5	37.9934	37.8941	38.0861	37.8809
			6	20.3539	20.3011	20.4033	20.2940
		8	5	49.8923	49.7873	49.9903	49.7733
			6	35.6377	35.5633	35.7074	35.5534
			7	19.1896	19.1498	19.2270	19.1445
B	30	10	6	49.9999	49.9998	49.9995	49.9998
			7	39.1304	39.1303	39.1301	39.1302
			8	27.2727	27.2726	27.2724	27.2726
			9	14.2857	14.2856	14.2855	14.2856
		11	7	47.4308	47.4307	47.4304	47.4306
			8	37.1900	37.1899	37.1898	37.1899
			9	25.9740	25.9739	25.9738	25.9739
			10	13.6363	13.6363	13.6362	13.6363
		12	8	45.4545	45.4544	45.4542	45.4544
			9	35.7142	35.7142	35.7140	35.7141
			10	24.9999	24.9999	24.9998	24.9999
			11	13.1578	13.1578	13.1578	13.1578
C	30	10	6	49.9999	49.9999	49.9998	49.9999
			7	39.1303	39.1304	39.1303	39.1304
			8	27.2726	27.2727	27.2726	27.2727
			9	14.2856	14.2857	14.2856	14.2857
		11	7	47.4307	47.4308	47.4307	47.4308
			8	37.1900	37.1900	37.1900	37.1900
			9	25.9739	25.9740	25.9739	25.9740
			10	13.6363	13.6363	13.6363	13.6363
		12	8	45.4545	45.4545	45.4544	45.4545
			9	35.7142	35.7142	35.7142	35.7142
			10	24.9999	24.9999	24.9999	24.9999
			11	13.1578	13.1578	13.1578	13.1578

TABLE 3: PRE_i of the considered estimators $\tau_i (i = 1, 2, 3, 4)$ over the estimator $\bar{y}_r$.

Data set	N	n	r	PRE_1	PRE_2	PRE_3	PRE_4
A	20	7	5	308.8149	343.1017	252.6360	347.0265
			6	308.5183	342.4508	252.6143	346.3249
		8	5	308.8149	343.1017	252.6360	347.0265
			6	308.5183	342.4508	252.6143	346.3249
			7	308.3082	341.9909	252.5989	345.8292
B	30	10	6	219.3318	225.7754	210.9029	226.5854
			7	219.3318	225.7752	210.9022	226.5851
			8	219.3318	225.7750	210.9018	226.5849
			9	219.3318	225.7749	210.9014	226.5847
		11	7	219.3318	225.7752	210.9022	226.5851
			8	219.3318	225.7750	210.9018	226.5849
			9	219.3318	225.7749	210.9014	226.5847
			10	219.3318	225.7748	210.9011	226.5845
		12	8	219.3318	225.7750	210.9018	226.5849
			9	219.3318	225.7749	210.9014	226.5847
			10	219.3318	225.7748	210.9011	226.5845
			11	219.3318	225.7747	210.9008	226.5844
C	30	10	6	200.8445	207.6942	192.8784	208.4871
			7	200.8444	207.6941	192.8782	208.4871
			8	200.8443	207.6941	192.8781	208.4871
			9	200.8443	207.6941	192.8780	208.4870
		11	7	200.8444	207.6941	192.8782	208.4871
			8	200.8443	207.6941	192.8781	208.4871
			9	200.8443	207.6941	192.8780	208.4870
			10	200.8442	207.6941	192.8780	208.4870
		12	8	200.8443	207.6941	192.8781	208.4871
			9	200.8443	207.6941	192.8780	208.4870
			10	200.8442	207.6941	192.8780	208.4870
			11	200.8442	207.6941	192.8779	208.4870

TABLE 4: PRE_i of the considered estimators $\tau_i (i = 1, 2, 3, 4)$ over the estimator $\bar{y}_{SD}$.

Data set	N	n	r	PRE_1	PRE_2	PRE_3	PRE_4
A	20	7	5	213.5177	237.2240	174.6750	239.9377
			6	257.5153	285.8383	210.8532	289.0719
		8	5	183.7374	204.1372	150.3124	206.4724
			6	219.2631	243.3788	179.5323	246.1321
			7	260.2804	288.7161	213.2494	291.9564
B	30	10	6	139.2002	143.2897	133.8508	143.8037
			7	156.6201	161.2212	150.6008	161.7995
			8	175.6236	180.7829	168.8736	181.4313
			9	196.4370	202.2076	188.8866	202.9329
		11	7	143.3176	147.5279	137.8095	148.0571
			8	159.7298	164.4221	153.5906	165.0119
			9	177.7050	182.9253	170.8746	183.5814
			10	197.4777	203.2787	189.8871	204.0078
		12	8	146.4849	150.7881	140.8547	151.3290
			9	162.0949	166.8567	155.8645	167.4551
			10	179.2660	184.5320	172.3754	185.1939
			11	198.2445	204.0680	190.6242	204.7999
C	30	10	6	126.2519	130.5577	121.2444	131.0561
			7	142.4676	147.3265	136.8169	147.8889
			8	160.1575	165.6197	153.8051	166.2520
			9	179.5321	185.6551	172.4112	186.3639
		11	7	130.0847	134.5212	124.9251	135.034
			8	145.3623	150.3199	139.5967	150.8938
			9	162.0949	167.6232	155.6656	168.2632
			10	180.5008	186.6569	173.3415	187.3695
		12	8	133.0329	137.5700	127.7564	138.0953
			9	147.5639	152.5966	141.7110	153.1792
			10	163.5480	169.1258	157.0610	169.7715
			11	181.2146	187.3950	174.0269	188.1104

Table 5: Percent relative losses PRL_i in the precision of the considered estimators $\tau_i(i = 1, 2, 3, 4)$ with respect to the another considered estimators $\xi_i(i = 1, 2, 3, 4)$ respectively for Simulation study of Population size $N = 5000$.

n	r	PRL_1	PRL_2	PRL_3	PRL_4
1700	1050	26.8230	26.8295	26.8215	26.8297
	1150	23.2859	23.2916	23.2846	23.2917
1750	1050	28.6132	28.6200	28.6116	28.6201
	1150	25.1627	25.1686	25.1612	25.1687
	1250	21.5280	21.5331	21.5268	21.5332
1800	1150	27.0042	27.0104	27.0027	27.0106
	1250	23.4590	23.4644	23.4577	23.4646
	1350	19.7196	19.7241	19.7185	19.7242
	1450	15.7695	15.7731	15.7686	15.7732
1850	1250	25.3542	25.3599	25.3528	25.3600
	1350	21.7073	21.7122	21.7061	21.7123
	1450	17.8551	17.8591	17.8541	17.8591
	1550	13.7794	13.7825	13.7787	13.7826
1900	1350	23.6585	23.6637	23.6573	23.6638
	1450	19.9022	19.9066	19.9012	19.9067
	1550	15.9282	15.9317	15.9273	15.9318
	1650	11.7169	11.7195	11.7163	11.7195

Table 6: PRE_i of the considered estimators $\tau_i(i = 1, 2, 3, 4)$ over the estimator $\bar{y}_r$ for Simulation study of Population size $N = 5000$.

n	r	PRE_1	PRE_2	PRE_3	PRE_4
1700	1050	255.9138	255.8986	255.9175	255.8983
	1150	238.7533	238.7428	238.7559	238.7426
1750	1050	255.9138	255.8986	255.9175	255.8983
	1150	238.7533	238.7428	238.7559	238.7426
	1250	224.3386	224.3320	224.3402	224.3318
1800	1150	238.7533	238.7428	238.7559	238.7426
	1250	224.3386	224.3320	224.3402	224.3318
	1350	212.0593	212.0561	212.0601	212.0560
	1450	201.4738	201.4734	201.4739	201.4734
1850	1250	224.3386	224.3320	224.3402	224.3318
	1350	212.0593	212.0561	212.0601	212.0560
	1450	201.4738	201.4734	201.4739	201.4734
	1550	192.2541	192.2562	192.2536	192.2562
1900	1350	212.0593	212.0561	212.0601	212.0560
	1450	201.4738	201.4734	201.4739	201.4734
	1550	192.2541	192.2562	192.2536	192.2562
	1650	184.1519	184.1562	184.1509	184.1563

TABLE 7: PRE_i of the considered estimators $\tau_i (i = 1, 2, 3, 4)$ over the estimator $\bar{y}_{SD}$ for simulation studies of $N = 5000$.

n	r	PRE_1	PRE_2	PRE_3	PRE_4
1700	1050	255.8172	255.8021	255.8210	255.8017
	1150	238.6752	238.6646	238.6778	238.6644
1750	1050	255.8128	255.7976	255.8165	255.7973
	1150	238.6705	238.6600	238.6731	238.6597
	1250	224.2720	224.2654	224.2736	224.2652
1800	1150	238.6661	238.6556	238.6687	238.6553
	1250	224.2673	224.2607	224.2690	224.2606
	1350	212.0027	211.9995	212.0035	211.9994
	1450	201.4308	201.4304	201.4309	201.4304
1850	1250	224.2630	224.2564	224.2646	224.2562
	1350	211.9981	211.9949	211.9989	211.9948
	1450	201.4259	201.4255	201.4260	201.4255
	1550	192.2189	192.2210	192.2184	192.2210
1900	1350	211.9938	211.9905	211.9946	211.9904
	1450	201.4214	201.4210	201.4215	201.4210
	1550	192.2141	192.2162	192.2136	192.2162
	1650	184.1238	184.1281	184.1227	184.1282

Acknowledgements

The authors want to express their sincere thanks to the anonymous referees and editors for their rigorous review and comments, which significantly improved our manuscript.

[Received: April 2022 — Accepted: December 2022]

References

Ahmed, M., Al-Titi, O., Al-Rawi, Z. & Abu-Dayyeh, W. (2006), 'Estimation of a population mean using different imputation methods', *Statistics in Transition* **7**(6), 1247–1264.

Diana, G. & Francesco Perri, P. (2010), 'Improved estimators of the population mean for missing data', *Communications in Statistics—Theory and Methods* **39**(18), 3245–3251.

Gira, A. A. (2015), 'Estimation of population mean with a new imputation methods', *Applied Mathematical Sciences* **9**(34), 1663–1672.

Heitjan, D. F. & Basu, S. (1996), 'Distinguishing "missing at random" and "missing completely at random"', *The American Statistician* **50**(3), 207–213.

Hyunshik Lee, E. R. & Särndal, C. E. (1994), 'Experiments with variance estimation from survey data with imputed values', *Journal of official Statistics* **10**(3), 231–243.

Kadilar, C. & Cingi, H. (2008), 'Estimators for the population mean in the case of missing data', *Communications in Statistics—Theory and Methods* **37**(14), 2226–2236.

Lee, H., Rancourt, E. & Sarndal, C. (1995), 'Variance estimation in the presence of imputed data for the generalized estimation system', *American Statistical Association (Social Survey Research Methods Section)* pp. 384–389.

Pandey, B. & Dubey, V. (1988), 'Modified product estimator using coefficient of variation of auxiliary variate', *Assam Statistical Rev* **2**(2), 64–66.

Prasad, S. (2017), 'Ratio exponential type estimators with imputation for missing data in sample surveys', *Model Assisted Statistics and Applications* **12**(2), 95–106.

Prasad, S. (2018*a*), 'Product exponential method of imputation in sample surveys', *Statistics in Transition. New Series* **19**(1), 159–166.

Prasad, S. (2018*b*), 'A study on new methods of ratio exponential type imputation in sample surveys', *Hacettepe Journal of Mathematics and Statistics* **47**(5), 1281–1301.

Prasad, S. (2019), 'Exponential method of imputation for non-response in sample surveys', *Pakistan Journal of Statistics* **35**(2), 97–107.

Prasad, S. (2021), 'Some compromised exponential ratio type imputation methods in simple random sampling', *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences* **91**(2), 337–349.

Sande, I. G. (1979), 'A personal view of hot-deck imputation procedures', *Survey Methodology* **5**(2), 238–258.

Singh, G., Maurya, S., Khetan, M. & Kadilar, C. (2016), 'Some imputation methods for missing data in sample surveys', *Hacettepe Journal of Mathematics and Statistics* **45**(6), 1865–1880.

Singh, G., Priyanka, K., Kim, J.-M. & Singh, S. (2010), 'Estimation of population mean using imputation techniques in sample surveys', *Journal of the Korean Statistical Society* **39**(1), 67–74.

Singh, R. & Mangat, N. S. (2013), *Elements of survey sampling*, Vol. 15, Springer Science & Business Media.

Singh, S. (2003), *Advanced Sampling Theory With Applications: How Michael Selected Amy*, Vol. 2, Springer Science & Business Media.

Singh, S. (2009), 'A new method of imputation in survey sampling', *Statistics* **43**(5), 499–511.

Singh, S. & Deo, B. (2003), 'Imputation by power transformation', *Statistical Papers* **44**(4), 555–579.

Singh, S. & Horn, S. (2000), 'Compromised imputation in survey sampling', *Metrika* **51**(3), 267–276.

Team, R. C. et al. (2021), 'R: A language and environment for statistical computing'. http://www. R-project. org/

Toutenburg, H., Srivastava, V. et al. (2008), 'Amputation versus imputation of missing values through ratio method in sample surveys', *Statistical Papers* **49**(2), 237–247.