Research Article

An Exponential-Cum-Sine-Type Hybrid Imputation Technique for Missing Data

D. Bhattacharyya, G.N. Singh, Taghreed M. Jawa, Neveen Sayed-Ahmed, and Awadhesh K. Pandey

1Department of Mathematics & Computing, Indian Institute of Technology (ISM), Dhanbad-826 004, Jharkhand, India
2Department of Mathematics and Statistics, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia
3Department of Mathematics, School of Physical Sciences, DIT University, Dehradun, Uttarakhand 248009, India

Correspondence should be addressed to Neveen Sayed-Ahmed; nevensayd@yahoo.com

Received 14 October 2021; Accepted 17 November 2021; Published 3 December 2021

Abstract

Missing data is a universal occurrence in sample surveys, leading to a decline in data quality and complications in making inferences. It is pivotal for survey statisticians to factor in the stochastic nature of incomplete data. This brings forth the question of what assumptions have to be made or which techniques have to be employed to handle the problem of ignorability of completeness mechanism. The mechanisms of missing data have been studied in detail in [9, 13], among others. Three missing data mechanisms are mostly of interest in the survey literature, namely, missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). MCAR is said to occur when data is missing randomly or by chance, MAR occurs when the missingness does not depend on the variable under study (which may be unobserved), but on some other variables (which is fully observed), and MNAR occurs when missingness depends on the variable under study.

Numerous statistical methods have been devised over the years to overcome the problem of missing data. Subsampling of nonrespondents in surveys via mail questionnaire was pioneered in [8]. Another commonly used method...
is imputation, in which the missing values are filled in by a suitable function of the available values, to ensure the structural completeness of the sample before analysis begins. Popular imputation techniques include mean imputation, regression imputation, hot deck imputation, cold deck imputation, and nearest neighbor method. Imputation techniques in the survey literature are from [3, 5, 14–21], among others. Some recent works in the area of imputation and estimation of population mean have been done in [22–29] and others.

Information from an auxiliary variable can be utilized to provide an improved estimate for population characteristics. Such information may be readily available as secondary data from previous surveys or census or may be collected during the survey procedure at little to no additional cost. Some examples of such auxiliary information include the lifetime of a previous batch of bulbs when studying the life of a current lot of bulbs, the speed of cars when studying the mileage of cars, etc.

In this manuscript, a new exponential-cum-sine-type hybrid imputation technique and corresponding point estimator have been proposed for estimation of population mean. Motivation for this estimator, its properties, and its uses have been discussed in the subsequent sections. The manuscript is henceforth divided into the following sections: Section 2 introduces the sample structure and notations used in the manuscript. Section 3 discusses some conventional estimators of population mean. Section 4 discusses the proposed estimator, including its existence, consistency, properties, and implementation in R. The simulation study has been presented in Section 5, the results and discussion in Section 6, and the conclusions in Section 7.

2. Sample Structure and Notations Used

Let the character of interest be denoted by $Y$. We consider the scenario in which complete information on a correlated auxiliary variable $X$ is available to the survey statisticians and its population mean is known. The sample structure and the notations used henceforth have been introduced in Table 1.

3. Some Conventional Estimators

Before the proposed estimator is introduced, it is important to examine some existing estimators for population mean and study their strengths and limitations. A few such estimators have been discussed in this section.

The mean estimator is a simple and traditional estimator, which makes use of the average of the responses to provide an estimate of the population mean. The ratio estimator tries to make an improvement over the mean estimator by incorporating auxiliary information into a correlated variable. Various other estimators that make innovative use of auxiliary information have been proposed, for instance, the estimator proposed in [30], regression-type estimators proposed in [10], and exponential type estimators in [31], among others.

The structures of some of these estimators have been given in Table 2, while the expressions for their respective variances ($V$) or mean square errors (MSEs) have been given in Table 3.

It is to be noted that most conventional estimators make use of simple functional forms, such as linear combinations, exponential functions, and chains. Combination of multiple mathematical functions is rarely seen. This can be attributed to computational limitations associated with such functions. However, with the advent of supercomputers and improvement in computational powers, such obstructions have been eliminated. It is worth exploring whether combinations of mathematical functions produce better estimates than traditional estimators. This has been the motivation behind the construction of the proposed estimator.

Two such functions have been used, namely, the exponential and sine functions. Such particular functions were selected based on their use in real-life situations. The exponential function is usually used to model growth and decay observed in nature, such as growth and decay of microorganisms like bacteria, human population, spread of pandemics, and compound interests. Sine function is commonly utilized for the purpose of modeling natural phenomena which are periodic in nature, such as sound waves, light waves, tides, sunlight intensity, and average temperature variations through the year, as well as ballistic trajectories, electrical currents, and GPS locations.

4. Formulation of the Proposed Estimator

Let $y_i$ and $x_i$ be the values of $Y$ and $X$, respectively, for the $i^{th}$ unit in the population. The following imputation method may be suggested to deal with the problem of missing data:

$$
y_i = \begin{cases} 
  y_i, & \text{if } i \in R, \\
  \frac{n}{n - r x_i} \exp \left[ \frac{\sin(x_n) - \sin(x_i)}{1 + \sin(x_n) + \sin(x_i)} \right] - \frac{r}{n - r} y_i, & \text{if } i \in R^c.
\end{cases}
$$

(1)

The point estimator under an imputation method is given in

$$
T = \frac{1}{n} \sum_{i \in S} y_i = \frac{1}{n} \left[ \sum_{i \in R} y_i + \sum_{i \in R^c} y_i \right].
$$

(2)

Using equation (2), under the imputation outlined in equation (1), the expression for the point estimator of $\bar{Y}$ is obtained as

$$
T = \bar{Y} \exp \left[ \frac{\sin(x_n) - \sin(x_i)}{1 + \sin(x_n) + \sin(x_i)} \right].
$$

(3)

4.1. Existence and Consistency of the Estimator. It is important to specify the domain of values for which an estimator exists, so that survey statisticians or those working in the field can determine whether an estimator can be reasonably used in a practical scenario.
The given estimator consists of two major functions: the trigonometrical function $\sin x$ and the exponential function $\exp x$. Both $\sin(x)$ and $\exp(x)$ exist in $\forall x \in \mathbb{R}$, so $y_i$ and $T$ exist in $\forall x \in \mathbb{R}$.

Hence, the proposed estimator can be used for all real values of the characters under study. For real-world scenarios, most, if not all, characters of interest take only real values. For example, measurements such as length, breadth, height, weight, diameter, currencies, and number of an item do not take nonreal values. Hence, the proposed estimator can be used in all practical scenarios.

It is to be noted that the structure of the estimator is consistent for large sample approximations. As $n \rightarrow \infty$, $\bar{y}_r \rightarrow \overline{y}$, $\bar{x}_r \rightarrow \overline{x}$, $\bar{x}_n \rightarrow \overline{x}$, and $\exp(0) = 1$. Hence, $T \rightarrow \overline{y}$.

4.2. Properties of the Proposed Estimator. The “goodness” of an estimator can be measured in terms of various properties. Two such properties, namely, bias and mean squared error (MSE), have been explored here. The bias gives an idea about the expected deviation from the true value of a parameter,
while MSE deals with the degree of spread. The expressions for the same have been derived under large sample assumptions up to the first order of approximations. Some transformations involving error terms have been used for the purpose, indicated as follows:

\[
\begin{align*}
\bar{y}_r &= \bar{Y}(1 + \eta_0), \\
\bar{x}_r &= \bar{X}(1 + \eta_1), \\
\bar{x}_n &= \bar{X}(1 + \eta_2), \\
\theta_1 &= \left(\frac{1}{r} - \frac{1}{N}\right), \\
\theta_2 &= \left(\frac{1}{n} - \frac{1}{N}\right), \\
\theta_3 &= \left(\frac{1}{r} - \frac{1}{n}\right).
\end{align*}
\]  

(4)

The error terms have the following expectations:

\[
E(\eta_0) = E(\eta_1) = E(\eta_2) = 0,
\]

(5)

\[
\begin{align*}
E(\eta_0^2) &= \theta_1 \bar{C}_Y, \\
E(\eta_1^2) &= \theta_1 \bar{C}_X, \\
E(\eta_2^2) &= \theta_2 \bar{C}_X, \\
E(\eta_0\eta_1) &= \theta_1 \rho \bar{C}_Y \bar{C}_X, \\
E(\eta_1\eta_2) &= \theta_2 \bar{C}_X, \\
E(\eta_2\eta_3) &= \theta_3 \rho \bar{C}_Y \bar{C}_X.
\end{align*}
\]

To obtain the expressions for bias and MSE, in the first step, algebraic expansion of the expression of the estimator given in equation (3) is done, using the following Taylor’s series:

\[
(1) \sin(x) = x - (x^3/3!) + (x^5/5!) - (x^7/7!) - \cdots
\]

(2) \exp(x) = 1 + x + (x^2/2!) + (x^3/3!) + (x^4/4!) + \cdots

(3) \( (1 + x)^{-1} = 1 - x + x^2 - x^3 + \cdots \)

The estimator takes the following form:

\[
T = \bar{y}_r \left[ 1 + \bar{x}_n - \bar{x}_r + 2\bar{x}_n^2 - 2\bar{x}_n\bar{x}_r \right].
\]  

(6)

In the second step, the transformations in equation (4) are applied to equation (6) to obtain the following form of the estimator:

\[
T = \bar{y}_r \left[ 1 + \bar{x}_n - \bar{x}_r + 2\bar{x}_n^2 - 2\bar{x}_n\bar{x}_r \right].
\]  

(7)

4.3. Implementation in R. In the current day and age, most computations are carried out using a suitable software environment. The following R [32] code snippet has been developed to carry out the proposed imputation on a data set of interest and calculate the value of the corresponding point estimator:

```r
# Import data of respondents from file
dfresp <- read.table(file.choose())

# Import data of nonrespondents from file
dfnonresp <- read.table(file.choose())

xbar <- mean(dfresp[, 1])

ybar <- -mean(dfresp[, 2])

xbarnonresp <- mean(dfnonresp[, 1])

r <- nrow(dfresp) # no. of respondents

nonresp <- nrow(dfnonresp) # no. of nonrespondents

n = r + nonresp # sample size

xbar = (r * xbar + nonresp * xbar) / n

num = sin(xbar) - sin(ybar)

deren = 1 + sin(xbar) + sin(xbar)
```
To this end, the bias of the estimator is calculated and the performance must be examined, in terms of its properties. Before an estimator can be used in practical scenarios, its importance in practical scenarios.

5. Simulation Study

Before an estimator can be used in practical scenarios, its performance must be examined, in terms of its properties. To this end, the bias of the estimator is calculated and the MSE is compared with that of the contemporary estimators given in Table 2 in terms of percentage relative efficiencies (PREs).

The PREs of the estimator with respect to the contemporary estimators are defined as follows:

\[
\begin{align*}
\text{PRE}_1 &= \frac{V(\overline{T}_m)}{M(T)} \times 100, \\
\text{PRE}_2 &= \frac{M(\overline{T}_{\text{RAT}})}{M(T)} \times 100, \\
\text{PRE}_3 &= \frac{M(T_{\text{TSS}})}{M(T)} \times 100, \\
\text{PRE}_4 &= \frac{M(T_{\text{SMKK}})}{M(T)} \times 100, \\
\text{PRE}_5 &= \frac{M(T_{\text{KC1}})}{M(T)} \times 100, \\
\text{PRE}_6 &= \frac{M(T_{\text{KC2}})}{M(T)} \times 100, \\
\text{PRE}_7 &= \frac{M(T_{\text{RAT}})}{M(T)} \times 100,
\end{align*}
\]

where the expression for the MSE of the proposed estimator \( T \) is given in equation (9), while that of the contemporary estimators is given in Table 3.

Using R [32], an extensive simulation study has been carried out on sufficiently large fictitious populations to compute the bias and the PREs defined above. Data is generated from three different probability distributions, namely, normal and Gamma distributions (continuous distributions) and Poisson distribution (discrete distribution). Some important properties of the distributions have been summarized in Table 4. Such distributions are chosen based on their occurrence in real-life situations.

Data from normal distribution is rampant in nature. It can be used to model heights of individuals, test scores of students, blood pressure, daily returns of any particular stock, weights of items produced by a manufacturing process, etc. Poisson distribution can be used to model the probability that a given number of events occur in a specific time interval, for example, the number of insurance claims filed per month, the number of network failures occurring per week, and the number of bulbs manufactured per minute. It also finds use by medical statisticians, such as for estimating the number of births that may be expected on a particular night, the number of patients with an infectious disease arriving at a clinic within a given hour, the number of mutations on a given strand of DNA per time unit, etc. Gamma distribution can be used for modeling wait time, reliability, service time in queuing theory, etc. For example, it can be used to model the amount of rainfall that accumulates in a given reservoir, the flow of items through manufacturing as well as distribution processes, the size of loan defaults, etc. Thus, these three distributions are chosen based on their importance in practical scenarios.

It is seen through trial and error that the estimator performs well when \( X \) and \( Y \) take small values and the variation in \( X \) is greater than that in \( Y \).

The steps of the simulation are as follows:

1. The sizes of the population, the sample, and the responding part of the sample are defined. For the purpose of the study, sufficiently large values of \( N = 100000, n = 40000, \) and \( r = 35000 \) have been chosen.
2. The parameters of the population are defined.
3. Simulation is conducted for various values of \( \rho \). For the purpose of the study, \( \rho \) is considered to be in the range \((0.1, 0.9)\); i.e., positively correlated variable \( X \) is considered.

The results of the simulation study related to the PREs have been presented in Tables 5–11, while the biases have been presented in Table 12.

6. Results and Discussion

The simulation study enables us to study the behavior of the proposed estimator under various scenarios involving various values of parameters. The chief conclusions are as follows:

1. From the values of \( \text{PRE}_1 \) in Table 5, it is seen that the proposed estimator is more efficient than \( \overline{T}_m \) for all values of \( \rho \) for normal data and for \( \rho \in (0.2, 0.9) \) for Gamma and Poisson data for the various values of response rates.
2. It is seen that the proposed estimator performs better than \( \overline{T}_{\text{RAT}} \) for all values of \( \rho \) for normal and Gamma data and for \( \rho \in (0.1, 0.8) \) for Poisson data for the various values of response rates from the values of \( \text{PRE}_2 \) in Table 6.
3. From the values of \( \text{PRE}_3 \) in Table 7, it is seen that the proposed estimator dominates \( T_{\text{TSS}} \) for all values of \( \rho \) for normal data and for \( \rho \in (0.1, 0.7) \) for Gamma and Poisson data for the various values of response rates.
Table 4: Some properties of normal, Poisson, and Gamma distributions.

| Distribution | Normal | Poisson | Gamma |
|--------------|--------|---------|-------|
| Parameters | \( \mu, \sigma^2 \) | \( \lambda > 0 \) | \( \alpha, \lambda \) |
| PdF | \( f(x) = \frac{1}{\sqrt{2\pi} \sigma^2} e^{-(x-\mu)^2/(2\sigma^2)} \) | \( f(x) = \lambda^\alpha e^{-\lambda x}/\Gamma(x) \) | \( f(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(x)} \) if \( x > 0 \), 0 otherwise |
| Mean \( \mu \) | \( \frac{\lambda^\alpha}{\Gamma(\alpha)} \) |
| Variance \( \sigma^2 \) | \( \frac{\lambda}{\Gamma(\alpha+1)} \) |

Table 5: Values of \( \text{PRE}_1 \) when \( \rho \in (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) \) and response rates are 75%, 80%, 85%, 90%, and 95% for data generated from normal, Gamma, and Poisson distributions.

| \( \rho \) | 75% | 80% | 85% | 90% | 95% |
|--------|-----|-----|-----|-----|-----|
|         | PRE\(_1\) when response rate is |         | PRE\(_1\) when response rate is |         |
|         | When data is generated from normal distribution |         | When data is generated from Gamma distribution |         |
| 0.1     | 100.1344 | 100.1106 | 100.0855 | 100.0587 | 100.0303 |
| 0.2     | 101.4571 | 101.1969 | 100.9224 | 100.6323 | 100.3254 |
| 0.3     | 102.6405 | 102.1645 | 101.6644 | 101.1383 | 100.5843 |
| 0.4     | 103.5638 | 102.9165 | 102.2389 | 101.5285 | 100.7832 |
| 0.5     | 104.4114 | 103.6048 | 102.7629 | 101.8832 | 100.9632 |
| 0.6     | 105.1428 | 103.9281 | 102.6678 | 101.3594 | 100.9632 |
| 0.7     | 106.7147 | 104.7474 | 103.4746 | 101.7637 |
| 0.8     | 107.6268 | 105.7225 | 103.1995 | 101.6262 |
| 0.9     | 109.2664 | 107.5084 | 105.7046 | 103.8532 | 101.9524 |

Table 6: Values of \( \text{PRE}_2 \) when \( \rho \in (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) \) and response rates are 75%, 80%, 85%, 90%, and 95% for data generated from normal, Gamma, and Poisson distributions.

| \( \rho \) | 75% | 80% | 85% | 90% | 95% |
|--------|-----|-----|-----|-----|-----|
|         | PRE\(_2\) when response rate is |         | PRE\(_2\) when response rate is |         |
|         | When data is generated from normal distribution |         | When data is generated from Gamma distribution |         |
| 0.1     | 101.3449 | 101.1676 | 100.8578 | 100.5843 | 100.3254 |
| 0.2     | 102.6405 | 102.1645 | 101.6644 | 101.1383 | 100.5843 |
| 0.3     | 103.5638 | 102.9165 | 102.2389 | 101.5285 | 100.7832 |
| 0.4     | 104.4114 | 103.6048 | 102.7629 | 101.8832 | 100.9632 |
| 0.5     | 105.1428 | 103.9281 | 102.6678 | 101.3594 | 100.9632 |
| 0.6     | 106.7147 | 104.7474 | 103.4746 | 101.7637 | 100.9632 |
| 0.7     | 107.6268 | 105.7225 | 103.1995 | 101.6262 | 100.9632 |
| 0.8     | 109.2664 | 107.5084 | 105.7046 | 103.8532 | 101.9524 |

(4) The values of \( \text{PRE}_4 \) in Table 8 show that the proposed estimator is more efficient than \( T_{\text{SMKK}} \) for all values of \( \rho \) for normal data and for \( \rho \in (0.1, 0.7) \) for Gamma and Poisson data for the various values of response rates.

(5) In Table 9, the values of \( \text{PRE}_5 \) show that the proposed estimator performs better than \( T_{\text{KCC}} \) for all values of \( \rho \) and for the various values of response rates for normal, Gamma, and Poisson data.
Table 7: Values of $\text{PRE}_3$ when $\rho \in (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$ and response rates are 75%, 80%, 85%, 90%, and 95% for data generated from normal, Gamma, and Poisson distributions.

| $\rho$ | 75%     | 80%     | 85%     | 90%     | 95%     |
|--------|---------|---------|---------|---------|---------|
|        | $\text{PRE}_3$ when response rate is |         |         |         |         |
|        | When data is generated from normal distribution |         |         |         |         |
| 0.1    | 470.6541 | 448.5989 | 405.0824 | 335.8116 | 235.9446 |
| 0.2    | 205.9296 | 200.4779 | 188.5543 | 168.8425 | 139.8731 |
| 0.3    | 188.8019 | 184.9377 | 175.3686 | 158.9122 | 134.2714 |
| 0.4    | 303.8193 | 294.1930 | 271.6763 | 233.7198 | 177.5226 |
| 0.5    | 229.1919 | 224.7408 | 211.4849 | 187.6201 | 151.1766 |
| 0.6    | 120.6402 | 122.2864 | 121.6535 | 118.1561 | 111.1646 |
| 0.7    | 133.2641 | 135.2763 | 133.8068 | 128.0301 | 117.0710 |
| 0.8    | 103.7266 | 108.0715 | 110.4939 | 110.3914 | 107.1219 |
| 0.9    | 105.5687 | 112.2012 | 114.2465 | 113.9394 | 109.4739 |
|        | When data is generated from Gamma distribution |         |         |         |         |
| 0.1    | 137.1073 | 135.0152 | 130.7378 | 123.8271 | 113.7744 |
| 0.2    | 134.0241 | 132.2356 | 128.3940 | 122.0727 | 112.7901 |
| 0.3    | 131.4353 | 129.9555 | 126.5143 | 120.6959 | 112.0335 |
| 0.4    | 125.7548 | 124.7915 | 122.1295 | 117.3962 | 110.1751 |
| 0.5    | 121.1861 | 120.7463 | 118.7798 | 114.9341 | 108.1887 |
| 0.6    | 115.3791 | 115.6415 | 114.5838 | 111.8737 | 107.1463 |
| 0.7    | 108.6115 | 109.7314 | 109.7580 | 108.3771 | 105.2479 |
| 0.8    | 97.5700  | 99.9578  | 101.6769 | 102.4550 | 102.0003 |
| 0.9    | 88.1678  | 91.8278  | 95.1090  | 97.7504  | 99.4773  |
|        | When data is generated from Poisson distribution |         |         |         |         |
| 0.1    | 121.7373 | 120.5093 | 118.0000 | 113.9484 | 108.0598 |
| 0.2    | 119.4690 | 118.4929 | 116.3248 | 112.7142 | 107.3709 |
| 0.3    | 116.3871 | 115.7496 | 114.0429 | 111.0313 | 106.4499 |
| 0.4    | 113.8979 | 113.5651 | 112.2512 | 109.7282 | 105.7403 |
| 0.5    | 108.9561 | 109.1837 | 108.6216 | 107.0625 | 104.2750 |
| 0.6    | 105.4578 | 106.1194 | 106.1137 | 105.2428 | 103.2868 |
| 0.7    | 99.9416  | 101.2359 | 102.0750 | 102.2825 | 101.6631 |
| 0.8    | 95.6267  | 97.4805  | 99.0220  | 100.0827 | 100.4770 |
| 0.9    | 87.5557  | 90.3922  | 93.2074  | 95.8558  | 98.1782  |

Table 8: Values of $\text{PRE}_4$ when $\rho \in (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$ and response rates are 75%, 80%, 85%, 90%, and 95% for data generated from normal, Gamma, and Poisson distributions.

| $\rho$ | 75%     | 80%     | 85%     | 90%     | 95%     |
|--------|---------|---------|---------|---------|---------|
|        | $\text{PRE}_4$ when response rate is |         |         |         |         |
|        | When data is generated from normal distribution |         |         |         |         |
| 0.1    | 1916.7801 | 1674.5493 | 1417.7633 | 1145.0696 | 854.9427 |
| 0.2    | 647.9025  | 569.8619  | 487.5278  | 400.5359  | 308.4795 |
| 0.3    | 582.5054  | 509.8201  | 433.4627  | 353.1480  | 268.5599 |
| 0.4    | 1197.0614 | 1032.4981 | 860.1941  | 679.5903  | 490.0720 |
| 0.5    | 844.3493  | 725.8484  | 602.1498  | 472.9041  | 337.7295 |
| 0.6    | 281.9458  | 244.0446  | 204.7487  | 163.9796  | 121.6530 |
| 0.7    | 380.2268  | 322.5477  | 263.1681  | 202.0115  | 138.9969 |
| 0.8    | 240.6409  | 203.1637  | 164.4878  | 124.5548  | 83.3023  |
| 0.9    | 288.7217  | 238.8024  | 187.5835  | 135.0138  | 81.0389  |
|        | When data is generated from Gamma distribution |         |         |         |         |
| 0.1    | 284.1391  | 259.9150  | 234.1570  | 206.7146  | 177.4171 |
| 0.2    | 272.9792  | 249.2994  | 224.1799  | 197.4853  | 169.0626 |
| 0.3    | 265.0104  | 241.3963  | 216.4054  | 189.9107  | 161.7754 |
| 0.4    | 242.3263  | 220.6217  | 197.7172  | 173.5102  | 147.8866 |
| 0.5    | 226.8768  | 205.7934  | 183.6276  | 160.2940  | 135.6978 |
| 0.6    | 207.6590  | 187.4203  | 166.2336  | 144.0308  | 120.7368 |
From the values of PRE6 in Table 10, it is seen that the proposed estimator dominates $T_{KC_p}$ for all values of $\rho$ and for the various values of response rates for normal, Gamma, and Poisson data.

It is seen that the proposed estimator is more efficient than $T_{KC_p}$ for all values of $\rho$ and for the various values of response rates for normal, Gamma, and Poisson data from the values of PRE7 in Table 11.

From Table 12, it is seen that the estimator is negatively biased. The bias is negligible, being of the order $10^{-3}$ and $10^{-7}$ for various values of the parameter $\rho$ and for various response rates, and hence, bias correction is not needed.
Table 10: Values of $\text{PRE}_6$ when $\rho \in (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$ and response rates are 75%, 80%, 85%, 90%, and 95% for data generated from normal, Gamma, and Poisson distributions.

| $\rho$ | 75%         | 80%         | 85%         | 90%         | 95%         |
|--------|-------------|-------------|-------------|-------------|-------------|
|        | 75%, 80%, 85%, 90%, 95% |            |             |             |             |
|        | When data is generated from normal distribution |            |             |             |             |
| 0.1    | 2062.182    | 2061.693    | 2061.175    | 2060.625    | 2060.039    |
| 0.2    | 757.255     | 755.313     | 753.264     | 751.099     | 748.808     |
| 0.3    | 728.352     | 724.973     | 721.425     | 717.692     | 713.760     |
| 0.4    | 1494.992    | 1485.649    | 1475.866    | 1465.612    | 1454.852    |
| 0.5    | 1138.649    | 1129.854    | 1120.672    | 1111.079    | 1101.046    |
| 0.6    | 471.604     | 466.406     | 461.018     | 455.428     | 449.623     |
| 0.7    | 653.137     | 643.694     | 633.972     | 623.959     | 613.642     |
| 0.8    | 491.045     | 484.524     | 477.794     | 470.845     | 463.667     |
| 0.9    | 615.439     | 605.537     | 595.377     | 584.948     | 574.243     |
|        | When data is generated from Gamma distribution |            |             |             |             |
| 0.1    | 297.869     | 298.238     | 298.631     | 299.050     | 299.497     |
| 0.2    | 298.423     | 298.449     | 298.477     | 298.506     | 298.538     |
| 0.3    | 303.239     | 302.923     | 302.592     | 302.398     | 301.865     |
| 0.4    | 294.366     | 292.551     | 291.796     | 290.998     | 290.153     |
| 0.5    | 294.795     | 292.187     | 290.824     | 289.362     | 287.241     |
| 0.6    | 296.138     | 293.739     | 290.497     | 287.502     | 284.380     |
| 0.7    | 273.078     | 269.764     | 266.336     | 262.787     | 259.112     |
| 0.8    | 262.412     | 258.184     | 253.846     | 249.394     | 244.823     |
| 0.9    | 181.441     | 180.130     | 178.761     | 177.328     | 175.828     |

Table 11: Values of $\text{PRE}_7$ when $\rho \in (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$ and response rates are 75%, 80%, 85%, 90%, and 95% for data generated from normal, Gamma, and Poisson distributions.

| $\rho$ | 75%         | 80%         | 85%         | 90%         | 95%         |
|--------|-------------|-------------|-------------|-------------|-------------|
|        | 75%, 80%, 85%, 90%, 95% |            |             |             |             |
|        | When data is generated from normal distribution |            |             |             |             |
| 0.1    | 2062.182    | 2061.693    | 2061.175    | 2060.625    | 2060.039    |
| 0.2    | 757.255     | 755.313     | 753.264     | 751.099     | 748.808     |
| 0.3    | 728.352     | 724.973     | 721.425     | 717.692     | 713.760     |
| 0.4    | 1494.992    | 1485.649    | 1475.866    | 1465.612    | 1454.852    |
| 0.5    | 1138.649    | 1129.854    | 1120.672    | 1111.079    | 1101.046    |
| 0.6    | 471.604     | 466.406     | 461.018     | 455.428     | 449.623     |
| 0.7    | 653.137     | 643.694     | 633.972     | 623.959     | 613.642     |
| 0.8    | 491.045     | 484.524     | 477.794     | 470.845     | 463.667     |
| 0.9    | 615.439     | 605.537     | 595.377     | 584.948     | 574.243     |
|        | When data is generated from Gamma distribution |            |             |             |             |
| 0.1    | 297.869     | 298.238     | 298.631     | 299.050     | 299.497     |
| 0.2    | 298.423     | 298.449     | 298.477     | 298.506     | 298.538     |
| 0.3    | 303.239     | 302.923     | 302.592     | 302.398     | 301.865     |
| 0.4    | 294.366     | 292.551     | 291.796     | 290.998     | 290.153     |
| 0.5    | 294.795     | 292.187     | 290.824     | 289.362     | 287.241     |
| 0.6    | 296.138     | 293.739     | 290.497     | 287.502     | 284.380     |
| 0.7    | 273.078     | 269.764     | 266.336     | 262.787     | 259.112     |
| 0.8    | 262.412     | 258.184     | 253.846     | 249.394     | 244.823     |
| 0.9    | 181.441     | 180.130     | 178.761     | 177.328     | 175.828     |
Table 12: Values of bias when $\rho \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ and response rates are 75%, 80%, 85%, 90%, and 95% for data generated from normal, Gamma, and Poisson distributions.

| $\rho$ | 75%   | 80%   | 85%   | 90%   | 95%   |
|--------|-------|-------|-------|-------|-------|
| 0.1    | -0.00000950000 | -0.00000900000 | -0.00000800000 | -0.00000750000 | -0.00000700000 |
| 0.2    | -0.00000930000 | -0.00000830000 | -0.00000780000 | -0.00000700000 | -0.00000650000 |
| 0.3    | -0.00001200000 | -0.00001090000 | -0.00001020000 | -0.00000970000 | -0.00000920000 |
| 0.4    | -0.00001140000 | -0.00001045900 | -0.00000989800 | -0.00000920000 | -0.00000869800 |
| 0.5    | -0.00001260000 | -0.00001130000 | -0.00000901000 | -0.00000820000 | -0.00000769000 |
| 0.6    | -0.00001350000 | -0.00001200000 | -0.00001010000 | -0.00000820000 | -0.00000769000 |
| 0.7    | -0.00001430000 | -0.00001260000 | -0.00001110000 | -0.00000869000 | -0.00000810000 |
| 0.8    | -0.00001540000 | -0.00001330000 | -0.00001150000 | -0.00000849000 | -0.00000909000 |
| 0.9    | -0.00001590000 | -0.00001360000 | -0.00001150000 | -0.00000970000 | -0.00000810000 |

When data is generated from normal distribution

| $\rho$ | 75%   | 80%   | 85%   | 90%   | 95%   |
|--------|-------|-------|-------|-------|-------|
| 0.1    | -0.00003035000 | -0.00003015000 | -0.00002990000 | -0.00002965000 | -0.00002940000 |
| 0.2    | -0.00002600000 | -0.00002580000 | -0.00002560000 | -0.00002540000 | -0.00002510000 |
| 0.3    | -0.00003000000 | -0.00002980000 | -0.00002960000 | -0.00002930000 | -0.00002900000 |
| 0.4    | -0.00002770000 | -0.00002755000 | -0.00002740000 | -0.00002725000 | -0.00002710000 |
| 0.5    | -0.00002755000 | -0.00002740000 | -0.00002725000 | -0.00002710000 | -0.00002695000 |
| 0.6    | -0.00003282000 | -0.00003217000 | -0.00003156000 | -0.00003106000 | -0.00003056000 |
| 0.7    | -0.00002955000 | -0.00002853000 | -0.00002780000 | -0.00002717000 | -0.00002659000 |
| 0.8    | -0.00002930000 | -0.00003015000 | -0.00003098000 | -0.00003087000 | -0.00003076000 |
| 0.9    | -0.00003481000 | -0.00003330000 | -0.00003230000 | -0.00003180000 | -0.00003130000 |

When data is generated from Gamma distribution

| $\rho$ | 75%   | 80%   | 85%   | 90%   | 95%   |
|--------|-------|-------|-------|-------|-------|
| 0.1    | -0.00000040300 | -0.00000039500 | -0.00000038700 | -0.00000038100 | -0.00000037500 |
| 0.2    | -0.00000044700 | -0.00000043000 | -0.00000041500 | -0.00000040200 | -0.00000039000 |
| 0.3    | -0.00000045500 | -0.00000043100 | -0.00000041100 | -0.00000039200 | -0.00000037600 |
| 0.4    | -0.00000051000 | -0.00000047300 | -0.00000044000 | -0.00000041100 | -0.00000038500 |
| 0.5    | -0.00000055200 | -0.00000050300 | -0.00000046000 | -0.00000042100 | -0.00000038700 |
| 0.6    | -0.00000063200 | -0.00000056800 | -0.00000051100 | -0.00000046100 | -0.00000041600 |
| 0.7    | -0.00000070300 | -0.00000062000 | -0.00000054700 | -0.00000048300 | -0.00000042500 |
| 0.8    | -0.00000077700 | -0.00000067500 | -0.00000058500 | -0.00000050500 | -0.00000043400 |
| 0.9    | -0.00000087100 | -0.00000074600 | -0.00000063500 | -0.00000053600 | -0.00000044800 |
7. Conclusion

The following trend in the PREs is noticed from the tables: $\text{PRE}_1$ increases with the increase in value of $\rho$, while $\text{PRE}_2$, $\text{PRE}_3$, $\text{PRE}_4$, $\text{PRE}_5$, and $\text{PRE}_7$ decrease with the increase in value of $\rho$.

The proposed estimator is seen to be consistent, exists for all real values of parameters, has negligible bias, and is more efficient than 7 other contemporary estimators. Hence, the proposed estimator may be recommended for use in field work.

Data Availability

The data used in the study are generated theoretically by the equations given in this paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Taif University Researchers Supporting Project number TURSP-2020/318, Taif University, Taif, Saudi Arabia.

References

[1] M. Mohiuddin, H. Al Bayatti, and R. Kannan, “A new generalization of garima distribution with application to real life data,” Applied Mathematics & Information Sciences, vol. 15, no. 5, pp. 577–592, 2021.
[2] A. El Sheikh, S. Barakat, and S. Mohamed, “New aspects on the modified group LASSO using the least angle regression and shrinkage algorithm,” Information Sciences Letters, vol. 10, no. 3, pp. 527–536, 2021.
[3] M. S. Ahmed, O. Al-Titi, Z. Al-Rawi, and W. Abu-Dayyeh, “Estimation of a population mean using different imputation methods,” Statistics in Transition, vol. 7, no. 6, pp. 1247–1264, 2006.
[4] A. M. Basheer, E. M. Almetwally, and H. M. Okasha, “Marshall-olkin alpha power inverse Weibull distribution: non bayesian and bayesian estimations,” Journal of Statistics Applications & Probability, vol. 10, no. 2, pp. 327–345, 2021.
[5] G. Diana and P. Francesco Perri, “Improved estimators of the population mean for missing data,” Communications in Statistics-Theoretical and Methods, vol. 39, no. 18, pp. 3245–3251, 2010.
[6] F. Daghestani, A. S. Sultan, and S. Al-Moisheer, “Mixture of Lindley and Weibull distributions: properties and estimation,” Journal of Statistics Applications & Probability, vol. 10, no. 2, pp. 301–313, 2021.
[7] M. Abu-Moussa, A. Abd-Elfattah, and E. Hafez, “Estimation of StressStrength parameter for Rayleigh distribution based on progressive type-II censoring,” Information Sciences Letters, vol. 10, no. 1, pp. 101–110, 2021.
[8] M. H. Hansen and W. N. Hurwitz, “The problem of nonresponse in sample surveys,” Journal of the American Statistical Association, vol. 41, no. 236, pp. 517–529, 1946.
[9] D. F. Heitjan and S. Basu, “Distinguishing “missing at random” and “missing completely at random,” The American Statistician, vol. 50, no. 3, pp. 207–213, 1996.
[10] C. Kadilar and H. Cingal, “Estimators for the population mean in the case of missing data,” Communications in Statistics-Theoretical and Methods, vol. 37, no. 14, pp. 2226–2236, 2008.
[11] M. O. Mohamed and E. Mohamed, “Estimation of parameters of Burr distribution under SSALT,” Applied Mathematics & Information Sciences, vol. 15, no. 3, pp. 293–298, 2021.
[12] R. M. EL-Sagheer and M. A. A. Khider, “Estimation in K-stage step-stress partially accelerated life tests for generalized pareto distribution with progressive type-I censoring,” Applied Mathematics & Information Sciences, vol. 15, no. 3, pp. 299–305, 2021.
[13] D. B. Rubin, “Inference and missing data,” Biometrika, vol. 63, no. 3, pp. 581–592, 1976.
[14] G. Kalton, D. Kasprzyk, and R. Santos, “Issues of nonresponse and imputation in the survey of income and program participation,” in Current Topics in Survey Sampling, pp. 455–480, Academic Press, New York, NY, USA, 1981.
[15] G. Kalton and D. Kasprzyk, “Imputing for missing survey responses,” in Proceedings of the Section on Survey Research Methods, American Statistical Association, American Statistical Association Cincinnati, Baltimore, Maryland, August 1982.
[16] G. Kalton and L. Kish, “Some efficient random imputation methods,” Communications in Statistics -Theory and Methods, vol. 13, no. 16, pp. 1919–1939, 1984.
[17] J. K. Kim and J. Shao, Statistical Methods for Handling Incomplete Data, CRC Press, Florita, FL, USA, 2013.
[18] H. Liu, S.-G. Li, H.-X. Wang, and G.-J. Li, “Adaptive fuzzy synchronization for a class of fractional-order neural networks,” Chinese Physics B, vol. 26, no. 3, Article ID 030504, 2017.
[19] I. G. Sande, “A personal view of hot-deck imputation procedures,” Survey Methodology, vol. 5, no. 2, pp. 238–258, 1979.
[20] S. Singh and S. Horn, “Compromised imputation in survey sampling,” Metrika, vol. 51, no. 3, pp. 267–276, 2000.
[21] S. Singh and B. Deo, “Imputation by power transformation,” Statistical Papers, vol. 44, no. 4, pp. 555–579, 2003.
[22] A. K. Pandey, M. Usman, and G. N. Singh, “Optimality of ratio and regression type estimators using dual of auxiliary variable under non response,” Alexandria Engineering Journal, vol. 60, no. 5, pp. 4461–4471, 2021.
[23] A. K. Pandey, G. N. Singh, N. Sayed-Ahmed, and H. Abu-Zinadah, “Improved estimators for mean estimation in presence of missing information,” Alexandria Engineering Journal, vol. 60, no. 6, pp. 5977–5990, 2021.
[24] U. Shahzad, N. H. Al-Noor, M. Hanif, I. Sajjad, and M. Muhammad Anas, “Imputation based mean estimators in case of missing data utilizing robust regression and variance-covariance matrices,” Communications in Statistics-Simulation and Computation, pp. 1–20, 2020.
[25] G. N. Singh, A. K. Pandey, and A. K. Sharma, “Some improved and alternative imputation methods for finite population mean in presence of missing information,” Communications in Statistics-Theoretical and Methods, vol. 50, pp. 1–27, 2020.
[26] M. U. Sohail, J. Shabbir, and S. Ahmed, “A class of ratio type estimators for imputing the missing values under rank set sampling,” Journal of Statistical Theory and Practice, vol. 12, no. 4, pp. 704–717, 2018.
[27] M. U. Sohail, J. Shabbir, and C. Kadilar, “Homogeneous imputation under two phase probability proportional to size sampling,” Hacettepe Journal of Mathematics and Statistics, vol. 48, no. 5, pp. 1522–1546, 2019.
[28] M. U. Sohail, J. Shabbir, and F. Sohli, “Imputation of missing values by using raw moments,” Statistics in Transition New Series, vol. 20, no. 1, pp. 21–40, 2019.
[29] M. U. Sohail, F. Sohil, and J. Shabbir, “Comparative study of different imputation methods,” *Communications in Statistics-Simulation and Computation*, pp. 1–23, 2021.

[30] H. Toutenburg and V. K. Srivastava, “Amputation versus imputation of missing values through ratio method in sample surveys,” *Statistical Papers*, vol. 49, no. 2, pp. 237–247, 2008.

[31] G. N. Singh, S. Maurya, M. Khetan, and C. Kadilar, “Some imputation methods for missing data in sample surveys,” *Hacettepe Journal of Mathematics and Statistics*, vol. 45, no. 6, pp. 1865–1880, 2016.

[32] R Core Team, *R: A Language and Environment for Statistical Computing*. R Found. Statistical Computing, Vienna, Austria, 2018, https://www.R-project.org/.