Revista Colombiana de Estadística - Applied Statistics
January 2026, volume 49, issue 1, pp. 307–324
http://doi.org/10.15446/rce.v49n1.120337

An Improved Estimation Procedure for Population Mean in the Presence of Non–Response

Un procedimiento de estimación mejorado para la media poblacional en presencia de no respuesta

Anoop Kumar[1,a], Chandraketu Singh[2,b], Rohit Agarwal[3,c], Varun Kumar Kashyap[4,d]

[1]Department of Statistics, Central University of Haryana, Mahendergarh, India

[2]Operations and Decision Sciences Area, Jaipuria Institute of Management, Lucknow, India

[3]Department of Computer Science, UiT The Arctic University of Norway, Tromso, Norway

[4]Department of Operational and Implementation Research , ICMR- NIRRCH, Mumbai, India

Abstract

This article proposes improved estimation procedures for the population mean in the presence of non-response using auxiliary information. Based on the Hansen-Hurwitz subsampling approach, two generalized exponential-type estimators are introduced for situations where non-response occurs either only on the study variable or on both the study and auxiliary variables. The proposed estimators incorporate tuning constants and an optimization parameter to minimize the mean square error (MSE) and generate optimum versions within each class. Expressions for the bias and MSE of the estimators are derived to the first order of approximation, and the efficiency comparisons of the proposed estimators with the existing estimators are established. A comprehensive empirical evaluation demonstrates that the proposed classes consistently provide more precise estimates than the traditional estimators. The results confirm that the proposed methodology provides an efficient alternative for mean estimation under non-response settings.

***Keywords*:** Bias; Empirical study; Mean square error; Non-response.

[a]Ph.D. E-mail: anoop.asy@gmail.com
[b]Ph.D. E-mail: chandraketu.lko@gmail.com
[c]Ph.D. E-mail: agarwal.102497@gmail.com
[d]Ph.D. E-mail: kashyapv@nirrch.res.in

Resumen

Este artículo propone procedimientos de estimación mejorados para la media poblacional en presencia de no respuesta, utilizando información auxiliar. Basándose en el enfoque de submuestreo de Hansen-Hurwitz, se introducen dos estimadores generalizados de tipo exponencial para situaciones en las que la no respuesta ocurre solo en la variable de estudio o en ambas, la variable de estudio y las variables auxiliares. Los estimadores propuestos incorporan constantes de ajuste y un parámetro de optimización para minimizar el error cuadrático medio (ECM) y generar versiones óptimas dentro de cada clase. Se derivan expresiones para el sesgo y el ECM de los estimadores hasta el primer orden de aproximación, y se establecen comparaciones de eficiencia entre los estimadores propuestos y los estimadores existentes. Una evaluación numérica exhaustiva demuestra que las clases propuestas proporcionan consistentemente estimaciones más precisas que los estimadores tradicionales. Los resultados confirman que la metodología propuesta ofrece una alternativa eficiente para la estimación de la media en contextos de no respuesta.

***Palabras clave*:** Error cuadrático medio; Estudio empírico; Falta de respuesta; Sesgo.

1. Introduction

The problem of non-response often occurs while conducting the sample surveys in the field of medical sciences, social sciences, and agriculture. It is defined as the failure to obtain information from all the units in the sample. It may occur due to various reasons like surveyor unable to contact the people, refusal by the people to participate in the survey, and participants unable to answer the questionnaire for some reason, etc. For example, in exit polls, participants refuse to answer the questionnaire of the surveyor. This may lead to non-response. It is a severe issue because it reduces the size of the sample, leads to poor quality of data, and might induce a bias when the non-respondents differ significantly from respondents in the sample. Thus, it becomes very important to tackle this situation.

Many times, a variable of interest known as the study variable is closely related to a variable that is not of concern termed as an auxiliary variable. It is well established that this auxiliary variable provides extra information in the estimation of the population mean. For example, the annual earnings of the family are closely related to the number of members in a family. Therefore, knowing the number of members in a family improves the estimation of the population mean of the annual earnings. Thus, it is customary to use the auxiliary variable while considering the estimators of the population mean.

Hansen & Hurwitz (1946) introduced the technique of sub-sampling from the non-respondents group to tackle the problem of non-response in the survey data. It is well-known that the auxiliary information aid in the estimation of the population parameter. Numerous other works have been done by several authors to study the problem of the estimation of population mean in presence of non-response using information on auxiliary variable. Rao (1986) developed the ratio and regression estimators under non-response. Khare & Srivastava (1993, 1995, 1997) and Khare & Rehman (2014) introduced several estimation procedures of population mean under non-response by utilizing auxiliary information.

Bhushan & Naqvi (2015) proposed the generalized efficient classes of estimators in presence of non-response utilizing two auxiliary variables. Bhushan & Kumar (2017) developed some cost efficient classes of estimators for population mean in presence of measurement errors and non-response simultaneously. Shahzad & Hanif (2019) suggested some imputation based new estimators of population mean under non-response. Bhushan & Pandey (2019) designed an efficient estimation procedure for the population mean under non-response. Bhushan & Pandey (2020) formulated a cost-effective computational approach with non response on two occasions. Audu et al. (2021) advocated some improved estimators for population mean under non-response. Pandey et al. (2021) suggested an improved class of estimators for population mean using auxiliary information under non-response. Singh et al. (2021) developed the estimation of population mean using auxiliary information under non-response. Ahmad et al. (2022) estimated the finite population mean using dual auxiliary variable for non-response using SRS. Bhushan & Pandey (2023) presented an optimal estimation of population mean in the presence of random non-response. Rather & Kadilar (2023) constructed some improved estimators for the population mean under non-response. Bhushan & Pandey (2025) proposed an optimal random non response framework for mean estimation on current occasion. Arsalan & Shabbir (2025) estimated the mean of finite population under double sampling stratification in the presence of non-response. Unal & Kadilar (2025) generated an exponential estimator under non-response cases.

In the present paper, we have proposed two estimators for estimating the population means of the finite population using auxiliary information in the presence of non-response. The bias and MSE of the proposed estimators are derived up to the first degree of approximation. The minimum value of the MSE is also obtained by finding the optimum value of the unknown scalar used in the estimator. We also compared the efficiency of the proposed estimators with Hansen & Hurwitz (1946), ratio, and exponential estimator using an empirical study.

2. Sample Structure and Symbols

Let $\mathcal{U} = \{1, 2, \ldots, N\}$ denote a finite population consisting of N units. Due to the presence of non-response, the population is conceptually partitioned into two mutually exclusive groups:

$$\mathcal{U} = \mathcal{U}_R \cup \mathcal{U}_{NR}, \qquad \mathcal{U}_R \cap \mathcal{U}_{NR} = \emptyset,$$

where $\mathcal{U}_R$ = set of responding units with size $N_1 = |\mathcal{U}_R|$, $\mathcal{U}_{NR}$ = set of non-responding units with size $N_2 = |\mathcal{U}_{NR}|$.

Thus,

$$N = N_1 + N_2, \qquad W_1 = \frac{N_1}{N}, \qquad W_2 = \frac{N_2}{N}.$$

Let y be the study variable and x be an auxiliary variable, with corresponding population means $\bar{Y}$ and $\bar{X}$, respectively.

From the population $\mathcal{U}$, a simple random sample without replacement (SRSWOR) of size n is drawn. Under non-response, the sample is similarly partitioned:

$$n = n_1 + n_2,$$

where n_1 = number of responding sampled units, n_2 = number of non-responding sampled units.

Following Hansen & Hurwitz (1946), a sub-sample of size h_2 is drawn at random from the n_2 non-responding units. Define

$$f = \frac{n_2}{h_2}, \qquad f \geq 1.$$

The symbols used throughout the paper are summarized below:

- N, n: population size and sample size, respectively.
- y, x: study variable and auxiliary variable.
- $\bar{y}, \bar{x}$: sample means of y and x, respectively.
- $\bar{Y}, \bar{X}$: population means of y and x, respectively.
- $\bar{y}^*, \bar{x}^*$: Hansen-Hurwitz adjusted sample means in the presence of non-response.
- S_y, S_x: population standard deviations of y and x, respectively.
- C_y, C_x: coefficients of variation of y and x, respectively.
- S_y^2, S_x^2: population variances of y and x, respectively.
- S_{2y}^2, S_{2x}^2: variances of y and x among non-responding units.
- C_{2y}^2, C_{2x}^2: coefficients of variation for non-responding units.
- S_{yx}: population covariance between y and x.
- ρ_{yx}: population correlation coefficient between y and x.
- ρ_{2yx}: correlation coefficient between y and x among non-responding units.
- W_1, W_2: proportions of responding and non-responding units in the population.
- $f = \dfrac{n_2}{h_2}, \quad f_1 = \left(\frac{1}{n} - \frac{1}{N}\right), \quad f_2 = W_2\left(\frac{f-1}{N}\right).$

3. Proposed Estimators

In this section, we introduce two estimators for the estimation of the population mean $\overline{Y}$ under two distinct cases of non-response. The first case corresponds to the situation where non-response is present only in the study variable y, while the auxiliary variable x is fully observed. For this case, we develop the estimator T_1. The second case arises when non-response affects both the study variable y and the auxiliary variable x. To address this more complex setting, we propose the estimator T_2.

The functional forms of the proposed estimators T_1 and T_2 are constructed to appropriately incorporate auxiliary information and to adjust for the Hansen–Hurwitz subsampling scheme.

Case - 1: When non-response occurs only on the study variable y
In this case, the estimator T_1 is given as follows

$$T_1 = \overline{y}^* \left(\frac{\overline{X}}{\alpha_1 \overline{x} + (1-\alpha_1)\overline{X}} \right) \exp\left(\frac{a(\overline{X} - \overline{x})}{a(\overline{X} + \overline{x}) + 2b} \right), \tag{1}$$

where,

- $\overline{y}^*$ is the Hansen–Hurwitz estimator of the population mean of study variable y,
- $\bar{x}$ is the sample mean of the auxiliary variable x,
- α_1 is an unknown scalar determining the optimum form of the estimator, whose value is calculated later in this section,
- a and b are known constants chosen at the design stage. These are either real numbers or known parameters of auxiliary variable such as mean, standard deviation, coefficient of variation, coefficient of skewness, coefficient of kurtosis, etc.

The term involving α_1 represents a convex linear combination between $\bar{x}$ and $\bar{X}$, thereby stabilizing the estimator when $\bar{x}$ is noisy due to sampling fluctuations. The exponential factor improves efficiency by incorporating the discrepancy between $\bar{X}$ and $\bar{x}$ in a nonlinear fashion.

Case - 2: When non-response occurs on the study variable y as well as the auxiliary variable x also

In this case, the estimator T_2 is given as follows

$$T_2 = \overline{y}^* \left(\frac{\overline{X}}{\alpha_2 \overline{x}^* + (1-\alpha_2)\overline{X}} \right) \exp\left(\frac{a(\overline{X} - \overline{x}^*)}{a(\overline{X} + \overline{x}^*) + 2b} \right) \tag{2}$$

- $\bar{x}^*$ is the Hansen–Hurwitz estimator of the population mean of x under non-response,

- α_2 is an unknown scalar which is determined later in this section. The value of α_2 gives the optimum estimator T_2,
- a and b are same as defined earlier.

***Observation* 1.** Let $\mathcal{T} = \{T(\alpha) : \alpha \in \mathbb{R}\}$ be a parametric family of estimators for the population mean $\bar{Y}$, where α is an unknown scalar constant. The estimator

$$T(\alpha^*) \in \mathcal{T}$$

is said to be the *optimum estimator* if it minimizes the MSE within the class $\mathcal{T}$, i.e.,

$$\alpha^* = \arg\min_{\alpha} \ \mathrm{MSE}\big(T(\alpha)\big).$$

The resulting estimator $T(\alpha^*)$ is referred to as the *optimum estimator*.

3.1. Properties of Proposed Estimators

To derive the bias, MSE, and the optimum MSE of the proposed estimators T_i $(i = 1, 2)$ up to the first order of approximation, we use the following transformations:

$$\overline{y}^* = \overline{Y}(1+\epsilon_0^*), \quad \overline{x} = \overline{X}(1+\epsilon_1), \quad \overline{x}^* = \overline{X}(1+\epsilon_1^*), \tag{3}$$

where ϵ_1, ϵ_0^* and ϵ_1^* are the error terms, such that

$$\epsilon_0^* = \frac{\overline{y}^* - \overline{Y}}{\overline{Y}}, \quad \epsilon_1 = \frac{\overline{x} - \overline{X}}{\overline{X}}, \quad \epsilon_1^* = \frac{\overline{x}^* - \overline{X}}{\overline{X}}. \tag{4}$$

Substituting the appropriate values from Equation (3) to Equation (1), we get

$$\begin{aligned}
T_1 &= \overline{Y}(1+\epsilon_0^*)\left\{\frac{\overline{X}}{\alpha_1\overline{X}(1+\epsilon_1)+(1-\alpha_1)\overline{X}}\right\}\exp\left\{\frac{a(\overline{X}-\overline{X}(1+\epsilon_1))}{a(\overline{X}+\overline{X}(1+\epsilon_1))+2b}\right\}\\
&= \overline{Y}(1+\epsilon_0^*)\left(\frac{1}{1+\alpha_1\epsilon_1}\right)\exp\left\{\frac{-a\overline{X}\epsilon_1}{2(a\overline{X}+b)+a\overline{X}\epsilon_1}\right\}\\
&= \overline{Y}(1+\epsilon_0^*)(1+\alpha_1\epsilon_1)^{-1}\exp\{(1+t\epsilon_1)^{-1}-1\},
\end{aligned} \tag{5}$$

where

$$t = \frac{a\overline{X}}{2(a\overline{X}+b)}. \tag{6}$$

Expanding the right-hand side of the Equation (5) and considering up to the first degree of approximation, i.e., neglecting the terms involving the powers of error terms greater than two, we have

$$T_1 = \overline{Y}[1+\epsilon_0^* - (\alpha_1+t)\epsilon_1 - (\alpha_1+t)\epsilon_0^*\epsilon_1 + (\alpha_1^2+\alpha_1 t+1.5t^2)\epsilon_1^2]. \tag{7}$$

Similarly, substituting the appropriate values from Equation (3) to Equation (2) and considering up to the first degree of approximation while expanding, we get

$$T_2 = \overline{Y}[1+\epsilon_0^* - (\alpha_2+t)\epsilon_1^* - (\alpha_2+t)\epsilon_0^*\epsilon_1^* + (\alpha_2^2+\alpha_2 t+1.5t^2)\epsilon_1^{*2}]. \tag{8}$$

To compute the bias and MSE of T_1 and T_2, we need to compute the expectation of the error terms, the expectation of the square of the error terms and the expectation of the multiplication of two error terms. Taking expectations on both sides of the Equation (4), we get

$$E(\epsilon_0^*) = E\left(\frac{\overline{y}^* - \overline{Y}}{\overline{Y}}\right) = \frac{E(\overline{y}^*) - E(\overline{Y})}{E(\overline{Y})} = \frac{\overline{Y} - \overline{Y}}{\overline{Y}} = 0, \tag{9}$$

$$E(\epsilon_1) = E\left(\frac{\overline{x} - \overline{X}}{\overline{X}}\right) = \frac{E(\overline{x}) - E(\overline{X})}{E(\overline{X})} = \frac{\overline{X} - \overline{X}}{\overline{X}} = 0, \tag{10}$$

$$E(\epsilon_1^*) = E\left(\frac{\overline{x}^* - \overline{X}}{\overline{X}}\right) = \frac{E(\overline{x}^*) - E(\overline{X})}{E(\overline{X})} = \frac{\overline{X} - \overline{X}}{\overline{X}} = 0, \tag{11}$$

Using the same technique as above, we obtain the following results.

$$E(\epsilon_0^{*2}) = f_1 C_y^2 + f_2 C_{2y}^2, \;\; E(\epsilon_1^2) = f_1 C_x^2, \;\; E(\epsilon_1^{*2}) = f_1 C_x^2 + f_2 C_{2x}^2, \tag{12}$$

$$E(\epsilon_0^* \epsilon_1) = f_1 \rho_{yx} C_y C_x, \;\; E(\epsilon_0^* \epsilon_1^*) = f_1 \rho_{yx} C_y C_x + f_2 \rho_{2yx} C_{2y} C_{2x}. \tag{13}$$

Using the above equations, we will derive the bias and the MSE of the estimators T_i $(i = 1, 2)$.

3.1.1. Bias of Proposed Estimators

In this section, we obtain the bias of the estimators T_1 and T_2 up to the first degree of approximation.

Theorem 1. *The biases of the estimators T_i $(i = 1, 2)$ up to the first degree of approximation are derived as*

$$B(T_1) = \overline{Y} f_1 C_x [(\alpha_1^2 + \alpha_1 t + 1.5t^2) C_x - (\alpha_1 + t)\rho_{yx} C_y] \tag{14}$$

$$B(T_2) = \overline{Y} \begin{bmatrix} f_1 C_x \{(\alpha_2^2 + \alpha_2 t + 1.5t^2) C_x - (\alpha_2 + t)\rho_{yx} C_y\} \\ + f_2 C_{2x} \{(\alpha_2^2 + \alpha_2 t + 1.5t^2) C_{2x} - (\alpha_2 + t)\rho_{2yx} C_{2y}\} \end{bmatrix}. \tag{15}$$

***Proof*.** The bias of the estimator T_i $(i = 1, 2)$ is given by

$$B(T_i) = E(T_i - \overline{Y}). \tag{16}$$

First, we will compute the bias of the estimator T_1. From Equation (7), we get

$$T_1 - \overline{Y} = \overline{Y}[\epsilon_0^* - (\alpha_1 + t)\epsilon_1 - (\alpha_1 + t)\epsilon_0^* \epsilon_1 + (\alpha_1^2 + \alpha_1 t + 1.5t^2)\epsilon_1^2]. \tag{17}$$

Taking expectations on both sides of the Equation (17), we get

$$E(T_1 - \overline{Y}) = \overline{Y}[E(\epsilon_0^*) - (\alpha_1 + t)E(\epsilon_1) - (\alpha_1 + t)E(\epsilon_0^* \epsilon_1) + (\alpha_1^2 + \alpha_1 t + 1.5t^2)E(\epsilon_1^2)] \tag{18}$$

□

Using the appropriate values from equations (9), (10), (12), and (13) and substituting in Equation (18), we have

$$B(T_1) = \overline{Y} f_1 C_x [(\alpha_1^2 + \alpha_1 t + 1.5t^2) C_x - (\alpha_1 + t)\rho_{yx} C_y].$$

Similarly, using Equation (8), we have

$$T_2 - \overline{Y} = \overline{Y}[\epsilon_0^* - (\alpha_2 + t)\epsilon_1^* - (\alpha_2 + t)\epsilon_0^*\epsilon_1^* + (\alpha_2^2 + \alpha_2 t + 1.5t^2)\epsilon_1^{*2}]. \quad (19)$$

Taking expectations on both sides of Equation (19) and substituting the appropriate values from equation (9), Equation (11), Equation (12) and Equation (13), we get

$$B(T_2) = \overline{Y} \left[\begin{array}{l} f_1 C_x \{(\alpha_2^2 + \alpha_2 t + 1.5t^2) C_x - (\alpha_2 + t)\rho_{yx} C_y\} \\ + f_2 C_{2x} \{(\alpha_2^2 + \alpha_2 t + 1.5t^2) C_{2x} - (\alpha_2 + t)\rho_{2yx} C_{2y}\} \end{array} \right].$$

3.1.2. MSE of Proposed Estimator

The MSE of an estimator gives the average squared difference between the estimator and the actual parameter. The MSE of the estimator T_1 and T_2 are given here.

Theorem 2. *The MSE of the estimators T_i $(i = 1, 2)$ up to the first degree of approximation is derived as*

$$MSE(T_1) = \overline{Y}^2 [f_1 \{C_y^2 + (\alpha_1 + t)^2 C_x^2 - 2(\alpha_1 + t)\rho_{yx} C_y C_x\} + f_2 C_{2y}^2] \quad (20)$$

$$MSE(T_2) = \overline{Y}^2 \left[\begin{array}{l} f_1 C_y^2 + f_2 C_{2y}^2 + (\alpha_2 + t)^2 (f_1 C_x^2 + f_2 C_{2x}^2) \\ -2(\alpha_2 + t)(f_1 \rho_{yx} C_y C_x + f_2 \rho_{2yx} C_{2y} C_{2x}) \end{array} \right]. \quad (21)$$

Proof. The MSE of the estimators T_i $(i = 1, 2)$ is given by

$$MSE(T_i) = E(T_i - \overline{Y})^2. \quad (22)$$

First, we will compute the MSE of the estimator T_1. Squaring Equation (17) and neglecting the power of error term greater than two, we get

$$(T_1 - \overline{Y})^2 = \overline{Y}^2 [\epsilon_0^{*2} + (\alpha_1 + t)^2 \epsilon_1^2 - 2(\alpha_1 + t)\epsilon_0^* \epsilon_1]. \quad (23)$$

Taking expectations on both sides of Equation (23), we get

$$E(T_1 - \overline{Y})^2 = \overline{Y}^2 [E(\epsilon_0^{*2}) + (\alpha_1 + t)^2 E(\epsilon_1^2) - 2(\alpha_1 + t) E(\epsilon_0^* \epsilon_1)]. \quad (24)$$

□

Substituting appropriate values from Equation (12) and Equation (13) in Equation (24), we get

$$MSE(T_1) = \overline{Y}^2 [f_1 \{C_y^2 + (\alpha_1 + t)^2 C_x^2 - 2(\alpha_1 + t)\rho_{yx} C_y C_x\} + f_2 C_{2y}^2].$$

Similarly, to compute the MSE of the estimator T_2, we square the Equation (19) as

$$(T_2 - \overline{Y})^2 = \overline{Y}^2[\epsilon_0^{*2} + (\alpha_2 + t)^2\epsilon_1^{*2} - 2(\alpha_2 + t)\epsilon_0^*\epsilon_1^*]. \tag{25}$$

Taking expectations on both sides of the Equation (25) and substituting appropriate values from Equation (12) and Equation (13), we get

$$MSE(T_2) = \overline{Y}^2 \left[\begin{array}{l} f_1C_y^2 + f_2C_{2y}^2 + (\alpha_2 + t)^2(f_1C_x^2 + f_2C_{2x}^2) \\ -2(\alpha_2 + t)(f_1\rho_{yx}C_yC_x + f_2\rho_{2yx}C_{2y}C_{2x}) \end{array} \right].$$

3.1.3. Optimum MSE of Proposed Estimators

The optimum MSE of an estimator can be achieved for an optimum value of the unknown scalars α_i $(i = 1, 2)$. To get the optimum value of α_i $(i = 1, 2)$, the MSE of the estimators T_i $(i = 1, 2)$ is minimized with respect to the scalars α_i $(i = 1, 2)$.

Theorem 3. *The optimum MSE of the estimators* T_i $(i = 1, 2)$ *at* $\alpha_{1(opt)} = (\rho_{yx}C_y/C_x) - t$ *and* $\alpha_{2(opt)} = \{(f_1\rho_{yx}C_yC_x + f_2\rho_{2yx}C_{2y}C_{2x})/(f_1C_x^2 + f_2C_{2x}^2)\} - t$ *is obtained as*

$$MSE(T_1)_{opt} = \overline{Y}^2[f_1C_y^2(1 - \rho_{yx}^2) + f_2C_{2y}^2] \tag{26}$$

$$MSE(T_2)_{opt} = \overline{Y}^2\left[f_1C_y^2 + f_2C_{2y}^2 - \frac{(f_1\rho_{yx}C_yC_x + f_2\rho_{2yx}C_{2y}C_{2x})^2}{f_1C_x^2 + f_2C_{2x}^2}\right], \tag{27}$$

where $t = a\overline{X}/2(a\overline{X} + b)$.

***Proof*.** To get the minimum MSE of the estimators T_i $(i = 1, 2)$, we take the partial derivative of the MSE of T_i $(i = 1, 2)$ with respect to α_i $(i = 1, 2)$ and equate it to 0. Therefore, taking the partial derivative of Equation (20) with respect to α_1 and equating it to 0, we get

$$\frac{\partial MSE(T_1)}{\partial \alpha_1} = \overline{Y}^2 f_1[2(\alpha_1 + t)C_x^2 - 2\rho_{yx}C_yC_x] = 0. \tag{28}$$

Solving Equation (28), we get

$$\alpha_{1(opt)} = \frac{\rho_{yx}C_y}{C_x} - t \tag{29}$$

□

Substituting the value of $\alpha_{1(opt)}$ from Equation (29) to Equation (20), we get

$$MSE(T_1)_{opt} = \overline{Y}^2[f_1C_y^2(1 - \rho_{yx}^2) + f_2C_{2y}^2].$$

Similarly, taking the partial derivative of the Equation (21) wrt α_2 and equating it to 0, we get

$$\frac{\partial MSE(T_2)}{\partial \alpha_2} = \overline{Y}^2[2(\alpha_2 + t)(f_1 C_x^2 + f_2 C_{2x}^2) - 2(f_1 \rho_{yx} C_y C_x + f_2 \rho_{2yx} C_{2y} C_{2x})] = 0. \tag{30}$$

Solving Equation (30), we get

$$\alpha_{2(opt)} = \frac{f_1 \rho_{yx} C_y C_x + f_2 \rho_{2yx} C_{2y} C_{2x}}{f_1 C_x^2 + f_2 C_{2x}^2} - t. \tag{31}$$

Substituting the value of $\alpha_{2(opt)}$ from Equation (31) in Equation (21), we get

$$MSE(T_2)_{opt} = \overline{Y}^2 \left[f_1 C_y^2 + f_2 C_{2y}^2 - \frac{(f_1 \rho_{yx} C_y C_x + f_2 \rho_{2yx} C_{2y} C_{2x})^2}{f_1 C_x^2 + f_2 C_{2x}^2} \right].$$

4. Efficiency Comparisons

In this section, we compare the efficiency of the proposed estimator with the well known existing estimators. The percent relative efficiency (PRE) of the proposed estimators is calculated with respect to the Hansen & Hurwitz (1946), ratio, and exponential estimators.

The PRE of the estimator T_1 with respect to (wrt) different estimators are given as:

(i) PRE of the estimator T_1 wrt Hansen & Hurwitz (1946) estimator ($\mathbf{\overline{y}_w^*}$):

$$\frac{Var(\overline{y}_w^*)}{MSE(T_1)_{opt}} * 100, \tag{32}$$

where $Var(\overline{y}_w^*) = f_1 S_y^2 + f_2 S_{2y}^2$.

(ii) PRE of the estimator T_1 wrt ratio estimator ($\mathbf{\overline{y}_R^*}$):

$$\frac{MSE(\overline{y}_R^*)}{MSE(T_1)_{opt}} * 100, \tag{33}$$

where $MSE(\overline{y}_R^*) = \overline{Y}^2[f_1(C_y^2 + C_x^2 - 2\rho_{yx} C_y C_x) + f_2 C_{2y}^2]$.

(iii) PRE of the estimator T_1 wrt exponential estimator ($\mathbf{\overline{y}_e^*}$):

$$\frac{MSE(\overline{y}_e^*)}{MSE(T_1)_{opt}} * 100, \tag{34}$$

where $MSE(\overline{y}_e^*) = \overline{Y}^2[f_1(C_y^2 + \frac{C_x^2}{4} - \rho_{yx} C_y C_x) + f_2 C_{2y}^2]$.

Now, we compare the estimator T_2 wrt different estimators and obtain the PRE as follow:

(i) PRE of the estimator T_2 wrt Hansen & Hurwitz (1946) estimator ($\overline{\mathbf{y}}^*_{\mathbf{w}}$):

$$\frac{Var(\overline{y}^*_w)}{MSE(T_2)_{opt}} * 100. \tag{35}$$

(ii) PRE of the estimator T_2 wrt ratio estimator ($\overline{\mathbf{y}}^{**}_{\mathbf{R}}$):

$$\frac{MSE(\overline{y}^{**}_R)}{MSE(T_2)_{opt}} * 100, \tag{36}$$

where $MSE(\overline{y}^{**}_R) = \overline{Y}^2[f_1(C_y^2+C_x^2-2\rho_{yx}C_yC_x)+f_2(C_{2y}^2+C_{2x}^2-2\rho_{2yx}C_{2y}C_{2x})]$.

(iii) PRE of the estimator T_2 wrt exponential estimator ($\overline{\mathbf{y}}^{**}_{\mathbf{e}}$):

$$\frac{MSE(\overline{y}^{**}_e)}{MSE(T_2)_{opt}} * 100, \tag{37}$$

where $MSE(\overline{y}^{**}_e) = \overline{Y}^2[f_1(C_y^2+\frac{C_x^2}{4}-\rho_{yx}C_yC_x)+f_2(C_{2y}^2+\frac{C_{2x}^2}{4}-\rho_{2yx}C_{2y}C_{2x})]$.

5. Empirical Study

In this section, we will compute the performance of our proposed estimators. To compare the efficiency of the proposed estimators, we have calculated the PRE of our proposed estimators with respect to Hansen & Hurwitz (1946), ratio, and exponential estimators on real and artificially generated datasets.

5.1. Study on Real Data

Population I: (Source: UCI Machine Learning Repository) The real estate valuation dataset contains historical market data of real estate valuation from Sindian Dist., New Taipei City, Taiwan across 2012 to 2013. The responding and non-responding units are chosen randomly, and even the sample from the population is chosen randomly. Here, the study and auxiliary variables are: $y =$ house price of unit area, and $x =$ distance to the nearest MRT station.

The different statistics of the data are as follows: $N = 414$, $n = 124$, $\overline{Y} = 37.9802$, $\overline{X} = 1083.8857$, $\rho_{yx} = -0.6736$, $S_y = 13.59$, and $S_x = 1260.584$.

The PRE of the proposed estimators T_1 and T_2 based on the population I is computed by varying W_2 and f, and is shown in Table 1 and Table 2, respectively.

Population II: (Source: UCI Machine Learning Repository) Boston Housing concerns housing values in the suburbs of Boston. The responding and non-responding units are chosen randomly, and even the sample from the population is chosen randomly. Here, the study and auxiliary variables are given as follows: $y =$ Lower status of the population, and $x =$ Index of accessibility to radial highways.

The different statistics of the data are: $N = 506$, $n = 151, \overline{Y} = 12.6531$, $\overline{X} = 9.5494$, $\rho_{yx} = 0.4887$, $S_y = 7.1340$, and $S_x = 8.6987$.

The PRE of the proposed estimators T_1 and T_2 on the population II is computed by varying W_2 and f, and is shown in Table 3 and Table 4, respectively.

Population III: (Khare & Sinha, 2007) This dataset contains data on the physical growth of an upper socio-economic group of 95 school children of Varanasi under an ICMR study, Department of Pediatrics, BHU, during 1983-1984. The first 1/4th (i.e., $W_2 = 0.25$) of the data is considered to be non-response. Here, the study and auxiliary variables are given as follows: $y =$ the weight in kg, and $x =$ the skull circumference in cm.

The different statistics of the data are: $N = 95$, $N_2 = 24$ $\overline{Y} = 19.49$, $\overline{X} = 51.17$, $\rho_{yx} = 0.32$, $\rho_{2yx} = 0.47$, $C_y = 0.15$, $C_x = 0.03$, $C_{2y} = 0.12$, and $C_{2x} = 0.02$.

The PRE of the proposed estimators T_1 and T_2 on the population III is computed by varying n and f, and is shown in Table 5 and Table 6, respectively.

Table 1: PRE of estimator T_1 with respect to $\overline{y}_w^*$, $\overline{y}_R^*$, $\overline{y}_e^*$ using Population I.

$\mathbf{W_2}$	f	$\overline{\mathbf{y}}_\mathbf{w}^*$	$\overline{\mathbf{y}}_\mathbf{R}^*$	$\overline{\mathbf{y}}_\mathbf{e}^*$
0.1	1.5	177.8645	2719.2822	998.9419
	2.0	173.3489	2546.1977	939.5390
	2.5	169.3931	2394.5705	887.5003
	3.0	165.8991	2260.6433	841.5363
0.2	1.5	170.7614	2508.3029	926.5334
	2.0	161.5836	2202.2092	821.4816
	2.5	154.4751	1965.1499	740.1245
	3.0	148.8076	1776.1373	675.2530
0.3	1.5	162.9004	2261.9607	841.9884
	2.0	150.3551	1853.4141	701.7746
	2.5	141.7976	1574.7332	606.1308
	3.0	135.5873	1372.4887	536.7201

Table 2: PRE of estimator T_2 with respect to $\overline{y}_w^*$, $\overline{y}_R^{**}$, $\overline{y}_e^{**}$ using Population I.

$\mathbf{W_2}$	f	$\overline{\mathbf{y}}_\mathbf{w}^*$	$\overline{\mathbf{y}}_\mathbf{R}^{**}$	$\overline{\mathbf{y}}_\mathbf{e}^{**}$
0.1	1.5	182.6002	2959.0123	1077.1886
	2.0	182.2476	2996.9705	1086.5648
	2.5	181.9831	3032.7329	1095.5102
	3.0	181.7867	3066.4405	1104.0288
0.2	1.5	184.8235	3066.4944	1114.0374
	2.0	186.3974	3194.6759	1154.5505
	2.5	187.8082	3306.9006	1190.0527
	3.0	189.0758	3405.9616	1221.4125
0.3	1.5	185.7558	3117.6530	1131.8069
	2.0	187.8927	3273.0873	1182.1700
	2.5	189.6277	3397.8757	1222.6199
	3.0	191.0625	3500.2636	1255.8183

Table 3: PRE of estimator T_1 with respect to $\overline{y}_w^*$, $\overline{y}_R^*$, $\overline{y}_e^*$ using Population II.

$\mathbf{W_2}$	f	$\overline{\mathbf{y}}_\mathbf{w}^*$	$\overline{\mathbf{y}}_\mathbf{R}^*$	$\overline{\mathbf{y}}_\mathbf{e}^*$
0.1	1.5	130.9725	254.0821	112.3564
	2.0	130.6296	243.1354	111.4785
	2.5	130.3321	233.6410	110.7171
	3.0	130.0716	225.3278	110.0505
0.2	1.5	127.1662	242.3486	111.4154
	2.0	124.0369	224.1263	109.9541
	2.5	121.6179	210.0400	108.8245
	3.0	119.6920	198.8249	107.9251
0.3	1.5	124.4455	228.6885	110.3200
	2.0	120.0971	204.7372	108.3992
	2.5	117.1134	188.3025	107.0812
	3.0	114.9390	176.3260	106.1208

Table 4: PRE of estimator T_2 with respect to $\overline{y}_w^*$, $\overline{y}_R^{**}$, $\overline{y}_e^{**}$ using Population II.

$\mathbf{W_2}$	f	$\overline{\mathbf{y}}_\mathbf{w}^*$	$\overline{\mathbf{y}}_\mathbf{R}^{**}$	$\overline{\mathbf{y}}_\mathbf{e}^{**}$
0.1	1.5	131.9634	269.0465	114.1724
	2.0	132.5066	271.0240	114.8872
	2.5	133.0639	272.8072	115.5348
	3.0	133.4668	274.4232	116.1238
0.2	1.5	131.0059	268.5115	113.8756
	2.0	130.7271	269.8384	114.2719
	2.5	130.5081	270.9180	114.5958
	3.0	130.3315	271.8137	114.8654
0.3	1.5	130.3985	264.6616	113.4767
	2.0	129.7422	263.1915	113.5462
	2.5	129.2700	262.1332	113.5983
	3.0	128.9140	261.3352	113.6389

Table 5: PRE of estimator T_1 with respect to $\overline{y}_w^*$, $\overline{y}_R^*$, $\overline{y}_e^*$ using Population III.

n	f	$\overline{\mathbf{y}}_\mathbf{w}^*$	$\overline{\mathbf{y}}_\mathbf{R}^*$	$\overline{\mathbf{y}}_\mathbf{e}^*$
15	1.5	110.3163	101.4507	104.8761
	2.0	109.4152	101.3240	104.4502
	2.5	108.6589	101.2177	104.0927
	3.0	108.0150	101.1271	103.7884
25	1.5	110.1772	101.4312	104.8103
	2.0	109.1860	101.2918	104.3418
	2.5	108.3707	101.1771	103.9565
	3.0	107.6883	101.0812	103.6339
35	1.5	109.9974	101.4059	104.7253
	2.0	108.8971	101.2512	104.2053
	2.5	108.0150	101.1271	103.7884
	3.0	107.2921	101.0254	103.4466

Table 6: PRE of estimator T_2 with respect to $\overline{y}_w^*$, $\overline{y}_R^{**}$, $\overline{y}_e^{**}$ using Population III.

n	f	$\overline{y}_w^*$	$\overline{y}_R^{**}$	$\overline{y}_e^{**}$
15	1.5	112.2732	101.9949	106.0411
	2.0	113.0567	102.3697	106.6378
	2.5	113.7692	102.7257	107.1869
	3.0	114.4195	103.0623	107.6929
25	1.5	112.3899	102.0495	106.1294
	2.0	113.2671	102.4734	106.7993
	2.5	114.0550	102.8724	107.4088
	3.0	114.7661	103.2458	107.9645
35	1.5	112.5429	102.1218	106.2456
	2.0	113.5390	102.6092	107.0088
	2.5	114.4194	103.0623	107.6929
	3.0	115.2028	103.4808	108.3081

5.2. Study on Artificial Data

Population IV: We have generated a normal population of size N= 450 such that the mean and the standard deviation of the study variable (y) are $\overline{Y} = 12$ and $S_y = 2.0402$, respectively, whereas the mean and the standard deviation of the auxiliary variable (x) are $\overline{X} = 28$ and $S_x = 18028$, respectively. The correlation coefficient between the variable y and variable x is fixed at $\rho_{xy} = 0.6812$, and the sample size is fixed at $n = 135$. The PRE of the proposed estimators T_1 and T_2 on the population IV is computed by varying W_2 and f, and shown in Table 7 and Table 8, respectively.

Population V: We have also investigated the impact of the correlation coefficient ρ_{xy} between the study variable and the auxiliary variable on the PRE of estimator T_1 and T_2. For this, we have generated a population of same size $N = 300$ such that the mean and the standard deviation of the study variable (y) are $\overline{Y} = 14$ and $S_y = 2.3$, respectively, whereas the mean and the standard deviation of the auxiliary variable (x) are $\overline{X} = 18$ and $S_x = 1.5$, respectively with varying correlation coefficient. We have fixed the value of $W_2 = 0.2$, $f = 1.5$, and sample size $n = 75$. The PRE of the proposed estimators T_1 and T_2 on population V is computed by varying ρ_{xy}, and shown in Table 9 and Table 10, respectively.

6. Interpretation of Results

The empirical results are shown in Tables 1-10 for real and artificial datasets. The following interpretations are made from the results obtained through an empirical study on different populations.

(i). From Tables Tables 1-10, it is observed that the PRE of the proposed estimators T_1 and T_2 is greater than 100 in all the situations, which shows that our estimators are better than Hansen & Hurwitz (1946), ratio, and exponential estimators.

(ii). From Tables 1, 3, and 7, it is observed that

Table 7: PRE of estimator T_1 with respect to $\overline{y}_w^*$, $\overline{y}_R^*$, $\overline{y}_e^*$ using Population IV.

$\mathbf{W_2}$	f	$\overline{\mathbf{y}}_{\mathbf{w}}^*$	$\overline{\mathbf{y}}_{\mathbf{R}}^*$	$\overline{\mathbf{y}}_{\mathbf{e}}^*$
0.1	1.5	170.4260	113.2432	136.2143
	2.0	164.4864	112.1394	133.1960
	2.5	159.4607	111.2055	130.6420
	3.0	155.1531	110.4050	128.4531
0.2	1.5	162.6781	111.7314	132.0804
	2.0	152.6510	109.8196	126.8523
	2.5	145.4341	108.4436	123.0895
	3.0	139.9912	107.4058	120.2516
0.3	1.5	159.1088	110.9505	129.9450
	2.0	148.0017	108.7724	123.9885
	2.5	140.5801	107.3169	120.0086
	3.0	135.2708	106.2757	117.1614

Table 8: PRE of estimator T_2 with respect to $\overline{y}_w^*$, $\overline{y}_R^{**}$, $\overline{y}_e^{**}$ using Population IV.

$\mathbf{W_2}$	f	$\overline{\mathbf{y}}_{\mathbf{w}}^*$	$\overline{\mathbf{y}}_{\mathbf{R}}^{**}$	$\overline{\mathbf{y}}_{\mathbf{e}}^{**}$
0.1	1.5	177.1156	114.5496	139.6928
	2.0	176.7237	114.5336	114.5336
	2.5	176.3712	114.5194	139.4492
	3.0	176.0523	114.5068	139.3450
0.2	1.5	176.7119	114.7290	139.5796
	2.0	176.0760	114.8663	139.3945
	2.5	175.5799	114.9844	139.2565
	3.0	175.1829	115.0869	139.1506
0.3	1.5	176.0051	113.8712	138.3832
	2.0	174.9099	113.3800	137.3544
	2.5	174.0946	113.0150	136.5879
	3.0	173.4641	112.7334	135.9948

Table 9: PRE of estimator T_1 with respect to $\overline{y}_w^*$, $\overline{y}_R^*$, $\overline{y}_e^*$ using Population V.

$\rho_{\mathbf{xy}}$	$\overline{\mathbf{y}}_{\mathbf{w}}^*$	$\overline{\mathbf{y}}_{\mathbf{R}}^*$	$\overline{\mathbf{y}}_{\mathbf{e}}^*$
-0.9	375.5196	786.5354	558.2098
-0.8	244.8205	495.5283	354.9308
-0.7	187.2284	364.3606	263.9723
-0.6	155.3389	289.4958	212.5416
-0.5	135.5204	240.9416	179.6135
-0.4	122.3860	206.7460	156.8379
0.4	112.6607	101.4632	101.3790
0.5	124.1488	100.0062	105.8433
0.6	142.3156	101.3274	114.6569
0.7	172.2964	106.686	130.7394
0.8	227.3337	119.6378	161.7542
0.9	354.1622	152.9114	234.7838

TABLE 10: PRE of estimator T_2 with respect to $\overline{y}_w^*$, $\overline{y}_R^{**}$, $\overline{y}_e^{**}$ using Population V

ρ_{xy}	$\overline{y}_w^*$	$\overline{y}_R^{**}$	$\overline{y}_e^{**}$
-0.9	559.9721	1250.7941	866.8216
-0.8	297.9323	642.9449	449.3450
-0.7	211.0722	437.6198	309.1646
-0.6	168.3218	333.8259	238.8759
-0.5	143.3677	270.8536	196.7217
-0.4	127.4096	228.2752	168.6860
0.4	113.6502	101.7332	101.4139
0.5	127.1503	100.0058	106.5901
0.6	149.4983	101.6729	117.3414
0.7	188.9680	108.6426	138.2684
0.8	270.7350	127.2276	183.5915
0.9	521.5821	189.5687	324.9960

a) The PRE of the estimator T_1 for a fixed value of W_2 decreases as the value of f increases.

b) The PRE of the estimator T_1 for a fixed value of f decreases as the value of W_2 increases.

(iii). From Tables 5 and 6, it is observed that

a) The PRE for a fixed value of n decreases for the estimator T_1 and increases for the estimator T_2 as the value of f increases.

b) The PRE for a fixed value of f decreases for the estimator T_1 and increases for the estimator T_2 as the value of f increases.

(iv). From Tables 9 and 10, it is observed that the PRE of the proposed estimator increases as the value of the correlation coefficient increases irrespective of its sign. It means that the efficiency of the proposed estimators is better for strongly correlated (positive or negative) data.

7. Conclusions

In this paper, we proposed two exponential-type estimators for estimating the finite population mean in the presence of non-response by efficiently incorporating auxiliary information. The first estimator accommodated the case where the auxiliary variable is fully observed, while the second estimator addressed a more general and practically relevant case in which both the study and auxiliary variables were subject to non-response. For each estimator, we formally derived the bias and MSE expressions up to the first order of approximation and obtained the optimal estimators by minimizing their respective MSEs with respect to the tuning constants.

The empirical results consistently show that the proposed estimators outperform the traditional Hansen–Hurwitz, ratio, and exponential estimators. The superiority of the proposed estimators is evident across varying levels of non-response, subsampling rates, and correlation structures, confirming their practical applicability. In particular, the gains in efficiency become more pronounced when the auxiliary variable exhibits a strong positive or negative correlation with the study variable. Given their theoretical soundness and empirical advantages, the proposed estimators are recommended for application in surveys affected by non-response. Future research may extend this framework to stratified sampling, multi-phase sampling, or situations involving measurement error, imputation procedures, or complex non-response models.

Acknowledgement

The authors would like to express their heartfelt thanks to the hon'ble reviewers for their constructive suggestions and to the editor-in-chief.

[Received: May 2025 — Accepted: December 2025]

References

Ahmad, S., Hussain, S., Aamir, M., Khan, F., Alshahrani, M. N. & Alqawba, M. (2022), 'Estimation of finite population mean using dual auxiliary variable for non-response using simple random sampling', *AIMS Mathematics* **7**(3), 4592–4613.

Arsalan, S. M. & Shabbir, J. (2025), 'Estimating the mean of finite population under double sampling stratification in the presence of non-response', *Austrian Journal of Statistics* **54**(5), 100–120.

Audu, A., Singh, R., Khare, S. & Dauran, N. (2021), 'Some improved estimators for population mean under non-response', *Journal of Statistics and Management Systems* **24**, 573–589.

Bhushan, S. & Kumar, A. (2017), 'On cost efficient classes of estimators for population mean in presence of measurement errors and non-response simultaneously', *International Journal of Statistics and Systems* **12**(1), 93–117.

Bhushan, S. & Naqvi, N. (2015), 'Generalized efficient classes of estimators in presence of non-response using two auxiliary variables', *Journal of Statistics and Management Systems* **18**(6), 573–602.

Bhushan, S. & Pandey, A. P. (2019), 'An efficient estimation procedure for the population mean under non-response', *Statistica* **79**(4), 363–378.

Bhushan, S. & Pandey, A. P. (2023), 'Optimal estimation of population mean in the presence of random non-response', *Thailand Statistician* **21**(3), 616–630.

Bhushan, S. & Pandey, S. (2020), 'A cost-effective computational approach with non-response on two occasions', *Communications in Statistics–Theory and Methods* **49**(20), 4951–4973.

Bhushan, S. & Pandey, S. (2025), 'Optimal random non-response framework for mean estimation on current occasion', *Communications in Statistics–Theory and Methods* **54**(4), 1205–1231.

Hansen, M. H. & Hurwitz, W. N. (1946), 'The problem of non-response in sample surveys', *Journal of the American Statistical Association* **41**, 517–529.

Khare, B. & Rehman, H. (2014), 'A class of estimators for population mean using auxiliary information under non-response', *Journal of Social and Economic Statistics* **3**, 1–9.

Khare, B. & Sinha, R. (2007), Some contributions to sampling theory, *in* 'Statistical techniques in life testing, reliability, sampling theory and quality control', Vol. 1, pp. 63–171.

Khare, B. & Srivastava, S. (1993), 'On non-response in survey sampling', *National Academy Science Letters (India)* **16**, 111–114.

Khare, B. & Srivastava, S. (1995), 'A general class of estimators for population mean in presence of non-response', *Proceedings of the National Academy of Sciences, India, Section A* **65**, 195–204.

Khare, B. & Srivastava, S. (1997), 'Estimation of population mean using auxiliary information in presence of non-response', *Communications in Statistics–Theory and Methods* **26**, 1779–1791.

Pandey, A., Usman, M. & Singh, G. (2021), 'An improved class of estimators for population mean using auxiliary information under non-response', *Alexandria Engineering Journal* **60**, 4461–4471.

Rao, P. S. R. S. (1986), 'Ratio and regression estimates under non-response', *Survey Methodology* **12**, 217–230.

Rather, K. U. I. & Kadilar, C. (2023), 'Improved estimators for the population mean under non-response', *Gazi University Journal of Science* pp. 1–1.

Shahzad, U. & Hanif, M. (2019), 'Some imputation based new estimators of population mean under non-response', *Journal of Statistics and Management Systems* **22**(8), 1381–1399.

Singh, C., Kamal, M., Singh, G. N. & Kim, J. M. (2021), 'Estimation of population mean using auxiliary information under non-response: A healthcare application', *Risk Management and Healthcare Policy* pp. 1595–1613.

Unal, C. & Kadilar, C. (2025), 'Exponential estimators under non-response cases', *Sinop Universitesi Fen Bilimleri Dergisi* **10**(1), 60–72.