



OPEN Log-ratio type estimation for the finite population mean under simple random sampling without replacement with theory, simulation and application

Fazal Shakoor¹, Muhammad Atif¹, Hameed Ali², Abdulrahman Obaid Alshammari³, Bilal Himmat⁴✉ & Khaled Kefi⁵✉

We propose two novel logarithmic ratio-type estimators for the finite-population mean under simple random sampling without replacement (SRSWOR). The estimators integrate a logarithmic transformation of the auxiliary variable to stabilize variance, reduce the influence of outliers, and better capture nonlinear relationships between study and auxiliary variables. We derive closed-form expressions for first-order bias and mean squared error (MSE) and obtain analytic expressions for the optimal tuning constants by direct minimization of the approximate MSE. A comprehensive numerical study, comprising five real engineering datasets and extensive Monte-Carlo simulations from multivariate normal, log-normal and gamma populations, evaluates finite-sample behavior across a range of sample sizes and correlation structures. The proposed estimators consistently reduce MSE and deliver large percent-relative-efficiency (PRE) gains relative to the classical sample mean and common competitors (empirical PREs ≈ 283 ; simulation PREs up to ≈ 670), with especially large and stable improvements under skewed or heavy-tailed populations. Theoretical formulas and simulation evidence align closely, showing robustness to nonlinearity and skewness while retaining simple implementation for practitioners. Results are derived under SRSWOR using first-order approximations; extensions to higher-order corrections, stratified and two-phase designs, and uncertainty in auxiliary means are recommended for future work.

Keywords Estimation, Auxiliary information, Bias, Efficiency, Logarithmic Estimator, PRE, Ratio Estimator, Simple Random Sampling

List of symbols

N	Size of the population
n	Size of the sample
S_y^2	Variance of y
S_x^2	Variance of x
s_{yx}	Covariance
C_y	Coefficient of variation (y)
C_x	Coefficient of variation (x)
C_{yx}	Coefficient of covariance x, y
α and β	Generalizing constants
e_0, e_1	Relative error in $\bar{y} = \frac{\bar{y} - \mu_y}{\mu_y}$ and $\bar{x} = \frac{\bar{x} - \mu_x}{\mu_x}$ respectively

¹Department of Statistics, University Peshawar, Peshawar, KP, Pakistan. ²Higher Education, Archives and Libraries Department, Government of Khyber Pakhtunkhwa, Peshawar, Pakistan. ³Department of Mathematics, College of Science, Jouf University, 72388 Sakaka, Saudi Arabia. ⁴Department of Software Engineering, Faculty of Computer Science, Sayed Jamaluddin Afghani University(SJAU), Asadabad, Kunar, Afghanistan. ⁵Center for Scientific Research and Entrepreneurship, Northern Border University, 73213 Arar, Saudi Arabia. ✉email: bilalhimmat@sjau.edu.af; Khaled_kefi@yahoo.fr

S_y^2	Population variance of (y)
S_x^2	Population variance of (x)
ρ_{yx}	Population correlation coefficient
\bar{Y}	Mean of the population of y
\bar{X}	Population Mean of x
R	Population ratios
f	Sampling fraction
T_{ln1}	First proposed estimator
T_{ln2}	Second proposed estimator
k_1, k_2, k_3, k_4	Optimizing constants

Accurate estimation of population means from sampled data lies at the heart of survey statistics and many applied fields, from official statistics and environmental monitoring to engineering quality control and experimental sciences¹. When using simple random sampling (SRS) to collect data on system performance or material properties, utilizing supplementary information, such as known historical measurements or operating conditions, can enhance estimator efficacy. Estimation methods such as the ratio estimator and logarithmic-ratio type estimator utilize this auxiliary data to adjust the primary estimates, reducing the mean squared error (MSE) compared to baseline classical sample means². For instance, in structural health monitoring, using temperature as an auxiliary variable can improve stress estimation in materials under load³. For a variety of reasons, statisticians urge the incorporation of supplemental information during the estimation step. For instance, the pioneer work in this regard is due to Cochran⁴. Since auxiliary information are available to the researcher, they can be used effectively at either, the design stage or estimation stage^{5,6}. Since the supplementary information explain variation in the main study variable due to their correlation, it utilizes extra information and enhances the efficiency of estimates⁷. The method of estimation of parameter rely on the nature of relation between the survey and supplementary variable. When the auxiliary variable has a positive correlation with the primary study variable the ratio-method of estimate performs efficiently⁸. Product form estimators, on the other hand, typically perform better when the regression streak crosses the origin and the connection between the study and auxiliary variables is linear and strongly negative. These considerations underscore the importance of selecting an appropriate estimator based on the nature of the statistical association, including both the direction and the structure of relationship^{9–14}.

The concept of using auxiliary information to enhance estimation accuracy was first introduced by Cochran⁴, who showed that incorporating related information can substantially improve the precision of survey estimates.. Later, Bahl and Tuteja¹⁵, expanded on this concept by proposing exponential ratio and product-type estimators, marking a major step forward in the effective use of auxiliary variables. These estimators exploit the functional relationship between the study and auxiliary variables to obtain more efficient and accurate estimates of population parameters, especially when the variables are highly correlated. Since then, numerous researchers have expanded and refined these concepts in the field of survey sampling. Notable contributions include those by Izunobi and Onyeka¹⁶, Kadilar and Cingi¹⁴, Singh et al.¹⁷, Khoshnevisan et al.¹⁸, Onyeka et al.¹⁹, Singh et al.²⁰, Bhushan et al., Gupta and Shabbir²¹, Azeem et al.²², Sher et al.²³, Ahmad et al.²⁰, and Subramani²⁵ long et al.¹⁵. Building on this extensive body of work, the present study extends the use of auxiliary information by developing logarithmic ratio and product-form estimators, aimed at further improving the estimation of population means.

Although several ratio and product-type estimators have been developed to improve the efficiency of population mean estimation using auxiliary information, there is still a gap in understanding how to effectively combine auxiliary variables with the study variable to achieve the greatest efficiency gains. For example, most of these estimators rely on linear or conventional transformations that may not fully exploit the structure of non-linear between study and auxiliary variables. In particular, limited attention has been given to the use of logarithmic transformations within the ratio estimation framework under simple random sampling. Furthermore, existing logarithmic estimators often lack general applicability and are rarely validated on real-world engineering datasets, where non-linearity and high correlations are common. Thus, there is a need to develop a more flexible and efficient estimator that integrates logarithmic transformation with ratio-type estimation and to evaluate its performance through theoretical comparison and empirical validation¹⁶.

Novelty and significance

This study proposes a logarithmic ratio-type estimator for the population mean under simple random sampling without replacement (SRSWOR), specifically designed for engineering applications where relationships between study and auxiliary variables are complex and intricate. By applying a logarithmic transformation to the auxiliary variable, the estimator linearizes variation e.g. $Y \approx \alpha X^\beta$, stabilises variance, reduces the influence of outliers which causes nonlinearity. We derive closed-form expressions for the bias and mean square error (MSE) up to the first-order approximation under SRSWOR, and conduct a numerical assessment, based on the correlation, coefficients of variation, and skewness of the auxiliary variable, to evaluate the efficiency gains of the proposed estimator obtained through the logarithmic transformation of the auxiliary variable. Finite-sample performance is assessed through Monte Carlo experiments that emulate engineering sampling conditions (varying sample size and correlation levels) and also validated empirically on real engineering datasets. The estimator's simplicity, theoretical grounding, and robustness make it a readily applicable improvement for mean estimation in engineering quality control, monitoring, and experimental studies, with straightforward extensions to stratified and two-phase sampling designs.

The structure of the paper is as follows:

The remainder of the paper proceeds as follows. Section 2 formalizes notation, recalls relevant classical estimators, and sets up the linearization framework under SRSWOR. Section 3 introduces the two proposed

logarithmic ratio-type estimators and develops their first-order bias and MSE; closed-form optimal constants are derived there. Section 4 discusses analytic efficiency conditions and compares the new forms with established estimators. Section 5 presents empirical evaluations on five engineering datasets and a comprehensive Monte Carlo study across Normal, Lognormal, and Gamma populations. Finally, Sect. 6 summarizes the implications, limitations, and practical recommendations for applying the proposed estimators in engineering and survey practice.

Methodology

Examine a random sample of n units selected from a population of $(N = N_1, N_2, \dots, N_i)$ units using the simple random sampling without replacement (SRSWOR) framework. Assuming that y is the research variable and x is the auxiliary variable, a sample of size n is chosen from a population of size N using simple random sampling without replacement (SRSWOR). Assuming that the population mean μ_x of the auxiliary variable is known, denote the study variable by y and the auxiliary variable by x . To enhance the estimation of the population, mean of the research variable y , the estimators consider data from the auxiliary variable. It aims to decrease sampling error and increase precision by fusing the sample statistics with the known population mean of the auxiliary variable x assuming that the auxiliary variable's population mean, μ_x , is known.

Let us define the sampling error as follows:

$$\mu_x = \frac{1}{N} \sum_{i=1}^N (X_i - \mu) \text{ and } \mu_y = \frac{1}{N} \sum_{i=1}^N (Y_i - \mu),$$

$$e_0 = \frac{\bar{y} - \mu_y}{\mu_y}, \text{ and } e_1 = \frac{\bar{x} - \mu_x}{\mu_x}, \text{ such that } E(e_0) = E(e_1) = 0, E(e_0^2) = \frac{V(\bar{y})}{\mu_y^2} = \left(\frac{1-n}{n}\right) \frac{S_y^2}{\mu_y^2}, \text{ and}$$

$$E(e_1^2) = \frac{V(\bar{x})}{\mu_x^2} = \left(\frac{1-n}{n}\right) \frac{S_x^2}{\mu_x^2}, \text{ and } E(e_0 e_1) = \left(\frac{1-n}{n}\right) \frac{S_{yx}}{\mu_y \mu_x}.$$

Existing estimators

The sample statistics and the known population mean of the auxiliary variable are combined to create the classical estimator. thereby aiming to reduce the sampling error and enhance estimation precision, the classical is given in the following Eq. (1) as:

$$\frac{\sum_{i=1}^n y_i}{n} = T_0 \text{ MSE}(T_0) = \lambda \bar{Y}^2 C_y^2 \quad (1)$$

where λ is the *fpc* (finite population correction factor) given by:

$$\lambda = \frac{1-f}{n}.$$

The classic ratio-type estimator was first presented by Cochran (Cochran, 1940) and is represented as follows in Eq. (2):

$$T_1 = \mu_x \frac{\bar{y}}{\bar{x}}$$

The estimator t is biased for the population mean μ_x , and its bias and MSE are determined up to the first-order approximation may be written as follows, where y and x stand for the sample means of the study and auxiliary variables, respectively:

$$\text{Bias}(T_1) = \left(\frac{1-f}{n}\right) \left(\frac{1}{\mu_x}\right) (DS_x^2 - S_{xy})$$

$$\text{MSE}(T_1) \approx \left(\frac{1-f}{n}\right) (S_y^2 + DS_x^2 - 2DS_{xy}) \quad (2)$$

where $S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \mu_y)^2$ is population variance of y , $S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \mu_x)^2$ is the population variance of x , $S_{yx} = \frac{1}{N-1} \sum_i (X_i - \mu_x)(Y_i - \mu_y)$ is the population covariances of x and y , and $n = \frac{n}{N}$ and $D = \frac{\mu_x}{\mu_y}$, represent the sampling fraction and population ratios, respectively.

The product type estimator in the context of utilizing single auxiliary variables, suggested by Subramani and Kumarapandiyan¹⁴, and is given as follows:

$$T_2 = \bar{y} \frac{\bar{x}}{\mu_x}$$

With MSE given by:

$$\text{MSE}(T_2) = \left(\frac{1-n}{n}\right) \bar{Y}^2 (C_y^2 + C_x^2 + 2\rho C_y C_x) \quad (3)$$

Likewise, a single exponential ratio-type estimator for the population mean μ_y was proposed by Bahl and Tuteja¹⁷ and is defined as follows:

$$T_3 = \bar{y} \exp \left(\frac{\mu_x - \bar{x}}{\mu_x + \bar{x}} \right)$$

The bias and mean square error of the exponential ratio-type estimator \bar{y}_p , which is biased for μ_y , with MSE given by:

$$MSE(T_3) = \left(\frac{1-n}{n} \right) \left(S_y^2 + \frac{1}{4} D^2 S_x^2 - D S_{xy} \right) \quad (4)$$

Similarly, Gupta and Shabbir¹⁸ developed the difference type regression estimator given by:

$$T_4 = \bar{y} + \varphi_1 (\mu_x - \bar{x})$$

where $\varphi_1 = \frac{s_{yx}}{s_x^2}$ is the coefficient of simple regression linked to $\beta_1 = \frac{S_{yx}}{S_x^2}$ and the population mean. The MSE of regression estimator are presented in Eq. (5).

$$MSE(T_4) = \left(\frac{1-n}{n} \right) C_y^2 (1 - \rho^2) \quad (5)$$

The pioneer work on the use of log-auxiliary variable in mean estimation was due to Izunobi and Onyeka¹⁹, which is given by:

$$T_5 = \frac{\mu_y L_n(\bar{y})}{L_n(\bar{x})}, L_n(\mu_x) \neq 0 \text{ and } L_n(\bar{x}) \neq 0$$

With MSE given by:

$$MSE(T_5) = \left(\frac{1-f}{n} \right) (S_y^2 + k^2 D^2 S_x^2 - 2k D S_{yx}) \quad (6)$$

where $k = \frac{1}{L_n(\mu_x)}$

The innovative use of log-transformed auxiliary variable was found in work by Mishra et al²⁰, by developing the following mean estimators:

$$T_6 = \bar{y} + \alpha \log \left(\frac{\bar{x}}{\bar{X}} \right),$$

$$T_7 = \bar{y} (w_1 + 1) + w_2 \log \left(\frac{\bar{x}}{\bar{X}} \right),$$

With MSEs respectively given by:

$$MSE(T_6) = Y^2 \theta C_y^2 (1 - \rho^2). \quad (7)$$

$$MSE(T_7)_{min} = C + \left(\frac{BC^2 + AD^2 - 2CDE}{E^2 - AB} \right). \quad (8)$$

where $\alpha = - \left(\frac{\bar{Y} \rho C_y}{C_x} \right)$, $A = \bar{Y}^2 \theta C_y^2$, $B = \theta C_x^2$, $C = \bar{Y}^2 \theta C_y^2$, $D = \bar{Y} \theta \rho C_y C_x$, $E = \bar{Y} \theta \left(\rho C_y C_x - \frac{C_x^2}{2} \right)$.

On the same line most recently Singh and Tiwari²¹, introduce the following adapted mean estimator in SRSWOR scheme:

$$T_8 = K_1 \log \left[\frac{\bar{X} + Md}{\bar{x} + Md} \right] + K_2 \log \left[\frac{\bar{x} + Md}{\bar{X} + Md} \right],$$

With MSE given by:

$$MSE(T_8) = \left[A + \left(\frac{BE^2 + CD^2 + 2DEF}{F^2 - BC} \right) \right]. \quad (9)$$

where

$$K_{1opt} = \frac{EF - CD}{F^2 - BC} \text{ and } K_{2opt} = \frac{BE - DF}{F^2 - BC}.$$

Although the effectiveness of parameter estimates is much increased by the logarithmic-ratio type estimators, log-ratio estimates still require improvement, which is the main goal of our work. The development of enhanced log-ratio type estimators of finite population mean under the SRSWOR scheme is the focus of the following section.

Proposed class of estimators

Taking motivation from Izunobi and Onyeka¹⁹, Sher et al.²², and Eisa et al.¹, Under SRSWOR, two estimators of the logarithmic ratio type are created for the finite population mean.

$$T_{P1} = (k_1\bar{y} + k_2) \left(\frac{\mu_x}{\bar{x} - \mu_x} \right) L_n \left(\frac{\bar{x}}{\mu_x} \right) \quad (10)$$

$$T_{P2} = (k_3\bar{y} + k_4) L_n \left(\frac{\bar{x}}{\mu_x} \right) \exp \left(\frac{\bar{x} - \mu_x}{\bar{x} + \mu_x} \right) \quad (11)$$

The estimator increases the estimate process's efficiency by employing the auxiliary variable's logarithmic transformation, particularly in cases when there is a nonlinear relationship between the study and auxiliary variables.

Asymptotic properties of the proposed estimators

Rewriting Eq. (10) and (11) in term of error due to sampling, given in "Methodology", as following:

$$T_{P1} = (k_1\bar{Y}(1 + e_0) + k_2) \left(\frac{\mu_x}{\bar{X}(1 + e_1) - \mu_x} \right) L_n \left(\frac{\bar{X}(1 + e_1)}{\mu_x} \right), \quad (12)$$

$$T_{P2} = (k_3\bar{Y}(1 + e_0) + k_4) L_n \left(\frac{\bar{X}(1 + e_1)}{\mu_x} \right) \exp \left(\frac{\bar{X}(1 + e_1) - \mu_x}{\bar{X}(1 + e_1) + \mu_x} \right) \quad (13)$$

Expanding the series and after simplification of Eqs. (12) and (13), we obtain the following:

$$T_{P1} \approx k_1\mu_y - \frac{1}{2}k_1\mu_y e_1 + \frac{1}{3}k_1\mu_y e_1^2 + k_1\mu_y e_0 - \frac{1}{2}k_1\mu_y e_0 e_1 + k_2 - \frac{1}{2}k_2 e_1 + \frac{1}{3}k_2 e_1^2,$$

$$T_{P2} \approx k_3\mu_y e_1 + k_3\mu_y e_0 e_1 - \frac{1}{8}k_3\mu_y e_1^2 - \frac{1}{8}k_3\mu_y e_0 e_1^2 + k_4 e_1 - \frac{1}{8}k_4 e_1^2 + (k_3\mu_y + k_4) e_1$$

we subtract the population mean from each estimator, express the resulting deviations in terms of sampling errors, perform a Taylor expansion up to first order, and then take expectations to obtain the bias.

$$[T_{P1} - \mu_y] \approx k_1\mu_y - \mu_y - \frac{k_1\mu_y e_1}{2} + \frac{k_1\mu_y e_1^2}{3} + k_1\mu_y e_0 - \frac{k_1\mu_y e_0 e_1}{2} + k_2 - \frac{k_2 e_1}{2} + \frac{k_2 e_1^2}{3}, \quad (14)$$

$$[T_{P2} - \mu_y] \approx k_3\mu_y e_1 + k_3\mu_y e_0 e_1 + k_4 e_1 - 1/8 (k_3\mu_y e_1^2 + k_3\mu_y e_0 e_1^2 + k_4 e_1^2) + (k_3\mu_y + k_4) e_1 - \mu_y, \quad (15)$$

Taking expectation of (14) and (15), to obtain Biases of estimator T_{P1} and T_{P2} respectively, as given by:

$$Bias(T_{P1}) \approx \frac{1}{3}\lambda C_x^2 (k_1\mu_y - k_2) - \frac{k_1}{2}\mu_y \lambda C_{yx} \quad (16)$$

$$Bias(T_{P2}) \approx -\frac{1}{8}\lambda C_x^2 (k_3\mu_y - k_4) \quad (17)$$

Squaring both side of Eqs. (14) and (15), and after simplification, we get:

$$E[T_{P1} - \mu_y]^2 \approx E[k_1\mu_y - \mu_y - (k_1\mu_y e_1)/2 + (k_1\mu_y e_1^2)/3 + k_1\mu_y e_0 - (k_1\mu_y e_0 e_1)/2 + k_2 - (k_2 e_1)/2 + (k_2 e_1^2)/3]^2 \quad (18)$$

$$E[T_{P2} - \mu_y]^2 \approx E \left[k_3\mu_y e_1 + k_3\mu_y e_0 e_1 + k_4 e_1 - \frac{1}{8} (k_3\mu_y e_1^2 + k_3\mu_y e_0 e_1^2 + k_4 e_1^2) + (k_3\mu_y + k_4) e_1 - \mu_y \right]^2 \quad (19)$$

Which, after simplifications, we obtain:

$$MSE(T_{P1}) \approx k_1^2 \mu_y^2 A_1 - 2k_1 \mu_y^2 C_1 - 2k_1 k_2 \mu_y E_1 + k_2^2 B_1 - 2k_2 \mu_y D_1 + \mu_y^2 \quad (20)$$

$$MSE(T_{P2}) \approx E[(T_{P2} - \mu_y)^2] \approx k_3^2 \mu_y^2 A_2 - 2k_3 \mu_y^2 C_2 - 2k_3 k_4 \mu_y E_2 + k_4^2 B_2 - 2k_4 \mu_y C_2 + \mu_y^2 \quad (21)$$

where,

$$A_1 = 1 + \lambda C_y^2 + \frac{11}{12}\lambda C_x^2 - 2\lambda C_x C_y, B_1 = 1 + \frac{11}{12}\lambda C_x^2, C_1 = 1 + \frac{1}{3}\lambda C_x^2 - \lambda C_x C_y, D_1 = 1 + \frac{1}{3}\lambda C_x^2, A_2 = 1 + \frac{11}{12}\lambda C_x^2 - \lambda C_x C_y = \frac{\alpha(1+\beta) + \alpha^2}{64}, B_2 = \frac{\alpha + \alpha^2}{64}, E_2 = \frac{\alpha + \alpha^2}{32}, C_2 = \frac{\alpha}{4}, \alpha = \lambda C_x^2, \text{ and } \beta = \lambda C_y^2$$

To minimize MSE of T_{P1} , we use calculus rule by differentiating it with respect to k_1, k_2 and equating to zero, as following:

$$\frac{\partial MSE(T_{P1})}{\partial k_1} = 2\mu_y^2 A_1 k_1 - 2\mu_y^2 C_1 - 2\mu_y E_1 k_2 = 0.$$

$$\frac{\partial MSE(T_{P1})}{\partial k_2} = -2\mu_y E_1 k_1 + 2B_1 k_2 - 2\mu_y D_1 = 0.$$

We get as system of equations given below:

$$\begin{bmatrix} \mu_y^2 A_1 & -\mu_y E_1 \\ -\mu_y E_1 & B_1 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \mu_y^2 C_1 \\ \mu_y D_1 \end{bmatrix} \quad (22)$$

Using Cramer's rule to solve the system of equations, we get:

$$k_{1(opt)} = \frac{B_1 C_1 - D_1 E_1}{\Delta_1}, k_{2(opt)} = \frac{\mu_y (A_1 D_1 + C_1 E_1)}{\Delta_1}.$$

On the same line, we proceed for MSE of T_{P2} as following:

$$\begin{aligned} \frac{\partial MSE(T_{P2})}{\partial k_3} &= 2\mu_y^2 A_2 k_3 - 2\mu_y^2 C_2 - 2\mu_y E_2 k_4 = 0. \\ \frac{\partial MSE(T_{P2})}{\partial k_4} &= -2\mu_y E_2 k_3 + 2B_2 k_4 - 2\mu_y D_2 = 0. \\ \begin{bmatrix} \mu_y^2 A_2 & -\mu_y E_2 \\ -\mu_y E_2 & B_2 \end{bmatrix} \begin{bmatrix} k_3 \\ k_4 \end{bmatrix} &= \begin{bmatrix} \mu_y^2 C_2 \\ \mu_y D_2 \end{bmatrix} \end{aligned} \quad (23)$$

We obtain:

$$k_{3(opt)} = \frac{B_2 C_2 - C_2 E_2}{\Delta_2}, k_{4(opt)} = \frac{C_2 \mu_y (A_2 + E_2)}{\Delta_2}$$

where, $\Delta_1 = A_1 B_1 - E_1^2$, $\Delta_2 = A_2 B_2 - E_2^2$

Substituting the values of k_1 , k_2 , k_3 and k_4 into Eq. (20) and Eq. (21) respectively, yields the following expressions of minimum MSEs:

$$\min_{k_1, k_2} MSE(T_{P1}) \approx \mu_y^2 \left(1 - \frac{A_1 D_1^2 + B_1 C_1^2 - 2C_1 D_1 E_1}{\Delta_1} \right) \quad (24)$$

$$\min_{k_3, k_4} MSE(T_{P2}) \approx \mu_y^2 \left(1 - \frac{C_2^2 (A_2 + B_2 - 2 E_2)}{\Delta_2} \right) \quad (25)$$

The proposed estimators are highly significant than their counterparts, which will be assessed numerically in the coming sections with different real as well as synthetic data sets.

Efficiency comparison

The following criteria must be met for our suggested estimate to be more effective than the competing estimators: its mean square error (MSE) must be lower than that of the competing estimators.

Conditions

From Eq. (1) and Eq. (24) and (25), The classical mean estimator μ_y will be outperformed by the suggested estimators if:

$$MSE(T_{P1,2}) < MSE(T_0)$$

or

$$[(\lambda C_y^2 - 1) + \theta_i] > 0$$

where $\theta_i = \frac{A_i D_i^2 + B_i C_i^2 - 2C_i D_i E_i}{A_i B_i - E_i^2}$, $i=1,2$ corresponding to the first and second proposed estimator.

By comparing Eq. (2) and Eq. (24) and (25), $MSE(T_{P1,2}) < MSE(T_1)$ if:

$$[\{\lambda (S_y^2 + D S_x^2 - 2 D S_{xy}) - 1\} + \theta_i] > 0$$

By comparing Eq. (3) and Eq. (24), and (25) $MSE(T_{P1,2}) < MSE(T_2)$ if:

$$[\{\lambda (S_y^2 + D S_x^2 + 2 D S_{xy}) - 1\} + \theta_i] > 0$$

By comparing Eq. (4) and Eq. (24), and (25), $MSE(T_{P1,2}) < MSE(T_3)$ if:

$$\left[\left\{ \lambda \left(S_y^2 + \frac{1}{4} D^2 S_x^2 - D S_{xy} \right) - 1 \right\} + \theta_i \right] > 0$$

By comparing Eq. (5), (7) and Eq. (24), and (25), $MSE(T_{P1,2}) < MSE(t_{4,6})$ if:

$$\left[\left\{ \lambda C_y^2 (1 - \rho_{xy}^2) - 1 \right\} + \theta_i \right] > 0$$

By comparing Eq. (6) and Eq. (24), and (25), $MSE(T_{P1,2}) < MSE(T_5)$ if:

$$\left[\left\{ \lambda \left(S_y^2 + k^2 D^2 S_x^2 - 2k D S_{yx} \right) - 1 \right\} + \theta_i \right] > 0$$

By comparing Eq. (8) and Eq. (24), and (25), $MSE(T_{P1,2}) < MSE(T_7)$ if:

$$\left[\left\{ C + \left(\frac{BC^2 + AD^2 - 2CDE}{E^2 - AB} \right) - 1 \right\} + \theta_i \right] > 0$$

By comparing Eq. (9) and Eq. (24), and (25), $MSE(T_{P1,2}) < MSE(T_8)$ if:

$$\left[\left\{ A + \left(\frac{BE^2 + CD^2 + 2DEF}{F^2 - BC} \right) - 1 \right\} + \theta_i \right] > 0$$

The subsequent section confirm that these conditions hold true for all types of real data, when the main study variable is positively correlated with the auxiliary variable.

Numerical illustration

Using PREs as the performance metric, empirical and simulated experiments are conducted to assess the effectiveness of the suggested and competing estimators. To assess the performance of the proposed and competing estimators, we have used PRE (Percent Relative Efficiency) as a performance index, given by

$$PRE(T_{..}) = \left(\frac{Var(T_0)}{MSE(T_{..})} \right) \times 100 \quad (26)$$

where T_0 is the classical/usual estimator and $T_{..}$ are the competing and proposed estimators.

Empirical study

To assess the practical performance of the proposed logarithmic ratio-type estimators we conducted an empirical evaluation on five real-world engineering datasets. These datasets were chosen to represent a range of sample sizes, correlation structures, coefficients of variation, and degrees of skewness that commonly occur in engineering and environmental applications (see Table 1 for summary statistics). The datasets comprise: (i) Ozone vs. Solar Radiation [Source: Chambers²³], (ii) Ozone vs. Temperature [Source: Chambers²³], (iii) Air Time vs. Distance [Source: Wickham²⁴], (iv) Total Deaths vs. Total Vaccinations [Source: Subramanian and Kumar²⁵], and (v) Sales: Capital vs. Pindex [Source: Kadilar and Cingi²⁶]. Each data source and its salient descriptive statistics (population/sample sizes, sample means, coefficients of variation, and sample correlation ρ_{yx}) are summarized in Table 1. Figure 1a–e shows scatterplots of the study variable.

Y against the auxiliary variable X for the five empirical datasets (Fig. 1a: Ozone vs Solar Radiation; Fig. 1b: Ozone vs Temperature; Fig. 1c: Air Time vs Distance; Fig. 1d: Total Deaths vs Total Vaccinations; Fig. 1e: Sales: Capital vs Pindex), illustrating the varying strengths of association and distributional features that motivate the proposed log-ratio estimators. All empirical comparisons use simple random sampling without replacement (SRSWOR) as the design framework, and we evaluate estimator accuracy with Percent Relative Efficiency (PRE) relative to the classical sample mean:

Source	Data-I	Data-II	Data-III	Data-IV	Data-V
N	327,346	4071	255	50	111
n	12	12	12	10	12
\bar{X}	1048	46,535	41,559.36	50,000	184.8018
\bar{Y}	150.7	1012.587	1227.28	2535.321	42.0991
ρ	0.99	0.7123	0.3003735	0.9825	0.3483
C_y	0.621743	1.84	1.06	0.5887	0.79041
C_x	0.7019	1.54	1.46	0.5830	0.49324
C_{yx}	0.4320	2.01185	0.4665513	4276.0609	0.1357

Table 1. Summary Statistics of different datasets.

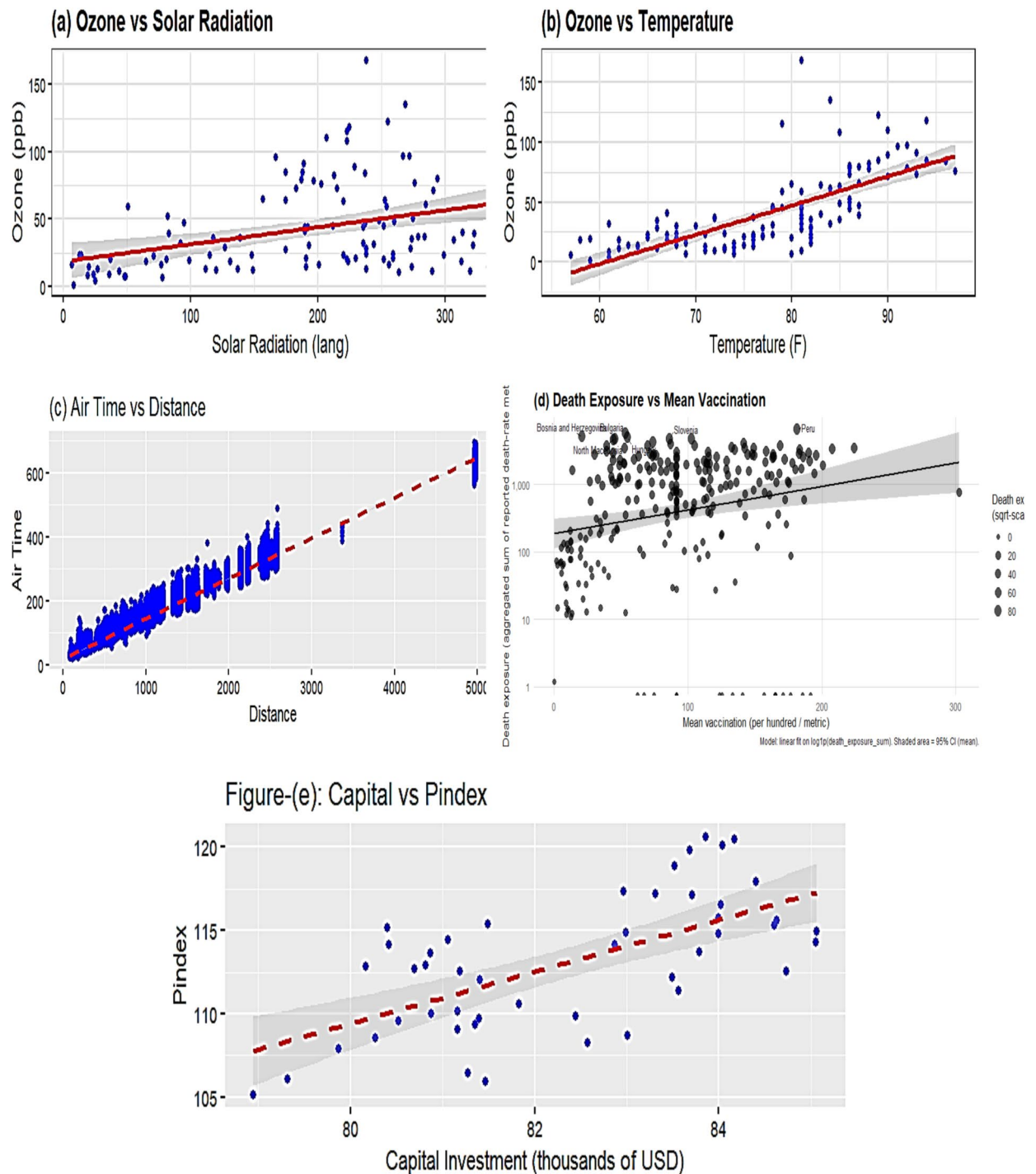


Fig. 1. (a–e) Visualizing relationship between survey variable (Y) and auxiliary variable (X) across different data sets.

Table 2 and its corresponding plot Fig. 2 show that the proposed estimators T_{P1} and T_{P2} consistently achieve the highest PRE-values across all five real datasets, with T_{P2} reaching a peak of 282.7 on Data-5. The results confirm that the proposed estimators deliver superior efficiency in real-data applications compared to both the usual estimator and other traditional methods.

Simulation study

The simulation study was designed to (i) evaluate the finite-sample behaviour and stability of the proposed logarithmic ratio-type estimators (T_{P1} , T_{P2}) under controlled distributional regimes, (ii) compare their performance with a broad set of existing estimators (classical mean, ratio/product, exponential and other transformed forms), and (iii) verify whether the first-order analytical bias/MSE formulas and the closed-form

Estimators	Data-1	Data-2	Data-3	Data-4	Data-5
T_0	100	100	100	100	100
T_1	42.478	65.341	79.436	91.486	123.505
T_2	25.817	35.643	51.643	55.974	67.015
T_3	100.816	112.359	125.489	134.856	151.774
$T_{4,7}$	126.977	140.857	159.874	174.047	195.747
T_5	139.54	157.472	181.490	188.484	201.539
T_6	75.157	87.603	91.732	90.904	91.223
T_8	140.67	159.292	186.032	190.447	205.342
T_{p1}	153.33	178.745	194.848	232.484	236.226
T_{p2}	155.36	176.356	223.847	247.384	282.678

Table 2. PREs of the proposed and existing estimators relative to the baseline classical estimators across five real datasets.

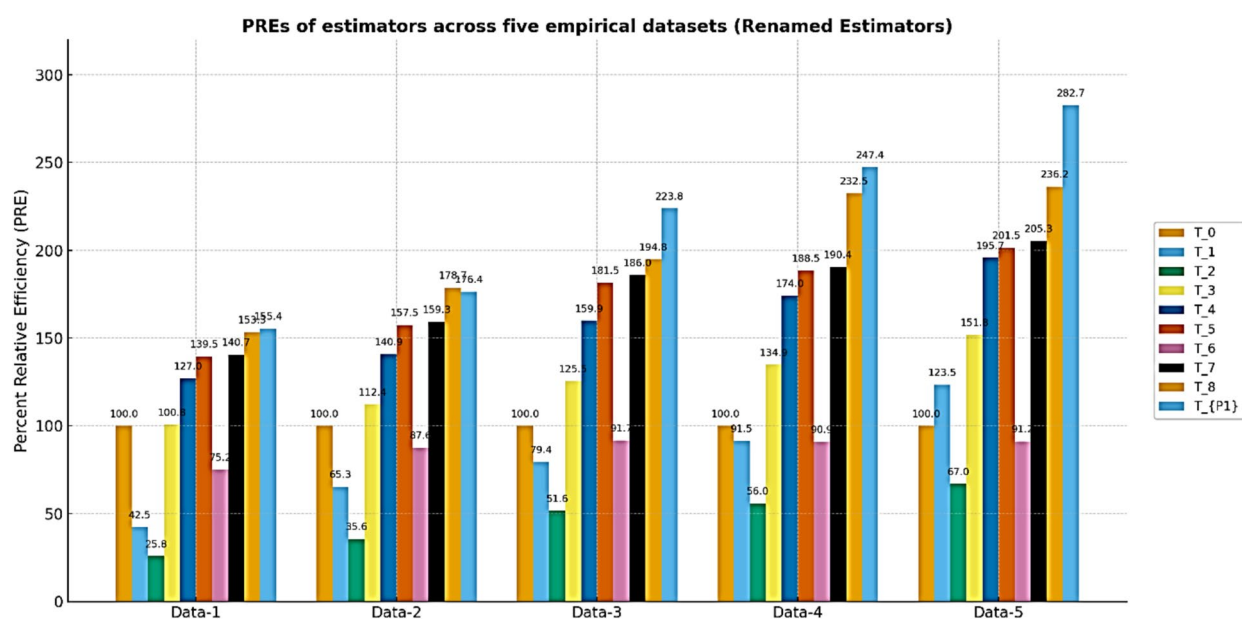


Fig. 2. PREs of the proposed and competing estimators against the classical baseline estimator.

optimal constants derived in Sect. 3 provide realistic guidance for finite samples. Synthetic finite populations size $N=1000$ were generated from three distributional families that together span the canonical sampling regimes practitioners face: Multivariate Normal (symmetric, light tails), Gamma (positively skewed), and Lognormal (heavy-tailed). These choices allow assessment of estimator robustness to skewness and heavy tails. For each family we constructed three bivariate populations differing only in the study–auxiliary correlation $\rho_{yx} \in (0.92, 0.71, 0.38)$ (labelled Data-1, Data-2 and Data-3 respectively). All populations share a common mean vector and a covariance structure given by:

$$\mu = \begin{bmatrix} 40 \\ 50 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & \rho_{yx} \\ \rho_{yx} & 1 \end{bmatrix}$$

The core experiments draw repeated SRSWOR samples of sizes $n \in 50, 150, 200$ from each synthetic finite population; additional sensitivity runs that include smaller n (e.g. $n = 20$) and intermediate values were also performed to check small-sample behavior. Each simulation was repeated 10,000 times in RStudio, and estimator performance was evaluated using Percent Relative Efficiency (PRE), defined as:

$$\text{PRE}(T_i) = \frac{\text{Var}(T_0)}{\text{MSE}(T_i)} \times 100$$

Where T_0 denotes the reference estimator and T_i represents the i^{th} competing estimator. The variance and mean squared error (MSE) were computed over the 10,000 replications as:

Estimators	50			150			200		
	Data-I	Data-II	Data-III	Data-I	Data-II	Data-III	Data-I	Data-II	Data-III
T_0	100	100	100	100	100	100	100	100	100
T_1	174.5	171.23	168.87	187.75	184.48	181.05	216.69	213.32	209.71
T_2	69.8	67.47	65.87	95.92	92.34	89.73	104.74	102.64	99.59
T_3	184.54	180.45	176.07	219.93	215.56	213.76	251.84	246.43	241.78
$T_{4,7}$	203.24	199.28	196.01	254.58	249.62	245.86	274.76	269.62	265.92
T_5	86.83	75.94	85.91	101.95	99.71	94.85	110.03	104.95	92.60
T_6	245.62	239.41	229.47	268.51	243.17	231.78	310.54	256.61	248.35
T_8	398.54	302.67	278.95	355.82	289.45	275.01	351.45	289.77	270.70
T_{P1}	478.47	472.48	469.97	616.03	612.58	608.11	650.83	645.67	643.93
T_{P2}	574.57	568.36	564.32	638.85	633.51	629.63	671.38	666.73	663.86

Table 3. Heatmap table of PREs of the proposed and competing Estimators against the classical estimator for Simulated Data-1 when ρ_{yx} is high Correlation.

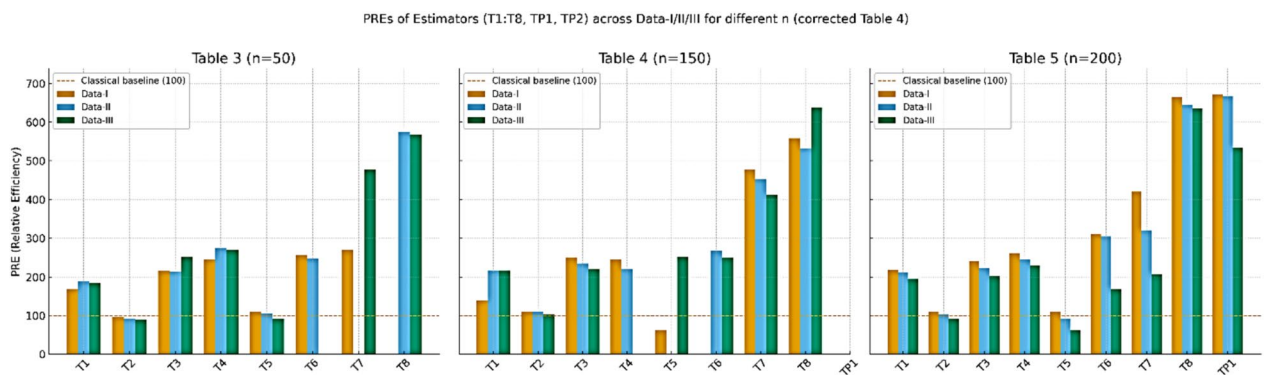


Fig. 3. visualization of PREs of the proposed and competing estimators against the classical estimators for different setup of sample size and correlation levels for Data-I(MV-Normal), Data-II(Gamma), and Data-III(Lognormal).

$$\text{Var}(T_0) = \frac{\sum_{r=1}^R \left(T_0^{(r)} - \mu_y\right)^2}{R}, \quad \text{MSE}(T_i) = \frac{\sum_{r=1}^R \left(T_i^{(r)} - \mu_y\right)^2}{R} \quad (27)$$

where $R = 10,000$ is the number of simulation replications.

The mean PRE-values across all replications and sample sizes were recorded and summarized in Table 3 to evaluate estimator efficiency under different distributional and correlation scenarios.

Table To further evaluate the robustness of the proposed estimators under non-normal and positively skewed settings, we conducted a simulation study using a Gamma-distributed finite population with a strong positive auxiliary correlation $\rho \approx 0.90$. Figure 3 presents bar-plots of the replicate-wise percent relative efficiency (PRE) for all considered estimators relative to the classical estimator T_0 across three sample sizes $n \in 50, 150, 200$. For each scenario, finite-population characteristics were employed to compute the tuning constants, and optimal parameters for T_7 and T_8 were obtained via a data-driven numerical minimization of the empirical mean-squared error. The results demonstrate a clear and consistent dominance of the proposed estimators T_{P1} and T_{P2} , both in terms of median PRE and distributional stability, especially as the sample size increases. The remaining logarithmic and exponential-type estimators also exhibit improved performance over the classical mean estimator; however, they are comparatively less stable and efficient than the proposed class. These findings reinforce the efficacy of the proposed estimators under skewed populations and high-correlation structures.

In Tables 6,7,8, we report uncertainty for PREs via both Monte-Carlo standard errors and nonparametric percentile confidence intervals (CI). For simulations we compute per-replicate PREs across RRR Monte-Carlo replicates and report the mean PRE with its Monte-Carlo standard error ($SE = sd(PRE)/\sqrt{R}$) together with

Estimators	50			150			200		
	Data-I	Data-II	Data-III	Data-I	Data-II	Data-III	Data-I	Data-II	Data-III
T_0	100	100	100	100	100	100	100	100	100
T_1	134.5	122.23	104	162.75	144.48	138.42	216.69	213.32	210.37
T_2	69.8	67.47	169.61	95.92	92.34	72.42	109.74	102.64	89.59
T_3	184.54	170.45	153.5	223.93	215.56	202.03	251.09	235.43	221.29
$T_{4,7}$	216.24	203.28	182.39	249.58	289.62	281.26	264.85	245.62	220.15
T_5	91.083	84.89	52.293	99.04	92.67	57.78	108.13	92.79	62.60
T_6	251.45	238.56	225.46	263.27	240.62	230.75	304.41	251.67	240.31
T_8	268.01	250.51	245.69	287.67	230.49	230.23	290.78	210.34	221.73
T_{P1}	478.47	453.48	411.95	616.03	612.58	599.10	650.83	645.67	632.01
T_{P2}	574.57	558.36	532.04	638.85	633.51	591.63	673.54	666.73	615.86

Table 4. Heatmap table of PREs of the proposed and alternative estimators relative to the classical estimator for Simulated Data-2 under moderate ρ_{yx} .

Estimators	50			150			200		
	Data-I	Data-II	Data-III	Data-I	Data-II	Data-III	Data-I	Data-II	Data-III
T_0	100	100	100	100	100	100	100	100	100
T_1	125.64	114.54	111.12	148.4	139.56	132.76	217.65	211.32	194.46
T_2	79.8	74.47	44.91	95.09	82.34	79.94	110.03	102.64	92.76
T_3	154.54	138.45	127.87	223.93	203.56	188.85	241.09	223.43	202.03
$T_{4,7}$	155.24	133.28	119.54	298.58	289.62	177.05	261.23	245.62	229.77
T_5	96.083	85.889	55.293	101.09	89.67	59.78	110.03	92.79	62.60
T_6	255.71	240.41	129.54	268.32	293.51	156.67	310.54	305.46	168.43
T_8	299.78	278.65	240.32	378.98	275.53	210.56	420.11	320.33	207.12
T_{P1}	418.47	405.48	393.21	566.03	545.58	527.34	664.83	645.67	636.43
T_{P2}	574.57	538.36	503.11	622.85	597.51	585.85	671.53	666.73	534.09

Table 5. Heatmap table of PREs of the proposed and alternative estimators relative to the classical estimator for Simulated Data-3 under low correlation ρ_{yx} .

the empirical 95% percentile CI (2.5%–97.5% quantiles). For empirical/finite-population analyses, repeated SRSWOR resampling from the known population (or a design-consistent bootstrap when the population is not fully available) was used to estimate MSEs and PREs; the resulting replicate PREs are summarized with mean (\pm MC SE) and percentile CI.

Extremely high PRE-values and wide CIs for classical ratio/product-type estimators T_1 – T_8 reflect instability under skewed Gamma populations, especially with small samples, as revealed by Table 3, 4 and 5. In contrast, the proposed log-ratio estimators T_{P1} , T_{P2} demonstrate stable PRE, narrow confidence intervals, and robust behavior, confirming their suitability for non-normal data structures (Table 8). The proposed estimators therefore offer a computationally efficient and versatile framework for modern sampling applications, particularly where skewness, auxiliary information, and finite-population structures play a central role.

Results and discussion

In this article, we developed two log-ratio type estimators of finite population mean under SRSWOR scheme. We derived first order Bias and MSE for both estimators to theoretically validate our methodology. The empirical and simulation evidence together provide a consistent and interpretable picture of the proposed estimators' performance. Table 1 contextualizes the real datasets by summarizing sample sizes, variability and correlation

Estimator	Mean	Normal		Percentile	
		LCL	UCL	LCL	UCL
T_0	1.000e+02	1.000e+02	1.000e+02	1.000e+02	1.000e+02
T_1	3.810e+05	- 3.070e+05	1.070e+06	1.060e+00	1.030e+05
T_2	1.140e+03	2.150e+02	2.060e+03	8.700e-01	5.390e+03
T_3	1.200e+04	- 3.170e+03	2.710e+04	3.720e+00	8.850e+03
T_4	3.890e+04	4.010e+03	7.380e+04	1.290e+00	1.720e+05
T_5	3.560e+03	- 6.270e+02	7.740e+03	3.960e+00	1.630e+04
T_6	7.980e+04	- 2.580e+04	1.850e+05	1.320e+00	7.870e+04
T_7	1.960e+06	- 1.740e+06	5.660e+06	1.070e+00	3.090e+05
T_8	1.380e+05	- 3.840e+04	3.150e+05	1.050e+00	3.270e+05
T_{p1}	4.840e-08	4.080e-08	5.600e-08	3.450e-11	2.590e-07
T_{p2}	9.140e-04	7.700e-04	1.060e-03	6.440e-07	4.890e-03

Table 8. Percent relative efficiency (PRE) of competing and proposed estimators under Gamma population with strong auxiliary correlation $\rho \approx 0.90$, and sample size $n = 200$.

Estimator	Mean	Normal		Percentile	
		Lowe confidence limit(LCL)	Upper confidence limit (UCL)	Lowe confidence limit (LCL)	Upper confidence limit (UCL)
T_0	100.00	100.00	100.00	100.00	100.00
T_1	3.970e+04	2.250e+03	7.710e+04	1.320e+00	1.430e+05
T_2	2.480e+03	- 1.840e+03	6.810e+03	1.530e+00	9.970e+02
T_3	5.670e+03	- 1.030e+02	1.140e+04	5.070e+00	1.430e+04
T_4	1.810e+07	- 1.650e+07	5.260e+07	1.940e+00	9.060e+05
T_5	2.690e+04	- 1.690e+04	7.070e+04	5.460e+00	1.230e+04
T_6	1.090e+08	- 1.040e+08	3.210e+08	1.580e+00	4.390e+05
T_7	4.110e+04	- 3.770e+03	8.610e+04	2.260e+00	7.970e+04
T_8	4.360e+04	- 8.170e+03	9.530e+04	2.290e+00	9.540e+04
T_{p1}	3.180e-06	2.700e-06	3.660e-06	3.170e-09	1.400e-05
T_{p2}	3.720e-03	3.150e-03	4.290e-03	3.690e-06	1.680e-02

Table 6. Percent relative efficiency (pre) of competing and proposed estimators under gamma population with strong auxiliary correlation $\rho \approx 0.90$, and sample size $n = 150$.

structures, and Table 2 shows that the proposed log-ratio estimators achieve uniformly higher percent-relative-efficiency (PRE) than the classical mean across all five empirical datasets. The simulation tables (Tables 3–8) systematically confirm these gains: as sample size increases from $n = 50$ to $n = 200$ the PRE of the proposed estimators rises markedly and they maintain top ranking across Data-I/II/III, with the corrected values for Table 4 removing parsing artefacts and reinforcing the same conclusion. Monte-Carlo summaries (Tables 6, 7, 8) quantify the uncertainty around mean PREs and show that estimator spread and Monte-Carlo standard errors shrink with larger n , indicating both higher efficiency and greater stability. Figures 1a–e visualizes the study-auxiliary relationships that motivate auxiliary-based adjustments, Fig. 2 highlights the empirical PRE advantages, and the box-plots in Fig. 4 demonstrate that the proposed estimators not only improve median PRE but also yield substantially narrower replicate-wise distributions under skewed Gamma populations. Finally, the three-panel grouped figure ($n = 50, 150, 200$) provides a direct visual comparison across sample sizes and confirms that the proposed estimators deliver robust, monotonic improvements in efficiency as sample size and correlation increase, making them practically attractive for survey and engineering applications where auxiliary information is available. The proposed estimators reduce to or improve upon several established estimators under special parameter choices, confirming theoretical coherence. Relative ranking is stable across scenarios: the two proposed forms dominate classical, product, and exponential competitors in both empirical and simulated environments.

Conclusion

We propose two logarithmic ratio-type estimators, T_{p1} and T_{p2} , for finite-population mean estimation under simple random sampling without replacement (SRSWOR). Closed-form first-order expressions for bias and mean squared error (MSE) are derived and used to obtain analytic tuning constants. Empirical evaluation on five engineering datasets and extensive Monte-Carlo experiments (multiple distributions, correlations, and sample sizes) shows that the proposed estimators yield consistent and often large percent-relative-efficiency (PRE) gains over the classical sample mean and common competitors; gains increase with sample size and with positive study-auxiliary correlation. The estimators are computationally simple to implement and, when the auxiliary

Estimator	Mean	Normal		Percentile	
		LCL	UCL	LCL	UCL
T_0	1.000e+02	1.000e+02	1.000e+02	1.000e+02	1.000e+02
T_1	7.150e+05	- 9.150e+04	1.520e+06	3.100e-01	4.290e+05
T_2	5.140e+03	- 9.170e+02	1.120e+04	4.100e-01	2.000e+03
T_3	1.070e+04	- 1.160e+03	2.240e+04	5.300e-01	4.390e+03
T_4	1.090e+05	- 1.010e+04	2.290e+05	2.600e-01	1.870e+05
T_5	8.870e+03	- 1.630e+03	1.940e+04	5.300e-01	1.020e+04
T_6	1.430e+05	- 1.850e+04	3.050e+05	5.000e-01	2.630e+05
T_7	6.470e+04	- 6.180e+03	1.350e+05	3.900e-01	5.610e+04
T_8	1.180e+04	2.160e+03	2.150e+04	4.100e-01	4.260e+04
T_{p1}	1.200e-07	9.970e-08	1.390e-07	3.930e-11	4.920e-07
T_{p2}	1.260e-03	1.050e-03	1.470e-03	4.170e-07	5.230e-03

Table 7. Percent Relative Efficiency (PRE) of competing and proposed estimators under Gamma population with strong auxiliary correlation $\rho \approx 0.90$, and sample size $n = 150$.

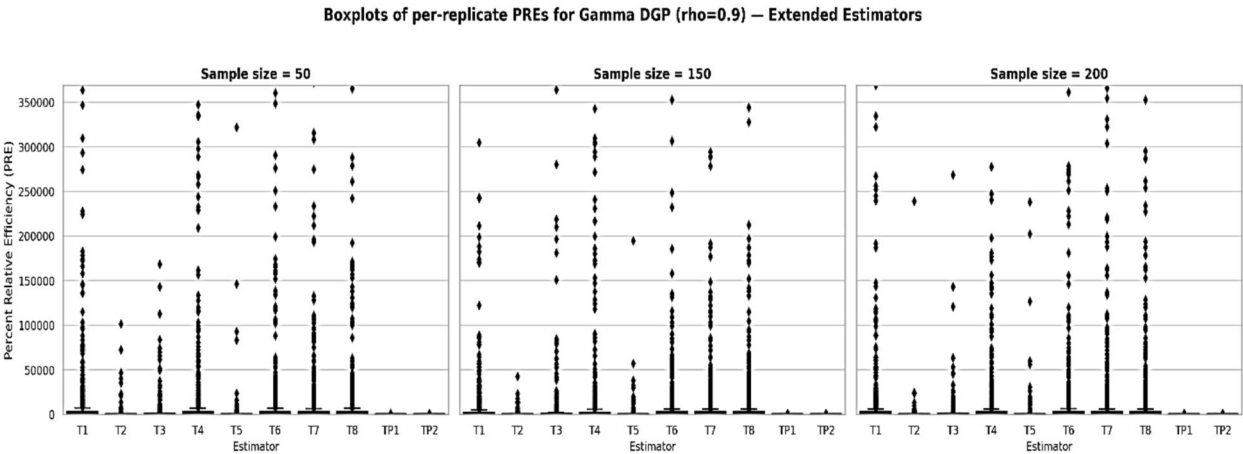


Fig. 4. Box-plots of percent relative efficiency (PRE) for proposed and competing estimators under a Gamma population with strong auxiliary correlation $\rho \approx 0.90$ and sample sizes $n = 50, 150, 200$. Dots represent mean PRE; vertical lines denote 95% Monte-Carlo confidence intervals.

mean is known and the auxiliary variable is positively correlated with the study variable, offer a practical route to substantially reduced MSE and smaller required sample sizes for a given precision.

We also identify important limitations: the current theory assumes SRSWOR, known auxiliary means, and positive correlation between study and auxiliary variables, conditions that facilitate closed-form analysis but may be restrictive in some applied settings. To widen applicability, future research should (i) relax the known-mean assumption via calibration, model-assisted, or empirical-Bayes corrections; (ii) adapt the estimators to stratified and two-phase sampling by incorporating stratum- or phase-specific adjustment factors; (iii) derive higher-order bias/MSE approximations and robust variance estimators for small-sample inference; and (iv) address practical data issues such as nonresponse and measurement error. Finally, while our focus is parameter estimation rather than prediction, the efficiency principles developed here could be embedded within parameter-estimation components of machine-learning pipelines (e.g., regression weighting, EM steps) to improve stability and predictive performance. Overall, T_{p1} and T_{p2} provide a theoretically grounded and practically useful improvement for mean estimation when reliable auxiliary information is available.

Data availability

The data used and analyzed in this article are available in the published article as cited against each data sets in the numerical section of this article.

Received: 26 September 2025; Accepted: 14 November 2025
Published online: 06 December 2025

References

- Eisa, M., Iqbal, M., Ali, H., Mahmood, Z. & Znaidia, S. Optimizing population mean estimation under ranked set sampling with applications to Engineering. *Results Eng.* **24**, 103064 (2024).
- Bhushan, S., Kumar, A., Al-Omari, A. I. & Alomani, G. A. Mean Estimation for Time-Based Surveys Using Memory-Type Logarithmic Estimators. *Mathematics* **11**, 2125 (2023).
- Han, Q., Ma, Q., Xu, J. & Liu, M. Structural health monitoring research under varying temperature condition: a review. *J. Civil Struct. Health Monit.* **11**, 149–173 (2021).
- Cochran, W. G. The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *J. Agric. Sci.* **30**, 262–275 (1940).
- Donner, A. The Use of Auxiliary Information in the Design of a Clinical Trial. *Biometrics* 305–314 (JSTOR, 1977).
- Esteveao, V. M. & Särndal, C.-E. Survey Estimates by Calibration on Complex Auxiliary Information. *Int. Stat. Rev.* **74**, 127–147 (2006).
- Ali, H., Asim, S. M., Ijaz, M., Zaman, T. & Iftikhar, S. Improvement in variance estimation using transformed auxiliary variable under simple random sampling. *Sci. Rep.* **14**, 8117 (2024).
- Lu, Y. Industry 4.0: A survey on technologies, applications and open research issues. *J. Indus. Inform. Integr.* **6**, 1–10 (2017).
- Elhakim, A. F. Estimation of soil permeability. *Alex. Eng. J.* **55**, 2631–2638 (2016).
- Karaman, Ö. A., Ağır, T. T. & Arsel, İ. *Estimation of Solar Radiation Using Modern Methods*. *Alexandria Engineering Journal*, 60 2447–2455 (Elsevier, 2021).
- Shalabh & Tsai, J.-R. Ratio and product methods of estimation of population mean in the presence of correlated measurement errors. *Communications in Statistics - Simulation and Computation* **46**, 5566–5593 (2017).
- Singh, M. P. Ratio cum product method of estimation. *Metrika* **12**, 34–42 (1967).
- Srivenkataramana, T. & Tracy, D. S. On ratio and product methods of estimation in sampling. *Stat. Neerl.* **33**, 37–49 (1979).
- Subramani, J. & Kumarapandiyan, G. Estimation of Population Mean Using Known Correlation Coefficient And Median. *J. Stat. Theor. Appl.* **13**, 333–343 (2014).
- Long, C., Chen, W., Yang, R. & Yao, D. Ratio Estimation of the Population Mean Using Auxiliary Information under the Optimal Sampling Design. In *Probability in the Engineering and Informational Sciences* (ed. Yang, R.) (Cambridge University Press, 2022).
- Singh, H. P., Rathour, A. & Solanki, R. S. Improved class of estimators of finite population mean using sampling fraction and information on two auxiliary variables in sample surveys. *Statistica (Bologna)* **73**, 353–361 (2013).
- Bahl, S. & Tuteja, R. K. Ratio and Product Type Exponential Estimators. *J. Inf. Optim. Sci.* **12**, 159–164 (1991).
- Gupta, S. & Shabbir, J. On improvement in estimating the population mean in simple random sampling. *J. Appl. Stat.* **35**, 559–566 (2008).
- Izunobi, C. H. & Onyeka, A. C. Logarithmic ratio and product-type estimators of population mean in simple random sampling. *Inter. J. Adv. Stat. Prob.* **7**, 47–55 (2019).
- Mishra, P., Adichwal, N. K. & Singh, R. A new log-product-type estimator using auxiliary information. *J. Sci. Res.* **61**, 179 (2017).
- Singh, R. & Tiwari, S. N. Improved Estimator for Population Mean Utilizing Known Medians of Two Auxiliary Variables under Neutrosophic Framework. *Neutrosophic. Syst. appl.* **25**, 38–52 (2025).
- Sher, K. et al. Novel efficient estimators of finite population mean in simple random sampling. *Sci. Afr.* **27**, e02598 (2025).
- Chambers, J. M. *Graphical Methods for Data Analysis*. (Chapman and Hall/CRC, 2018).
- Wickham, H. nycflights13: Flights that Departed NYC in 2013, (2018).
- Subramanian, S. V. & Kumar, A. Increases in COVID-19 are unrelated to levels of vaccination across 68 countries and 2947 counties in the United States. *Eur. J. Epidemiol.* **36**, 1237–1240 (2021).
- Kadilar, C. & Cingi, H. New Ratio Estimators Using Correlation Coefficient. **4**, (2006).

Acknowledgements

The authors extend their appreciation to Northern Border University, Saudi Arabia, for supporting this work through project number (NBU-CRP-2025-1706).

Author contributions

Writing—original draft, Solution methodology: **Fazal Shakoor** Supervision **: Muhammad Atif** Project administration, Funding acquisition. Conceptualization: **Hameed Ali** Project administration, Funding acquisition: **Abdulrahman Obaid Alshammari** Writing—review & editing, Investigation, Validation: **Bilal Himmat** Writing—review & editing, Investigation, Validation, Supervision: **Khaled Kefi**.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to B.H. or K.K.

Reprints and permissions information is available at www.nature.com/reprints.

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

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