

Log-transformed approaches to variance estimation using auxiliary data

Poonam Singh ^{a,b}, Prayas Sharma ^{c,*}, Anjali Singh ^a, Tolga Zaman ^d, Walid Emam ^e

^a Department of Statistics, Banaras Hindu University, Varanasi, India

^b School of Mathematical and Physical Sciences, University of Technology, Sydney, Australia

^c Department of Statistics, BabaSaheb Bhimrao Ambedkar University, Lucknow, India

^d Department of Mathematics Engineering, Graduate Education Institute, Gumushane University, Gumushane, Turkey

^e Department of Statistics and Operations Research, Faculty of Science, King Saud University, Riyadh, Saudi Arabia

ARTICLE INFO

Keywords:

Bias
Variance estimation
Mean square error (MSE)
Logarithmic estimators
Simulation

ABSTRACT

Reliable statistical inference requires accurate population variance estimate, especially in domains where measurement limitations and intrinsic variability impact data. Applying traditional estimators to skewed or heavy-tailed populations frequently results in inefficient results. We provide a novel class of generalized logarithmic variance estimators that use logarithmic transformations and auxiliary data to stabilize variance and improve estimator performance in order to overcome this constraint. We calculate the suggested estimators' bias and Mean Squared Error (MSE) expressions and assess their effectiveness using comprehensive Monte Carlo simulations with different sample sizes. The suggested estimator, T_{p1-} , produces a significant reduction in MSE up to 57% increase in efficiency (PRE) at $n=600$ when compared to the usual estimator. The suggested methodologies resilience and suitability for high-variability situations are further confirmed using real-world datasets. In every context, the findings show that the suggested estimators perform better than the current ones in terms of lower MSE and greater PRE. This research demonstrates how logarithmic transformations may be used to create variance estimators that are more accurate and effective, especially when auxiliary variables are provided.

1. Introduction

Sampling is essential to research and monitoring efforts because it allows researchers to evaluate conditions, distributions, and possible hazards without resorting to detailed and sometimes unfeasible full-scale surveys. Sampling offers a productive substitute for measuring conditions over wide regions or populations when measures are expensive, time-consuming, or involve possible risks. By choosing representative samples, researchers may minimize resource usage and staff risk while inferring more general traits and patterns. In order to guarantee that the data gathered appropriately reflects the underlying distribution of the variable of interest across various contexts or groups, effective sampling techniques such as stratified and systematic sampling are essential (Murthy, 1967).

An auxiliary variable is a secondary variable that is connected with the main variable of interest and is used to improve estimate. An auxiliary variable is called an additional variable. Auxiliary data in this study is essential for increasing the effectiveness of the suggested variance estimators. Based on their empirical correlation with the study variable and theoretical relevance, the auxiliary variables used

in the real-world datasets were carefully chosen. For example, auxiliary variables like enrollment size, average test scores, age, or income level were used in applications pertaining to health and education surveys. These variables function as trustworthy stand-ins that have a good correlation with the main variable of interest and are usually available through administrative records or previous surveys. Their incorporation into the estimation framework helps to increase precision and reduce variance (Murthy, 1967).

When direct measurement is expensive, challenging, or unfeasible, auxiliary variables are frequently utilized in a variety of industries to increase estimation efficiency. For instance, in agriculture, considerable field sampling is not necessary for estimating crop yields across wide areas using satellite imaging or vegetation indices. Without testing whole populations, demographic information like age or income levels aids in the refinement of illness prevalence estimates in public health. Similar to how purchase histories are used in market research to predict consumer behavior, student attendance records are utilized in education to estimate changes in academic performance. These auxiliary variables improve the precision and economy of data gathering and processing in a variety of fields (Singh, 2003).

* Corresponding author.

E-mail addresses: poonamsingh@bhu.ac.in (P. Singh), prayassharma02@gmail.com (P. Sharma), anjali.singh24@bhu.ac.in (A. Singh), zamantolga@gmail.com (T. Zaman).

<https://doi.org/10.1016/j.jrras.2025.101913>

Received 16 June 2025; Received in revised form 12 August 2025; Accepted 25 August 2025

Available online 6 September 2025

1687-8507/© 2025 The Authors. Published by Elsevier B.V. on behalf of The Egyptian Society of Radiation Sciences and Applications. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Nomenclature		
Symbols		
Symbol	Description	SI Unit/Dimension
s	Standard deviation	[unit] (e.g., mGy)
S	Population standard deviation	[unit]
T	Estimator	Same as y , e.g., mGy ²
y	Study variable	[unit] (e.g., mGy)
x	Auxiliary variable	[unit] (e.g., mGy)
MSE	Mean squared error	[unit] ²
PRE	Percent relative efficiency	Dimensionless (%)
τ	Tuning parameter	Depends on form (typically dimensionless)
a, b	Constants used in transformation	Dimensionless
θ	Tuning parameter for log transformation	Dimensionless
n	Sample size	Dimensionless
$\log$	Natural logarithm	Dimensionless
Dimensionless quantities		
The following variables are considered dimensionless: a, b, θ, τ (if not directly multiplied by dimensional terms), PRE , and all logarithmic expressions.		
Subscripts		
x	Refers to the auxiliary variable	
y	Refers to the study variable	
0	Indicates reference or known population value	
$r1, r2$	Denotes different estimator families	
Superscripts		
$+$	Estimator form corresponding to $\theta = 1$	
$-$	Estimator form corresponding to $\theta = -1$	
$*$	Adjusted or modified value	

Estimation makes use of auxiliary data to enhance the reliability of statistical judgments, particularly in cases when direct measurements are impractical, incomplete, or too expensive (Singh, 2003). Auxiliary information was originally used to improve demographic parameter estimate in agricultural studies by Cochran (1940). Ratio estimation boosts population estimates using a correlated auxiliary variable, which he invented. This approach has higher efficiency and lower variation than simple random sampling. He pioneered improved estimating methods in environmental, social, and radiation sciences. Murthy (1964) introduced the product estimation approach as an alternative to the ratio method, especially for negatively correlated auxiliary variables. Bahl and Tuteja presented the exponential estimator (Bahl & Tuteja, 1991) as an alternative to ratio and regression estimators. It is especially effective when the research variable and auxiliary variable are nonlinear. In exponential growth or decay data, the exponential estimator reduces bias and improves efficiency. Regression outperforms ratio and product estimators. It uses auxiliary data in a linear regression model to reduce variance and improve estimate (Singh, 2003). Kadilar and Cingi (2004) suggested improved ratio estimation methods, whereas Gupta and Shabbir (2007) used transformed auxiliary variables for population parameter estimation. Perri (2007), Semaary et al. (2024), Sharma, Singh, Kumari, and Singh (2024), Singh, Sharma and Maurya (2024), Singh, Sharma, Singh, Aloraini and Akhtar (2024), and Singh, Singh, Sharma, and Aloraini (2025) estimated population parameters employing auxiliary data.

Radiological exposure estimates for big populations may benefit greatly from supplementary data in the field of radiation studies. An auxiliary variable, such as the distance from a radiation source or building shielding factors, may be used by researchers to circumvent the costly and intrusive process of directly measuring radiation exposure for every person. Research may benefit from statistical methods such as regression estimation or ratio estimation, which increase the efficiency

of estimates while decreasing sampling mistakes, if a known association between the auxiliary variable and radiation exposure is established.

Variance estimation is essential in radiation sciences because it quantifies the uncertainty and reliability of measurements, ensuring accurate risk assessments and decision-making. Since radiation exposure and contamination levels can fluctuate due to environmental factors, measurement errors, or individual differences, understanding variance helps determine the precision of estimates. For example, in population radiation exposure studies, a high variance indicates significant differences in individual doses, requiring more refined estimation methods, while a low variance suggests more reliable and consistent data. Accurate variance estimation is crucial for regulatory compliance, environmental monitoring, and optimizing radiation protection strategies, ensuring that safety measures are based on statistically sound and precise data. The first to estimate variance using auxiliary information was Isaki and Cary (1983). Several writers, including, Adichwal, Sharma, and Singh (2017), Cebrián, Arcos, García, and Rueda (1997), Singh and Pal (2017), Yaqub and Shabbir (2016) and Yadav, Ahmad, and Riaz (2022) have contributed to variance estimation.

Bhushan and Kumar (2022) and Bhushan, Kumar, Al-Omari, and Alomani (2023) worked on estimation using logarithmic relationships. To linearize nonlinear data, stabilize variance, and increase estimate accuracy – particularly for skewed distributions and additive effects – radiation sciences use logarithmic relationships. Despite the fact that authors such as Kumari and Thakur (2020) and Triveni, Danish, and Albalawi (2024) have worked on variance estimation using log type estimators, this issue needs more attention, and the hunt for more efficient estimators in this area is what encouraged us to look for new ones.

A family of generalized logarithmic-type variance estimators is proposed and evaluated in this paper. These estimators use auxiliary information to increase accuracy in the presence of skewed distributions, high variability, and measurement limitations. For trustworthy

inference, especially in intricate observational contexts, precise variance estimation is essential. Because they stabilize variance and lessen the impact of outliers, logarithmic transformations – which have their roots in transformation theory – are appropriate for data with high tails. With theoretical qualities obtained by Taylor expansions, the suggested estimators expand on conventional ratio and regression techniques into a nonlinear context. The study's objectives are to: (i) create effective estimators with the use of auxiliary data; (ii) obtain bias and MSE expressions; and (iii) verify performance using simulations and actual datasets. According to our hypothesis, these estimators are more efficient and produce a lower MSE than traditional techniques, particularly under non-normal circumstances.

Although Due to the exponential trends of many radiation-related variables, including dose–response and radioactive decay, log transformations are helpful in clarifying intricate connections. Additionally, they improve the accuracy of statistical models, decrease heteroscedasticity, and normalize data. Log-based estimators, for instance, stabilize variability across exposure levels, improving dose assessments in radiation dosimetry and producing more objective and trustworthy estimates. The reason that led us to suggest an estimate approach for estimating variance by using a logarithmic connection between the research variable and the auxiliary variable was the driving force behind our decision. Using four real-life data sets gathered from radiological sciences, we have produced MSE and bias expressions, which are further verified using empirical research undertaken showing the efficacy and applicability of our suggested work. After that, we carried out a simulation research in order to demonstrate the efficacy of estimators in a variety of situations, some of which would not have been within the scope of the empirical investigation. The purpose of the suggested study is to demonstrate its usefulness, and this is accomplished via both empirical and simulation research.

The remainder of article is organized as follows: Section 3 offers an extensive analysis of current methodologies in the literature, laying the groundwork for comparison with our suggested strategy. Section 4 presents our variance estimate methods, including the computation of its MSE and Bias to evaluate its statistical characteristics. Section 5 delineates the empirical and simulation investigations undertaken to assess the practical application of our methodology. Section 6 provides a comprehensive analysis of the findings, highlighting their importance and ramifications. Ultimately, Section 7 summarizes the study by encapsulating principal results and proposing prospective avenues for further research.

2. Notations

Consider a finite population $a_1, a_2, \dots, a_N$ of size N , from which a random sample of size n is extracted using the Simple Random Sampling Without Replacement technique. Let $y_{(i)}$ be the i th sample observation of the data, and correspondingly for the auxiliary variable $x_{(i)}$. Let $\bar{Y}$ and $\bar{X}$ be the population means of the study variable and the auxiliary variable, respectively. Let S_x^2 and s_x^2 denote the population and sample variance of the auxiliary variable, respectively, whereas S_y^2 and s_y^2 represent the population and sample variance of the study variable, respectively. Let C_y represent the coefficient of variation, and ρ_{xy} denote the correlation between Y and X . Furthermore, the parameter $\beta_{2(x)}$ pertains to the auxiliary variable X . To compute the MSE and Bias expressions of the estimators, we have maintained the following assumptions throughout the manuscript.

$$e_y = \frac{s_y^2 - S_y^2}{S_y^2} \quad (2.1)$$

$$e_x = \frac{s_x^2 - S_x^2}{S_x^2} \quad (2.2)$$

such that

$$E(e_y) = E(e_x) = 0 \quad (2.3)$$

$$E(e_y^2) = \zeta(\theta_{40} - 1) \quad (2.4)$$

$$E(e_x^2) = \zeta(\theta_{04} - 1) \quad (2.5)$$

$$E(e_y e_x) = \zeta(\theta_{22} - 1) \quad (2.6)$$

where,

$$\zeta = \left(\frac{1}{n} - \frac{1}{N} \right)$$

and

$$\theta_{mn} = \frac{\gamma_{mn}}{\gamma_{20}^{m/2} \gamma_{02}^{n/2}},$$

m and n be any non-negative integer, $\gamma_{mn} = \frac{1}{N-1} \sum (Y_i - \bar{Y})^m (X_i - \bar{X})^n$

3. Existing estimation techniques for variance estimation

In order to assess the efficacy of our proposed estimator, we have analyzed multiple established estimators that are thoroughly documented in the existing literature. The estimators function as benchmarks for comparison, enabling the assessment of the relative efficiency and accuracy of the approach.

3.1. Simple variance estimator

The simple variance estimator is given as

$$T_0 = s_y^2 \quad (3.1)$$

Variance of simple variance estimator is given as

$$V(T_0) = \zeta(\theta_{40} - 1) S_y^4 \quad (3.2)$$

3.2. Isaki's ratio estimator (1983)

Isaki and Cary (1983) introduced a ratio estimator for estimating the variance of a finite population in the presence of subsidiary information.

$$T_r = \frac{s_y^2}{s_x^2} S_y^2 \quad (3.3)$$

The MSE of the estimator T_r , calculated up to the first order of approximation, is presented as follows:

$$MSE(T_r) = \zeta S_y^4 [(\theta_{40} - 1) + (\theta_{04} - 1) - 2(\theta_{22} - 1)] \quad (3.4)$$

3.3. Isaki's regression estimator (1983)

The Regression Estimator, as presented by Isaki and Cary (1983), is utilized for the estimation of population variance and is defined as follows:

$$T_{reg} = s_y^2 + b(S_x^2 - s_x^2) \quad (3.5)$$

where, b is the sample regression coefficient between s_y^2 and s_x^2 . To the first order of approximation, MSE of the estimator is given by

$$MSE(T_{reg}) = \zeta S_x^4 \left\{ (\theta_{40} - 1) - \frac{(\theta_{22} - 1)^2}{(\theta_{04} - 1)} \right\} \quad (3.6)$$

3.4. Kadilar's ratio type estimators (2006)

Kadilar and Cingi (2006) proposed a variance estimator utilizing the coefficient of variation C_x , the standard deviation S_x , and the population kurtosis $\beta_{2(x)}$.

$$T_{K1} = \frac{s_y^2}{s_x^2 - C_x} [S_x^2 - C_x] \quad (3.7)$$

$$T_{K2} = \frac{s_y^2}{s_x^2 - \beta_{2(x)}} [S_x^2 - \beta_{2(x)}] \quad (3.8)$$

$$T_{K3} = \frac{s_y^2}{s_x^2 \beta_{2(x)} - C_x} [S_x^2 \beta_{2(x)} - C_x] \quad (3.9)$$

$$T_{K4} = \frac{s_y^2}{s_x^2 C_x - \beta_{2(x)}} [S_x^2 C_x - \beta_{2(x)}] \quad (3.10)$$

MSE of the above estimator up to the first order of approximation is given by

$$MSE(T_{Kj}) = \zeta S_y^4 \left\{ (\vartheta_{40} - 1) - 2A_j(\vartheta_{22} - 1) + A_j^2(\vartheta_{04} - 1) \right\} \quad (3.11)$$

$$\text{where, } A_1 = \frac{S_x^2}{S_x^2 - C_x}, A_2 = \frac{S_x^2}{S_x^2 - \beta_{2(x)}}, A_3 = \frac{S_x^2 \beta_{2(x)}}{S_x^2 \beta_{2(x)} - C_x}, A_4 = \frac{S_x^2 C_x}{S_x^2 C_x - \beta_{2(x)}}$$

3.5. Kumar's exponential estimator (2011)

The exponential type estimator utilizing the auxiliary variable is presented as follows, as referenced by [Kumar, Singh, Sawan, and Chauhan \(2011\)](#):

$$T_{exp} = s_y^2 \exp \left[\frac{S_x^2 - s_x^2}{S_x^2 - s_x^2} \right] \quad (3.12)$$

MSE of T_{exp} up to the first order of approximation are given as

$$MSE(T_{exp}) = \zeta_N S_y^4 \left[(\vartheta_{40} - 1) + \frac{(\vartheta_{04} - 1)}{4} - (\vartheta_{22} - 1) \right] \quad (3.13)$$

3.6. Sharma's estimator (2014)

[Sharma and Singh \(2014\)](#) introduced a regression-type variance estimator defined as follows:

$$T_{ss} = \omega_1 s_y^2 + \omega_2 (S_x^2 - s_x^2) \quad (3.14)$$

Minimum MSE of the estimator T_{ss} is given as

$$MSE(T_{ss})_{min} = \frac{S_y^4}{D_1} \left[(D_1 - 1) - \frac{2\zeta_N(\vartheta_{22} - 1)^2}{(\vartheta_{04} - 1)D_1} \right] \quad (3.15)$$

$$\text{where, } D_1 = 1 + \zeta_N(\vartheta_{40} - 1) + \zeta_N \frac{(\vartheta_{22} - 1)^2}{(\vartheta_{04} - 1)}$$

3.7. Cekim's estimator (2020)

[Cekim and Kadilar \(2020\)](#) proposed a log type estimator for estimating variance through the use of an auxiliary variable:

$$T_{ck} = s_y^2 \ln \left(\frac{ts_x^2 + k}{ts_x^2 + k} + \frac{tS_x^2 + k}{ts_x^2 + k} \right) \quad (3.16)$$

MSE of the estimators is given as

$$MSE(T_{ck}) = S_y^2 \left\{ (\ln(2) - 1)^2 + (\ln(2))^2 \zeta(\vartheta_{40} - 1) \right\} \quad (3.17)$$

3.8. Prabhakar's estimators (2024)

[Mishra, Sharma, Adichwal, Rai, and Singh \(2024\)](#) proposed the following log-product-type estimators utilizing an auxiliary variable as follows

$$T_{Ms1} = s_y^2 + \alpha \log \left[\frac{s_x^2}{S_x^2} \right] \quad (3.18)$$

$$T_{Ms2} = s_y^2(w_1 + 1) + w_2 \log \left(\frac{s_x^2}{S_x^2} \right) \quad (3.19)$$

MSE of T_{Ms1} and T_{Ms2} up to first-order of approximation are given by

$$MSE(T_{Ms1}) = S_y^4 A + \alpha^2 B + 2S_y^2 \alpha C \quad (3.20)$$

$$MSE(T_{Ms2}) = C_1^2 + w_1^2 A_1 + w_2^2 B_1 + 2w_1 C_1 + 2w_2 D_1 + 2w_1 w_2 E_1 \quad (3.21)$$

where,

$$A = \zeta(\vartheta_{40} - 1)$$

$$B = \zeta(\vartheta_{04} - 1)$$

$$C = \zeta(\vartheta_{22} - 1)$$

$$A_1 = S_y^4 \{1 + \zeta(\vartheta_{40} - 1)\}$$

$$B_1 = \zeta(\vartheta_{04} - 1)$$

$$C_1 = S_y^4 \zeta(\vartheta_{40} - 1)$$

$$D_1 = S_y^4 \zeta(\vartheta_{22} - 1)$$

$$E_1 = S_y^2 \zeta \left\{ (\vartheta_{22} - 1) - \frac{(\vartheta_{04} - 1)}{2} \right\}$$

So, Minimum $MSE(T_{Ms1})$ at optimum value of α is given by:

$$MSE(T_{Ms1})_{min} = S_y^4 \left[A - \frac{C^2}{B} \right] \quad (3.22)$$

where $\alpha_{opt} = -\frac{S_y^2 C}{B}$ and at optimum value of w_1 and w_2 , the minimum $MSE(T_{Ms2})$ expression is given by:

$$MSE(T_{Ms2})_{min} = C_1 + \frac{B_1 C_1^2 + A_1 D_1^2 - 2C_1 D_1 E_1}{E_1^2 - A_1 B_1} \quad (3.23)$$

where, optimum value of w_1 and w_2 is

$$w_1 = \frac{C_1 B_1 - D_1 E_1}{E_1^2 - A_1 B_1} = w_{1opt} \text{ (say) and } w_2 = \frac{A_1 D_1 - C_1 E_1}{E_1^2 - A_1 B_1} = w_{2opt} \text{ (say)}$$

4. Proposed techniques for variance estimation

We have proposed following generalized log type estimators for estimating the population variance S_y^2 of the study variable Y using auxiliary variable X .

$$T_{r1} = (\tau_1 + \tau_2 s_y^2) \left[1 + \theta_1 \log \left(\frac{as_x^2 + b}{aS_x^2 + b} \right) \right] \quad (4.1)$$

$$T_{r2} = (\tau_3 + \tau_4 s_y^2) \left[1 + \log \left(\frac{as_x^2 + b}{aS_x^2 + b} \right) \right]^{\theta_2} \quad (4.2)$$

here, $a \neq 0$ and b are either real numbers or function of the known parameters of auxiliary variable X such as coefficient of kurtosis $\beta_2(x)$, correlation coefficient ρ of the population, standard deviation S_x and coefficient of variation C_x . $\tau_1, \tau_2, \tau_3, \tau_4$ are suitably chosen constants for the estimators T_{r1} and T_{r2} respectively, to be determined such that MSE of proposed estimators is minimum. Different values a and b are given in [Table 1](#).

For the first estimator, the values $\theta_1 = 1$ and $\theta_1 = -1$ are considered independently, resulting in two distinct forms T_{r1+} and T_{r1-} .

$$T_{r1+} = (\tau_1 + \tau_2 s_y^2) \left[1 + \log \left(\frac{as_x^2 + b}{aS_x^2 + b} \right) \right] \quad (4.3)$$

$$T_{r1-} = (\tau_1 + \tau_2 s_y^2) \left[1 - \log \left(\frac{as_x^2 + b}{aS_x^2 + b} \right) \right] \quad (4.4)$$

Similarly, for the second estimator, θ_2 is independently assigned the values 1 and -1 to obtain two corresponding versions as T_{r2+} and T_{r2-} .

$$T_{r2+} = (\tau_3 + \tau_4 s_y^2) \left[1 + \log \left(\frac{as_x^2 + b}{aS_x^2 + b} \right) \right] \quad (4.5)$$

$$T_{r2-} = (\tau_3 + \tau_4 s_y^2) \left[1 + \log \left(\frac{as_x^2 + b}{aS_x^2 + b} \right) \right]^{-1} \quad (4.6)$$

Note that T_{r1+} and T_{r2+} take the same functional form.

Table 1

Members of generalized proposed estimator T_{r_1} .

Estimators (T_{r_1})	a	b
$T_{r_{1_1}} = (\tau_{1_1} + \tau_{2_1} S_y^2) \left[1 + \theta_{1_1} \log \left(\frac{S_x^2 + C_x}{S_y^2 + C_x} \right) \right]$	1	C_x
$T_{r_{1_2}} = (\tau_{1_2} + \tau_{2_2} S_y^2) \left[1 + \theta_{1_2} \log \left(\frac{\beta_2(x) S_x^2 + C_x}{\beta_2(x) S_y^2 + C_x} \right) \right]$	$\beta_2(x)$	C_x
$T_{r_{1_3}} = (\tau_{1_3} + \tau_{2_3} S_y^2) \left[1 + \theta_{1_3} \log \left(\frac{C_x S_x^2 + \beta_2(x)}{C_x S_y^2 + \beta_2(x)} \right) \right]$	C_x	$\beta_2(x)$
$T_{r_{1_4}} = (\tau_{1_4} + \tau_{2_4} S_y^2) \left[1 + \theta_{1_4} \log \left(\frac{S_x^2 + S_x}{S_y^2 + S_x} \right) \right]$	1	S_x
$T_{r_{1_5}} = (\tau_{1_5} + \tau_{2_5} S_y^2) \left[1 + \theta_{1_5} \log \left(\frac{\beta_2(x) S_x^2 + S_x}{\beta_2(x) S_y^2 + S_x} \right) \right]$	$\beta_2(x)$	S_x
$T_{r_{1_6}} = (\tau_{1_6} + \tau_{2_6} S_y^2) \left[1 + \theta_{1_6} \log \left(\frac{S_x^2 + \rho}{S_y^2 + \rho} \right) \right]$	1	ρ
$T_{r_{1_7}} = (\tau_{1_7} + \tau_{2_7} S_y^2) \left[1 + \theta_{1_7} \log \left(\frac{S_x^2 + \beta_2(x)}{\rho X + \beta_2(x)} \right) \right]$	1	$\beta_2(x)$
$T_{r_{1_8}} = (\tau_{1_8} + \tau_{2_8} S_y^2) \left[1 + \theta_{1_8} \log \left(\frac{C_x S_x^2 + \rho}{C_x + \rho} \right) \right]$	C_x	ρ
$T_{r_{1_9}} = (\tau_{1_9} + \tau_{2_9} S_y^2) \left[1 + \theta_{1_9} \log \left(\frac{\rho S_x^2 + C_x}{\rho S_y^2 + C_x} \right) \right]$	ρ	C_x
$T_{r_{1_{10}}} = (\tau_{1_{10}} + \tau_{2_{10}} S_y^2) \left[1 + \theta_{1_{10}} \log \left(\frac{\beta_2(x) S_x^2 + \rho}{\beta_2(x) X + \rho} \right) \right]$	$\beta_2(x)$	ρ
$T_{r_{1_{11}}} = (\tau_{1_{11}} + \tau_{2_{11}} S_y^2) \left[1 + \theta_{1_{11}} \log \left(\frac{\rho S_x^2 + \beta_2(x)}{\rho X + \beta_2(x)} \right) \right]$	ρ	$\beta_2(x)$
$T_{r_{1_{12}}} = (\tau_{1_{12}} + \tau_{2_{12}} S_y^2) \left[1 + \theta_{1_{12}} \log \left(\frac{\rho S_x^2 + S_x}{\rho X + S_x} \right) \right]$	ρ	S_x
$T_{r_{1_{13}}} = (\tau_{1_{13}} + \tau_{2_{13}} S_y^2) \left[1 + \theta_{1_{13}} \log \left(\frac{S_x^2 + \beta_2(x)}{\rho X + \beta_2(x)} \right) \right]$	1	$\beta_2(x)$

4.1. Properties of the proposed techniques

We use a first-order Taylor series expansion to determine the approximate MSE of the suggested estimator in order to assess its performance. We are taking into account approximations made in Eqs. (2.1) and (2.2) for this. Using these we write estimator T_{r_1} as:

$$T_{r_1} = (\tau_1 + \tau_2 S_y^2) \left[1 + \theta_1 \log \left(\frac{aS_x^2 + b}{aS_y^2 + b} \right) \right] \quad (4.7)$$

$$= [\tau_1 + \tau_2 S_y^2 (1 + \epsilon_y)] \left[1 + \theta_1 \log \left\{ \frac{aS_x^2 (1 + \epsilon_x) + b}{aS_y^2 + b} \right\} \right] \quad (4.8)$$

$$= [\tau_1 + \tau_2 S_y^2 (1 + \epsilon_y)] [1 + \theta_1 \log (1 + \lambda \epsilon_x)] \quad (4.9)$$

where $\lambda = \frac{aS_x^2}{aS_y^2 + b}$

$$T_{r_1} = \tau_1 (1 + \theta_1 \lambda S_y^2 - \frac{\theta_1}{2} \lambda^2 \epsilon_x^2) + \tau_2 (S_y^2 + \lambda S_y^2 \epsilon_x - \theta_1 \lambda^2 S_y^2 \frac{\epsilon_x^2}{2} + S_y^2 \epsilon_y + \theta_1 \lambda S_y^2 \epsilon_x \epsilon_y) \quad (4.10)$$

to calculate bias of the proposed estimator, subtract S_y^2 from the both sides of Eq. (4.10)

$$T_{r_1} - S_y^2 = \tau_1 \left(1 + \theta_1 \lambda S_y^2 - \frac{\theta_1}{2} \lambda^2 \epsilon_x^2 \right) + \tau_2 \left(S_y^2 + \theta_1 S_y^2 \epsilon_x - \theta_1 \lambda^2 S_y^2 \frac{\epsilon_x^2}{2} + S_y^2 \epsilon_y + \theta_1 \lambda S_y^2 \epsilon_y \epsilon_x \right) - S_y^2 \quad (4.11)$$

Taking expectation on both side of Eq. (4.11), we get

$$E(T_{r_1} - S_y^2) = Bias(T_{r_1}) = \tau_1 \left\{ 1 + \theta_1 \lambda S_y^2 - \theta_1 \lambda^2 \frac{E(\epsilon_x^2)}{2} \right\} + \tau_2 \left\{ S_y^2 - \theta_1 \lambda^2 S_y^2 \frac{E(\epsilon_x^2)}{2} + \theta_1 \lambda S_y^2 E(\epsilon_x \epsilon_y) \right\} - S_y^2 \quad (4.12)$$

Squaring and taking expectation of Eq. (4.12)

$$MSE(T_{r_1}) = \tau_1^2 A + \tau_2^2 B + 2\tau_1 C + 2\tau_2 D + 2\tau_1 \tau_2 E + S_y^4 \quad (4.13)$$

where,

$$A = 1 + \theta_1^2 \lambda^2 E(\epsilon_x^2) - \theta_1 \lambda^2 E(\epsilon_x^2)$$

$$B = S_y^4 + S_y^4 E(\epsilon_y^2) + \theta_1^2 \lambda^2 S_y^4 E(\epsilon_x^2) + 4\lambda \theta_1 S_y^4 E(\epsilon_x \epsilon_y) - \theta_1 \lambda^2 S_y^4 E(\epsilon_x^2)$$

$$C = -S_y^2 + \theta_1 \lambda^2 S_y^2 \frac{E(\epsilon_x^2)}{2}$$

$$D = -S_y^4 + \theta_1 \lambda^2 S_y^4 \frac{E(\epsilon_x^2)}{2} - \lambda \theta_1 S_y^4 E(\epsilon_x \epsilon_y)$$

$$E = S_y^2 + 2\lambda \theta_1 E(\epsilon_x \epsilon_y) - \lambda^2 \theta_1 S_y^2 E(\epsilon_x^2) + \lambda^2 \theta_1^2 S_y^2 E(\epsilon_x \epsilon_y)$$

or

$$A = 1 + \theta_1^2 \lambda^2 \zeta(\theta_{04} - 1) - \theta_1 \lambda^2 (\theta_{04} - 1)$$

$$B = S_y^4 + S_y^4 \zeta(\theta_{40} - 1) + \theta_1^2 \lambda^2 S_y^4 \zeta(\theta_{04} - 1) + 4\lambda \theta_1 S_y^4 \zeta(\theta_{22} - 1) - \theta_1 \lambda^2 S_y^4 \zeta(\theta_{04} - 1)$$

$$C = -S_y^2 + \theta_1 \lambda^2 S_y^2 \frac{\zeta(\theta_{04} - 1)}{2}$$

$$D = -S_y^4 + \theta_1 \lambda^2 S_y^4 \frac{\zeta(\theta_{04} - 1)}{2} - \lambda \theta_1 S_y^4 \zeta(\theta_{22} - 1)$$

$$E = S_y^2 + 2\lambda \theta_1 \zeta(\theta_{22} - 1) - \lambda^2 \theta_1 S_y^2 \zeta(\theta_{04} - 1) + \lambda^2 \theta_1^2 S_y^2 \zeta(\theta_{22} - 1)$$

Differentiating Eq. (4.13) with regard to τ_1 and τ_2 and equating it to zero yields the optimum values of τ_1 and τ_2 .

$$\tau_1 = \frac{D\mathcal{F} - B\mathcal{C}}{A\mathcal{B} - \mathcal{F}^2} = \tau_{1_{opt}} \text{ (say) and } \tau_2 = \frac{C\mathcal{F} - A\mathcal{D}}{A\mathcal{B} - \mathcal{F}^2} = \tau_{2_{opt}} \text{ (say)} \quad (4.14)$$

Putting the optimum value of $\tau_{1_{opt}}$ and $\tau_{2_{opt}}$, we get the minimum MSE of estimator T_{R_1} , as under

$$MSE(T_{r_1})_{min} = S_y^4 + \left[\frac{2C\mathcal{D}\mathcal{F} - B\mathcal{C}^2 - A\mathcal{D}^2}{A\mathcal{B} - \mathcal{F}^2} \right] \quad (4.15)$$

Bias and MSE of the estimator T_{r_2} up to first order of approximation

Using a first-order Taylor series expansion on the means of the relevant variables, we use the same approximation process as for T_{r_1} to get the bias and MSE of the estimator T_{r_2} . This enables us to derive simplified expressions for performance evaluation and analytical comparison.

$$T_{r_2} = (\tau_3 + \tau_4 S_y^2) \left[1 + \log \left(\frac{aS_x^2 + b}{aS_y^2 + b} \right) \right]^{\theta_2} \quad (4.16)$$

$$= [\tau_3 + \tau_4 S_y^2 (1 + \epsilon_y)] \left[1 + \log \left\{ \frac{aS_x^2 (1 + \epsilon_x) + b}{aS_y^2 + b} \right\} \right]^{\theta_2} \quad (4.17)$$

$$= [\tau_3 + \tau_4 S_y^2 (1 + \epsilon_y)] [1 + \theta_2 \log (1 + \lambda \epsilon_x)]^{\theta_2} \quad (4.18)$$

$$T_{r_2} = \tau_3 (1 + \theta_2 \lambda S_y^2 - \theta_2 \lambda^2 \epsilon_x^2 + \frac{\theta_2^2}{2} \lambda^2 \epsilon_x^2) + \tau_4 (S_y^2 + \theta_2 \lambda S_y^2 \epsilon_x + S_y^2 \epsilon_y + \theta_2 \lambda S_y^2 \epsilon_x \epsilon_y - \theta_2 S_y^2 \lambda^2 \epsilon_x^2 + \theta_2^2 \lambda^2 S_y^2 \frac{\epsilon_x^2}{2}) \quad (4.19)$$

$$T_{r_2} - S_y^2 = \tau_3 (1 + \theta_2 \lambda S_y^2 - \theta_2 \lambda^2 \epsilon_x^2 + \frac{\theta_2^2}{2} \lambda^2 \epsilon_x^2) + \tau_4 (S_y^2 + \theta_2 \lambda S_y^2 \epsilon_x + S_y^2 \epsilon_y + \theta_2 \lambda S_y^2 \epsilon_x \epsilon_y - \theta_2 S_y^2 \lambda^2 \epsilon_x^2 + \theta_2^2 \lambda^2 S_y^2 \frac{\epsilon_x^2}{2}) - S_y^2 \quad (4.20)$$

Taking expectation on both side of Eq. (4.20), we get the bias of estimator T_{r_2} as follows-

$$Bias(T_{r_2}) = \tau_3 (1 + \theta_2 \lambda S_y^2 - \theta_2 \lambda^2 E(\epsilon_x^2) + \frac{\theta_2^2}{2} \lambda^2 E(\epsilon_x^2)) + \tau_4 (S_y^2 + \theta_2 \lambda S_y^2 E(\epsilon_x) + S_y^2 E(\epsilon_y) + \theta_2 \lambda S_y^2 E(\epsilon_x \epsilon_y) - \theta_2 S_y^2 \lambda^2 E(\epsilon_x^2) + \theta_2^2 \lambda^2 S_y^2 \frac{E(\epsilon_x^2)}{2}) - S_y^2 \quad (4.21)$$

Squaring and taking expectation of Eq. (4.20), we get the MSE -

$$MSE(T_{r_2}) = \tau_3^2 A' + \tau_4^2 B' + 2\tau_3 C' + 2\tau_4 D' + 2\tau_3 \tau_4 E' + S_y^4 \quad (4.22)$$

Table 2

Members of generalized proposed estimator T_{r_2} .

Estimators (T_{r_2})	a	b
$T_{r_{21}} = (\tau_{31} + \tau_{41} s_y^2) \left[1 + \log \left(\frac{s_x^2 + C_x}{S_x^2 + C_x} \right) \right]^{\theta_{21}}$	1	C_x
$T_{r_{22}} = (\tau_{32} + \tau_{42} s_y^2) \left[1 + \log \left(\frac{\beta_2(x)s_x^2 + C_x}{\beta_2(x)S_x^2 + C_x} \right) \right]^{\theta_{22}}$	$\beta_2(x)$	C_x
$T_{r_{23}} = (\tau_{33} + \tau_{43} s_y^2) \left[1 + \log \left(\frac{C_x s_x^2 + \beta_2(x)}{C_x S_x^2 + \beta_2(x)} \right) \right]^{\theta_{23}}$	C_x	$\beta_2(x)$
$T_{r_{24}} = (\tau_{34} + \tau_{44} s_y^2) \left[1 + \log \left(\frac{s_x^2 + S_x}{S_x^2 + S_x} \right) \right]^{\theta_{24}}$	1	S_x
$T_{r_{25}} = (\tau_{35} + \tau_{45} s_y^2) \left[1 + \log \left(\frac{\beta_2(x)s_x^2 + S_x}{\beta_2(x)S_x^2 + S_x} \right) \right]^{\theta_{25}}$	$\beta_2(x)$	S_x
$T_{r_{26}} = (\tau_{36} + \tau_{46} s_y^2) \left[1 + \log \left(\frac{s_x^2 + \rho}{S_x^2 + \rho} \right) \right]^{\theta_{26}}$	1	ρ
$T_{r_{27}} = (\tau_{37} + \tau_{47} s_y^2) \left[1 + \log \left(\frac{s_x^2 + \beta_2(x)}{S_x^2 + \beta_2(x)} \right) \right]^{\theta_{27}}$	1	$\beta_2(x)$
$T_{r_{28}} = (\tau_{38} + \tau_{48} s_y^2) \left[1 + \log \left(\frac{C_x s_x^2 + \rho}{C_x S_x^2 + \rho} \right) \right]^{\theta_{28}}$	C_x	ρ
$T_{r_{29}} = (\tau_{39} + \tau_{49} s_y^2) \left[1 + \log \left(\frac{\rho s_x^2 + C_x}{\rho S_x^2 + C_x} \right) \right]^{\theta_{29}}$	ρ	C_x
$T_{r_{210}} = (\tau_{310} + \tau_{410} s_y^2) \left[1 + \log \left(\frac{\beta_2(x)s_x^2 + \rho}{\beta_2(x)S_x^2 + \rho} \right) \right]^{\theta_{210}}$	$\beta_2(x)$	ρ
$T_{r_{211}} = (\tau_{311} + \tau_{411} s_y^2) \left[1 + \log \left(\frac{\rho s_x^2 + \beta_2(x)}{\rho S_x^2 + \beta_2(x)} \right) \right]^{\theta_{211}}$	ρ	$\beta_2(x)$
$T_{r_{212}} = (\tau_{312} + \tau_{412} s_y^2) \left[1 + \log \left(\frac{\rho s_x^2 + S_x}{\rho S_x^2 + S_x} \right) \right]^{\theta_{212}}$	ρ	S_x
$T_{r_{213}} = (\tau_{313} + \tau_{413} s_y^2) \left[1 + \log \left(\frac{s_x^2 + \beta_2(x)}{\rho S_x^2 + \beta_2(x)} \right) \right]^{\theta_{213}}$	1	$\beta_2(x)$

where,

$$\begin{aligned} A' &= 1 + 2\theta_2^2 \lambda^2 E(\epsilon_x^2) - 2\theta_2 \lambda^2 E(\epsilon_x^2) \\ B' &= S_y^4 + S_y^4 E(\epsilon_y^2) + 2\theta_2^2 \lambda^2 S_y^4 E(\epsilon_x^2) + 4\lambda\theta_2 S_y^4 E(\epsilon_x \epsilon_y) - 2\theta_2 \lambda^2 S_y^4 E(\epsilon_x^2) \\ C' &= -S_y^2 + \theta_2 \lambda^2 S_y^2 E(\epsilon_x^2) - \theta_2 \lambda^2 S_y^2 \frac{E(\epsilon_x^2)}{2} \\ D' &= -S_y^4 - \lambda\theta_2 S_y^4 E(\epsilon_x \epsilon_y) + \theta_2 \lambda^2 S_y^4 \frac{E(\epsilon_x^2)}{2} \\ E' &= S_y^2 - 2\lambda^2 \theta_2 S_y^2 E(\epsilon_x^2) + 2\lambda^2 \theta_2^2 S_y^2 E(\epsilon_x^2) + 2\lambda\theta_2 E(\epsilon_x \epsilon_y) \end{aligned}$$

or

$$\begin{aligned} A' &= 1 + 2\theta_2^2 \lambda^2 \zeta(\theta_{04} - 1) - 2\theta_2 \lambda^2 \zeta\theta_{04} \\ B' &= S_y^4 + S_y^4 \zeta(\theta_{40} - 1) + 2\theta_2^2 \lambda^2 S_y^4 \zeta(\theta_{04} - 1) + 4\lambda\theta_2 S_y^4 \zeta(\theta_{22} - 1) \\ &\quad - 2\theta_2 \lambda^2 S_y^4 (\theta_{04} - 1) \\ C' &= -S_y^2 + \theta_2 \lambda^2 S_y^2 \zeta(\theta_{04} - 1) - \theta_2 \lambda^2 S_y^2 \frac{\zeta(\theta_{04} - 1)}{2} \\ D' &= -S_y^4 - \lambda\theta_2 S_y^4 \zeta(\theta_{22} - 1) + \theta_2 \lambda^2 S_y^4 \frac{\zeta(\theta_{04} - 1)}{2} \\ E' &= S_y^2 - 2\lambda^2 \theta_2 S_y^2 \zeta(\theta_{04} - 1) + 2\lambda^2 \theta_2^2 S_y^2 \zeta(\theta_{04} - 1) + 2\lambda\theta_2 \zeta(\theta_{22} - 1) \end{aligned}$$

Optimum Value of τ_3 and τ_4 is

$$\tau_3 = \frac{D'E' - B'C'}{A'B' - E^2} = \tau_{3_{opt}} \text{ (say) and } \tau_4 = \frac{C'E' - A'D'}{A'B' - E^2} = \tau_{4_{opt}} \text{ (say)} \quad (4.23)$$

Putting the optimum value of $\tau_{3_{opt}}$ and $\tau_{4_{opt}}$, we get the minimum MSE of estimator T_{r_2} , as under

$$MSE(T_{r_2})_{min} = S_y^4 + \left[\frac{2C'D'E' - B'C'^2 - A'D'^2}{A'B' - E^2} \right] \quad (4.24)$$

4.2. Some members of the class of proposed estimator T_{r_1} and T_{r_2}

Taking the constant $a(\neq 0)$ and b are either real numbers or function of the known parameters of auxiliary variable X such as coefficient of kurtosis $\beta_2(x)$, correlation coefficient ρ of the population, standard deviation S_x , coefficient of variation C_x . We can generate the following class of proposed estimators (Table 1):-

5. Application based analysis

5.1. Empirical analysis

To estimate the variance of Adult Stegobium paniceum under Different Irradiation Treatments, we have considered three data sets. These three data sets were used by [Abdelfattah and Sayed \(2022\)](#) to study different radiation exposures (microwave or gamma) influence the generation time, survival, and reproductive success of adult beetles and their offspring, which can provide insights into the effects of such treatments on the population dynamics of Stegobium paniceum.

5.2. Data set 1

This data set represents an experiment on the feeding behavior of drugstore beetles (stegobium paniceum) and their F1 offspring under different treatments. The experiment evaluates how beetles respond to peppermint packets that were either microwaved for 1,2, or 3 min or exposed to gamma radiation at 6,8, or 10 kGy. First set of values corresponds to adult beetles exposed to different treatments: Second set of values corresponds to F1 offspring (next generation) beetles under similar conditions. Here to estimate the variance of adult beetles using the F1 offspring as an auxiliary variable, we have considered Study Variable (Y) as adult beetles and Auxiliary Variable (X) as F1 offspring data

5.3. Data set 2

This study focuses on estimating the variance of adult Stegobium paniceum populations after exposure to different irradiation treatments, including microwave (1, 2, and 3 min) and gamma radiation (6, 8, and 10 kGy). The aim is to analyze the variability in adult beetle counts across different treatments and assess the reproductive patterns in the F1 generation. The primary objective of this study is to estimate the variance of adult beetles (study variable) using the F1 offspring count as an auxiliary variable to improve the accuracy of the estimation. Data Description The dataset consists of adult beetle counts under different treatment conditions. Two sets of data are considered: Adult beetle counts from parents exposed to irradiated peppermint packets and Log-transformed F1 offspring counts divided by parental numbers Study Variable (Y): The count of adult Stegobium paniceum beetles in each treatment group. Auxiliary Variable (X): The log-transformed count of F1 adults divided by the number of parents.

5.4. Data set 3

Stored product pests such as Stegobium paniceum (drugstore beetle) pose a significant threat to food security. Various control measures, including microwave and gamma radiation, have been used to manage beetle populations. Understanding the variance in adult beetle counts after exposure to different irradiation treatments can provide insights into the effectiveness of these methods. This study aims to estimate the variance of adult beetle counts using F1 offspring as an auxiliary variable, enhancing the precision of variance estimation in experimental entomology. The dataset consists of F1 and F2 adult beetle counts subjected to different irradiation treatments. The treatments include:

Microwave radiation (1, 2, and 3 min) Gamma radiation (6, 8, and 10 kGy) Control (non-irradiated) Study Variables:

Response Variable (Y): Number of F2 adult beetles emerging under different treatments. Auxiliary Variable (X): Number of F1 adult beetles, assumed to be correlated with F2 counts, aiding in variance estimation.

Table 3
MSE and PRE table for the empirical study conducted for four data sets.

Est.	Data set 1		Data set 2		Data set 3		Data set 4	
	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
T_u	182826.75	100	1.72E-04	100	5619.491	100	590.859	100.000
T_r	144201.34	126.785	3.27E-05	526.486	22819.275	24.626	873.552	67.639
T_{reg}	130588.28	140.002	2.59E-05	664.030	5489.587	102.366	590.400	100.078
T_{k1}	144157.73	126.824	1.72E-04	100.008	22813.091	24.632	874.077	67.598
T_{k2}	144658.1	126.385	1.72E-04	99.998	22812.32	24.633	867.989	68.072
T_{k3}	144231.33	126.759	1.72E-04	99.988	22809.102	24.637	873.204	67.666
T_{k4}	147724.22	123.762	1.72E-04	99.999	22806.409	24.639	837.005	70.592
T_{exp}	133717.69	136.725	7.98E-05	215.431	9104.284	61.723	655.608	90.124
T_{is}	123911.17	147.546	7.83E-05	219.723	4928.851	114.012	583.900	101.192
T_{ck}	206647.31	88.472	1.57E-04	109.571	7234.324	77.678	5277.103	11.197
T_{pr1}	130588.28	140.002	2.59E-05	664.030	5489.587	102.366	590.400	100.078
T_{pr2}	181523.88	100.717	3.98E-06	1332.612	5629.425	99.823	591.405	99.908
T_{r1}	85066.53	214.922	1.70E-04	1637.984	4289.52	131.005	305.711	193.273
T_{r2}	85066.53	214.922	1.05E-05	1637.984	4289.52	131.005	305.711	193.273

5.5. Data set 4

Solar Radiation and Meteorological Data set is taken from [Kaggle: Solar Radiation Dataset](#). A particular emphasis is placed on the correlation between the angle of the sun zenith and the relative humidity in this dataset, which is comprised of daily averages of meteorological data collected throughout 2019. At the surface of the Earth, the quantity of solar radiation that is received is impacted by the solar zenith angle, which is a representation of the sun's location in relation to a vertical axis. The dataset is generated from the 3rd version of the Physical Solar Model (PSM-v3), which offers high-resolution solar radiation data at intervals of 15 min, with a resolution of 4 km by 4 km. These data are then aggregated to daily averages. This dataset, which contains 365 observations, is intended to investigate the ways in which solar location influences the amounts of atmospheric moisture, with the goal of providing insights into seasonal fluctuations and regional climate trends.

Both the sun zenith angle (X) as the auxiliary variable and the relative humidity (Y) as the study variable are considered to be the most important factors. As the zenith angle increases, which means the sun is lower in the sky, the amount of solar radiation that is received reduces, which may have an effect on the humidity levels.

X: solar zenith angle

Y: Relative humidity

N = 365

A lower MSE indicates that the proposed estimators yield more accurate and consistent estimates of population variance, which is critical when assessing radiation dose levels where precision directly affects safety protocols and medical decision-making. Similarly, a higher PRE reflects that the proposed estimators outperform existing ones, allowing for improved inference even with limited or costly sample data in sensitive applications like radiation risk studies.

5.6. Simulation study

For simulation study we have used the concept of [Raghav \(2023\)](#). Following steps are used for the simulation study-

- Generate auxiliary variable X follows normal distribution as $X \sim N(\mu_x, \sigma_x^2)$, $\mu_x = 5$, $\sigma_x^2 = 2$
- Generate study variable using the model $Y = X - 6.2e$, where $e \sim N(0, 1)$.

The parameters for our simulation study are listed in [Table 4](#).

Further, we have drawn samples of size $n(n_1 = 200, n_2 = 400$ and $n_3 = 600)$ from the population of size $N = 1000$ by the method of simple random sampling without replacement. With the help of these samples, we have calculated MSEs of the proposed estimators. The whole process of getting MSEs of the estimators through the simple

Table 4

Description of the parameters for the estimation of variance under SRSWOR for simulation study.

Parameters	Values	Parameters	Values
N	1000	S_y	6.36416
$\bar{Y}$	4.9384	C_y	1.2887
$\bar{X}$	5.05578	C_x	0.40429
S_x	2.04402	ρ_{xy}	0.2334306

Table 5

Depicting results (MSE & PRE) of simulation analysis for different sample sizes.

Estimators	n = 200		n = 400		n = 600	
	MSE	PRE	MSE	PRE	MSE	PRE
T_u	12.415	100.000	4.667	100.000	2.086	100.000
T_r	25.049	49.564	9.417	49.563	4.203	49.635
T_{reg}	12.361	100.435	4.658	100.193	2.084	100.110
T_{k1}	22.870	54.285	8.598	54.283	3.838	54.350
T_{k2}	25.197	49.272	9.404	49.630	4.190	49.786
T_{k3}	25.231	49.206	9.321	50.074	4.023	51.854
T_{k4}	26.573	46.721	9.564	48.802	4.208	49.578
T_{exp}	15.453	80.344	5.805	80.406	2.593	80.464
T_{is}	12.959	95.806	4.926	94.752	2.202	94.728
T_{ck}	12.361	100.435	4.658	100.193	2.084	100.110
T_{pr1}	12.621	98.369	4.661	100.131	2.072	100.671
T_{pr2}	13.562	91.544	4.221	110.576	1.982	105.252
T_{r1+}	8.892	139.627	3.373	138.363	1.507	138.400
T_{r1-}	6.352	195.4501	2.172	214.871	0.892	233.856
T_{r2+}	8.892	139.627	3.373	138.363	1.507	138.400
T_{r2-}	25.22	49.226	9.23	50.563	4.008	52.045

random sampling method is repeated 5000 times. The results of MSEs and PREs for the proposed estimators are shown in [Tables 5](#). All simulations and data analysis were conducted using R software and for line graph we have used MATLAB (see [Fig. 4](#)).

6. Discussion

We performed both an empirical analysis based on real-life applications and a simulation study to look at the characteristics of the estimators under different circumstances in order to verify our suggested variance estimation approach for radiation investigations.

- We infer from [Table 3](#) that our suggested estimating approach is the most efficient option in the empirical analysis described in [Section 5.1](#) as it consistently beats current methods on all four real datasets.
- Members of the proposed generalized class of estimators for T_{r1} and T_{r2} , respectively, are shown in [Tables 1](#) and [2](#). The MSE and PRE values for each member of these classes stay the same,

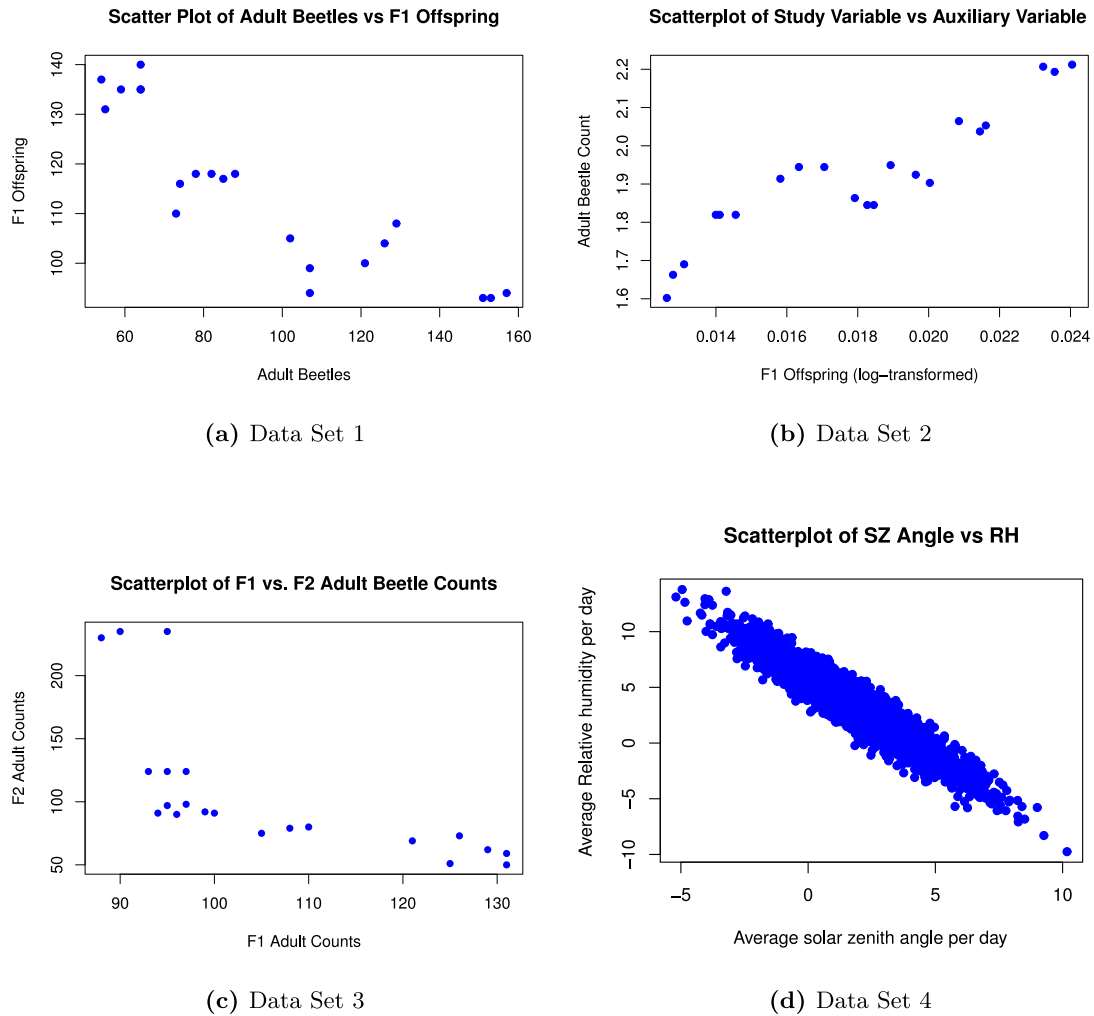


Fig. 1. Graphical depiction of data sets used for empirical study.

according to our study. For this reason, we have only included the MSE and PRE for T_{r1} and T_{r2} , which are two typical classes. Alternate information from the population must be available before a particular estimator class is chosen for variance estimation. There are situations in which full details on the auxiliary variable might not be available. The most efficient variance estimate in such cases should be ensured by selecting an estimator based on the population's available values for parameters a and b .

- The findings of the simulation, which were ran for three different sample sizes, are shown in Table 5. The simulation results are likewise comparable to the empirical results. Validating the suggested estimators by simulation for three distinct sample sizes $n = 200, 400$, and 600 guarantees a comprehensive assessment. It assists in evaluating their robustness, consistency, and stability across a range of data sizes and confirms their asymptotic behavior, which states that an effective estimator should exhibit higher efficiency and a lower MSE as sample size increases.
- In Fig. 1, the graphical representation of four data sets that were taken into consideration for the empirical study is shown. The explanation of these data sets may be found in Section 6. Within the data set, there is a negligible correlation between two variables that were taken into consideration for the study. In the second data set, both variables are positively associated with one another. In the third data set, both the study variables and the auxiliary variable are negatively connected with one another. In the fourth data set, the variables are similarly correlated with

another variable. An examination of the robustness of the suggested technique of estimate of variance is carried out by using four distinct data sets, each of which contains connections that are either positively or negatively linked.

- The MSE of the mentioned estimators for each of the four data sets used in the empirical investigation is shown in Fig. 2. The minimal MSE of our suggested estimators T_{r1+} , T_{r1-} and T_{r2+} is vividly seen. However T_{r2-} is not good estimator. It should not be considered for variance estimation. The PRE of each of the suggested and current estimators taken into consideration for variance estimate is shown in Fig. 3. It clearly demonstrates comparable outcomes and indicates that the suggested approach is worthwhile for variance estimation.
- The MSE of estimators calculated using simulation research, as described in Section 5.6, is shown in Fig. 4 and Fig. 5. The PRE of the estimators calculated using simulation studies for the various sample sizes $n = 200, n = 400$, and $n = 600$ is plotted on a graph in Fig. 6 to demonstrate that the nature of the estimators does not change when sample sizes are changed.

The proposed estimators T_{r1+} , T_{r1-} , T_{r2+} , and T_{r2-} involve logarithmic terms of the form

$$\log\left(\frac{as_x^2 + b}{aS_x^2 + b}\right).$$

While this transformation offers flexibility and improved efficiency under certain conditions, it introduces several important limitations:

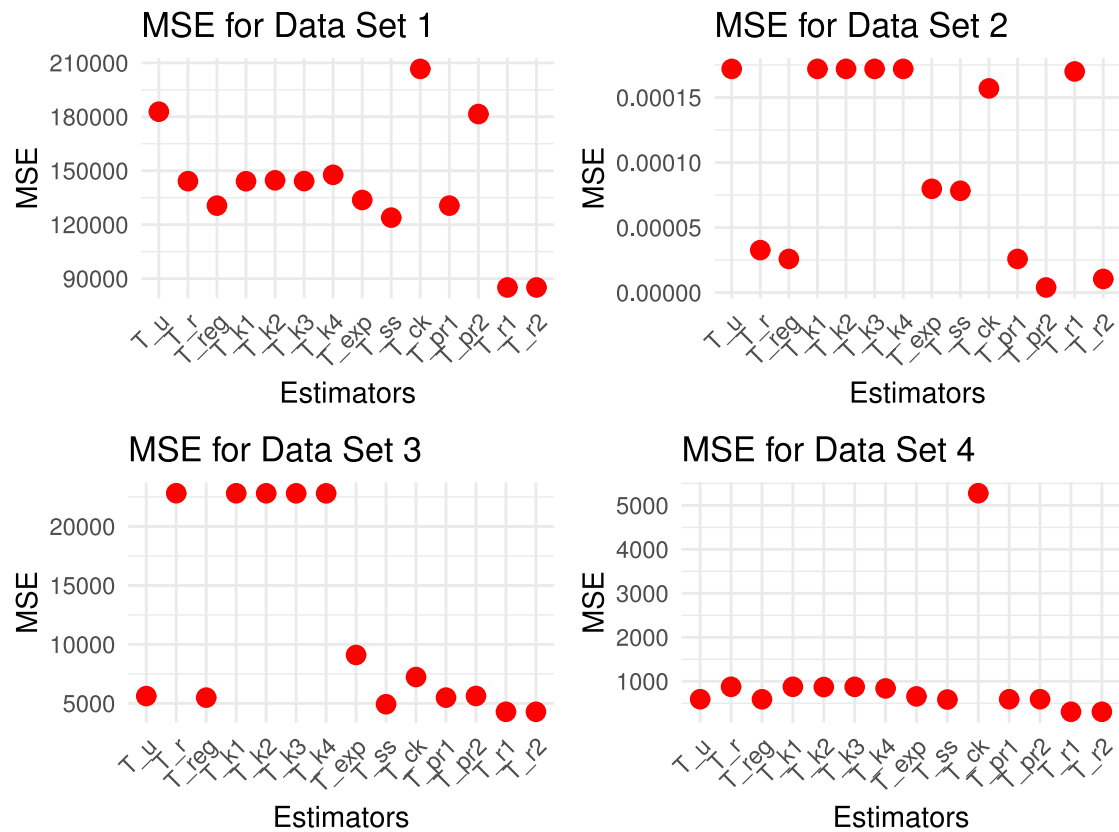


Fig. 2. Graph depicting MSE for all four data sets.

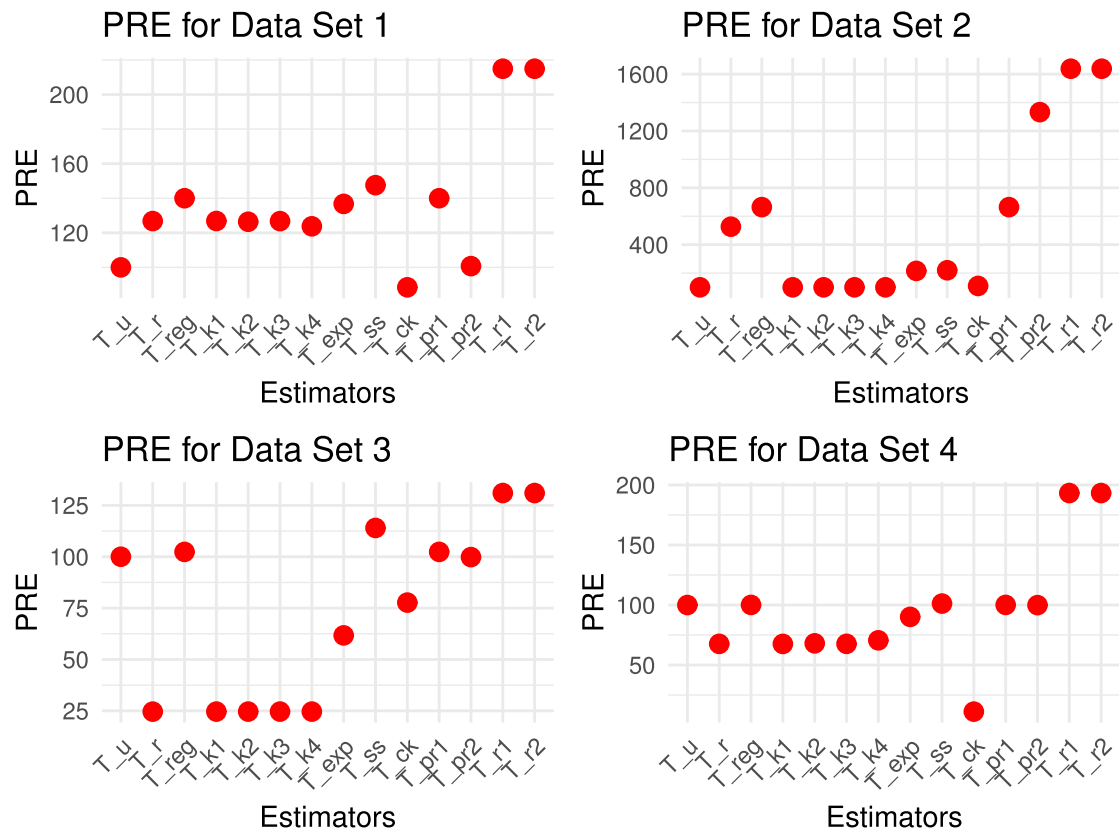


Fig. 3. Graph depicting PRE for all four data sets used in empirical analysis.

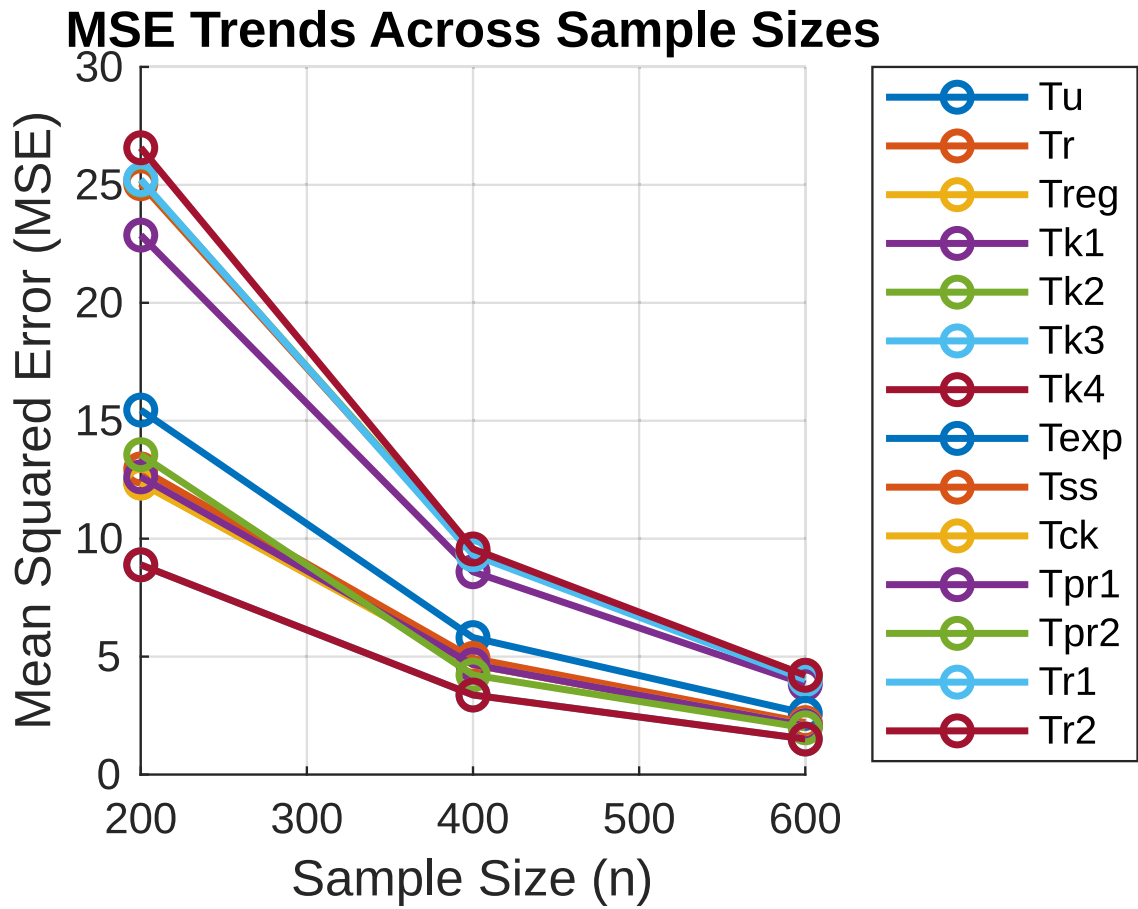


Fig. 4. Graph depicting for different sample sizes.

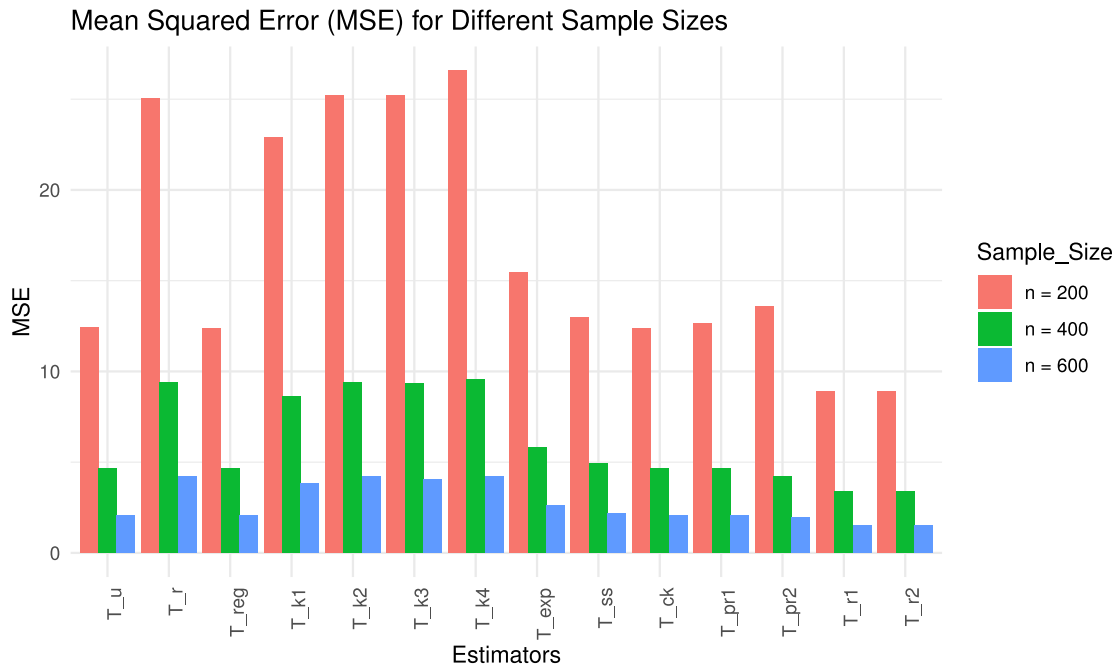


Fig. 5. Graph depicting MSE obtained from simulation analysis.

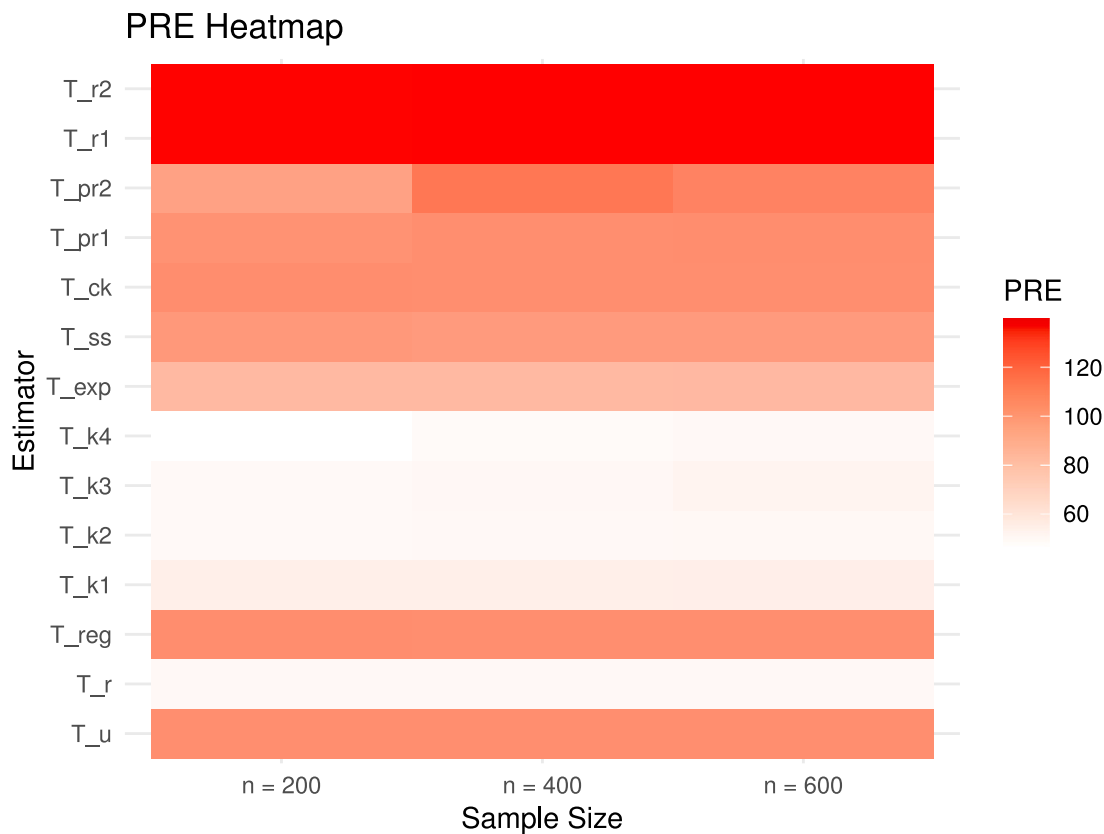


Fig. 6. Graph depicting PRE obtained from simulation analysis.

The logarithmic function is undefined for zero or negative arguments. Therefore, if either $as_x^2 + b \leq 0$ or $aS_x^2 + b \leq 0$, the estimator becomes invalid and may return undefined (NaN) or complex values. When the argument of the logarithmic term approaches zero from the positive side, the log-value becomes large in magnitude (negative), potentially inflating the estimator or increasing its variance. This issue is especially critical in small samples.

7. Conclusion

By utilizing logarithmic transformation techniques in conjunction with auxiliary variables, this work presents an innovative approach for estimating variance in radiation exposure investigations. The suggested approach continuously outperforms conventional estimators in terms of accuracy and efficiency, as shown by empirical and simulated experiments. Our method provides a more accurate calculation of variance by using auxiliary information, especially in situations when complete population data might not be easily accessible. The findings show that the suggested estimators are robust and flexible enough to accommodate different sample sizes, which qualifies them for practical use in radiation exposure studies. All things considered, the methodology of variance estimation has advanced significantly with this study, giving researchers a strong instrument to increase the accuracy and dependability of their research.

This study's future prospects for variance estimation in radiation exposure investigations using logarithmic transformation are promising. An additional avenue for investigation would be the expansion of this methodology to manage more intricate data structures, including multivariate or hierarchical data, which are frequently seen in actual radiation exposure studies.

CRediT authorship contribution statement

Poonam Singh: Supervision, Methodology. **Prayas Sharma:** Writing – original draft, Supervision, Formal analysis. **Anjali Singh:** Writing – original draft, Validation. **Tolga Zaman:** Validation, Project administration, Funding acquisition. **Walid Emam:** Project administration, Funding acquisition, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The study was funded by Researchers Supporting Project number (RSPD2025R749), King Saud University, Riyadh, Saudi Arabia.

Data availability

Data used for the research is the mentioned in the study.

References

Abdelfattah, N. A., & Sayed, R. M. (2022). Comparative efficacy of gamma and microwave radiation in protecting peppermint from infestation by drugstore beetle (*Stegobium paniceum*) L.. *International Journal of Tropical Insect Science*, 42(2), 1367–1372.

Adichwal, N. K., Sharma, P., & Singh, R. (2017). Generalized class of estimators for population variance using information on two auxiliary variables. *International Journal of Applied and Computational Mathematics*, 3, 651–661.

- Bahl, S., & Tuteja, R. K. (1991). Ratio and product type exponential estimators. *Journal of Information and Optimization Sciences*, 12(1), 159–164.
- Bhushan, S., & Kumar, A. (2022). Novel log type class of estimators under ranked set sampling. *Sankhya B. Applied and Interdisciplinary Statistics*, 84(1), 421–447.
- Bhushan, S., Kumar, A., Al-Omari, A. I., & Alomani, G. A. (2023). Mean estimation for time-based surveys using memory-type logarithmic estimators. *Mathematics*, 11(9), 21–25.
- Cebrián, Arcos, A., García, & Rueda, M. (1997). Variance estimation using auxiliary information: An almost unbiased multivariate ratio estimator. *Metrika*, 45(1), 171–178.
- Cekim, H. O., & Kadilar, C. (2020). Ln-type variance estimators in simple random sampling. *Pakistan Journal of Statistics and Operation Research*, 16(4), 689–696.
- Cochran, W. G. (1940). The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *The Journal of Agricultural Science*, 30(2), 262–275.
- Gupta, S., & Shabbir, J. (2007). On the use of transformed auxiliary variables in estimating population mean by using two auxiliary variables. *Journal of Statistical Planning and Inference*, 137(5), 1606–1611.
- Isaki, & Cary, T. (1983). Variance estimation using auxiliary information. *Journal of the American Statistical Association*, 78(381), 117–123.
- Kadilar, C., & Cingi, H. (2004). Ratio estimators in simple random sampling. *Applied Mathematics and Computation*, 151(3), 893–902.
- Kadilar, C., & Cingi, H. (2006). Ratio estimators for the population variance in simple and stratified random sampling. *Applied Mathematics and Computation*, 173(2), 1047–1059.
- Kumar, M., Singh, R., Sawan, N., & Chauhan, P. (2011). Exponential ratio method of estimation in the presence of measurement errors. *International Journal of Agricultural and Statistical Sciences*, 7(2), 457–461.
- Kumari, C., & Thakur, R. K. (2020). An advanced class of log-type estimators for population variance using an attribute and a variable. *International Journal of Industrial Engineering Research and Development*, 11(1), 1–7.
- Mishra, P., Sharma, A., Adichwal, N. K., Rai, S., & Singh, R. (2024). Log-product-type estimator for estimation of population variance using auxiliary information. *Thailand Statistician*, 22(3), 610–617.
- Murthy, M. N. (1964). Product method of estimation. *Sankhyā: The Indian Journal of Statistics Series A*, 69–74.
- Murthy, M. N. (1967). Sampling theory and methods.
- Perri, P. F. (2007). Improved ratio-cum-product type estimators. *Statistics in Transition*, 8(1), 51–69.
- Raghav, Yashpal Singh (2023). Neutrosophic generalized exponential robust ratio type estimators. *International Journal of Analysis and Applications*, 21, 1–18.
- Semary, H. E., Ahmad, S., Hamdi, W. A., Albalawi, O., Elbatal, I., Chesneau, C., et al. (2024). Improved novel estimation for estimation of population distribution function using auxiliary information under stratified sampling strategy. *Journal of Radiation Research and Applied Sciences*, 17(3), 1–10.
- Sharma, P., & Singh, R. (2014). Improved dual to variance ratio type estimators for population variance. *Chilean Journal of Statistics*, 5(2), 45–54.
- Sharma, P., Singh, P., Kumari, M., & Singh, R. (2024). Estimation procedures for population mean using EWMA for time scaled survey. *Sankhya B. Applied and Interdisciplinary Statistics*, 86(2), 1–26.
- Singh, S. (2003). vol. 2, *Advanced sampling theory with applications: how michael selected amy*.
- Singh, H. P., & Pal, S. K. (2017). Improvement over variance estimation using auxiliary information in sample surveys. *Communications in Statistics. Theory and Methods*, 46(15), 7732–7750.
- Singh, P., Sharma, P., & Maurya, P. (2024). Enhancing accuracy in population mean estimation with advanced memory type exponential estimators. *Journal of Reliability and Statistical Studies*, 17(2), 417–434.
- Singh, P., Sharma, P., Singh, R., Aloraini, B., & Akhtar, A. (2024). Optimizing population mean estimation in stratified sampling using linear cost: A simulation study. *Heliyon*, 10(24), Article 1-13.
- Singh, A., Singh, P., Sharma, P., & Aloraini, B. (2025). Estimation of population mean using neutrosophic exponential estimators with application to real data. *International Journal of Neutrosophic Science*, 25(3), 322–338.
- Triveni, G. R. V., Danish, F., & Albalawi, O. (2024). Optimizing variance estimation in stratified random sampling through a log-type estimator for finite populations. *Symmetry*, 16(5), 5–40.
- Yadav, S. K., Ahmad, S., & Riaz, M. (2022). On enhanced estimation and monitoring of population variance using robust auxiliary parameters based estimators. *Concurrency and Computation: Practice and Experience*, 34(24), 1–20.
- Yaqub, M., & Shabbir, J. (2016). An improved class of estimators for finite population variance. *Hacetatepe Journal of Mathematics and Statistics*, 45(5), 1641–1660.