Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 8 (August) (2025)

Optimized Variance Estimation in Simple Random Sampling with Auxiliary Information

Fazal Shakoor
Department of Statistics, University Peshawar KP, Pakistan.
Email: fazalstate47@gmail.com

Muhammad Atif
Department of Statistics, University Peshawar KP, Pakistan.
Email: dratif96@gmail.com

Khazan sher
Higher Education, Archives and Libraries Department,
Government of Khyber Pakhtunkhwa, Pakistan.
Email: khazan680@gmail.com

Sajid Khan
Department of Statistics, University Peshawar KP, Pakistan.
Email: sajidkhan2022@uop.edu.pk

Shumaila Wazir
Department of Statistics, University Peshawar KP, Pakistan.

Muhammad Farooq
Department of Statistics, University Peshawar KP, Pakistan.

ABSTRACT
In this study, we introduce an enhanced estimator for the population variance of the primary study variable under the simple random sampling without replacement (SRSWOR) framework. This estimator leverages both the correlation coefficient between the study variable and an auxiliary variable, as well as the interquartile range of the auxiliary variable. We derive the expressions for its Bias and Mean Square Error (MSE) up to the first order of approximation. Furthermore, the efficiency of the proposed estimator is evaluated by comparing its Percentage Relative Efficiency (PRE) against several existing estimators of population variance. The theoretical analysis supported by numerical illustrations using real secondary data demonstrates that the proposed estimator consistently outperforms the other estimators considered in terms of lower Bias, reduced MSE, and higher PRE.

Key Words: Ratio Estimator; Quartile; MSE; PRE; Efficiency.

Introduction
In survey sampling theory, the use of auxiliary information has become a powerful tool for enhancing the accuracy of population parameter estimates related to the primary study variable. Numerous researchers have highlighted its significance in improving both the design and estimation phases of sampling. Auxiliary information contributes to the development of more efficient sampling strategies such as stratified, systematic, and probability proportional to size (PPS) methods and enhances the precision of estimators for key population characteristics, like the mean or variance of the main variable of

interest. In the present study, auxiliary information is applied solely during the estimation phase. The auxiliary variable (X) is assumed to be strongly associated with the main study variable (Y). When the two variables exhibit a strong positive correlation and the regression line of Y on X passes through the origin, ratio-type estimators are typically employed to improve estimation efficiency. Conversely, if the correlation is negative, product-type estimators are more appropriate. Regression-type estimators are considered when the regression line does not intersect the origin.

Estimating the population variance is a central concern in survey sampling, and numerous efforts have been devoted to enhancing the precision of such estimates. Within the scope of sampling theory, a wide range of methods have been developed that incorporate auxiliary information particularly through ratio, product, and regression techniques to achieve more accurate estimations of the population variance. Several existing estimators have been developed in the past, incorporating auxiliary information to improve efficiency.

Sohaib et al [1] introduced an improved estimator for the finite population variance by incorporating double auxiliary information under simple random sampling. Muhammad et al [2] proposed estimators for both the mean and variance within the same sampling framework. Gupta et al [3] extended existing variance estimators through a linear scrambling model, demonstrating that their approach achieved greater precision than Isaki's [4] estimator under a linear response setting. More recently, Salem et al. [5] put forward a new randomized response technique and developed an innovative variance estimator that utilizes two auxiliary variables, showing substantial gains in efficiency compared to existing estimators.

Zaman et al. [6] proposed new variance estimators by combining mathematical functions with auxiliary parameters. In a separate work, Zaman et al. [7] introduced another variance estimator that makes use of a covariance factor. Additionally, in a more recent study, Zaman et al. [8] designed a randomized method that enhances the efficiency of estimates through group discussion techniques.

Consider a finite population comprising N well-defined and distinguishable units. A simple random sample without replacement (SRSWOR) of size n is drawn, resulting in a bivariate sample (x_i, y_i), for i =1, 2…n corresponding to the auxiliary variable X and the study variable Y. Let $\overline{X}$ and $\overline{Y}$ represent the population means of X and Y, while $\bar{x}$ and $\bar{y}$ denote their respective sample means, which serve as unbiased estimators of the population means. Assume that ρ denotes the correlation coefficient between the two variables, and Q_r stands for the interquartile range of the auxiliary variable X. In this study, we introduce a ratio-type estimator for the population mean of the study variable Y, utilizing the known values of ρ and Q_r. It is further assumed that a dependable prior estimate of the correlation coefficient ρ is available.

This study introduces a refined ratio-type estimator aimed at enhancing the estimation of population variance. The primary objective is to construct a more efficient estimator that builds upon the traditional ratio estimation approach. To achieve this, a generalized form of the population variance estimator is proposed, from which several existing estimators can be derived as specific cases by assigning particular values to a defining scalar parameter.

Literature Based Variance Estimators

Numerous variance estimators have been developed in the literature, particularly those utilizing auxiliary information to enhance precision. These estimators often based on ratio, product, and regression techniques form the foundation for ongoing improvements

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 8 (August) (2025)
in population variance estimation. The following section reviews key contributions in this domain.

The most commonly used method for estimating population variance is the sample variance, which serves as the foundational estimator in survey sampling.

$$t_0 = s_y^2 \tag{2.1}$$

The unbiased and variance up to the first order of approximation as:

$$V(t_0) = \theta s_y^2(\lambda_{20} - 1) \tag{2.2}$$

In 1983, Isaki [9] introduced a ratio-type estimator for estimating population variance by incorporating auxiliary information.

$$t_r = s_y^2\left(\frac{S_x^2}{s_x^2}\right) \tag{2.3}$$

Where,

$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$, $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$, and $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$

The approximate expressions for the bias and mean square error (MSE) of the estimator in equation Eq. (2.3), calculated up to the first-order, are derived as follow

$$\text{Bias}(t_r) = \theta S_y^2[(\lambda_{02} - 1) - (\lambda_{11} - 1)] \tag{2.4}$$

$$\text{MSE}(t_r) = \theta S_y^4[(\lambda_{20} - 1) + (\lambda_{02} - 1) - 2(\lambda_{11} - 1)] \tag{2.5}$$

Numerous researchers have utilized auxiliary information through known population parameters of the auxiliary variable to develop various estimators for the population variance of the study variable. These estimators are often presented along with their corresponding expressions for bias, mean square error (MSE), and associated constants.

Kadilar and Cingi [10] proposed an enhanced estimator for the population variance. The corresponding expressions for its bias and mean square error (MSE) are presented in Equation (2.6).

$$s_1^2 = s_y^2\left[\frac{S_x^2 + C_x}{s_x^2 + C_x}\right] \tag{2.6}$$

$$\text{Bias}(s_1^2) = \theta S_y^2 K_1[K_1(\lambda_{02} - 1) - (\lambda_{11} - 1)] \tag{2.7}$$

$$\text{MSE}(s_1^2) = \theta S_y^4[(\lambda_{20} - 1) + K_1^2(\lambda_{02} - 1) - 2K_1(\lambda_{11} - 1)] \tag{2.8}$$

Where, $K_1 = \frac{S_x^2}{S_x^2 + C_x}$ is constant terms in the above estimator.

Similarly, Upadhyaya and Singh [11] introduced another estimator for the population variance by utilizing auxiliary variable information. The bias and mean square error (MSE) of this estimator are derived and examined in their study.

$$s_2^2 = s_y^2\left[\frac{S_x^2 + \varphi_{2(x)}}{s_x^2 + \varphi_{2(x)}}\right] \tag{2.9}$$

$$\text{Bias}(s_2^2) = \theta S_y^2 K_2[K_2(\lambda_{02} - 1) - (\lambda_{11} - 1)] \tag{2.10}$$

$$\text{MSE}(s_2^2) = \theta S_y^4[(\lambda_{20} - 1) + K_2^2(\lambda_{02} - 1) - 2K_2(\lambda_{11} - 1)]$$

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 8 (August) (2025)

(2.11)

Where, $K_2 = \frac{S_x^2}{S_x^2 + C_x}$ is the constant terms.

Subramani and Kumarapandiyan [12] developed a variance estimator based on quartiles and functions of an auxiliary variable. The bias and mean square error (MSE) expressions for their proposed estimator are presented below.

$$s_3^2 = s_y^2 \left[\frac{S_x^2 + Q_1}{s_x^2 + Q_1}\right] \quad (2.12)$$

$$\text{Bias}(s_3^2) = \theta S_y^2 K_3 [K_3(\lambda_{02} - 1) - (\lambda_{11} - 1)] \quad (2.13)$$

$$\text{MSE}(s_3^2) = \theta S_y^4 [(\lambda_{20} - 1) + K_3^2(\lambda_{02} - 1) - 2K_3(\lambda_{11} - 1)] \quad (2.14)$$

Where, $K_3 = \left[\frac{S_x^2}{S_x^2 + Q_1}\right]$ are the constant.

Similarly, Khan and Shabbir (2013) [13].

$$s_4^2 = s_y^2 \left[\frac{S_x^2 p + Q_3}{s_x^2 p + Q_3}\right] \quad (2.15)$$

$$\text{Bias}(s_4^2) = \theta S_y^2 K_4 [K_4(\lambda_{02} - 1) - (\lambda_{11} - 1)] \quad (2.16)$$

$$\text{MSE}(s_4^2) = \theta S_y^4 [(\lambda_{20} - 1) + K_4^2(\lambda_{02} - 1) - 2K_4(\lambda_{11} - 1)] \quad (2.17)$$

Where, $K_4 = \left[\frac{S_x^2 p}{S_x^2 p + Q_3}\right]$ are the constant.

Yadav et al (2014) [14].

$$s_5^2 = s_y^2 \left[\frac{S_x^2 p + Q_r}{s_x^2 p + Q_r}\right] \quad (2.18)$$

$$\text{Bias}(s_5^2) = \theta S_y^2 K_5 [K_5(\lambda_{02} - 1) - (\lambda_{11} - 1)] \quad (2.19)$$

$$\text{MSE}(s_5^2) = \theta S_y^4 [(\lambda_{20} - 1) + K_5^2(\lambda_{02} - 1) - 2K_5(\lambda_{11} - 1)] \quad (2.20)$$

Where, $K_5 = \left[\frac{S_x^2 p}{S_x^2 p + Q_r}\right]$ are the constant.

Where, $Q_i = 1,2,3$ are the Quartile which divide the whole data into four equal parts. Where the function used for inter quartile range is $Q_r = Q_3 - Q_1$, the semi quartile range $Q_d = \frac{Q_3 - Q_1}{2}$, and Average of the Quartile $Q_a = \frac{Q_3 + Q_1}{2}$, thus the MSE of the above Eq. (2.6) to Eq. (2.20) is written as,

$$\text{MSE}(s_i^2) = \theta S_y^4 [(\lambda_{20} - 1) + K_i^2(\lambda_{02} - 1) - 2K_i(\lambda_{11} - 1)], \text{where } i = 1,2, \dots,5 \quad (2.21)$$

Proposed Estimator

Motivated by Shahzad et al [15] ratio exponential-type estimator, this study aims to develop improved estimators by incorporating auxiliary information through exponential transformations. Such approaches are known to enhance the efficiency of population parameter estimation, especially when there is a strong correlation between study and auxiliary variables.

$$s_p^2 = (k_1 + k_2 s_y^2)\left(\frac{a + S_x^2}{a + s_x^2}\right) \exp\left\{\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2}\right\}$$

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

Vol. 3 No. 8 (August) (2025)

(3.1)

$$s_p^2 = \left(k_1 + k_2 S_y^2 + k_2 S_y^2 e_o\right)\left(1 + \frac{S_x^2 e_1}{a + S_x^2}\right)^{-1} \exp\left(\frac{-e_1}{2}\left(1 + \frac{e_1}{2}\right)\right)^{-1}$$

(3.2)

Here, a is a constant determined by minimizing the mean square error (MSE) of the proposed estimator S_p^2. Some key properties of the proposed estimator S_p^2 are outlined below.

Let we define $s_y^2 = S_y^2(1 + e_0)$, and $s_x^2 = S_x^2(1 + e_1)$ also, $E(e_0) = E(e_1) = 0$, and $E(e_0^2) = \frac{1-f}{n}(\lambda_{20} - 1)$, $E(e_1^2) = \frac{1-f}{n}(\lambda_{02} - 1)$, and $E(e_0 e_1) = \frac{1-f}{n}(\lambda_{11} - 1)$, than let we have

$$s_p^2 = \left(k_1 + k_2 S_y^2 + k_2 S_y^2 e_o\right)(1 + \emptyset e_1)^{-1} \exp\left(\frac{-e_1}{2}\left(1 + \frac{e_1}{2}\right)\right)^{-1}$$

(3.3)

Where $\emptyset = \frac{S_x^2}{a + S_x^2}$

By applying taller series and exponential series up to the first order and ignoring high order of approximation than we have,

$$s_p^2 = \left(k_1 + k_2 S_y^2 + k_2 S_y^2 e_o\right)(1 - \emptyset e_1 + \emptyset^2 e_1)\left(1 - \frac{e_1}{2} + \frac{3}{8} e_1^2\right)$$

(3.4)

By simplifying equation Eq. (3.4) we have

$$s_p^2 = \left(k_1 + k_2 S_y^2 + k_2 S_y^2 e_o\right)\left(1 - \left(\emptyset + \frac{1}{2}\right)\right) e_1 \left(\emptyset^2 + \frac{\emptyset}{2} + \frac{3}{8}\right) e_1^2$$

(3.5)

Let $\Psi_1 = \left(\emptyset + \frac{1}{2}\right)$, and $\Psi_2 = \left(\emptyset^2 + \frac{\emptyset}{2} + \frac{3}{8}\right)$ than the above equation can be written as:

$$s_p^2 = \left(k_1 + k_2 S_y^2 + k_2 S_y^2 e_o\right)(1 - \Psi_1 e_1 + \Psi_2 e_1^2)$$

(3.6)

After multiplying both terms in Eq. (3.6) and Subtract S_y^2 from both sides we have,

$$s_p^2 - S_y^2 = k_1 + k_2 S_y^2 - S_y^2 + k_2 S_y^2 e_o - k_1 \Psi_1 e_1 - k_2 \Psi_1 S_y^2 e_1 - k_2 \Psi_1 S_y^2 e_0 e_1 + k_1 \Psi_2 e_1^2 + k_2 \Psi_2 S_y^2 e_1^2$$

(3.7)

To obtained the bias of the proposed estimator we applying expectation on Eq. (3.7) as:

$$E\left(s_p^2 - S_y^2\right) = k_1 + k_2 S_y^2 - S_y^2 - k_2 \Psi_1 S_y^2 e_0 e_1 + k_1 \Psi_2 e_1^2 + k_2 \Psi_2 S_y^2 e_1^2$$

(3.8)

To obtained the MSE taking square on both side on Eq. (3.7).

$$\left(s_p^2 - S_y^2\right)^2 = k_1^2 + k_2^2 S_y^4 + S_y^4 + k_2^2 S_y^4 e_0^2 + k_1^2 \Psi_1^2 e_1^2 + k_2^2 \Psi_1^2 S_y^4 e_1^2 + 2k_1 k_2 S_y^2 - 2k_1 S_y^2 - 2k_1 k_2 \Psi_1 S_y^2 e_0 e_1 + 2k_1^2 \Psi_2 e_1^2 + 2k_1 k_2 \Psi_2 S_y^2 e_1^2 - 2k_2 S_y^4 - 2k_2^2 \Psi_1 S_y^4 e_0 e_1 + 2k_1 k_2 \Psi_2 S_y^2 e_1^2 + 2k_1^2 \Psi_2 S_y^4 e_1^2 + 2k_2 \Psi_1 S_y^4 e_0 e_1 - 2k_1 \Psi_2 S_y^2 e_1^2 - 2k_2 \Psi_2 S_y^4 e_1^2 - 2k_1 k_2 \Psi_1 S_y^4 e_0 e_1 - 2k_2^2 \Psi_1 S_y^4 e_0 e_1 + 2k_1 k_2 \Psi_1^2 S_y^4 e_1^2$$

(3.9)

The simplified form of Eq. (3.9) is written as,

$$\left(s_p^2 - S_y^2\right)^2 = S_y^4 + \{1 + (\Psi_1^2 + 2\Psi_2) e_1^2\} k_1^2 + \{1 + e_0^2 + (\Psi_1^2 + 2\Psi_2) e_1^2 - 4\Psi_1 e_0 e_1\} k_2^2 S_y^4 - 2(1 + \Psi_2 e_1^2) k_1 S_y^2 - 2(1 + \Psi_2 e_1^2 - \Psi_1 e_0 e_1) k_2 S_y^4 + 2\{1 + (\Psi_1^2 + 2\Psi_2) e_1^2 - 2\Psi_1 e_0 e_1\} k_1 k_2 S_y^2 \quad (3.10)$$

By taking the expectation of Equation (3.10), we obtain the expression for the Mean Square Error (MSE) of the proposed estimator.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 8 (August) (2025)

$\mathrm{MSE}(s_p^2) = S_y^4 + \{1 + (\Psi_1^2 + 2\Psi_2)\sigma(\lambda_{02} - 1)\}k_1^2 + [1 + \sigma\{(\lambda_{20} - 1) + (\Psi_2^2 + 2\Psi_2)(\lambda_{02} - 1) - 4\Psi_1((\lambda_{11} - 1))\}]k_2^2 S_y^4 - 2\{1 + \Psi_2\sigma(\lambda_{02} - 1)\}k_1 S_y^2 - 2\{1 + \Psi_2\sigma(\lambda_{02} - 1) - \Psi_1\sigma(\lambda_{11} - 1)\}k_2 S_y^4 + 2[1 + \sigma\{(\Psi_1^2 + 2\Psi_2)(\lambda_{02} - 1) - 2\Psi_1(\lambda_{11} - 1)\}]k_1 k_2 S_y^2$ (3.11)

Where, $E(e_0^2) = \sigma(\lambda_{20} - 1), E(e_1^2) = \sigma(\lambda_{02} - 1)$ and $E(e_0 e_1) = \sigma(\lambda_{11} - 1)$, and $\theta_{rs} = \frac{\mu_{rs}}{\mu_{10}^{\frac{r}{2}}\mu_{01}^{\frac{s}{2}}}$, $\mu_{rs} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^r (X_i - \bar{X})^s$, and $rs = 0,1$

After simplifying Equation (3.11), the resulting expression for the MSE of the proposed estimator is given below.

$$\mathrm{MSE}(s_p^2) = S_y^4 + k_1^2 A + k_2^2 S_y^4 B - 2k_1 S_y^2 C - 2k_2 S_y^4 D + 2k_1 k_2 S_y^2 E \quad (3.12)$$

Where, $A = 1 + (\Psi_1^2 + 2\Psi_2)\sigma(\lambda_{02} - 1)$, $B = 1 + \sigma\{(\lambda_{20} - 1) + (\Psi_2^2 + 2\Psi_2)(\lambda_{02} - 1) - 4\Psi_1((\lambda_{11} - 1))\}$, $C = -2\{1 + \Psi_2\sigma(\lambda_{02} - 1)\}$, $D = -2\{1 + \Psi_2\sigma(\lambda_{02} - 1) - \Psi_1\sigma(\lambda_{02} - 1)\}$, and $E = 2[1 + \sigma\{(\Psi_2^2 + 2\Psi_2)(\lambda_{02} - 1) - 2\Psi_1(\lambda_{11} - 1)\}]$

To differentiate Equation (3.12) with respect to k_1, and k_2 and equate to zero, we need to see the explicit form of Equation (3.13) as:

$$2Ek_2 S_y^2 - 2CS_y^2 + 2Ak_1 = 0, \text{ and } 2Bk_2 S_y^4 + 2Ek_1 S_y^2 + 2DS_y^4 = 0 \quad (3.13)$$

The optimum values of k_1, and k_2 are in equation (3.13) is given bellow.

$$k_1 = \frac{S_y^2(BC - DE)}{AB - E^2}, \text{ and } k_2 = \frac{AD - CE}{AB - E^2} \quad (3.14)$$

Substituting the optimal values of k_1, and k_2 into Equation (3.12), we obtain the resulting MSE in the form given below. And the final and Simplified form of the proposed estimator MSE is given bellow in Eq. (3.15).

$$\mathrm{MSE}(S_p^2) = S_y^4(1 - \frac{AD^2 + BC^2 - 2CDE}{AB - E^2} \quad (3.15)$$

Efficiency Comparison

In this section, we compare our proposed estimator $\mathbf{S_p^2}$ given in equation (3.14) with the existing estimator $\mathbf{t_0}$ in equation (2.2). The proposed estimator outperforms the existing estimator if the following condition is satisfied.

$$V(t_0) - \mathrm{MSE}(S_p^2) > 0, \text{ if } (\lambda_{11} - 1) > 0 \quad (4.1)$$

As we compare equation (3.14) to the Existence estimator in equation (2.5), we have the proposed estimator S_p^2 is more efficient than the estimator t_r, if

$$\mathrm{MSE}(t_r) - \mathrm{MSE}(S_p^2) > 0, \text{ if } (\lambda_{11} - 1) > (\lambda_{02} - 1) \quad (4.2)$$

Similarly, if we compare equation (3.14) to the Estimators in equation (2.36), the proposed one gives us better performance if,

$$\mathrm{MSE}(s_i^2) - \mathrm{MSE}(S_p^2) > 0, \text{ if } (\lambda_{11} - 1) > (\lambda_{02} - 1), i = 1,2, \ldots 11 \quad (4.3)$$

Empirical Illustration

To validate the theoretical findings and assess the performance of the proposed estimators, this chapter presents a detailed numerical illustration. Using a real or hypothetical dataset, the efficiency, bias, and mean square error (MSE) of the estimators are computed and compared to demonstrate their practical utility.

Data-1: Waste data published by Italy's Agency for Environmental Protection (APAT) (2004) [16]

X: Total number of residents in Italy during 2003
Y: Volume (tons) of recyclable materials gathered across Italy in 2003

$$N = 103, n = 40, \bar{Y} = 626.2123, \bar{X} = 556.5594, \rho = 0.7298, S_y = 913.5498, C_y = 1.4588, S_x = 610.1643, C_x = 1.0963, \lambda_{02} = 17.8738, \lambda_{20} = 37.1279, \lambda_{11} = 17.2220, Q_1 = 259.3830, Q_3 = 628.0235, Q_r = 368.6405, Q_d = 184.3293, Q_a = 443.7033$$

Data-2: According to Murthy (1967) [17].
X: Amount of fixed capital invested.
Y: Production output from 80 factories within the region.

$$N = 80, n = 20, \bar{Y} = 51.8264, \bar{X} = 11.2646, \rho = 0.9413, S_y = 18.3549, C_y = 0.3542, S_x = 8.4563, C_x = 0.7507, \lambda_{02} = 2.8664, \lambda_{20} = 2.2667, \lambda_{11} = 2.2209, Q_1 = 5.1500, Q_3 = 16.975, Q_r = 11.825, Q_d = 5.9125, Q_a = 11.0625$$

Data-3: According to Singh and Chaudhary (1986) [18].

$$N = 70, n = 25, \bar{Y} = 96.7000, \bar{X} = 175.2671, \rho = 0.7293, S_y = 60.7140, C_y = 0.6254, S_x = 140.8572, C_x = 0.8037, \lambda_{02} = 7.0952, \lambda_{20} = 4.7596, \lambda_{11} = 4.6038, Q_1 = 80.1500, Q_3 = 225.0250, Q_r = 144.8750, Q_d = 72.4375, Q_a = 152.5875$$

Table:1 Proposed Estimator MSE's performance on real data sets

Estimators	Data-1	Data-2	Data-3
t_0	384782187275	5391.544	1313625
t_r	218950915258	2942.428	924946.5
s_1^2	218950874376	2886.199	924876
s_2^2	218950249143	2742.454	924324.4
s_3^2	218937357134	2322.655	912437.8
s_4^2	218919853126	2157.976	898785.4
s_5^2	218932433255	2298.715	907897.1
$\mathbf{S_p^2}$	**1329114669**	**197.7509**	**34632.47**

Table:2 PRE's of the proposed estimator on real data sets

Estimators	Data-1	Data-2	Data-3
t_0	100	100	100
t_r	175.739	183.2345	142.0218
s_1^2	175.7391	186.8043	142.0326
s_2^2	175.7396	196.5956	142.1173
s_3^2	175.7499	232.1285	143.9687
s_4^2	175.764	249.8426	146.1556
s_5^2	175.7539	234.546	144.6888
$\mathbf{S_p^2}$	**28950.26**	**2726.432**	**3793.045**

Conclusion

In this paper, we have proposed a novel generalized ratio-type estimator aimed at improving the estimation of the population mean under simple random sampling without replacement (SRSWOR). The estimator leverages auxiliary information through the correlation coefficient between the study variable and an auxiliary variable, thereby enhancing the precision of the estimate. To evaluate the statistical properties of the proposed estimator, we derived explicit expressions for its bias and mean square error (MSE) up to the first-order approximation. This theoretical analysis enables practitioners to understand how the estimator behaves asymptotically and under finite samples.

Moreover, we determined the optimal value of the scalar constant incorporated in the estimator to minimize the MSE, which provides guidance on how to achieve the best possible efficiency when applying the estimator in practice. The comparative analysis presented in Section-4 and the empirical results summarized in Table-1 and Table-2 demonstrate that the proposed estimator consistently outperforms the existing estimators listed in these tables, showing lower MSE values in Table-1, higher PRE-values in Table-2, and yielding smaller bias and reduced MSE across various scenarios, which confirms its superior efficiency and reliability for estimating the population variance. It should be noted that the application of the proposed estimator requires prior knowledge of the correlation coefficient (ρ) between the primary study variable and the auxiliary variable. In practice, this information is often accessible from historical data, previous surveys, or a preliminary study conducted on a small sub-sample of the target population. In instances where the true value of ρ is unavailable, it can be replaced by its sample estimate. Previous studies and our empirical findings suggest that substituting an estimated correlation coefficient does not significantly compromise the performance of the estimator, making it practical for real-world survey applications.

Hence, the proposed estimator should be preferred over the estimators given in Table-1 for estimating the population variance of the main variable under study. Considering the theoretical derivations and empirical evidence presented in this paper, the proposed estimator is strongly recommended for practical use in survey sampling when precise and efficient estimation of the population mean is required. Overall, this study highlights the estimator's potential to improve the accuracy and reliability of population mean estimates, making it a valuable contribution to the field of survey sampling.

References

Ahmad, S., Adichwal, N. K., Aamir, M., Shabbir, J., Alsadat, N., Elgarhy, M., & Ahmad, H. (2023). An enhanced estimator of finite population variance using two auxiliary variables under simple random sampling. Scientific Reports, 13(1), 21444.

Alomair, M. A., & Gardazi, S. A. H. S. (2024). Hybrid class of robust type estimators for variance estimation using mean and variance of auxiliary variable. Heliyon, 10(10).

Gupta, S., Qureshi, M. N., & Khalil, S. (2020). Variance estimation using randomized response technique. REVSTAT-Statistical Journal, 18(2), 165-176.

Isaki, C. T. (1983). Variance estimation using auxiliary information. Journal of the American statistical association, 78(381), 117-123.

Saleem, I., Sanaullah, A., Al-Essa, L. A., Bashir, S., & Al Mutairi, A. (2023). Efficient estimation of population variance of a sensitive variable using a new scrambling response model. Scientific Reports, 13(1), 19913.

Zaman, T., & Bulut, H. (2022). A new class of robust ratio estimators for finite population variance. Scientia Iranica.

Zaman, T., & Bulut, H. (2023). An efficient family of robust-type estimators for the population variance in simple and stratified random sampling. Communications in Statistics-Theory and Methods, 52(8), 2610-2624.

Zaman, Q., Ijaz, M., & Zaman, T. (2023). A randomization tool for obtaining efficient estimators through focus group discussion in sensitive surveys. Communications in Statistics-Theory and Methods, 52(10), 3414-3428.

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 8 (August) (2025)

C. T.Isaki, Variance estimation using auxiliary information, Journal of American Statistical Association, 78, 117- 123 (1983).

C. Kadilar and H. Cingi, Improvement in variance estimation using auxiliary information, Hacettepe Journal of mathematics and Statistics,35, 111-15 (2006).

L. N. Upadhyaya and H. P. Singh, Use of auxiliary information in the estimation of population variance, mathematical forum4, 33-36 (1983).

J. Subramani and G. Kumarapandiyan, Variance estimation using quartiles and their functions of an auxiliary variable, International Journal of Statistics and Applications,2,67-72 (2012).

M. Khan and J. Shabbir, A Ratio Type Estimator for the Estimation of Population Variance using Quartiles of an Auxiliary Variable, Journal of Statistics Applications and Probability,2, No.3,319- 325 (2013).

Yadav, S. K., Mishra, S. S., Shukla, A. K., & Tiwari, V. (2015). Improvement of estimator for population variance using correlation coefficient and quartiles of the auxiliary variable. Journal of Statistics Applications & Probability, 4(2), 259.

Shahzad, U., Hanif, M., Koyuncu, N., & Sanaullah, A. (2018). On the estimation of population variance using auxiliary attribute in absence and presence of non-response. Electronic Journal of Applied Statistical Analysis, 11(2), 608-621.

Mazzanti, M., & Montini, A. (Eds.). (2009). Waste and environmental policy (Vol. 15). London: Routledge.

M. N. Murthy, Sampling Theory and Methods, Statistical Publishing Society Calcutta, India, (1967).

D. Singh, and F. S. Chaudhary, Theory and analysis of sample survey designs, New-Age International Publisher, (1986).