Thailand Statistician
October 2026; 24(4): 803-813
http://statassoc.or.th
Contributed paper

Novel Estimators for Enhanced Finite Population Variance Estimation with Auxiliary Data

Housila P. Singh [a] and Shakti Prasad* [b]
[a] Retired Professor, School of Studies in Statistics, Samrat Vikramaditya University, Ujjain, Madhya Pradesh, India
[b] Department of Mathematics, National Institute of Technology Jamshedpur, Jamshedpur, Jharkhand, India
*Corresponding author; e-mail: shakti.pd@gmail.com

Received: 18 October 2024
Revised: 22 June 2026
Accepted: 9 July 2026

Abstract

This paper proposes two novel classes of efficient estimators for estimating the finite population variance under the Simple Random Sampling Without Replacement (SRSWOR) scheme by utilizing auxiliary information in sample surveys. The bias and mean squared error (MSE) expressions of the proposed classes of estimators are derived up to the first order of approximation. The optimum conditions under which the proposed estimators achieve minimum MSE are also obtained. Furthermore, regions of preference are established to determine the conditions under which the proposed estimators perform better than the existing estimators available in the literature. An empirical study based on three real data sets is carried out to evaluate the efficiency and practical applicability of the proposed estimators.

Keywords: Population variance, mean squared error, bias, auxiliary variable, study variable.

1. Introduction

The use of auxiliary information at the estimation stage is well known to improve the precision of estimating population characteristics related to both the study and auxiliary variables, including the mean, variance, coefficient of variation, and correlation coefficient. In practical applications, variability is often difficult to control, which makes the estimation of finite population variance an important aspect of sample survey theory. This issue is particularly prominent in disciplines such as environmental science and ecology, where experimental observations are often influenced by uncontrollable sources of variation, thereby affecting the reliability of the results. Das and Tripathi (1978) introduced various estimators for the estimation of finite population variance. Mukhopadhyay (1982) contributed substantially to the estimation of population variance under super-population models. Wu (1982) exploited the correlation between the study and auxiliary variables to improve variance estimation. Isaki (1983) proposed ratio- and regression-type estimators for estimating population variance. Biradar and Singh (1998), Yadav and Kadilar (2013) and Singh and Solanki (2013) investigated variance estimators under various sampling methods. Sharma and Singh (2014) studied the estimation of population variance using auxiliary attributes in sampling. Bhushan et al. (2023) introduced an efficient class of estimators for variance estimation under simple random sampling. Mishra et al. (2024)

examined efficient estimators for the population variance is survey samplng. Various authors, including Yadav et al. (2015), Niaz et al. (2022), Yadav and Prasad (2024), Singh and Prasad (2025), Daraz et al. (2025), Djebar et al. (2026) and Swain et al. (2026), have proposed estimators for population variance and investigated their properties using auxiliary variable at the estimation stage. Motivated by the works of Kumar et al. (2011) and Mishra et al. (2024), we propose two new classes of estimators, by combining exponential, ratio-type, and logarithmic transformations of auxiliary information to achieve improved estimation of the population variance in sample surveys.

The rest of this paper is structured as follows. Section 2 introduces the preliminary concepts and notation used throughout the study. Section 3 reviews some existing estimators available in the literature. we propose two novel classes of efficient estimators for the finite population variance and obtain their statistical properties up to the first-order approximation. Section 5 provides efficiency comparisons between the proposed and existing estimators. Section 6 demonstrates the effectiveness of the proposed estimators through an empirical analysis of three real population datasets and discusses the resulting insight. Finally, concluding remarks are summarized in Section 7.

2. Preliminary Concepts and Notation

Consider a finite population $\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_N)$, where N is the population size. Let y and x represent the study and auxiliary variables, respectively. The observed values for the i-th unit Ω_i are denoted by (y_i, x_i), where $i = 1, 2, \ldots, N$. The population means of the variables y and x are given by $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ and $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$. The corresponding population variances are defined as $S_y^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2$ and $S_x^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^2$. Suppose a sample of size n is selected from the population using Simple Random Sampling Without Replacement (SRSWOR). Let $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ be the sample means corresponding to the study variable y and auxiliary variable x, respectively. The sample variances are then expressed as $s_y^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$ and $s_x^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$. It is assumed that the population variance S_x^2 of the auxiliary variable x is known in advance. The objective is to estimate the unknown population variance S_y^2 of the study variable y by utilizing the sample information (s_y^2, s_x^2) along with the known value of S_x^2. For deriving the expressions of bias and mean squared error (MSE) of various estimators for the population variance S_y^2, let us define the relative error terms $e_0 = \frac{s_y^2 - S_y^2}{S_y^2}$ and $e_1 = \frac{s_x^2 - S_x^2}{S_x^2}$. Accordingly, the sample variances may be expressed as $s_y^2 = S_y^2(1+e_0)$ and $s_x^2 = S_x^2(1+e_1)$. Under SRSWOR, the error terms satisfy $E(e_0) = E(e_1) = 0$. Further, retaining only terms up to the first order of approximation, the following results are obtained: $E(e_0^2) = \xi(\lambda_{40} - 1)$, $E(e_1^2) = \xi(\lambda_{04} - 1)$, $E(e_0 e_1) = \xi(\lambda_{22} - 1)$, where$\xi = \frac{1-f}{n}$, $f = \frac{n}{N}$. The quantities λ_{rs} are defined by $\lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{r/2}\,\mu_{02}^{s/2}}$ with $\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{X})^s(y_i - \bar{Y})^r$, where r and s are integers greater than or equal to zero.

3. A Brief Review of Existing Population Variance Estimators

This section reviews some existing estimators of the population variance S_y^2 of y using information on the population variance S_x^2 of the auxiliary variable x. The usual unbiased variance estimator s_y^2 of y is defined by

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2. \tag{1}$$

The variance (Var) of s_y^2 to the first degree of approximation is obtained as

$$Var\left(s_y^2\right) = \xi(\lambda_{40} - 1)S_y^4. \tag{2}$$

Using information on population variance S_x^2 of the auxiliary information x, Isaki (1983) considered a ratio estimator for the population variance S_y^2 of study variable y as follows:

$$I_R = s_y^2 \left(\frac{S_x^2}{s_x^2} \right). \tag{3}$$

The MSE of I_R is found as

$$MSE\left(I_R\right) = S_y^4 \xi \left[(\lambda_{40} - 1) - 2(\lambda_{22} - 1) + (\lambda_{04} - 1) \right]. \tag{4}$$

Kumar et al. (2011) introduced the exponential ratio estimator for S_y^2 of study variable y, given by

$$K_{Re} = s_y^2 exp \left[\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2} \right]. \tag{5}$$

The MSE of K_{Re} is obtained as

$$MSE\left(K_{Re}\right) = S_y^4 \xi \left[(\lambda_{40} - 1) - (\lambda_{22} - 1) + \frac{1}{4}(\lambda_{04} - 1) \right]. \tag{6}$$

Das and Tripathi (1978) suggested the following family of estimators for S_y^2 as

$$DT_1 = s_y^2 \left(\frac{S_x^2}{s_x^2} \right)^{\alpha_1}, \tag{7}$$

$$DT_2 = s_y^2 \left(\frac{S_x^2}{S_x^2 + \alpha_2 (s_x^2 - S_x^2)} \right), \tag{8}$$

$$DT_3 = s_y^2 + \alpha_3 (S_x^2 - s_x^2), \tag{9}$$

where α_1, α_2 and α_3 are appropriately selected constants.

The minimum MSE of DT_i is derived as:

$$\begin{aligned} MSE_{min}\left(DT_i\right) &= S_y^4 \xi \left[(\lambda_{40} - 1) - \frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)} \right] \forall i = 1, 2, 3. \\ &= S_y^4 \xi (\lambda_{40} - 1)(1 - \rho^{*^2}), \end{aligned} \tag{10}$$

where $\rho^{*^2} = \frac{(\lambda_{22}-1)^2}{(\lambda_{40}-1)(\lambda_{04}-1)}$.

Singh et al. (1988) proposed difference type estimator for S_y^2 of y, given by

$$S_D = w_1 s_y^2 + w_2 (S_x^2 - s_x^2), \tag{11}$$

where w_1 and w_2 are appropriately selected constants.

The minimum MSE of S_D is obtained as

$$MSE_{min}\left(S_D\right) = \frac{S_y^4 \xi (1 - \rho^{*^2})(\lambda_{40} - 1)}{1 + \xi (1 - \rho^{*^2})(\lambda_{40} - 1)}. \tag{12}$$

Yadav and Kadilar (2013) introduced a exponential ratio type estimators for the population variance S_y^2 of study variable y, given by

$$YK_{Re} = s_y^2 exp \left[\frac{S_x^2 - s_x^2}{S_x^2 + (\phi - 1) s_x^2} \right], \tag{13}$$

where ϕ is suitable chosen constant.

The minimum MSE of YK_{Re} is expressed as

$$MSE_{min}\left(YK_{Re}\right) = S_y^4 \xi (\lambda_{40} - 1)(1 - \rho^{*^2}). \tag{14}$$

From (2), (4), (6), (10) and (14), we obtain

$$MSE\left(s_y^2\right) - MSE_{min}\left(t\right) = S_y^4 \xi {\rho^*}^2 (\lambda_{40} - 1) \geq 0, \tag{15}$$

$$MSE\left(I_R\right) - MSE_{min}\left(t\right) = S_y^4 \xi \frac{\{(\lambda_{04} - 1) - (\lambda_{22} - 1)\}^2}{(\lambda_{04} - 1)} \geq 0, \tag{16}$$

$$MSE\left(K_{Re}\right) - MSE_{min}\left(t\right) = S_y^4 \xi \frac{\{\frac{1}{2}(\lambda_{04} - 1) - (\lambda_{22} - 1)\}^2}{(\lambda_{04} - 1)} \geq 0, \tag{17}$$

where $t = DT_1, DT_2, DT_3$ and YK_{Re}.

It can be observed from (15) to (17) that, under their optimum conditions, the estimators $DT_i (i = 1, 2, 3)$ and YK_{Re} outperform s_y^2, I_R and K_{Re} in terms of efficiency.

Further, from (10), (12) and (14), we obtain

$$MSE_{min}\left(t\right) - MSE_{min}\left(S_D\right) = S_y^4 \xi^2 \frac{(\lambda_{40} - 1)^2 (1 - {\rho^*}^2)^2}{1 + \xi(\lambda_{40} - 1)(1 - {\rho^*}^2)} \geq 0 \quad \forall t = DT_1, DT_2, DT_3, YK_{Re}. \tag{18}$$

It follows from (15), (16), (17) and (18) that the estimator S_D, considered by Singh et al. (1988), is the most efficient (in the sense of having the smallest minimum MSE) among the estimators s_y^2, I_R, K_{Re}, DT_1, DT_2, DT_3, YK_{Re} and S_D.

Mishra et al. (2024) considered the following family of estimators for S_y^2 of y, given by

$$M_1 = s_y^2 + \eta log\left(\frac{s_x^2}{S_x^2}\right), \tag{19}$$

$$M_2 = s_y^2 (w_1 + 1) + w_2 \eta log\left(\frac{s_x^2}{S_x^2}\right), \tag{20}$$

where η, w_1, and w_2 are appropriately selected constants.

The minimum values of the MSEs of M_1 and M_2 can be obtained by

$$MSE_{min}\left(M_1\right) = S_y^4 \xi (\lambda_{40} - 1)(1 - {\rho^*}^2), \tag{21}$$

$$MSE_{min}\left(M_2\right) = S_y^4 \left[A_0 - \frac{A_2 A_4^2 - 2A_3 A_4 A_5 + A_1 A_5^2}{A_1 A_2 - A_3^2}\right], \tag{22}$$

where $A_0 = \xi(\lambda_{40} - 1)$, $A_1 = 1 + \xi(\lambda_{40} - 1)$, $A_2 = \frac{\xi(\lambda_{04} - 1)}{R^2 S_x^4}$, $A_3 = \frac{\xi}{R S_x^2}\{(\lambda_{22} - 1) - \frac{1}{2}(\lambda_{04} - 1)\}$, $A_4 = \xi(\lambda_{40} - 1)$, $A_5 = \frac{\xi}{R S_x^2}(\lambda_{22} - 1)$.

It can be inferred from (10), (14), and (21) that the estimator M_1 performs comparably to the estimators DT_i (i = 1,2,3) and YK_{Re} i.e.,

$$MSE_{min}\left(DT_i\right) = MSE_{min}\left(YK_{Re}\right) = MSE_{min}\left(M_1\right) \quad \forall i = 1, 2, 3. \tag{23}$$

Thus from (18) and (23), it is evident that S_D exhibits superior performance compared to the estimators DT_i (i=1,2,3), YK_{Re} and M_1, and consequently is superior to s_y^2, I_R and K_{Re}.

From (12) and (22) that $MSE_{min}\left(M_2\right) < MSE_{min}\left(S_D\right)$ if

$$A_0 < \left\{\frac{\xi(\lambda_{40} - 1)(1 - {\rho^*}^2)}{1 + \xi(\lambda_{40} - 1)(1 - {\rho^*}^2)} + \frac{A_2 A_4^2 - 2A_3 A_4 A_5 + A_1 A_5^2}{A_1 A_2 - A_3^2}\right\}. \tag{24}$$

Thus, the estimator M_2 achieves higher efficiency than the enhanced difference estimator S_D provided that condition (24) is satisfied.

4. Development of Proposed Estimators

Motivated by the works of Kumar et al. (2011) and Mishra et al. (2024), we propose two novel classes of efficient estimators for the population variance S_y^2 of y, as follows:

$$SP_1 = w_1 s_y^2 \left(\frac{S_x^2}{s_x^2}\right)^{\alpha} exp\left[\frac{\eta(S_x^2 - s_x^2)}{S_x^2 + s_x^2}\right] + w_2(S_x^2 - s_x^2), \tag{25}$$

$$SP_2 = \delta_1 s_y^2 \left(\frac{S_x^2}{s_x^2}\right)^{\alpha} exp\left[\frac{\eta(S_x^2 - s_x^2)}{S_x^2 + s_x^2}\right] + \delta_2 log\left(\frac{s_x^2}{S_x^2}\right), \tag{26}$$

where (w_1, w_2) and (δ_1, δ_2) are suitable chosen constants selected to minimize the mean squared errors (MSEs) of SP_1 and SP_2, respectively, while α and η are suitably chosen real scalars.

Expressing the estimators SP_1 and SP_2 in terms of (e_0, e_1), we found

$$SP_1 = S_y^2 \left[w_1(1+e_0)(1+e_1)^{-\alpha} exp\left\{\frac{-\eta e_1}{2}\left(1+\frac{e_1}{2}\right)^{-1}\right\} - w_2(1/R)e_1\right], \tag{27}$$

$$SP_2 = S_y^2 \left[\delta_1(1+e_0)(1+e_1)^{-\alpha} exp\left\{\frac{-\eta e_1}{2}\left(1+\frac{e_1}{2}\right)^{-1}\right\} + \delta_2(1/R)log(1+e_1)\right]. \tag{28}$$

Expanding the equations (27) and (28), and neglecting the terms involving powers of e greater than two, we obtain

$$\begin{aligned} SP_1 &\cong S_y^2\left[w_1\left\{1+e_0-\theta e_1(1+e_0)+\frac{\theta(\theta+1)}{2}e_1^2\right\} - w_2(1/R)e_1\right], \\ \Longrightarrow \left(SP_1 - S_y^2\right) &\cong S_y^2\left[w_1\left\{1+e_0-\theta e_1(1+e_0)+\frac{\theta(\theta+1)}{2}e_1^2\right\} - w_2(1/R)e_1 - 1\right], \end{aligned} \tag{29}$$

$$\begin{aligned} SP_2 &\cong S_y^2\left[\delta_1\left\{1+e_0-\theta e_1(1+e_0)+\frac{\theta(\theta+1)}{2}e_1^2\right\} + \delta_2\frac{1}{RS_x^2}\left(e_1 - \frac{1}{2}e_1^2\right)\right], \\ \Longrightarrow \left(SP_2 - S_y^2\right) &\cong S_y^2\left[\delta_1\left\{1+e_0-\theta e_1(1+e_0)+\frac{\theta(\theta+1)}{2}e_1^2\right\} + \delta_2\frac{1}{RS_x^2}\left(e_1 - \frac{1}{2}e_1^2\right) - 1\right], \end{aligned} \tag{30}$$

where $R = (S_y^2/S_x^2)$ and $\theta = \alpha + \frac{\eta}{2}$.

By taking the expectation of Equations (29) and (30), we can obtain the mathematical expressions for the biases of SP_1 and SP_2, which are given below:

$$Bias\left(SP_1\right) = E\left(SP_1 - S_y^2\right) = S_y^2\Big[w_1 B_4 - 1\Big], \tag{31}$$

$$Bias\left(SP_2\right) = E\left(SP_2 - S_y^2\right) = S_y^2\Big[\delta_1 B_4 + \delta_2 B_5 - 1\Big], \tag{32}$$

where $B_4 = \left[1 + \xi\left\{(\lambda_{04}-1)\frac{\theta(\theta+1)}{2} - \theta(\lambda_{22}-1)\right\}\right]$, $B_5 = -\frac{\xi}{2RS_x^2}(\lambda_{04}-1)$.

Squaring both sides of (29) and (30) and omitting terms of e's having power larger than two, we get

$$\begin{aligned} \left(SP_1 - S_y^2\right)^2 \cong S_y^4\Big[1 &+ w_1^2\left\{1 + 2e_0 - 2\theta e_1 + e_0^2 - 4\theta e_0 e_1 + \theta(2\theta+1)e_1^2\right\} \\ &+ w_2^2\frac{1}{R^2}e_1^2 + 2w_1w_2\frac{1}{R}\left(\theta e_1^2 - e_0e_1 - e_1\right) \\ &- 2w_1\left\{1 + e_0 - \theta e_1 - \theta e_0 e_1 + \frac{\theta(\theta+1)}{2}e_1^2\right\} + 2w_2\frac{1}{R}e_1\Big] \end{aligned} \tag{33}$$

$$\left(SP_2 - S_y^2\right)^2 \cong S_y^4\Big[1 + \delta_1^2\left\{1 + 2e_0 - 2\theta e_1 + e_0^2 - 4\theta e_0 e_1 + \theta(2\theta+1)e_1^2\right\} + \delta_2^2\frac{1}{R^2S_x^4}e_1^2 + 2\delta_1\delta_2\frac{1}{RS_x^2}(e_1 + e_0e_1 - \frac{2\theta+1}{2}e_1^2) - 2\delta_1\left\{1 + e_0 - \theta e_1 - \theta e_0 e_1 + \frac{\theta(\theta+1)}{2}e_1^2\right\} - 2\delta_2\frac{1}{RS_x^2}\left(e_1 - \frac{1}{2}e_1^2\right)\Big]. \quad (34)$$

Taking expectation of Equations (33) and (34), we obtain mean squared errors of SP_1 and SP_2 to the first degree of approximation, which are respectively given by

$$MSE\left(SP_1\right) = E\left(SP_1 - S_y^2\right)^2 = S_y^4\Big[1 + w_1^2B_1 + w_2^2B_2 + 2w_1w_2B_3 - 2w_1B_4\Big], \quad (35)$$

$$MSE\left(SP_2\right) = E\left(SP_2 - S_y^2\right)^2 = S_y^4\Big[1 + \delta_1^2B_1 + \delta_2^2B_2^* + 2\delta_1\delta_2B_3^* - 2\delta_1B_4 - 2\delta_2B_5\Big], \quad (36)$$

where $B_1 = \Big[1 + \xi\{(\lambda_{40} - 1) - 4\theta\lambda_{22} + 4\theta + (\lambda_{04} - 1)\theta(2\theta+1)\}\Big]$, $B_2 = \frac{\xi}{R^2}(\lambda_{04} - 1)$, $B_2^* = \frac{\xi}{R^2S_x^4}(\lambda_{04} - 1)$, $B_3 = \frac{\xi}{R}\{(\lambda_{04} - 1)\theta - \lambda_{22} + 1\}$, $B_3^* = \frac{\xi}{RS_x^2}\{(\lambda_{22} - 1) - \frac{(2\theta+1)}{2}(\lambda_{04} - 1)\}$, $B_4 = \Big[1 + \xi\{(\lambda_{04} - 1)\frac{\theta(\theta+1)}{2} - \theta(\lambda_{22} - 1)\}\Big]$, $B_5 = -\frac{\xi}{2RS_x^2}(\lambda_{04} - 1)$.

Setting $\frac{\partial}{\partial w_1}\left(MSE\left(SP_1\right)\right) = 0$ and $\frac{\partial}{\partial w_2}\left(MSE\left(SP_2\right)\right) = 0$, we have

$$\begin{bmatrix} B_1 & B_3 \\ B_3 & B_2 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} B_4 \\ 0 \end{bmatrix}. \quad (37)$$

Solving (37) for (w_1, w_2), we obtain the optimum values of (w_1, w_2) as follows:

$$w_{1_{opt}} = \frac{B_2B_4}{B_1B_2 - B_3^2}, w_{2_{opt}} = -\frac{B_3B_4}{B_1B_2 - B_3^2}. \quad (38)$$

Thus, the resulting minimum MSE of SP_1 is given by

$$MSE_{min}\left(SP_1\right) = S_y^4\left[1 - \frac{B_2B_4^2}{B_1B_2 - B_3^2}\right], \quad (39)$$

which is non negative, if

$$0 < \frac{B_2B_4^2}{B_1B_2 - B_3^2} < 1 \quad and \quad (B_1B_2 - B_3^2) > 0. \quad (40)$$

We are now in a position to state the following theorem.

Theorem 1 *The MSE of SP_1 is always greater than or equal to its minimum MSE, under the non-negativity condition (40):*

$$MSE\left(SP_1\right) \geq MSE_{min}\left(SP_1\right) = S_y^4\left[1 - \frac{B_4^2B_2}{B_1B_2 - B_3^2}\right], \quad (41)$$

with equality holding, if $w_1 = \frac{B_2B_4}{B_1B_2 - B_3^2}$ and $w_2 = -\frac{B_3B_4}{B_1B_2 - B_3^2}$.

Further, setting $\frac{\partial}{\partial \delta_1}\left(MSE\left(SP_2\right)\right) = 0$ and $\frac{\partial}{\partial \delta_2}\left(MSE\left(SP_2\right)\right) = 0$, we get

$$\begin{bmatrix} B_1 & B_3^* \\ B_3^* & B_2^* \end{bmatrix}\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} B_4 \\ B_5 \end{bmatrix}. \quad (42)$$

Simplifying (42) for (δ_1, δ_2), we obtain the optimum values of (δ_1, δ_2) respectively as

$$\delta_{1_{opt}} = \frac{B_2^* B_4 - B_3^* B_5}{B_1 B_2^* - B_3^{*^2}} \quad and \quad \delta_{2_{opt}} = \frac{B_1 B_5 - B_3^* B_4}{B_1 B_2^* - B_3^{*^2}}. \tag{43}$$

Thus the resulting minimum MSE of SP_2 is obtained as follows:

$$MSE_{min}\,(SP_2) = S_y^4 \left[1 - \frac{B_2^* B_4^2 - 2B_3^* B_4 B_5 + B_1 B_5^2}{B_1 B_2^* - B_3^{*^2}}\right], \tag{44}$$

which is non negative if

$$0 < \frac{B_2^* B_4^2 - 2B_3^* B_4 B_5 + B_1 B_5^2}{B_1 B_2^* - B_3^{*^2}} < 1 \quad and \quad (B_1 B_2^* - B_3^{*^2}) > 0. \tag{45}$$

We now present the key theorem.

Theorem 2 *The MSE of SP_2 is always larger than or equal to the minimum MSE of SP_2, under the non-negative condition (45):*

$$MSE\,(SP_2) > MSE_{min}\,(SP_2) = S_y^4 \left[1 - \frac{B_2^* B_4^2 - 2B_3^* B_4 B_5 + B_1 B_5^2}{B_1 B_2^* - B_3^{*^2}}\right],$$

with equality holding, if

$$\delta_1 = \frac{B_2^* B_4 - B_3^* B_5}{B_1 B_2^* - B_3^{*^2}} \quad and \quad \delta_2 = \frac{B_1 B_5 - B_3^* B_4}{B_1 B_2^* - B_3^{*^2}}.$$

5. Theoretical Efficiency Comparisons

From Section 3, it is evident that the difference estimator S_D is superior among the estimators s_y^2, I_R, K_{Re}, DT_i (i = 1, 2, 3), YK_{Re} and M_1. Hence, the performance of the considered estimators SP_1 and SP_2 is assessed in comparison with S_D and M_2.

- From (12) and (39), we have that $MSE_{min}\,(SP_1) < MSE_{min}\,(S_D)$ if

$$\frac{1}{1 + \xi(\lambda_{40} - 1)(1 - \rho^{*^2})} < \frac{B_2 B_4^2}{B_1 B_2 - B_3^2}. \tag{46}$$

- From (22) and (39), we have that $MSE_{min}\,(SP_1) < MSE_{min}\,(M_2)$ if

$$\left[\frac{A_2 A_4^2 - 2A_3 A_4 A_5 + A_1 A_5^2}{A_1 A_2 - A_3^2} + 1\right] < \left[A_0 + \frac{B_2 B_4^2}{B_1 B_2 - B_3^2}\right]. \tag{47}$$

- From (12) and (44), we note that $MSE_{min}\,(SP_2) < MSE_{min}\,(S_D)$ if

$$\frac{1}{1 + \xi(\lambda_{40} - 1)(1 - \rho^{*^2})} < \frac{B_2^* B_4^2 - 2B_3^* B_4 B_5 + B_1 B_5^2}{B_1 B_2^* - B_3^{*^2}}. \tag{48}$$

- From (22) and (44), it is observed that $MSE_{min}\,(SP_2) < MSE_{min}\,(M_2)$ if

$$\left[\frac{A_2 A_4^2 - 2A_3 A_4 A_5 + A_1 A_5^2}{A_1 A_2 - A_3^2} + 1\right] < \left[A_0 + \frac{B_2^* B_4^2 - 2B_3^* B_4 B_5 + B_1 B_5^2}{B_1 B_2^* - B_3^{*^2}}\right]. \tag{49}$$

It follows from (46) and (47) that the proposed estimator SP_1 is more superior in efficiency to the estimators S_D and M_2, respectively, provided these conditions are satisfied. Similarly, the proposed estimator SP_2 is also superior in efficiency to the estimators S_D and M_2 under the conditions (48) and (49).

6. Empirical Analysis

To assess the performance of the proposed estimators SP_1 and SP_2 in comparison with other existing estimators, an empirical investigation is carried out using three real-life data sets. The descriptive statistics of the data sets used in the analysis are given below:

Data Set I Source: Cochran (1977), p.325)

y (study variable): Number of persons per block.

x (auxiliary variable): Number of rooms per block.

N = 100, n = 10, $S_y^2 = 214.69$, $S_x^2 = 56.76$, $\lambda_{40} = 2.2387$, $\lambda_{04} = 2.2523$, $\lambda_{22} = 1.5432$, ρ (Correlation coefficient between x and y) = 0.6515.

Data Set II Source: Yasser H. (2022)

y : the Body Mass Index (BMI) of an individual based on gender.

x : the weight of an individual (according to gender) measured in kilograms.

N = 500, n = 75, $S_y^2 = 1.8362$, $S_x^2 = 1048.633$, $\lambda_{40} = 2.961475$, $\lambda_{04} = 1.802638$, $\lambda_{22} = 1.680199$, ρ = 0.8046.

Data Set III Source: Directorate of Economics and Statistics, Department of Agriculture & Farmers Welfare, Government of India (2024)

y (study variable): Rice production (in million tonnes) in All India during the period 195051 to 202425.

x (auxiliary variable): Area under rice cultivation (in million hectares) in All India during the period 195051 to 202425.

N = 75, n = 12, $S_y^2 = 1094.227$, $S_x^2 = 22.5625$, $\lambda_{40} = 2.20207$, $\lambda_{04} = 2.595149$, $\lambda_{22} = 1.9858$, ρ = 0.921167.

The PRE of different estimators with respect to the usual unbiased variance estimator S_y^2 is obtained using the following expression:

$$PRE(T^*, s_y^2) = \frac{MSE(s_y^2)}{MSE_{min}(T^*)} \times 100,$$

where $T^* = s_y^2$, I_R, K_{Re}, $DT_i (i = 1, 2, 3)$, S_D, YK_{Re}, M_1, M_2, SP_1 and SP_2.

The percent relative efficiencies (PREs) of the existing estimators s_y^2, I_R, K_{Re}, $DT_i (i = 1, 2, 3)$, S_D, YK_{Re}, M_1 and M_2 w.r.t. the usual unbiased variance estimator s_y^2 are presented in Table 1, whereas $PRE(SP_1, s_y^2)$ and $PRE(SP_2, s_y^2)$ for different choices of (α, η) are reported in Table 2.

Table 1 PREs of the estimators s_y^2, I_R, K_{Re}, $DT_i (i = 1, 2, 3)$, S_D, YK_{Re}, M_1 and M_2 w.r.t. unbiased variance estimator s_y^2 for different data sets

Estiamator	s_y^2	I_R	K_{Re}	$DT_i (i = 1, 2, 3)$	YK_{Re}	M_1	S_D	M_2
				Data Set I				
$PRE(., s_y^2)$	100.000	88.188	122.817	123.489	123.489	123.489	134.638	142.141
				Data Set II				
$PRE(., s_y^2)$	100.000	139.735	132.359	141.619	141.619	141.619	143.842	145.079
				Data Set III				
$PRE(., s_y^2)$	100.000	145.596	195.441	202.762	202.762	202.762	211.177	232.678

From Tables 1 - 2 and Figure 1, it is found that the suggested estimators SP_1 and SP_2 are more efficient than the usual unbiased variance estimator s_y^2, the estimator I_R due to Isaki (1983), the estimator K_{Re} proposed by Kumar et al. (2011), the estimators $DT_i (i = 1, 2, 3)$ due to Das and Tripathi (1978), the estimator S_D proposed by Singh et al. (1986), the estimator YK_{Re} due to Yadav and Kadilar (2013) and the estimators M_1 and M_2 considered by Mishra et al. (2024). The proposed estimators show substantial gains in efficiency for suitably chosen values of the real scalars (α, η).

For (α, η) = (2, 0), (1.5, 1), (1, 2), the proposed estimators SP_1 yields largest PRE of 538.689%, followed by the proposed estimator SP_2, which achieves a PRE of 235.765% for Data Set I.

Table 2 PREs of the proposed estimators SP_1 and SP_2 with respect to unbiased variance estimator s_y^2 for various values of (α, η) across data sets

		Data Set I		Data Set II		Data Set III	
α	η	$PRE(SP_1, s_y^2)$	$PRE(SP_2, s_y^2)$	$PRE(SP_1, s_y^2)$	$PRE(SP_2, s_y^2)$	$PRE(SP_1, s_y^2)$	$PRE(SP_2, s_y^2)$
2	0	538.689	235.765	148.538	146.139	4469.569	405.800
	-0.5	291.943	190.235	146.880	145.237	573.602	300.172
	-1	213.821	166.263	145.678	144.623	348.091	255.319
	-1.5	177.317	152.614	144.828	144.239	272.791	233.506
1.5	1	538.689	235.765	148.538	146.139	4469.569	405.800
	0.5	291.943	190.235	146.880	145.237	573.602	300.172
	0	213.821	166.263	145.678	144.623	348.091	255.319
	-0.5	177.317	152.614	144.828	144.239	272.791	233.506
	-1	157.506	144.751	144.253	144.048	238.950	222.828
1	2	538.689	235.765	148.538	146.139	4469.569	405.800
	1.5	291.943	190.235	146.880	145.237	573.602	300.172
	1	213.821	166.263	145.678	144.623	348.091	255.319
	0.5	177.317	152.614	144.828	144.239	272.791	233.506
	0	157.506	144.751	144.253	144.048	238.950	222.828
0.5	2	213.821	166.263	145.678	144.623	348.091	255.319
	1.5	177.317	152.614	144.828	144.239	272.791	233.506
	1	157.506	144.751	144.253	144.048	238.950	222.828
0	2	157.506	144.751	144.253	144.048	238.950	222.828
	-1.5	142.157	175.808	145.463	148.721	237.682	396.616
	-2	150.702	207.743	146.580	150.922	270.160	749.761
-0.5	-0.5	142.157	175.808	145.463	148.721	237.682	396.616
	-1	150.702	207.743	146.580	150.922	270.160	749.761
-1	0.5	142.157	175.808	145.463	148.721	237.682	396.616
	0	150.702	207.743	146.580	150.922	270.160	749.761
-1.5	1.5	142.157	175.808	145.463	148.721	237.682	396.616
	1	150.702	207.743	146.580	150.922	270.160	749.761
-2	2	150.702	207.743	146.580	150.922	270.160	749.761

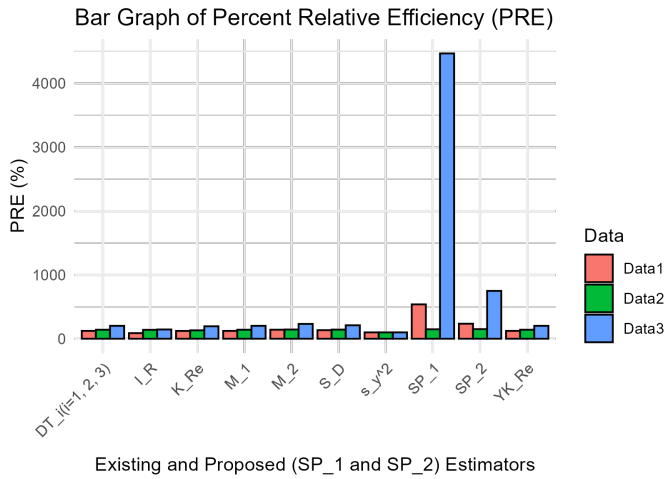


Figure 1 PREs of existing and proposed (SP_1 & SP_2) estimators for various data sets

For (α, η) = (0, -2), (-0.5, -1), (-1, 0), (-1.5, 1), (-2, 2), the proposed estimators SP_2 yields largest PRE of 150.922% for Data Set II. This is followed by the proposed estimator SP_1, which achieves a PRE of 148.538% at (α, η) = (2, 0), (1.5, 1), (1, 2).

Similarly, for Data Set III, the proposed estimators SP_1 yields largest PRE value of 4469.569% for (α, η) = (2, 0), (1.5, 1), (1, 2). This is followed by the proposed estimator SP_2, which achieves a PRE of 749.761% at (α, η) = (0, -2), (-0.5, -1), (-1, 0), (-1.5, 1), (-2, 2).

Table 2 clearly indicates that there is enough scope of selecting the values of scalars (α, η) for obtaining estimators from the suggested estimators SP_1 and SP_2 better than the existing estimators s_y^2, I_R, K_{Re}, $DT_i (i = 1, 2, 3)$, S_D, YK_{Re}, M_1 and M_2. Hence, it may be concluded that the suggested classes of estimators SP_1 and SP_2 are highly efficient and suitable for practical applications.

7. Conclusions

This research investigates the estimation of the population variance S_y^2 under the SRSWOR scheme in sample surveys. Two classes of estimators of S_y^2 of y utilizing information on S_x^2 of x have been proposed. The bias, MSE and minimum MSE of the suggested classes of estimators have been derived mathematically up to the approximation of first order. Conditions under which the proposed classes of estimators SP_1 and SP_2 outperform the existing estimators have also been established. An empirical study has been conducted to validate the proposed methodology. The empirical results reveal that the proposed classes of estimators SP_1 and SP_2 exhibit superior performance compared to the other existing estimators, with appreciable gains in efficiency. Hence, the proposed estimators SP_1 and SP_2 may be recommended for practical use in population variance estimation.

The scope of the present study may be extended to situations involving non-response and other complex sampling schemes such as stratified, systematic, and two-phase sampling designs. Further investigation may also be carried out to assess the performance of the proposed estimators under different practical survey conditions.

Acknowledgements

The authors would like to express heartfelt appreciation to the Editor-in-Chief and the anonymous reviewers for their valuable comments and insightful suggestions, which greatly enhanced the quality and clarity of this study.

References

Bhushan S, Kumar A, Alsubie A, Lone SA. Variance estimation under an efficient class of estimators in simple random sampling. Ain Shams Eng. J. 2023; 14(8): 102–012.

Biradar RS, Singh HP. Predictive estimation of finite population variance. Calcutta Stat Assoc Bull. 1998; 48 (3 - 4): 229–236.

Cochran WG. Sampling Techniques (3rd edition). New York: John Wiley and Sons; 1977.

Daraz U, Wu J, Agustiana D, Emam W. Finite population variance estimation using Monte Carlo simulation and real life application. Symmetry. 2025; 17(1): 84.

Das AK, Tripathi TP. Use of auxiliary information in estimating the finite population variance. Sankhy Ser C. 1978; 40(2): 139-148.

Directorate of Economics and Statistics, Department of Agriculture & Farmers Welfare, Government of India. Agricultural statistics at a glance 2024; https://www.agriwelfare.gov.in

Djebar A, Alghamdi AS, Mustafa MS, Ahmad S. Computation of population variance estimation in simple random sampling structures by developing generalized estimator. Mathematics. 2026; 14(2): 375.

Isaki CT. Variance estimation using auxiliary information. J Am Stat Assoc. 1983; 78(381); 117–123.

Kumar M, Singh R, Sawan N, Chauhan P. Exponential ratio method of estimation in the presence of measurement errors. Int J Agric Stat Sci. 2011;7(2):457-461.

Mishra P, Sharma A, Adichwal NK, Rai S, Singh R. Log-product-type estimator for estimation of population variance using auxiliary information. Thail Stat. 2024; 22(3): 610-617.

Mukhopadhyay P. Optimum strategies for estimating the variance of a finite population under a super population model. Metrika. 1982; 29: 143-158.

Niaz I, Sanaullah A, Saleem I, Shabbir J. An improved efficient class of estimators for the population variance. Concurr Comput Pract Exp. 2022; 34(4): e6620. doi:10.1002/cpe.6620.

Sharma P, Singh R. Improved dual to variance ratio type estimation for population variance. Chil J Stat. 2014; 5(2): 45-54.

Singh HP, Prasad S. Hybrid estimators for finite population variance: combining auxiliary data with traditional methods. Lobachevskii J Math. 2025; 46(4): 1716-1731.

Singh HP, Solanki RS. A new procedure for variance estimation in simple random sampling using auxiliary information. Stat Pap. 2013; 54(2): 479-497.

Singh HP, Upadhyaya LN, Namjoshi UD. Estimation of finite population variance. Curr Sci. 1988; 57(24): 1331-1334.

Swain AKPC, Dihidar K, Mitra A. On estimating the finite population variance by utilizing auxiliary information. Commun Stat Theory. 2026; 1-32.

Wu CF. Estimation of variance of the ratio estimator. Biometrika. 1982; 69(1): 183-189.

Yadav SK, Kadilar C. Improved exponential type ratio estimator of population variance. Rev Colomb Estad. 2013; 36(1): 145-152.

Yadav SK, Kadilar C, Shabbir J, Gupta S. Improved family of estimators of population variance in simple random sampling. J Stat Theory Pract. 2015; 9: 219-226.

Yadav VK, Prasad S. Some exponential estimators in sample survey using robust regression method in the presence of outliers. Lobachevskii J Math. 2024; 45(4): 1674-1690.

Yasser H. BMI dataset. 2022; https://www.kaggle.com/datasets/yasserh/bmidataset/data