AIMS Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 9(11): 33139–33160.
DOI: 10.3934/math.20241582
Received: 17 August 2024
Revised: 30 October 2024
Accepted: 01 November 2024
Published: 21 November 2024

Research article

Estimation techniques utilizing dual auxiliary variables in stratified two-phase sampling

Olayan Albalawi*

Department of Statistics, Faculty of Science, University of Tabuk, Tabuk, Saudi Arabia

* **Correspondence:** Email: oalbalwi@ut.edu.sa.

Abstract: In this research paper, an improved set of estimators for finding the finite population variance of a study variable under a stratified two-phase sampling design is introduced. These estimators rely on information about extreme values and the ranks of an auxiliary variable. We examined the properties of these estimators using first-order approximation, focusing on biases and mean squared errors (MSEs). Additionally, we conducted an extensive simulation study to evaluate their performance and validate our theoretical insights. Furthermore, in the application section, we employed some datasets to further assess the performances of our estimators as compared to other existing estimators. The results demonstrated that $S^2_{Q_2}$ was the best-performing estimator, and significantly outperforms existing estimators, achieving a percent relative efficiency (PRE) in the exponential distribution as high as 385.467. The percent relative efficiency values were continuously higher than 100 in a variety of situations, with values as high as 353.129 in other distributions like the uniform and gamma. The suggested estimators are superior to the conventional estimators, as demonstrated by empirical assessments using datasets, where percent relative efficiency improvements ranged from 115.026 to 139.897. These results highlight the robustness and applicability of the proposed class of estimators in real-world sampling.

Keywords: stratified two-phase sampling; exponential estimators; variance estimation; outliers; ranks; percent relative efficiency
Mathematics Subject Classification: 62D

1. Introduction

It is standard procedure in sampling theory to include auxiliary variables with the study variable in order to improve design and increase the efficiency of the estimator by utilizing their relationship. Although information about auxiliary variables is sometimes unavailable in practical circumstances prior to conducting a survey, in such instances, a two-phase sampling procedure is preferable. Two steps are used in two-phase sampling, sometimes referred to as double sampling, to choose a sample from a population. Since two-phase sampling is an economical sampling strategy, it is frequently employed in sample surveys when supplementary data is not available ahead of time. A brief summary of two-phase sampling was initially introduced by [1]. Such works were not explored further after that, until the works of [2]. Due to its low-cost variable screening qualities, two-phase sampling has received a lot of interest in recent years. For estimation of finite population mean under two-phase sampling schemes, different estimators proposed by [3–5]. In order to estimate the finite population variance, different estimators suggested by [6–8]. For more information, see [9–11] and references therein.

Since variation occurs naturally, estimating finite population variance is a serious problem. The utilization of auxiliary information to estimate the population variance was initially introduced by [12] and then expanded upon by [13]. Employing supplementary information in an informed strategy can improve estimators performance. In order to determine the population variance, [14] proposed exponential estimators based on ratios and products. By using the different transformations, [15–17] introduced some new estimators to improve the variance estimation. Under simple random sampling and stratified random sampling, different families of estimators obtained by [18–20]. For more details about different estimators and methods for estimating the finite population variances, we refer to [21–23].

In the sample survey data, there may be unusual observations. When the sample contains outlier values, the results may be distorted. In this regard, several researchers have focused on outlier values and presented various methods to estimate population characteristics. The researchers in [24] used a linear transformation to obtain two estimators based on the auxiliary minimum and maximum observations. After that, these works were not investigated until [25]. The researchers employed numerous finite population mean estimators, as well as the concept of using extreme values in them. For calculating the finite population mean, [26, 27] introduced different transformations methods to handle the outliers. [28] used stratified random sampling to improve the estimate of the limited population mean under extreme values. [29] provided effective estimators for estimating population variance using extreme value transformations. The work [30] proposed novel estimators that use extreme values to estimate population variance with the least mean squared errors (MSE). [31] proposed double exponential ratio estimators that use extreme values of the auxiliary variable to evaluate their effectiveness in estimating population variance. To improve estimator accuracy, [32] developed efficient estimators that leverage auxiliary variables under simple random sampling. For further information, readers can read [33, 34].

Several important considerations motivated the development of a new method to estimate the finite population variance:

- Traditional estimators for finite population variance frequently neglect extreme values (outliers) and rankings of auxiliary variables. Outliers are often considered challenging, resulting in skewed conclusions or inflated MSE. The inefficiency of stratified two-phase sampling designs highlights the need for a more efficient approach that addresses these problems.
- Existing estimators often struggle with stratified two-phase sampling due to its complicated data structure. These issues emphasize the need for more robust and efficient estimators.
- In most cases, two-phase sampling is more economical than one-phase sample, particularly when

dealing with large populations. It lowers total expenditures by enabling researchers to gather preliminary data with a smaller sample before selecting a second sample.

- Two-phase sampling enables researchers to choose certain clusters or strata that reflect the whole population, it helps guarantee that varied sub-populations are effectively represented.
- Two-phase sampling is a useful method in a variety of research situations because it offers more accurate representation and control of variability along with cost savings, enhanced precision, and flexibility.

In this article, our main objective is to properly utilize the information about the outlier values of the auxiliary variable, which are used as supplementary information to increase the accuracy of the proposed class of estimators. It is well known that outlier information is often removed from sample data, and therefore classical estimators generally decrease its significance as MSE increases. When there is a relationship between the two variables, the ranks of the auxiliary variable are linked to the study variable. Consequently, these rankings provide a useful tool to improve the accuracy of the estimators. We apply the transformations technique, motivated by [29–32], to provide a new class of estimators using the ranks of the auxiliary variable and the known information on the outlier values to estimate the finite population variance in two-phase stratified sampling. The new suggested method is particularly useful in economic surveys, public health examinations, and environmental evaluations, where similar sample strategies are often used. The new estimators are ideal for disciplines like market research and agricultural surveys that frequently meet extreme values, as they can efficiently include outlier information without distorting results.

This article is organized as follows: In Section 2, we introduce the foundational concepts and notations. In Section 3, we discuss various established estimators. Our proposed class of estimators is detailed in Section 4. A theoretical comparison is presented in Section 5. In Section 6, we conduct simulations on six different artificial populations with varying probability distributions to evaluate the theoretical results discussed in Section 5. This section also provides numerical examples to validate our theoretical findings. Finally, in Section7, we offer a discussion of the results and suggestions for future research.

2. Concepts and notations

Let us consider a finite population

$$\phi = (\phi_1, \phi_2, \ldots, \phi_N)$$

of size N units. This population is divided into L strata, each of which is $N_h (h = 1, 2, \ldots, L)$, with the property that

$$\sum_{h=1}^{L} N_h = N.$$

Let y_{hi}, x_{hi}, and r_{hi} be the values of the study variable (Y), the auxiliary variable (X), and the ranks of the auxiliary variable R in the hth stratum of the $ith(i = 1, 2, \ldots, N_h)$ unit, respectively. We define the population variances for these variables in the hth stratum as

$$S_{yh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(Y_{hi} - \bar{Y}_h\right)^2,$$

$$S_{xh}^{2} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(X_{hi} - \bar{X}_h\right)^2,$$
$$S_{rh}^{2} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(R_{hi} - \bar{R}_h\right)^2,$$

where

$$\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} Y_{hi},$$
$$\bar{X}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} X_{hi}$$

and

$$\bar{R}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} R_{hi}$$

denote the population means of the study variable (Y), auxiliary variable (X), and the ranks of the auxiliary variable (R) in the *hth* stratum that corresponding to the population means

$$\bar{Y} = \frac{1}{N_h} \sum_{h=1}^{L} W_h \bar{Y}_h,$$
$$\bar{X} = \frac{1}{N_h} \sum_{h=1}^{L} W_h \bar{X}_h,$$
$$\bar{R} = \frac{1}{N_h} \sum_{h=1}^{L} W_h \bar{R}_h,$$

respectively, where W_h is the stratum weight and defined by

$$W_h = \frac{N_h}{N}.$$

The population coefficients of variations in the *hth* stratum, are defined as

$$C_{yh} = \frac{S_{yh}}{\bar{Y}_h},$$
$$C_{xh} = \frac{S_{xh}}{\bar{X}_h}$$

and

$$C_{rh} = \frac{S_{rh}}{\bar{R}_h}$$

where S_{yh}, S_{xh}, and S_{rh} are the population standard deviations of (Y, X, R) in the *hth* stratum, respectively.

Furthermore, define the population correlation coefficients between (Y, X), (Y, R), and (X, R) in the *hth* stratum as follows:

$$\rho_{yxh} = \frac{S_{yxh}}{S_{yh} S_{xh}},$$
$$\rho_{yrh} = \frac{S_{yrh}}{S_{yh} S_{rh}},$$
$$\rho_{xrh} = \frac{S_{xrh}}{S_{xh} S_{rh}},$$

where

$$S_{yxh} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(Y_{hi} - \bar{Y}_h\right)\left(X_{hi} - \bar{X}_h\right),$$
$$S_{yrh} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(Y_{hi} - \bar{Y}_h\right)\left(R_{hi} - \bar{R}_h\right)$$

and

$$S_{xrh} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(x_{hi} - \bar{X}_h\right)\left(R_{hi} - \bar{R}_h\right),$$

are the population co-variances, respectively.

In this paper, we provide a set of estimators to estimate the finite population variance S_y^2 of Y in the presence of the auxiliary variable X. The definition of the two-phase sampling scheme is:

(1) A sample of size $(\acute{n}_h < N_h)$ from the first phase is chosen in order to estimate the population variance S_{xh}^2.
(2) For the second phase, a sample size of $(n_h < \acute{n}_h)$ is chosen in order to observe both y and x, respectively.

We define the following concepts in order to calculate the biases and mean square errors for different estimators:

$$\xi_{0h} = \left(\frac{s_{yh}^2 - S_{yh}^2}{S_{yh}^2}\right), \quad \xi_{1h} = \left(\frac{s_{xh}^2 - S_{xh}^2}{S_{xh}^2}\right), \quad \xi_{2h} = \left(\frac{\acute{s}_{xh}^2 - S_{xh}^2}{S_{xh}^2}\right), \quad \xi_{3h} = \left(\frac{s_{rh}^2 - S_{rh}^2}{S_{rh}^2}\right), \quad \xi_{4h} = \left(\frac{\acute{s}_{rh}^2 - S_{rh}^2}{S_{rh}^2}\right),$$

such that

$$E\left(\xi_{ih}\right) = 0$$

for $i = 0, 1, 2, 3, 4$.

$$E\left(\xi_{0h}^2\right) = \eta_h \Delta_{400h}^*, \qquad E\left(\xi_{1h}^2\right) = \eta_h \Delta_{040h}^*, \qquad E\left(\xi_{2h}^2\right) = \eta_h' \Delta_{040h}^*,$$
$$E\left(\xi_{3h}^2\right) = \eta_h \Delta_{004h}^*, \qquad E\left(\xi_{4h}^2\right) = \eta_h' \Delta_{004h}^*, \qquad E\left(\xi_{0h}\xi_{1h}\right) = \eta_h \Delta_{220h}^*,$$
$$E\left(\xi_{0h}\xi_{2h}\right) = \eta_h' \Delta_{220h}^*, \qquad E\left(\xi_{0h}\xi_{3h}\right) = \eta_h \Delta_{202h}^*, \qquad E\left(\xi_{0h}\xi_{4h}\right) = \eta_h' \Delta_{202h}^*,$$
$$E\left(\xi_{1h}\xi_{2h}\right) = \eta_h' \Delta_{040h}^*, \qquad E\left(\xi_{1h}\xi_{3h}\right) = \eta_h \Delta_{022h}^*, \qquad E\left(\xi_{1h}\xi_{4h}\right) = \eta_h' \Delta_{022h}^*,$$
$$E\left(\xi_{2h}\xi_{3h}\right) = \eta_h' \Delta_{022h}^*, \qquad E\left(\xi_{2h}\xi_{4h}\right) = \eta_h' \Delta_{022h}^*, \qquad E\left(\xi_{3h}\xi_{4h}\right) = \eta_h' \Delta_{004h}^*,$$

where

$$\Delta^*_{400h} = (\Delta_{400h} - 1),\ \ \Delta^*_{040h} = (\Delta_{040h} - 1),\ \ \Delta^*_{004h} = (\Delta_{004h} - 1),\ \ \Delta^*_{220h} = (\Delta_{220h} - 1),$$
$$\Delta^*_{202h} = (\Delta_{202h} - 1),\ \ \Delta^*_{022h} = (\Delta_{022h} - 1),\ \ \eta_h = \left(\frac{1}{n_h} - \frac{1}{N_h}\right),\ \ \eta'_h = \left(\frac{1}{n'_h} - \frac{1}{N_h}\right),\ \ \eta''_h = \left(\frac{1}{n_h} - \frac{1}{n'_h}\right).$$

Also

$$\Delta_{lqsh} = \frac{\varphi_{lqsh}}{\varphi^{l/2}_{200h}\varphi^{q/2}_{020h}\varphi^{s/2}_{002h}},$$

where

$$\varphi_{lqsh} = \frac{\sum_{i=1}^{N_h}\left(Y_{hi} - \bar{Y}_h\right)^l \left(X_{hi} - \bar{X}_h\right)^q \left(R_{hi} - \bar{R}_h\right)^s}{N_h - 1}.$$

Here,

$$\Delta_{400h} = \beta_{2(yh)},\quad \Delta_{040h} = \beta_{2(xh)},\quad \text{and}\quad \Delta_{004h} = \beta_{2(rh)}$$

are the population coefficients of kurtosis.

3. Literature review

Next, we review the other estimators of the finite population variances while comparing them with the estimators in our proposed class.

The variance of the usual estimator

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$$

in stratified random sampling is defined as follows:

$$Var(\bar{y}_{st}) = \sum_{h=1}^{L} \eta_h W_h^2 S^2_{yh} = S^2_{yst}.$$

The unbiased estimator $\hat{S}^2_{T_1}$ of S^2_{yst}, is defined as

$$\hat{S}^2_{T_1} = \sum_{h=1}^{L} \eta_h W_h^2 s^2_{yh}.$$

The usual variance estimator of $\hat{S}^2_{T_1}$ for population variance is given by

$$Var(\hat{S}^2_{T_1}) = \sum_{h=1}^{L} \eta_h^3 W_h^4 S^4_{yh} \Delta^*_{400h}. \tag{3.1}$$

A ratio estimator for population variance $\hat{S}^2_{T_2}$, proposed by [13], is given by

$$\hat{S}^2_{T_2} = \sum_{h=1}^{L} \eta_h W_h^2 s^2_{yh} \left(\frac{s'^2_{xh}}{s^2_{xh}}\right). \tag{3.2}$$

The following equations represent the bias and MSE of $\hat{S}^2_{T_2}$;

$$Bias\left(\hat{S}^2_{T_2}\right) \cong \sum_{h=1}^{L} \eta''^2_h W_h^2 S^2_{yh} \left(\Delta^*_{040h} - \Delta^*_{220h}\right) \tag{3.3}$$

and

$$MSE\left(\hat{S}^2_{T_2}\right) \cong \sum_{h=1}^{L} W_h^4 S^4_{yh} \left(\eta^3_h \Delta^*_{400h} + \eta''^3_h \Delta^*_{040h} - 2\eta''^3_h \Delta^*_{220h}\right). \tag{3.4}$$

According to [35], the linear regression estimator $\hat{S}^2_{T_3}$, is defined as

$$\hat{S}^2_{T_3} = \sum_{h=1}^{L} \eta_h W_h^2 \left[s^2_{yh} + b_{(s^2_{yh}, s^2_{xh})}\left(\acute{s}^2_{xh} - s^2_{xh}\right)\right], \tag{3.5}$$

where

$$b_{(s^2_{yh}, s^2_{xh})} = \frac{s^2_{yh} \hat{\Delta}^*_{220h}}{s^2_{xh} \hat{\Delta}^*_{040h}}$$

is the sample regression coefficient.

The following equation represents a MSE of $\hat{S}^2_{T_3}$;

$$MSE\left(\hat{S}^2_{T_3}\right) \cong \sum_{h=1}^{L} S^4_{yh} W_h^4 \Delta^*_{400h} \left(\eta^3_h - \eta''^3_h \rho^{*2}_{yxh}\right), \tag{3.6}$$

where

$$\rho^*_{yxh} = \frac{\Delta^*_{220h}}{\sqrt{\Delta^*_{400h}}\sqrt{\Delta^*_{040h}}}.$$

An exponential ratio type estimator $\hat{S}^2_{T_4}$, presented by [14], is defined as follows

$$\hat{S}^2_{T_4} = \sum_{h=1}^{L} \eta_h W_h^2 s^2_{yh} \exp\left(\frac{\acute{s}^2_{xh} - s^2_{xh}}{\acute{s}^2_{xh} + s^2_{xh}}\right). \tag{3.7}$$

The following equations represent the bias and MSE of $\hat{S}^2_{T_4}$;

$$Bias\left(\hat{S}^2_{T_4}\right) \cong \frac{1}{2} \sum_{h=1}^{L} \eta''^2_h W_h^2 S^2_{yh} \left(\frac{3\Delta^*_{040h}}{4} - \Delta^*_{220h}\right) \tag{3.8}$$

and

$$MSE\left(\hat{S}^2_{T_4}\right) \cong \sum_{h=1}^{L} W_h^4 S^4_{yh} \left[\eta^3_h \Delta^*_{400h} + \eta''^3_h \left(\frac{\Delta^*_{040h}}{4} - \Delta^*_{220h}\right)\right]. \tag{3.9}$$

By employing the kurtosis of an auxiliary variable, [15] proposed a ratio-type estimator $\hat{S}^2_{T_5}$, is defined as

$$\hat{S}^2_{T_5} = \sum_{h=1}^{L} \eta_h W_h^2 s^2_{yh} \left(\frac{\acute{s}^2_{xh} + \Delta_{040h}}{s^2_{xh} + \Delta_{040h}}\right). \tag{3.10}$$

The following equations represent the bias and MSE of $\hat{S}^2_{T_5}$;

$$Bias\left(\hat{S}^2_{T_5}\right) \cong \sum_{h=1}^{L} \eta_h''^2 g_h W_h^2 S^2_{yh} \left(g_h \Delta^*_{040h} - \Delta^*_{220h}\right) \tag{3.11}$$

and

$$MSE\left(\hat{S}^2_{T_5}\right) \cong \sum_{h=1}^{L} W_h^4 S^4_{yh} \left[\eta_h^3 \Delta^*_{400h} + \eta_h''^3 \left(g_h^2 \Delta^*_{040h} - 2 g_h \Delta^*_{220h}\right)\right], \tag{3.12}$$

where

$$g_h = \frac{S^2_{xh}}{S^2_{xh} + \Delta_{040h}}.$$

The classifications of some ratio estimators is given in [17], which are defined as

$$\hat{S}^2_{T_6} = \sum_{h=1}^{L} \eta_h W_h^2 s^2_{yh} \left(\frac{\hat{s}^2_{xh} + C_{xh}}{s^2_{xh} + C_{xh}}\right), \tag{3.13}$$

$$\hat{S}^2_{T_7} = \sum_{h=1}^{L} \eta_h W_h^2 s^2_{yh} \left(\frac{\Delta_{040h} \hat{s}^2_{xh} + C_{xh}}{\Delta_{040h} s^2_{xh} + C_{xh}}\right) \tag{3.14}$$

and

$$\hat{S}^2_{T_8} = \sum_{h=1}^{L} \eta_h W_h^2 s^2_{yh} \left(\frac{C_{xh} \hat{s}^2_{xh} + \Delta_{040h}}{C_{xh} s^2_{xh} + \Delta_{040h}}\right). \tag{3.15}$$

The following equations represent the bias and MSE of $\hat{S}^2_{T_i}$;

$$Bias\left(\hat{S}^2_{T_i}\right) \cong \sum_{h=1}^{L} \eta_h''^2 t_{ih} W_h^2 S^2_{yh} \left(t_{ih} \Delta^*_{040h} - \Delta^*_{220h}\right) \tag{3.16}$$

and

$$MSE\left(\hat{S}^2_{T_i}\right) \cong \sum_{h=1}^{L} W_h^4 S^4_{yh} \left[\eta_h^3 \Delta^*_{400h} + \eta_h''^3 \left(t_{ih}^2 \Delta^*_{040h} - 2 t_{ih} \Delta^*_{220h}\right)\right], \tag{3.17}$$

where

$$t_{1h} = \frac{S^2_{xh}}{S^2_{xh} + C_{xh}}, \quad t_{2h} = \frac{\Delta_{040h} S^2_{xh}}{\Delta_{040h} S^2_{xh} + C_{xh}}, \quad \text{and} \quad t_{3h} = \frac{C_{xh} S^2_{xh}}{C_{xh} S^2_{xh} + \Delta_{040h}}.$$

4. Proposed class of estimators

In this section, we present an improved class of estimators inspired by prior works [29–32]. These estimators employ minimum and maximum values of auxiliary variables, along with their ranks, in two-phase sampling to estimate the variance of the finite population. The suggested estimator is defined as

$$\hat{S}^2_Q = \sum_{h=1}^{L} \eta_h W_h^2 s^2_{yh} \exp\left[\theta_{1h} \left\{\frac{\gamma_{1h}\left(\hat{s}^2_{xh} - s^2_{xh}\right)}{\gamma_{1h}\left(\hat{s}^2_{h} + s^2_{xh}\right) + 2\gamma_{2h}}\right\}\right] \exp\left[\theta_{2h} \left\{\frac{\gamma_{3h}\left(\hat{s}^2_{rh} - s^2_{rh}\right)}{\gamma_{3h}\left(\hat{s}^2_{rh} + s^2_{rh}\right) + 2\gamma_{4h}}\right\}\right], \tag{4.1}$$

where $(\theta_{ih}, i = 1, 2)$ are known constants values either (1 or 2), and $(\gamma_{ih}, i = 1, 2, 3, 4)$ are the parameters of the auxiliary variables. The minimum and maximum values (outliers) of the auxiliary variable are denoted by (x_{mh}, x_{Mh}), while the minimum and maximum values (outliers) of the ranks of the auxiliary variable are denoted by (R_{mh}, R_{Mh}). The known values of γ_{1h}, γ_{2h} are given in Table 1,

$$\gamma_{3h} = 1$$

and

$$\gamma_{4h} = R_M - R_m.$$

We can derive the various classes of the suggested estimator from (4.1), which are listed in Table 1.

Table 1. Some classes of the proposed estimator.

Subsets of the proposed estimator $\hat{S}^2_Q$	γ_{1h}	γ_{2h}
$\hat{S}^2_{Q_1} = \sum_{h=1}^{L} \eta_h W_h^2 s_{yh}^2 \exp\left[\theta_{1h}\left\{\frac{-\beta_{2(xh)}\left(\acute{s}^2_{xh}-s^2_{xh}\right)}{-\beta_{2(xh)}\left(\acute{s}^2_{xh}+s^2_{xh}\right)+2(x_{Mh}-x_{mh})}\right\}\right]\exp\left[\theta_{2h}\delta_h\right]$	$-\beta_{2(xh)}$	$x_{Mh} - x_{mh}$
$\hat{S}^2_{Q_2} = \sum_{h=1}^{L} \eta_h W_h^2 s_{yh}^2 \exp\left[\theta_{1h}\left\{\frac{c_{xh}\left(\acute{s}^2_{xh}-s^2_{xh}\right)}{c_{xh}\left(\acute{s}^2_{xh}+s^2_{xh}\right)+2(x_{Mh}-x_{mh})}\right\}\right]\exp\left[\theta_{2h}\delta_h\right]$	c_{xh}	$x_{Mh} - x_{mh}$
$\hat{S}^2_{Q_3} = \sum_{h=1}^{L} \eta_h W_h^2 s_{yh}^2 \exp\left[\theta_{1h}\left\{\frac{(x_{Mh}-x_{mh})\left(\acute{s}^2_{xh}-s^2_{xh}\right)}{(x_{Mh}-x_{mh})\left(\acute{s}^2_{xh}+s^2_{xh}\right)+2c_{xh}}\right\}\right]\exp\left[\theta_{2h}\delta_h\right]$	$x_{Mh} - x_{mh}$	c_{xh}
$\hat{S}^2_{Q_4} = \sum_{h=1}^{L} \eta_h W_h^2 s_{yh}^2 \exp\left[\theta_{1h}\left\{\frac{(x_{Mh}-x_{mh})\left(\acute{s}^2_{xh}-s^2_{xh}\right)}{(x_{Mh}-x_{mh})\left(\acute{s}^2_{xh}+s^2_{xh}\right)-2c_{xh}}\right\}\right]\exp\left[\theta_{2}\delta_h\right]$	$x_{Mh} - x_{mh}$	$-c_{xh}$
$\hat{S}^2_{Q_5} = \sum_{h=1}^{L} \eta_h W_h^2 s_{yh}^2 exp\left[\theta_{1h}\left\{\frac{(x_{Mh}-x_{mh})\left(\acute{s}^2_{xh}-s^2_{xh}\right)}{(x_{Mh}-x_{mh})\left(\acute{s}^2_{xh}+s^2_{xh}\right)+2\beta_{2(xh)}}\right\}\right]\exp\left[\theta_{2h}\delta_h\right]$	$x_{Mh} - x_{mh}$	$\beta_{2(xh)}$
$\hat{S}^2_{Q_6} = \sum_{h=1}^{L} \eta_h W_h^2 s_{yh}^2 \exp\left[\theta_{1h}\left\{\frac{\beta_{2(xh)}\left(\acute{s}^2_{xh}-s^2_{xh}\right)}{\beta_{2(xh)}\left(\acute{s}^2_{xh}+s^2_{xh}\right)+2(x_{Mh}-x_{mh})}\right\}\right]\exp\left[\theta_{2h}\delta_h\right]$	$\beta_{2(xh)}$	$x_{Mh} - x_{mh}$
$\hat{S}^2_{Q_6} = \sum_{h=1}^{L} \eta_h W_h^2 s_{yh}^2 \exp\left[\theta_{1h}\left\{\frac{(x_{Mh}-x_{mh})\left(\acute{s}^2_{xh}-s^2_{xh}\right)}{(x_{Mh}-x_{mh})\left(\acute{s}^2_{xh}+s^2_{xh}\right)-2\beta_{2(xh)}}\right\}\right]\exp\left[\theta_{2}\delta\right]$	$x_{Mh} - x_{mh}$	$-\beta_{2(xh)}$
$\hat{S}^2_{Q_8} = \sum_{h=1}^{L} \eta_h W_h^2 s_{yh}^2 \exp\left[\theta_{1h}\left\{\frac{-c_{xh}\left(\acute{s}^2_{xh}-s^2_{xh}\right)}{-c_{xh}\left(\acute{s}^2_{xh}+s^2_{xh}\right)+2(x_{Mh}-x_{mh})}\right\}\right]\exp\left[\theta_{2h}\delta_h\right]$	$-c_{xh}$	$x_{Mh} - x_{mh}$

where

$$\delta_h = \left(\frac{\acute{s}^2_{rh} - s^2_{rh}}{\acute{s}^2_{rh} + s^2_{rh} + 2(R_{Mh} - R_{mh})}\right).$$

Now, we discuss the properties of the new proposed class of estimators, we rewrite (4.1) in terms of errors to get the bias and the MSE of $\hat{S}^2_Q$, i.e.,

$$\begin{aligned}\hat{S}^2_Q = \sum_{h=1}^{L} \eta_h W_h^2 S^2_{yh} (1 + \xi_{0h}) \exp\left[\frac{b_{1h}(\xi_{2h} - \xi_{1h})}{2}\left(1 + \frac{b_{1h}}{2}(\xi_{1h} + \xi_{2h})\right)^{-1}\right] \\ \times \exp\left[\frac{b_{2h}(\xi_{4h} - \xi_{3h})}{2}\left(1 + \frac{b_{2h}}{2}(\xi_{3h} + \xi_{4h})\right)^{-1}\right],\end{aligned} \tag{4.2}$$

where

$$\theta_{1h} = \theta_{2h} = 1, \quad b_{1h} = \frac{\gamma_{1h} S^2_{xh}}{\gamma_{1h} S^2_{xh} + \gamma_{2h}}, \quad \text{and} \quad b_{2h} = \frac{S^2_{rh}}{S^2_{rh} + \gamma_{4h}}.$$

Applying the Taylor series to the first approximation order, we obtain

$$\begin{aligned}\hat{S}_Q^2 - \textstyle\sum_{h=1}^{L} \eta_h W_h^2 S_{yh}^2 \cong \textstyle\sum_{h=1}^{L} \eta_h W_h^2 S_{yh}^2 \Big[&\xi_{0h} - \tfrac{b_{1h}}{2}\left(\xi_{1h} - \xi_{2h}\right) - \tfrac{b_{2h}}{2}\left(\xi_{3h} - \xi_{4h}\right) + \tfrac{3b_{1h}^2}{8}\xi_{1h}^2 \\ &- \tfrac{b_{1h}^2}{8}\xi_{2h}^2 + \tfrac{b_{2h}^2}{8}\xi_{3h}^2 - \tfrac{b_{2h}^2}{8}\xi_{4h}^2 - \tfrac{\xi_{1h}}{2}\xi_{0h}\xi_{1h} + \tfrac{b_{1h}}{2}\xi_{0h}\xi_{2h} - \tfrac{b_{2h}}{2}\xi_{0h}\xi_{3h} \\ &+ \tfrac{b_{2h}}{2}\xi_{0h}\xi_{4h} - \tfrac{b_{1h}^2}{2}\xi_{1h}\xi_{2h} + \tfrac{b_{1h}b_{2h}}{4}\xi_{1h}\xi_{3h} - \tfrac{b_{1h}b_{2h}}{4}\xi_{1h}\xi_{4h} - \tfrac{b_{1h}b_{2h}}{4}\xi_{2h}\xi_{3h} \\ &+ \tfrac{b_{1h}b_{2h}}{4}\xi_{2h}\xi_{4h} - \tfrac{b_{2h}^2}{2}\xi_{3h}\xi_{4h}\Big].\end{aligned} \tag{4.3}$$

Using (4.3), the bias of $\hat{S}_T^2$ is given by

$$\begin{aligned}Bias\left(\hat{S}_Q^2\right) \cong &\textstyle\sum_{h=1}^{L} \eta_h^2 W_h^2 S_{yh}^2 \left[\tfrac{3b_{1h}^2}{8}\Delta_{040h}^* + \tfrac{3b_{2h}^2}{8}\Delta_{004h}^* - \tfrac{b_{1h}}{2}\Delta_{220h}^* - \tfrac{b_{2h}}{2}\Delta_{202h}^* + \tfrac{b_{1h}b_{2h}}{2}\Delta_{022h}^*\right] \\ &- \textstyle\sum_{h=1}^{L} \eta_h^{'2} W_h^2 S_{yh}^2 \left[\tfrac{3b_{1h}^2}{8}\Delta_{040h}^* + \tfrac{3b_{2h}^2}{8}\Delta_{004h}^* - \tfrac{b_{1h}}{2}\Delta_{220h}^* - \tfrac{b_{2h}}{2}\Delta_{202h}^* + \tfrac{b_{1h}b_{2h}}{2}\lambda_{022h}^*\right].\end{aligned}$$

After the simple simplifications, we get

$$Bias\left(\hat{S}_Q^2\right) \cong \sum_{h=1}^{L} \eta_h^{''2} W_h^2 S_{yh}^2 \left[\frac{3}{8}\left(b_{1h}^2\Delta_{040h}^* + b_{2h}^2\Delta_{004h}^*\right) - \frac{1}{2}\left(b_{1h}\Delta_{220h}^* + b_{2h}\Delta_{202h}^* - b_{1h}b_{2h}\Delta_{022h}^*\right)\right], \tag{4.4}$$

where

$$\eta_h^{''} = \eta_h - \eta_h^{'}.$$

The Eq (4.3) is squared and the expected value is taken to obtain a first-order approximation of the MSE, which is represented by the following equation

$$\begin{aligned}MSE\left(\hat{S}_Q^2\right) \cong &\textstyle\sum_{h=1}^{L} \eta_h^3 W_h^4 S_{yh}^4 \left[\Delta_{400h}^* + \tfrac{b_{1h}^2}{4}\Delta_{040h}^* + \tfrac{b_{2h}^2}{4}\Delta_{004h}^* - b_{1h}\Delta_{220h}^* - b_{2h}\Delta_{202h}^* + \tfrac{b_{1h}b_{2h}}{2}\Delta_{022h}^*\right] \\ &- \textstyle\sum_{h=1}^{L} \eta_h^{''3} W_h^4 S_{yh}^4 \left[\tfrac{b_{1h}^2}{4}\Delta_{040h}^* + \tfrac{b_{2h}^2}{4}\Delta_{004h}^* - b_{1h}\Delta_{220h}^* - b_{2h}\Delta_{202h}^* + \tfrac{b_{4h}b_{5h}}{2}\lambda_{022h}^*\right].\end{aligned}$$

After the simplification, we get

$$MSE\left(\hat{S}_Q^2\right) \cong \sum_{h=1}^{L} W_h^4 S_{yh}^4 \left[\eta_h^3\Delta_{400h}^* + \frac{\eta_h^{''3}}{4}\left(b_{1h}^2\Delta_{040h}^* + b_{2h}^2\Delta_{004h}^* - 4b_{1h}\Delta_{220h}^* - 4b_{2h}\Delta_{202h}^* + 2b_{1h}b_{2h}\Delta_{022h}^*\right)\right]. \tag{4.5}$$

5. Mathematical comparison

The proposed class of estimators $\hat{S}_Q^2$ is compared in this section to other existing estimators, including $\hat{S}_{T_1}^2, \hat{S}_{T_2}^2, \hat{S}_{T_3}^2, \hat{S}_{T_4}^2, \hat{S}_{T_5}^2$, and $\hat{S}_{T_i}^2$.

Condition (i): By (3.1) and (4.5)

$$Var(\hat{S}_{T_1}^2) > MSE\left(\hat{S}_Q^2\right),$$

if

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left(\eta_h^3 - \eta_h'^3\right)\left(b_{1h}^2\Delta_{040h}^* + b_{2h}^2\Delta_{004h}^* - 4b_{1h}\Delta_{220h}^* - 4b_{2h}\Delta_{202h}^* + 2b_{1h}b_2\Delta_{022h}^*\right) < 0.$$

For

$$\eta_h^3 - \eta_h'^3 > 0,$$

that is,

$$\eta_h^3 > \eta_h'^3,$$

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left(b_{1h}^2\Delta_{040h}^* + b_{2h}^2\Delta_{004h}^* - 4b_{1h}\Delta_{220h}^* - 4b_{2h}\Delta_{202h}^* + 2b_{1h}b_{2h}\Delta_{022h}^*\right) < 0. \tag{5.1}$$

Similarly

$$\eta_h - \eta_h' < 0,$$

that is,

$$\eta < \eta',$$

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left(b_{1h}^2\Delta_{040h}^* + b_{2h}^2\Delta_{004h}^* - 4b_{1h}\Delta_{220h}^* - 4b_{2h}\Delta_{202h}^* + 2b_{1h}b_{2h}\Delta_{022h}^*\right) > 0. \tag{5.2}$$

If condition (5.1) or (5.2) holds true, the suggested estimator $\hat{S}_Q^2$ demonstrates higher efficiency in comparison to $MSE(\hat{S}_{T_1}^2)$.

Condition (ii): By (3.4) and (4.5)

$$MSE(\hat{S}_{T_2}^2) > MSE\left(\hat{S}_Q^2\right),$$

if

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left(\eta_h^3 - \eta_h'^3\right)\left[(4 - b_{1h}^2)\Delta_{040h}^* - b_{2h}^2\Delta_{004h}^* + 4(b_{1h} - 2)\Delta_{220h}^* + 4b_{2h}\Delta_{202h}^* - 2b_{1h}b_{2h}\Delta_{022h}^*\right] > 0.$$

For

$$\eta_h^3 - \eta_h'^3 < 0,$$

that is,

$$\eta_h^3 < \eta_h'^3,$$

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left[(4 - b_{1h}^2)\Delta_{040h}^* - b_{2h}^2\Delta_{004h}^* + 4(b_{1h} - 2)\Delta_{220h}^* + 4b_{2h}\Delta_{202h}^* - 2b_{1h}b_{2h}\Delta_{022h}^*\right] > 0. \tag{5.3}$$

Similarly,

$$\eta_h^3 - \eta_h'^3 > 0,$$

that is,

$$\eta_h^3 > \eta_h'^3,$$

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left[(4 - b_{1h}^2)\Delta_{040h}^* - b_{2h}^2\Delta_{004h}^* + 4(b_{1h} - 2)\Delta_{220h}^* + 4b_{2h}\Delta_{202h}^* - 2b_{1h}b_{2h}\Delta_{022h}^*\right] < 0. \tag{5.4}$$

If condition (5.3) or (5.4) holds true, the suggested estimator $\hat{S}_Q^2$ demonstrates higher efficiency in comparison to $MSE(\hat{S}_{T_2}^2)$.

Condition (iii): By (3.6) and (4.5)

$$MSE(\hat{S}_{T_3}^2) > MSE\left(\hat{S}_Q^2\right),$$

if

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left(\eta_h^3 - \eta_h'^3\right)\left[\rho_{yxh}^{*2} + \frac{1}{4}\left(b_{1h}^2 \Delta_{040h}^* + b_2^{2h} \Delta_{004h}^* - 4b_{1h}\Delta_{220h}^* - 4b_{2h}\Delta_{202h}^* + 2b_{1h}b_{2h}\Delta_{022h}^*\right)\right] < 0.$$

For

$$\eta_h^3 - \eta_h'^3 > 0,$$

that is,

$$\eta_h^3 > \eta_h'^3,$$

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left[\rho_{yxh}^{*2} + \frac{1}{4}\left(b_{1h}^2 \Delta_{040h}^* + b_{2h}^2 \Delta_{004h}^* - 4b_{1h}\Delta_{220h}^* - 4b_{2h}\Delta_{202h}^* + 2b_{1h}b_{2h}\Delta_{022h}^*\right)\right] < 0. \qquad (5.5)$$

Similarly,

$$\eta_h^3 - \eta_h'^3 < 0,$$

that is,

$$\eta_h^3 < \eta_h'^3,$$

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left[\rho_{yxh}^{*2} + \frac{1}{4}\left(b_{1h}^2 \Delta_{040h}^* + b_{2h}^2 \Delta_{004h}^* - 4b_{1h}\Delta_{220h}^* - 4b_{2h}\Delta_{202h}^* + 2b_{1h}b_{2h}\Delta_{022h}^*\right)\right] > 0. \qquad (5.6)$$

If condition (5.5) or (5.6) holds true, the suggested estimator $\hat{S}_Q^2$ demonstrates higher efficiency in comparison to $MSE(\hat{S}_{T_3}^2)$.

Condition (iv): By (3.9) and (4.5)

$$MSE(\hat{S}_{T_4}^2) > MSE\left(\hat{S}_Q^2\right),$$

if

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left(\eta_h^3 - \eta_h'^3\right)\left[(b_{1h}^2 - 1)\Delta_{040h}^* + b_{2h}^2 \Delta_{004h}^* - 4(b_{1h} - 1)\Delta_{220h}^* - 4b_{2h}\Delta_{202h}^* + 2b_{1h}b_{2h}\Delta_{022h}^*\right] < 0.$$

For

$$\eta_h^3 - \eta_h'^3 > 0,$$

that is,

$$\eta_h^3 > \eta_h'^3,$$

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left[(b_{1h}^2 - 1)\Delta_{040h}^* + b_{2h}^2 \Delta_{004h}^* - 4(b_{1h} - 1)\Delta_{220h}^* - 4b_{2h}\Delta_{202h}^* + 2b_{1h}b_{2h}\Delta_{022h}^*\right] < 0. \qquad (5.7)$$

Similarly,

$$\eta_h^3 - \eta_h'^3 < 0,$$

that is,

$$\eta_h^3 < \eta_h'^3,$$

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left(\eta_h^3 - \eta_h'^3\right) \left[(b_{1h}^2 - 1)\Delta_{040h}^* + b_{2h}^2\Delta_{004h}^* - 4(b_{1h} - 1)\Delta_{220h}^* - 4b_{2h}\Delta_{202h}^* + 2b_{1h}b_{2h}\Delta_{022h}^*\right] > 0. \tag{5.8}$$

If condition (5.7) or (5.8) holds true, the suggested estimator $\hat{S}_Q^2$ demonstrates higher efficiency in comparison to $MSE(\hat{S}_{T_4}^2)$.

Condition (v): By (3.12) and (4.5)

$$MSE(\hat{S}_{T_5}^2) > MSE\left(\hat{S}_Q^2\right),$$

if

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left(\eta_h^3 - \eta_h'^3\right) \left[(b_{1h}^2 - 4g_h^2)\Delta_{040h}^* + b_{2h}^2\Delta_{004h}^* - 4(b_{1h} - g_h)\Delta_{220h}^* - 4b_{2h}\Delta_{202h}^* + 2b_{1h}b_{2h}\Delta_{022h}^*\right] < 0.$$

For

$$\eta_h^3 - \eta_h'^3 > 0,$$

that is,

$$\eta_h^3 > \eta_h'^3,$$

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left[(b_{1h}^2 - 4g_h^2)\Delta_{040h}^* + b_{2h}^2\Delta_{004h}^* - 4(b_{1h} - g_h)\Delta_{220h}^* - 4b_{2h}\Delta_{202h}^* + 2b_{1h}b_{2h}\Delta_{022h}^*\right] < 0. \tag{5.9}$$

Similarly,

$$\eta_h^3 - \eta_h'^3 < 0,$$

that is,

$$\eta_h^3 < \eta_h'^3,$$

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left[(b_{1h}^2 - 4g_h^2)\Delta_{040h}^* + b_{2h}^2\Delta_{004h}^* - 4(b_{1h} - g_h)\Delta_{220h}^* - 4b_{2h}\Delta_{202h}^* + 2b_{1h}b_{2h}\Delta_{022h}^*\right] > 0. \tag{5.10}$$

If condition (5.9) or (5.10) holds true, the suggested estimator $\hat{S}_Q^2$ demonstrates higher efficiency in comparison to $MSE(\hat{S}_{T_5}^2)$.

Condition (vi): By (3.17) and (4.5)

$$MSE(\hat{S}_{T_i}^2) > MSE\left(\hat{S}_Q^2\right),$$

if

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left(\eta_h^3 - \eta_h'^3\right) \left[(b_{1h}^2 - 4t_{ih}^2)\Delta_{040h}^* + b_{2h}^2\Delta_{004h}^* - 4(b_{1h} - t_{ih})\Delta_{220h}^* - 4b_{2h}\Delta_{202h}^* + 2b_{1h}b_{2h}\Delta_{022h}^*\right] < 0.$$

For

$$\eta_h^3 - \eta_h'^3 > 0,$$

that is,

$$\eta_h^3 > \eta_h'^3,$$

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left[(b_{1h}^2 - 4t_{ih}^2)\Delta_{040h}^* + b_{2h}^2\Delta_{004h}^* - 4(b_{1h} - t_{ih})\Delta_{220h}^* - 4b_{2h}\Delta_{202h}^* + 2b_{1h}b_{2h}\Delta_{022h}^*\right] < 0. \quad (5.11)$$

Similarly,

$$\eta_h^3 - \eta_h'^3 < 0,$$

that is,

$$\eta_h^3 < \eta_h'^3,$$

$$\sum_{h=1}^{L} W_h^4 S_{yh}^4 \left[(b_{1h}^2 - 4t_{ih}^2)\Delta_{040h}^* + b_{2h}^2\Delta_{004h}^* - 4(b_{1h} - t_{ih})\Delta_{220h}^* - 4b_{2h}\Delta_{202h}^* + 2b_{1h}b_{2h}\Delta_{022h}^*\right] > 0. \quad (5.12)$$

If condition (5.11) or (5.12) holds true, the suggested estimator $\hat{S}_Q^2$ demonstrates higher efficiency in comparison to $MSE(\hat{S}_{T_i}^2)$.

6. Numerical comparison

In this part, we examine the performance of the proposed class of estimators as compared to other estimators using percent relative efficiency (*PREs*). This examination is carried out using both simulated and three separate real data sets.

6.1. Simulation study

To confirm the theoretical results reported in Section 5, we use the methods proposed by [30–32] to undertake a simulation study. The goal is to evaluate the performance of the suggested class of estimators using the known minimum and maximum values of the auxiliary variable, as well as its ranks within the context of two-phase stratified sampling. The following probability distributions can possibly be used to artificially produce six distinct populations for the auxiliary variable X:

- Population 1: $X \sim Exponential\ (1)$;
- Population 2: $X \sim Exponential\ (3)$;
- Population 3: $X \sim Uniform\ (1, 3)$;
- Population 4: $X \sim Uniform\ (1, 2)$;
- Population 5: $X \sim Gamma\ (1, 4)$;
- Population 6: $X \sim Gamma\ (2, 5)$.

The variable of interest, Y, is computed as

$$Y = r_{yx} \times X + e,$$

where

$$r_{yx} = 0.80$$

indicates the correlation coefficient between the study and the auxiliary variables, and $e \sim N(0, 1)$ signifies the error term.

To compute the *PREs*, we used the following algorithms in *R*:

Step 1: We first use the various probability distributions mentioned above to generate a population of size 2000. In order to compute distinct values for the existing and suggested class of estimators, this population is split into two strata using stratified random sampling techniques.

Step 2: To collect a first phase sample of size $\acute{n}_h$ from a population of size N_h, use the simple random sampling without replacement (*SRSWOR*) technique.

Step 3: Using the *SRSWOR* technique, obtain the second phase sample size n_h from the first phase sample.

Step 4: We calculate the population total and the extreme values of the auxiliary variables from the above steps.

Step 5: For each population, we use *SRSWOR* approach to generate distinct sample sizes for each stratum. The sample sizes are specified as $20\%, 30\%$, and 40%.

Step 6: Obtained the *PREs* values for each sample size using all of the estimators presented in this article. This step ensures that the relative efficiency of each estimator is evaluated across different sample sizes.

Step 7: Steps 5 and 6 are then repeated 50,000 times to ensure the robustness of the results. The outcomes for artificial populations are presented in Table 2, which provides a comprehensive analysis of the estimators performance under simulated conditions.

Table 2. Percent relative efficiency (*PRE*) using the artificial populations.

Estimator	*Exp* (1)	*Exp* (3)	*Uni* $(1,3)$	*Uni* $(1,2)$	*Gam* $(1,4)$	*Gam* $(2,5)$
$\hat{S}^2_{T_1}$	100	100	100	100	100	100
$\hat{S}^2_{T_2}$	110.789	109.760	113.025	115.247	120.126	118.587
$\hat{S}^2_{T_3}$	120.970	123.638	115.188	120.315	122.505	120.765
$\hat{S}^2_{T_4}$	126.486	127.946	118.344	123.200	124.425	123.078
$\hat{S}^2_{T_5}$	128.145	128.108	122.670	126.526	128.772	127.589
$\hat{S}^2_{T_6}$	135.980	129.964	125.520	128.789	131.405	1131.164
$\hat{S}^2_{T_7}$	135.112	129.123	125.345	128.002	131.408	1131.664
$\hat{S}^2_{T_8}$	136.304	130.245	126.156	130.328	134.528	133.589
$\hat{S}^2_{Q_1}$	194.668	180.356	150.712	160.139	147. 547	153.329
$\hat{S}^2_{Q_2}$	353.129	385.467	320.225	333.167	296.724	280.289
$\hat{S}^2_{Q_3}$	259.189	256.578	270.667	259.369	220.949	214.345
$\hat{S}^2_{Q_4}$	162.707	168.689	180.576	165.508	157.333	142.148
$\hat{S}^2_{Q_5}$	142.837	148.790	160.031	143.625	140.495	140.279
$\hat{S}^2_{Q_6}$	190.456	202.098	200.321	178.353	182.132	162.399
$\hat{S}^2_{Q_7}$	170.065	192.987	190.401	150.323	152.369	158.950
$\hat{S}^2_{Q_8}$	1966.825	210.876	210.677	220.688	200. 712	192.952

Step 8: Furthermore, obtain the *MSEs* and *PREs* for each estimator over all replications using the

following formulas:

$$MSE(\hat{S}_l^2)_{\min} = \frac{\sum_{i=1}^{50000}\left(\hat{S}_l^2 - S_i^2\right)^2}{50000}$$

and

$$PRE = \frac{V(\hat{S}_{T_1}^2)}{MSE(\hat{S}_l^2)_{min}} \times 100,$$

where l is one of $T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_{Q_k}(k = 1, 2, \ldots, 8)$.

6.2. Numerical examples

To evaluate the effectiveness of the suggested estimators, we examine the *PREs* of several estimators on three real data sets. The data sets descriptions are given below, while the summary statistics are given in Tables 3–5.

Table 3. Summary statistics for data 1.

Descriptive statistics				
$N_1 = 18$	$\bar{X}_1 = 415$	$\bar{Y}_1 = 85572$	$\bar{R}_1 = 9.500$	$X_{M_1} = 2055$
$X_{m1} = 24$	$R_{M_1} = 18$	$R_{m_1} = 1$	$S_{x1} = 52.675$	$S_{y1} = 248216$
$S_{r1} = 5.338$	$C_{x1} = 1.258$	$C_{y1} = 2.901$	$C_{r1} = 0.562$	$\rho_{yx1} = 0.337$
$\rho_{yr1} = 0.304$	$\rho_{xr1} = 0.709$	$\Delta_{4001} = 3270$	$\Delta_{0401} = 3345$	$\Delta_{0041} = 1.692$
$\Delta_{2201} = 2398$	$\Delta_{2021} = 1267$	$\Delta_{0221} = 944$	$\eta_1' = 0.144$	$\eta_1'' = 0.056$
$N_2 = 18$	$\bar{X}_2 = 257$	$\bar{Y}_2 = 19293.610$	$\bar{R}_2 = 27.500$	$X_{M_2} = 1674$
$X_{m_2} = 52$	$R_{m_2} = 19$	$R_{M_2} = 36$	$S_{x2} = 365.696$	$S_{y2} = 37979$
$S_{r2} = 5.338$	$C_{x2} = 1.423$	$C_{y2} = 1.969$	$C_{r2} = 0.194$	$\rho_{yx2} = 0.976$
$\rho_{yr2} = 0.565$	$\rho_{xr2} = 0.786$	$\Delta_{4002} = 2542$	$\Delta_{0402} = 2388$	$\Delta_{0042} = 1.622$
$\Delta_{2202} = 2246$	$\Delta_{2022} = 739$	$\Delta_{0222} = 988$	$\eta_2' = 0.144$	$\eta_2'' = 0.056$

Table 4. Summary statistics for data 2.

Descriptive statistics				
$N_1 = 18$	$\bar{X}_1 = 962$	$\bar{Y}_1 = 162979$	$\bar{R}_1 = 9.500$	$X_{M_1} = 1530$
$X_{m_1} = 388$	$R_{M_1} = 36$	$R_{m_1} = 19$	$S_{x1} = 308$	$S_{y1} = 255887$
$S_{r1} = 5.338$	$C_{x1} = 0.320$	$C_{y1} = 1.571$	$C_{r1} = 0.562$	$\rho_{yx1} = 0.145$
$\rho_{yr1} = 0.135$	$\rho_{xr1} = 0.802$	$\Delta_{4001} = 2625$	$\Delta_{0401} = 3237$	$\Delta_{0041} = 1.692$
$\Delta_{2201} = 1568$	$\Delta_{2021} = 1548$	$\Delta_{0221} = 1298$	$\eta_1' = 0.144$	$\eta_1'' = 0.056$
$N_2 = 18$	$\bar{X}_2 = 1146$	$\bar{Y}_2 = 134458$	$\bar{R}_2 = 27.500$	$X_{M_2} = 2370$
$X_{m_2} = 58$	$R_{m_2} = 19$	$R_{M_2} = 36$	$S_{x2} = 469.931$	$S_{y2} = 50236$
$S_{r2} = 5.338$	$C_{x2} = 0.409$	$C_{y2} = 0.374$	$C_{r2} = 0.194$	$\rho_{yx2} = 0.787$
$\rho_{yr2} = 0.657$	$\rho_{xr2} = 889$	$\Delta_{4002} = 2240$	$\Delta_{0402} = 2558$	$\Delta_{0042} = 1.622$
$\Delta_{2202} = 1807$	$\Delta_{2022} = 2049$	$\Delta_{0222} = 1200$	$\eta_2' = 0.144$	$\eta_2'' = 0.056$

Table 5. Summary statistics for data 3.

Descriptive statistics				
$N_1 = 18$	$\bar{X}_1 = 72.550$	$\bar{Y}_1 = 27.490$	$\bar{R}_1 = 9.500$	$X_{M_1} = 95$
$X_{m1} = 28$	$R_{M1} = 18$	$R_{m1} = 1$	$S_{x1} = 10.580$	$S_{y1} = 10.130$
$S_{r1} = 5.338$	$C_{x1} = 0.155$	$C_{y1} = 0.376$	$C_{r1} = 0.562$	$\rho_{yx1} = 0.337$
$\rho_{yr1} = 0.284$	$\rho_{xr1} = 0.557$	$\Delta_{4001} = 2.550$	$\Delta_{0401} = 2.845$	$\Delta_{0041} = 1.692$
$\Delta_{2201} = 3.158$	$\Delta_{2021} = 4.544$	$\Delta_{0221} = 4.542$	$\eta'_1 = 0.144$	$\eta''_1 = 0.056$
$N_2 = 18$	$\bar{X}_2 = 60.870$	$\bar{Y}_2 = 20.820$	$\bar{R}_2 = 27.500$	$X_{M_2} = 75$
$X_{m_2} = 15$	$R_{m_2} = 19$	$R_{M_2} = 36$	$S_{x2} = 8.980$	$S_{y2} = 12.750$
$S_{r2} = 5.338$	$C_{x2} = 0.142$	$C_{y2} = 0.269$	$C_{r2} = 0.194$	$\rho_{yx2} = 0.496$
$\rho_{yr2} = 0.297$	$\rho_{xr2} = 0.756$	$\Delta_{4002} = 4.142$	$\Delta_{0402} = 3.934$	$\Delta_{0042} = 1.622$
$\Delta_{2202} = 1.384$	$\Delta_{2022} = 1.239$	$\Delta_{0222} = 2.488$	$\eta'_2 = 0.144$	$\eta''_2 = 0.056$

- **Data 1.** (**Source:** ([36, p.226]))
  Y: The employment levels recorded by the different departments for 2012, which represents the overall number of workers.
  X: The total number of factories that these departments officially registered in 2012, which gives information on industrial activity.
  R: The rankings assigned to each department based on the total number of factories they registered in 2012, offering a comparative view of industrial engagement across departments.
  Two distinct groups have been created from the data-set:
  Group 1: The Gujranwala, Rawalpindi, Sargodha, and Lahore divisions are included in this group; they all contribute to the examination of employment and industrial registration.
  Group 2: This group represents another aspect of the information for comparison analysis and is made up of the divisions of Bahawalpur, Faisalabad, Multan, Sahiwal, and Khan.

- **Data 2.** (**Source:** [36, p.135])
  Y: Represents the total number of students attended at educational institutions in 2012.
  X: Represents the overall number of government-funded schools in 2012.
  R: Represents the order of government-funded schools in 2012 according to the number of schools they had in that year.
  Two distinct groups have been generated from the data-set:
  Group 1: The Gujranwala, Rawalpindi, Sargodha, and Lahore divisions are included in this group; they all contribute to the examination of employment and industrial registration.
  Group 2: This group represents another aspect of the information for comparison analysis and is made up of the divisions of Bahawalpur, Faisalabad, Multan, Sahiwal, and Khan.

- **Data 3.** (**Source:** [37, p.24])
  Y: The expenses incurred on food by the family, directly related to their employment.
  X: The total weekly income earned by the family, reflecting their financial resources for that period.
  R: The ranking of families based on their weekly income, providing a comparative measure of their earnings.

For efficiency comparisons, we use the following formula:

$$PRE = \frac{V(\hat{S}^2_{T_1})}{MSE(\hat{S}^2_l)} \times 100,$$

where l is one of $T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_{Q_k} (k = 1, 2, \ldots, 8)$.

Additionally, Table 6 presents a summary of the findings for real data-sets.

Table 6. Percent relative efficiency using empirical data-sets.

Estimator	Data 1	Data 2	Data 3
$\hat{S}^2_{T_1}$	100	100	100
$\hat{S}^2_{T_2}$	111.535	101.057	100.991
$\hat{S}^2_{T_3}$	113.964	109.260	109.302
$\hat{S}^2_{T_4}$	112.526	109.250	106.626
$\hat{S}^2_{T_5}$	111.735	102.009	102.023
$\hat{S}^2_{T_6}$	111.535	101.048	128.081
$\hat{S}^2_{T_7}$	111.535	101.047	128.089
$\hat{S}^2_{T_8}$	111.695	103.703	129.687
$\hat{S}^2_{Q_1}$	117.202	116.082	138.417
$\hat{S}^2_{Q_2}$	119.177	117.150	139.897
$\hat{S}^2_{Q_3}$	117.384	115.726	134.367
$\hat{S}^2_{Q_4}$	114.751	116.077	137.845
$\hat{S}^2_{Q_5}$	117.201	115.027	135.217
$\hat{S}^2_{Q_6}$	117.202	115.026	138.091
$\hat{S}^2_{Q_7}$	117.202	115.025	133.450
$\hat{S}^2_{Q_8}$	117.227	114.851	134.986

7. Conclusions

A class of efficient estimators for estimating the finite population variance was introduced in this article. These estimators accounted for both the rankings and the auxiliary variable's extreme values. The theoretical prerequisites outlined in Section 5 show how the suggested class of estimators is more efficient than others, allowing for a comparison with those that already exist. To verify these limits, we conducted a simulation study and examined three empirical data sets. The outcomes, displayed in Table 2, demonstrate that the suggested class of estimators consistently performs better in terms of $PREs$ than the other existing estimators. The theoretical results in Section 5 are further confirmed by the empirical data shown in Table 6. We draw the conclusion that, in comparison to the other estimators under consideration, the suggested class of estimators $\hat{S}^2_{Q_i}$ $(i = 1, 2, 3, \ldots, 8,)$ exhibits superior efficiency based on both simulation and empirical data. Because it has the lowest MSE of these suggested estimators, $\hat{S}^2_{Q_2}$ is particularly preferable.

There are some advantages of this study in practical applications are given below:

- **Improved accuracy and efficiency:** Using extreme values and rankings of auxiliary variables, the novel approach improves precision and efficiency when calculating population variance. The suggested estimators outperform previous approaches, achieving *PRE* values of up to 385.467 in simulated experiments. This improved performance is especially valuable in real-world applications where survey data may contain outliers.

- **Applicability in stratified two-phase sampling:** The suggested approach is especially designed to support stratified two-phase sampling, which is usual in large-scale surveys when supplementary information may be unavailable until later stages. This makes the approach particularly useful in economic surveys, public health examinations, and environmental evaluations, where similar sample strategies are often used.
- **Handling of outliers in practical contexts:** The new estimators are ideal for disciplines like market research and agricultural surveys that frequently meet extreme values, as they can efficiently include outlier information without distorting results.

Benchmark analysis

For this study, a thorough benchmark analysis was conducted using the procedures listed below:

Selection of competing methods:

- Find and choose the estimators that are employed in stratified two-phase sampling to estimate the finite population variance. Regression-based estimators, exponential ratio estimators, and conventional variance estimators like these are a few examples.
- For an extensive comparison basis, use the most widely utilized techniques from the review of the literature, such as those put out by Isaki (1983) [13], Bahl and Tuteja (1991) [14], and Upadhyaya and Singh (1999) [15].

Performance metrics:

- *PRE*, which shows the improvement in effectiveness over a standard approach, should be the main tool used to assess how well various estimators perform.
- To examine estimators performance in entirety, take into account other measures, including bias, adaptability to outliers, and MSE.
- Analyze the computational efficiency of the suggested and existing estimators, particularly for large data sets.

Simulation study and real life data sets

- The research encompassed practical stratification situations and included a variety of artificial populations with varying probability distributions, including exponential, uniform, and gamma. Comparing the results using practical problems variables, such as industrial activity and employment levels, provided valuable insights into practical application, while several replications guaranteed statistical robustness.
- According to the findings, the suggested estimators continuously performed better than conventional techniques in terms of *PRE*, with appreciable gains over a range of sample sizes and distributions. The investigation was made more detailed by the use of statistical tests to validate the significance of the according to efficiency increases. The advantages of the novel methodology were established by this thorough study, which also supported its consideration by proving its superiority over other methods.

Moreover, we investigated the characteristics of the suggested efficient class of estimators using a two-phase stratified sampling technique. It is also conceivable to propose some novel estimators utilizing the non-response sampling approach, and our findings can be useful in determining the more efficient estimators with the lowest $MSEs$. It is also an appropriate topic for future investigation.

Conflict of interest

The author declares no conflict of interest.

References

1. J. Neyman, Contribution to the theory of sampling human population. *J. Amer. Stat. Assoc.*, **33** (1938), 101–116. https://doi.org/10.2307/2279117
2. B. V. Sukhatme, Some ratio-type estimators in two-phase sampling, *J. Amer. Stat. Assoc.*, **57** (1962), 628–632. https://doi.org/10.2307/2282400
3. G. K. Vishwakarma, S. M. Zeeshan, Generalized ratio-cum-product estimator for finite population mean under two-phase sampling scheme, *J. Mod. Appl. Stat. Meth.*, **19** (2021), 2–16. https://doi.org/10.22237/jmasm/1608553320
4. T. Zaman, C. Kadilar, New class of exponential estimators for finite population mean in two-phase sampling, *Commun. Stat.*, **50** (2021), 874–889. https://doi.org/10.1080/03610926.2019.1643480
5. A. Y.Erinola, R. V. K. Singh, A. Audu, T. James, Modified class of estimator for finite population mean under two-phase sampling using regression estimation approach, *Asian. J. Prob. Stat.*, **4** (2021), 52–64. https://doi.org/10.9734/ajpas/2021/v14i430338
6. M. N. Qureshi, M. U. Tariq, M. Hanif, Memory-type ratio and product estimators for population variance using exponentially weighted moving averages for time-scaled surveys, *Commun. Stat. Simul. Comput.*, **53** (2024), 1484–1493. https://doi.org/10.1080/03610918.2022.2050390
7. A. Sanaullah, M. Hanif, A. Asghar, Generalized exponential estimators for population variance under two-phase sampling, *Int. J. Appl. Comput. Math.*, **2** (2016), 75–84. https://doi.org/10.1007/s40819-015-0047-5
8. H. P. Singh, S. Singh, J. M. Kim, Efficient use of auxiliary variables in estimating finite population variance in two-phase sampling, *Commun. Korean Stat. Soc.*, **17** (2010), 165–181. https://doi.org/10.5351/CKSS.2010.17.2.165
9. M. A. Alomair, U. Daraz, Dual transformation of auxiliary variables by using outliers in stratified random sampling, *Mathematics*, **12** (2024), 2839. https://doi.org/10.3390/math12182839
10. U. Daraz, M. A. Alomair, O. Albalawi, A. S. Al Naim, New techniques for estimating finite population variance using ranks of auxiliary variable in two-stage sampling, *Mathematics*, **12** (2024), 2741. https://doi.org/10.3390/math12172741
11. M. Jabbar, Z. Javid, A. Zaheer, R. Zainab, Ratio type exponential estimator for the estimation of finite population variance under two-stage sampling, *Res. J. Appl. Sci. Eng. Technol.*, **7** (2024), 4095–4099. https://doi.org/0.19026/rjaset.7.772

12. A. K. Das, T. P. Tripathi, Use of auxiliary information in estimating the finite population variance, *Sankhya*, **40** (1978), 39–148.

13. C. T. Isaki, Variance estimation using auxiliary information, *J. Am. Stat. Assoc.*, **78** (1983), 117–123. https://doi.org/10.2307/2287117

14. S. Bahl, R. Tuteja, Ratio and product type exponential estimators, *J. Inf. Optim. Sci.*, **12** (1991), 159–164. https://doi.org/10.1080/02522667.1991.10699058

15. L. Upadhyaya, H. Singh, An estimator for population variance that utilizes the kurtosis of an auxiliary variable in sample surveys, *Vikram Math. J.*, **19** (1999), 14–17.

16. V. Dubey, H. Sharma, On estimating population variance using auxiliary information, *Stat. Transit. New Ser.*, **9** (2008), 7–18.

17. C. Kadilar, H. Cingi, Ratio estimators for the population variance in simple and stratified random sampling, *Appl. Math. Comput.*, **173** (2006), 1047–1059. https://doi.org/10.1016/j.amc.2005.04.032

18. H. Singh, P. Chandra, An alternative to ratio estimator of the population variance in sample surveys, *J. Transp. Stat.*, **9** (2008), 89–103.

19. J. Shabbir, S. Gupta, Some estimators of finite population variance of stratified sample mean, *Commun. Stat.*, **39** (2010), 3001–3008. https://doi.org/10.1080/03610920903170384

20. H. P. Singh, R. S. Solanki, A new procedure for variance estimation in simple random sampling using auxiliary information, *Stat. Papers*, **54** (2013), 479–497. https://doi.org/10.1007/s00362-012-0445-2

21. S. K. Yadav, C. Kadilar, J. Shabbir, S. Gupta, Improved family of estimators of population variance in simple random sampling, *J. Stat. Theory Practice*, **9** (2015), 219–226. https://doi.org/10.1080/15598608.2013.856359

22. J. Shabbir, S. Gupta, Using rank of the auxiliary variable in estimating variance of the stratified sample mean, *Int. J. Comput. Theor. Stat.*, **6** (2019), 171–181. http://doi.org/10.12785/IJCTS/060207

23. T. Zaman, H. Bulut, An efficient family of robust-type estimators for the population variance in simple and stratified random sampling, *Commun. Stat.*, **52** (2023), 2610–2624. https://doi.org/10.1080/03610926.2021.1955388

24. S. Mohanty, J. Sahoo, A note on improving the ratio method of estimation through linear transformation using certain known population parameters, *Sankhyā Indian J. Stat.*, **57** (1995), 93–102.

25. M. Khan, J. Shabbir, Some improved ratio, product, and regression estimators of finite population mean when using minimum and maximum values, *Sci. World J.*, **2013** (2013), 431868. https://doi.org/10.1155/2013/431868

26. G. S. Walia, H. Kaur, M. Sharma, Ratio type estimator of population mean through efficient linear transformation, *Amer. J. Math. Stat.*, **5** (2015), 144–149. https://doi.org/10.5923/j.ajms.20150503.06

27. M. Khan, Improvement in estimating the finite population mean under maximum and minimum values in double sampling scheme, *J. Stat. Appl. Probab. Lett.*, **2** (2015), 115–121. https://doi.org/10.12785/jsapl/020203

28. U. Daraz, J. Shabbir, H. Khan, Estimation of finite population mean by using minimum and maximum values in stratified random sampling, *J. Mod. Appl. Stat. Methods*, **17** (2018), 20. https://doi.org/10.22237/jmasm/1532007537

29. U. Daraz, M. Khan, Estimation of variance of the difference-cum-ratio-type exponential estimator in simple random sampling, *Res. Math. Stat.*, **8** (2021), 1899402. https://doi.org/10.1080/27658449.2021.1899402

30. U. Daraz, J. Wu, O. Albalawi, Double exponential ratio estimator of a finite population variance under extreme values in simple random sampling, *Mathematics*, **12** (2024), 1737. https://doi.org/10.3390/math12111737

31. U. Daraz, J. Wu, M. A. Alomair, L. A. Aldoghan, New classes of difference cum-ratio-type exponential estimators for a finite population variance in stratified random sampling, *Heliyon*, **10** (2024), e33402. https://doi.org/10.1016/j.heliyon.2024.e33402

32. U. Daraz, M. A. Alomair, O. Albalawi, Variance estimation under some transformation for both symmetric and asymmetric data, *Symmetry*, **16** (2024), 957. https://doi.org/10.3390/sym16080957

33. H. O. Cekim, H. Cingi, Some estimator types for population mean using linear transformation with the help of the minimum and maximum values of the auxiliary variable, *Hacet. J. Math. Stat.*, **46** (2017), 685–694. https://doi.org/10.15672/hjms.201510114186

34. S. Chatterjee, A. S. Hadi, *Regression analysis by example*, John Wiley & Sons, Inc., 2013. https://doi.org/10.1002/0470055464

35. D. J. Watson, The estimation of leaf area in field crops, *J. Agric. Sci.*, **27** (1937), 474–483. https://doi.org/10.1017/S002185960005173X

36. Bureau of Statistics, *Punjab development statistics government of the Punjab, Lahore, Pakistan*, 2013.

37. W. B. Cochran, *Sampling techniques*, John Wiley & Sons, Inc., 1963.

AIMS Press