

Article

Twofold Auxiliary Information Under Two-Phase Sampling: An Improved Family of Double-Transformed Variance Estimators

Umer Daraz ^{1,*} , Della Agustiana ², Jinbiao Wu ¹  and Walid Emam ³ 

¹ School of Mathematics and Statistics, Central South University, Changsha 410017, China; wujinbiao@csu.edu.cn

² School of Metallurgy and Environment, Central South University, Changsha 410083, China; dellaagustiana99@gmail.com or agustianadella@csu.edu.cn

³ Department of Statistics and Operations Research, Faculty of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia; wemam.c@ksu.edu.sa

* Correspondence: umerdaraz099@gmail.com or 22023161@csu.edu.cn

Abstract: Outlier values and rankings are important for emphasizing data distribution variability, which improves the accuracy and effectiveness of variance estimations. To enhance the estimation of finite population variance in a two-phase sampling framework, this study presents an improved class of double exponential-type estimators by utilizing the outlier values and ranks of an auxiliary variable. A theoretical analysis is conducted to derive the biases and mean squared errors (MSEs) of these estimators using first-order approximations. A comprehensive simulation study is then performed to analyze the performance of the proposed estimators. The results clearly show that the new estimators provide more precise estimates, achieving a higher percentage relative efficiency (PRE) across all simulated scenarios. Furthermore, three data sets are analyzed to further confirm the efficiency of the proposed estimators as compared to other existing estimators. These results emphasize the potential of the proposed class of estimators to optimize variance estimation techniques, making it a more cost-effective and accurate choice for researchers using two-phase sampling in a variety of domains.

Keywords: double exponential estimators; variance estimation; two-phase sampling; outlier values; ranks; bias; percent relative efficiency

MSC: 62D05



Academic Editor: Hari Mohan Srivastava

Received: 16 December 2024

Revised: 9 January 2025

Accepted: 13 January 2025

Published: 16 January 2025

Citation: Daraz, U.; Agustiana, D.; Wu, J.; Emam, W. Twofold Auxiliary Information Under Two-Phase Sampling: An Improved Family of Double-Transformed Variance Estimators. *Axioms* **2025**, *14*, 64. <https://doi.org/10.3390/axioms14010064>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In sampling theory, it is usual to employ auxiliary variables along with the study variable to enhance design, as well as to increase the efficiency of estimators by taking advantage of the connection between the auxiliary and study variables. In certain cases, data on many auxiliary variables are readily available. For example, in order to examine the public health and welfare situation in a whole country or state, it may be necessary to know the number of beds in various hospitals, the number of physicians and other staff members, and the total amount of money available for medical treatment. When this type of information is lacking, two-phase sampling is preferred since it requires one to gather a large preliminary sample from which the auxiliary variable is calculated. This strategy is frequently employed when it is more cost-effective or efficient to undertake an initial preparation phase prior to the main phase. In two-phase sampling, the first and second samples are referred to as the first and second phase samples, respectively. A cost-effective technique for sample surveys is two-phase sampling, especially in situations when

supplementary data are not available. Two-phase sampling was initially suggested by [1], with further investigations conducted by [2], who proposed a general ratio-type estimator, and this was later expanded upon by [3–5]. For more details about two-phase sampling, readers can look at [6,7].

A major challenge in measuring variability is the unpredictable nature of the results, necessitating the precise control and careful selection of auxiliary data for improving the accuracy of the estimators. The concept of population variance has been extensively studied and improved over time, with various estimators introduced by various researchers. The initial discussion on auxiliary information in population variance calculations was initiated by [4]. The use of ratio and product type exponential estimators has also been suggested by [8] for improving estimation. By using transformation, Ref. [9] proposed an improved variance estimator. As discussed in various studies [10–15], researchers have introduced different variance type estimators to provide better estimates for population variance.

The presence of outlier (extreme) values can greatly affect the accuracy of estimations and lead to significantly underestimated outcomes in populations. The work by [16] presented two estimators that employed a transformation of the auxiliary variable known largest and smallest observations. Nevertheless, this method was not investigated further until the study by [17], who used the idea of the largest and smallest values for various estimators. By including extreme values into various transformations, the estimation of the finite population mean by [18] was significantly improved. To estimate population variance with least mean squared errors, Ref. [19,20] presented a novel family of estimators based on extreme values. To estimate the population variance, Ref. [21] recently introduced a number of effective estimators by transforming extreme values. To improve the mean estimation methods under extreme values, Refs. [22–24] introduced different transformed estimators.

Recently, some researchers obtained different classes of estimators by using transformation on a single auxiliary variable. In order to provide the more precise estimators while taking extreme values into consideration, we introduce a new family of estimators by employing the double transformation method on two auxiliary variables and show that our new double-transformed exponential estimators provide more precise results as compared to other variance estimators under two-phase sampling. These double-transformed estimators are based on the largest and smallest observations of the auxiliary variable as well as the ranking of the smallest and largest observation of the auxiliary variable. Therefore, this ranking method can possibly be helpful for improving the performance and precision of the estimators.

Practical applications of the proposed estimators:

The improved estimators have significant potential for application in various fields and specific domains, such as medicine, industry, retail, and transportation. These estimators can be directly applied to real-world scenarios where accurate predictions are crucial.

- **Medicine:** In the medical field, we illustrate how our proposed method can be applied to improve the accuracy of early-stage cancer detection, particularly in the analysis of medical imaging data. By using our method with existing diagnostic tools, such as MRI or CT scans, it can help healthcare professionals more accurately identify malignant tissues, enabling earlier intervention and improving patient outcomes. For instance, the method could enhance the detection of tumors in breast cancer patients by analyzing patterns in mammography images, leading to fewer false negatives and earlier treatment.
- **Industry:** In the industrial sector, we explore the application of our method in optimizing predictive maintenance within manufacturing processes. Specifically, our approach can be used to monitor and predict equipment failure in an automotive

manufacturing plant. By analyzing sensor data from machinery, our method can predict when a component is likely to fail, allowing for pre-emptive maintenance that minimizes downtime and reduces repair costs. This can lead to increased operational efficiency and cost savings, as well as fewer disruptions in the production line.

- **Retail:** In retail, our method can enhance inventory management by predicting product demand more accurately. This would allow retailers to stock the right amount of products at the right time, reducing overstocking and understocking issues. We believe that including these practical examples not only demonstrates the versatility of our proposed method but also illustrates its potential to drive meaningful improvements in both healthcare and industrial operations.
- **Transportation:** In the transportation sector, our method can improve route optimization for delivery services. By analyzing traffic patterns and predicting delays, it can help logistics companies plan the most efficient routes, reducing fuel consumption and delivery times.

The following are the arrangements of the next sections of this article: The notations and an overview of various existing estimators are presented in Section 2. We explore our double-transformed family of estimators in depth in Section 3. A mathematical comparison of these estimators is demonstrated in Section 4. A simulation analysis is presented in Section 5 that validates the theoretical results from Section 4 by generating simulated populations based on various probability distributions. In order to demonstrate our theoretical findings, this section also provides numerical examples. In Section 6, we provide conclusions and discuss future research directions as well.

2. Notations and Existing Estimators

A finite population of N units is represented as $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_N)$. The study variable is denoted by Y , the auxiliary variable is X , and R represents the ranks of the auxiliary variable for each i th unit of y_i, x_i , and r_i . Let the population means be defined below:

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i,$$

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i,$$

$$\bar{R} = \frac{1}{N} \sum_{i=1}^N R_i.$$

Additionally, population variances are defined as follows (without replacement sampling):

$$S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2,$$

$$S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$$

and

$$S_r^2 = \frac{1}{N-1} \sum_{i=1}^N (R_i - \bar{R})^2.$$

Let C_y, C_x , and C_r be defined as population coefficients of variation, which are calculated as

$$C_y = \frac{S_y}{\bar{Y}},$$

$$C_x = \frac{S_x}{\bar{X}}$$

and

$$C_r = \frac{S_r}{\bar{R}},$$

respectively.

The population correlation coefficients between (Y, X) , (Y, R) , and (X, R) are defined as

$$\rho_{yx} = \frac{S_{yx}}{S_y S_x},$$

$$\rho_{yr} = \frac{S_{yr}}{S_y S_r}$$

and

$$\rho_{xr} = \frac{S_{xr}}{S_x S_r},$$

respectively.

Consider the following sample variances:

$$s_y^2 = \hat{S}_y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{y})^2,$$

$$s_x^2 = \hat{S}_x^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{x})^2,$$

$$s_r^2 = \hat{S}_r^2 = \frac{1}{n-1} \sum_{i=1}^n (R_i - \bar{r})^2,$$

where

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n Y_i,$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n X_i$$

and

$$\bar{r} = \frac{1}{n} \sum_{i=1}^n R_i,$$

are the sample means of Y , X , and R . Further, the sample coefficients of variation are defined as follows:

$$c_y = \frac{s_y}{\bar{y}},$$

$$c_x = \frac{s_x}{\bar{x}}$$

and

$$c_r = \frac{s_r}{\bar{r}}.$$

This study examines a double exponential family of estimators that use different transformations to estimate the finite population variance of study variable Y when the auxiliary variable X is present. The two-phase sampling strategy is defined as follows:

1. For obtaining an accurate estimate of a population variance S_x^2 , a first-phase sample m ($m \subset N$) is taken from a population of size N to measure only auxiliary s_x^2 .

2. Further, a second-phase sample size denoted by n ($n \subset m$) is obtained from a first-phase sample of size m to observe both y and x , respectively.

Moreover, let $\hat{s}_x^2$ and $\hat{s}_r^2$ represent the sample variances obtained from the first-phase sample with a size of m , while s_y^2 , s_x^2 , and s_r^2 denote the sample variances derived from the second-phase sample with a size of n .

We now construct the error terms for calculating the biases and mean square errors of the various existing estimators that are used for comparison in this study:

$$e_0 = \left(\frac{s_y^2 - S_y^2}{S_y^2} \right), e_1 = \left(\frac{s_x^2 - S_x^2}{S_x^2} \right), e_2 = \left(\frac{s_x^2 - S_x^2}{S_x^2} \right), e_3 = \left(\frac{s_r^2 - S_r^2}{S_r^2} \right), e_4 = \left(\frac{s_r^2 - S_r^2}{S_R^2} \right)$$

such that $E(e_i) = 0$ for $i = 0, 1, 2, 3, 4$.

$$\begin{aligned} E(e_0^2) &= \theta \lambda_{400}^*, & E(e_1^2) &= \theta \lambda_{040}^*, & E(e_2^2) &= \theta' \lambda_{040}^*, \\ E(e_3^2) &= \theta \lambda_{004}^*, & E(e_4^2) &= \theta' \lambda_{004}^*, & E(e_0 e_1) &= \theta \lambda_{220}^*, \\ E(e_0 e_2) &= \theta' \lambda_{220}^*, & E(e_0 e_3) &= \theta \lambda_{202}^*, & E(e_0 e_4) &= \theta' \lambda_{202}^*, \\ E(e_1 e_2) &= \theta' \lambda_{040}^*, & E(e_1 e_3) &= \theta \lambda_{022}^*, & E(e_1 e_4) &= \theta' \lambda_{022}^*, \\ E(e_2 e_3) &= \theta' \lambda_{022}^*, & E(e_2 e_4) &= \theta' \lambda_{022}^*, & E(e_3 e_4) &= \theta' \lambda_{004}^*, \end{aligned}$$

where

$$\begin{aligned} \lambda_{400}^* &= (\lambda_{400} - 1), \lambda_{040}^* = (\lambda_{040} - 1), \lambda_{004}^* = (\lambda_{004} - 1), \\ \lambda_{220}^* &= (\lambda_{220} - 1), \lambda_{202}^* = (\lambda_{202} - 1), \lambda_{022}^* = (\lambda_{022} - 1) \end{aligned}$$

and

$$\begin{aligned} \theta &= \left(\frac{1}{n} - \frac{1}{N} \right), \\ \theta' &= \left(\frac{1}{m} - \frac{1}{N} \right), \\ \theta'' &= \left(\frac{1}{n} - \frac{1}{m} \right), \end{aligned}$$

θ , θ' , and θ'' are sampling fraction differences or sampling correction terms used to adjust for varying sample sizes at each phase: θ for the second phase, θ' for the first phase, and θ'' for the inter-phase. They are important in designing and analyzing estimators to ensure unbiasedness and efficiency in complex survey sampling, particularly in two-phase sampling schemes.

Further,

$$\lambda_{lqs} = \frac{\mu_{lqs}}{\mu_{200}^{l/2} \mu_{020}^{q/2} \mu_{002}^{s/2}}$$

and

$$\mu_{lqs} = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^l (X_i - \bar{X})^q (R_i - \bar{R})^s}{N - 1},$$

where λ_{lqs} represents the population central moment with orders (l, q, s) , and $(\mu_{200}, \mu_{020}, \mu_{002})$ denotes the standard deviation of (Y, X, R) . Here, $\lambda_{400} = \beta_{2(y)}$, $\lambda_{040} = \beta_{2(x)}$, and $\lambda_{004} = \beta_{2(r)}$ are the population coefficients of kurtosis.

We know that the estimator for variance is expressed as:

$$T_1 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{Y})^2.$$

The variance of T_1 can be calculated as:

$$\text{Var}(T_1) = \theta S_y^4 \lambda_{400}^*. \quad (1)$$

The ratio estimator is denoted as T_2 : introduced by [4], which is given by

$$T_2 = s_y^2 \left(\frac{s_x^2}{s_x^2} \right). \quad (2)$$

For T_2 , the following equations give the bias and MSE expressions:

$$\text{Bias}(T_2) \cong \theta'' S_y^4 (\lambda_{040}^* - \lambda_{220}^*) \quad (3)$$

and

$$\text{MSE}(T_2) \cong S_y^4 (\theta \lambda_{400}^* + \theta'' \lambda_{040}^* - 2\theta'' \lambda_{220}^*). \quad (4)$$

We rewrite the linear regression estimator in terms of two-phase sampling, denoted by T_3 , proposed by [25], which can be expressed as

$$\hat{S}_{lr}^2 = s_y^2 + b_{(s_y^2, s_x^2)} (s_x^2 - s_x^2), \quad (5)$$

where

$$b_{(s_y^2, s_x^2)} = \frac{s_y^2 \hat{\lambda}_{220}^*}{s_x^2 \hat{\lambda}_{040}^*}$$

represents a sample regression coefficient.

For T_3 , the following equation gives the MSE expression:

$$\text{MSE}(T_3) \cong S_y^4 \lambda_{400}^* (\theta - \theta'' \rho_{yx}^*), \quad (6)$$

where

$$\rho_{yx}^* = \frac{\lambda_{220}^*}{\sqrt{\lambda_{400}^*} \sqrt{\lambda_{040}^*}}.$$

We rewrite the exponential ratio-type estimator T_4 proposed by [8] under two-phase sampling as follows:

$$T_4 = s_y^2 \exp \left(\frac{s_x^2 - s_x^2}{s_x^2 + s_x^2} \right). \quad (7)$$

For T_4 , the following equations give the bias and MSE expressions:

$$\text{Bias}(T_4) \cong \frac{1}{2} \theta'' S_y^2 \left(\frac{3\lambda_{040}^*}{4} - \lambda_{220}^* \right) \quad (8)$$

and

$$\text{MSE}(T_4) \cong S_y^4 \left[\theta \lambda_{400}^* + \theta'' \left(\frac{\lambda_{040}^*}{4} - \lambda_{220}^* \right) \right]. \quad (9)$$

We rewrite an improved ratio-type estimator T_5 proposed by [9] under two-phase sampling as follows:

$$T_5 = s_y^2 \left(\frac{s_x^2 + \lambda_{040}^*}{s_x^2 + \lambda_{040}^*} \right). \quad (10)$$

For T_5 , the following equations give the bias and MSE expressions:

$$\text{Bias}(T_5) \cong \theta'' g S_y^2 (g \lambda_{040}^* - \lambda_{220}^*) \quad (11)$$

and

$$\text{MSE}(T_5) \cong S_y^4 \left[\theta \lambda_{400}^* + \theta'' (g^2 \lambda_{040}^* - 2g \lambda_{220}^*) \right], \quad (12)$$

where

$$g = \frac{S_x^2}{S_x^2 + \lambda_{040}}.$$

We rewrite the exponential ratio-type estimators introduced by [11] under two-phase sampling as follows:

$$T_6 = s_y^2 \left(\frac{\hat{s}_x^2 + C_x}{s_x^2 + C_x} \right), \quad (13)$$

$$T_7 = s_y^2 \left(\frac{\lambda_{040} \hat{s}_x^2 + C_x}{\lambda_{040} s_x^2 + C_x} \right) \quad (14)$$

and

$$T_8 = s_y^2 \left(\frac{C_x \hat{s}_x^2 + \lambda_{040}}{C_x s_x^2 + \lambda_{040}} \right). \quad (15)$$

For $T_i (i = 6, 7, 8)$, the following equations give the bias and MSE expressions:

$$\text{Bias}(T_i) \cong \theta'' t_i S_y^2 (t_j \lambda_{040}^* - \lambda_{220}^*), \quad j = 0, 1, 2 \quad (16)$$

and

$$\text{MSE}(T_i) \cong S_y^4 \left[\theta \lambda_{400}^* + \theta'' (t_j^2 \lambda_{040}^* - 2t_i \lambda_{220}^*) \right], \quad (17)$$

where

$$\begin{aligned} t_0 &= \frac{S_x^2}{S_x^2 + C_x}, \\ t_1 &= \frac{\lambda_{040} S_x^2}{\lambda_{040} S_x^2 + C_x}, \\ t_2 &= \frac{C_x S_x^2}{C_x S_x^2 + \lambda_{040}}. \end{aligned}$$

3. Proposed Estimator

In this section, a class of double-transformed exponential estimators is established by utilizing known maximum and minimum data points of the auxiliary variable along with the ranks of the auxiliary variable in a two-phase sampling strategy. The following defines the suggested estimator:

$$\hat{S}_H^2 = s_y^2 \exp \left[v_1 \left\{ \frac{d_1 (\hat{s}_x^2 - s_x^2)}{d_1 (\hat{s}_x^2 + s_x^2) + 2d_2} \right\} \right] \exp \left[v_2 \left\{ \frac{d_3 (\hat{s}_r^2 - s_r^2)}{d_3 (\hat{s}_r^2 + s_r^2) + 2d_4} \right\} \right], \quad (18)$$

where $(d_i, i = 1, 2, 3, 4)$ are known parameters of variable X , and (v_1, v_2) are constant values of either 1 or 2. Notations (X_m, X_{M_h}) indicate the minimum and maximum auxiliary variable values, whereas (R_m, R_M) represent the minimum and maximum values of the auxiliary variable ranks. The transformation values of $(d_i, i = 1, 2, 3, 4)$ are described in Table 1, while Table 2 presents estimators using (18).

Table 1. Different known parameters of the auxiliary variables.

d_1	d_2	d_3	d_4
1	$X_M - X_m$	1	$R_M - R_m$
ρ_{yx}	$X_M - X_m$	ρ_{yr}	$R_M - R_m$
$X_M - X_m$	ρ_{yx}	$R_M - R_m$	ρ_{yr}
$X_M - X_m$	$\beta_{2(x)}$	$R_M - R_m$	$\beta_{2(r)}$
$X_M - X_m$	1	$R_M - R_m$	1
$\beta_{2(x)}$	$X_M - X_m$	$\beta_{2(r)}$	$R_M - R_m$
$X_M - X_m$	C_x	$R_M - R_m$	C_r
C_x	$X_M - X_m$	C_r	$R_M - R_m$

Table 2. Different classes of double exponential estimators.

$\hat{S}_{H1}^2 = s_y^2 \exp \left[v_1 \left\{ \frac{s_x^2 - s_x^2}{s_x^2 + s_x^2 + 2(X_M - X_m)} \right\} \right] \exp \left[v_2 \left\{ \frac{s_r^2 - s_r^2}{s_r^2 + s_r^2 + 2(R_M - R_m)} \right\} \right]$
$\hat{S}_{H2}^2 = s_y^2 \exp \left[v_1 \left\{ \frac{\rho_{yx}(s_x^2 - s_x^2)}{\rho_{yx}(s_x^2 + s_x^2) + 2(X_M - X_m)} \right\} \right] \exp \left[v_2 \left\{ \frac{\rho_{yr}(s_r^2 - s_r^2)}{\rho_{yr}(s_r^2 + s_r^2) + 2(R_M - R_m)} \right\} \right]$
$\hat{S}_{H3}^2 = s_y^2 \exp \left[v_1 \left\{ \frac{(X_M - X_m)(s_x^2 - s_x^2)}{(X_M - X_m)(s_x^2 + s_x^2) + 2\rho_{yx}} \right\} \right] \exp \left[v_2 \left\{ \frac{(R_M - R_m)(s_r^2 - s_r^2)}{(R_M - R_m)(s_r^2 + s_r^2) + 2\rho_{yr}} \right\} \right]$
$\hat{S}_{H4}^2 = s_y^2 \exp \left[v_1 \left\{ \frac{(X_M - X_m)(s_x^2 - s_x^2)}{(X_M - X_m)(s_x^2 + s_x^2) + 2\beta_{2(x)}} \right\} \right] \exp \left[v_2 \left\{ \frac{(R_M - R_m)(s_r^2 - s_r^2)}{(R_M - R_m)(s_r^2 + s_r^2) + 2\beta_{2(r)}} \right\} \right]$
$\hat{S}_{H5}^2 = s_y^2 \exp \left[v_1 \left\{ \frac{(X_M - X_m)(s_x^2 - s_x^2)}{(X_M - X_m)(s_x^2 + s_x^2) + 2} \right\} \right] \exp \left[v_2 \left\{ \frac{(R_M - R_m)(s_r^2 - s_r^2)}{(R_M - R_m)(s_r^2 + s_r^2) + 2} \right\} \right]$
$\hat{S}_{H6}^2 = s_y^2 \exp \left[v_1 \left\{ \frac{\beta_{2(x)}(s_x^2 - s_x^2)}{\beta_{2(x)}(s_x^2 + s_x^2) + 2(X_M - X_m)} \right\} \right] \exp \left[v_2 \left\{ \frac{\beta_{2(r)}(s_r^2 - s_r^2)}{\beta_{2(r)}(s_r^2 + s_r^2) + 2(R_M - R_m)} \right\} \right]$
$\hat{S}_{H7}^2 = s_y^2 \exp \left[v_1 \left\{ \frac{(X_M - X_m)(s_x^2 - s_x^2)}{(X_M - X_m)(s_x^2 + s_x^2) + 2C_x} \right\} \right] \exp \left[v_2 \left\{ \frac{(R_M - R_m)(s_r^2 - s_r^2)}{(R_M - R_m)(s_r^2 + s_r^2) + 2C_r} \right\} \right]$
$\hat{S}_{H8}^2 = s_y^2 \exp \left[v_1 \left\{ \frac{C_x(s_x^2 - s_x^2)}{C_x(s_x^2 + s_x^2) + 2(X_M - X_m)} \right\} \right] \exp \left[v_2 \left\{ \frac{C_r(s_r^2 - s_r^2)}{C_r(s_r^2 + s_r^2) + 2(R_M - R_m)} \right\} \right]$

3.1. Properties of the Proposed Estimator

The bias and mean squared error (MSE) of $\hat{S}_H^2$ are obtained by rewriting (18) in terms of errors in order to describe the characteristics of the double exponential class of estimators. For example,

$$\hat{S}_H^2 = s_y^2(1 + e_0) \exp \left[\frac{t_4(e_2 - e_1)}{2} \left(1 + \frac{t_4}{2}(e_1 + e_2) \right)^{-1} \right] \exp \left[\frac{t_5(e_4 - e_3)}{2} \left(1 + \frac{t_5}{2}(e_3 + e_4) \right)^{-1} \right] \quad (19)$$

where

$$v_1 = v_2 = 1,$$

$$t_4 = \frac{d_1 S_x^2}{d_1 S_x^2 + d_2}$$

and

$$t_5 = \frac{d_3 S_r^2}{d_3 S_r^2 + d_4}.$$

By using the Taylor series expansion up to the first order, we obtain

$$\begin{aligned} \hat{S}_H^2 - S_y^2 \cong S_y^2 \left[e_0 - \frac{t_4}{2}(e_1 - e_2) - \frac{t_5}{2}(e_3 - e_4) + \frac{3t_4^2}{8}e_1^2 - \frac{t_4^2}{8}e_2^2 \right. \\ \left. + \frac{3t_5^2}{8}e_3^2 - \frac{t_5^2}{8}e_4^2 - \frac{t_4}{2}e_0e_1 + \frac{t_4}{2}e_0e_2 - \frac{t_5}{2}e_0e_3 + \frac{t_5}{2}e_0e_4 - \frac{t_4^2}{2}e_1e_2 \right. \\ \left. + \frac{t_4t_5}{4}e_1e_3 - \frac{t_4t_5}{4}e_1e_4 - \frac{t_4t_5}{4}e_2e_3 + \frac{t_4t_5}{4}e_2e_4 - \frac{t_5^2}{2}e_3e_4 \right]. \end{aligned} \quad (20)$$

Using (20), the bias of $\hat{S}_{due}^2$ is given by

$$\begin{aligned} Bias(\hat{S}_H^2) \cong \theta S_y^2 \left[\frac{3t_4^2}{8}\lambda_{040}^* + \frac{3t_5^2}{8}\lambda_{004}^* - \frac{t_4}{2}\lambda_{220}^* - \frac{t_5}{2}\lambda_{202}^* + \frac{t_4t_5}{2}\lambda_{022}^* \right] \\ - \theta' S_y^2 \left[\frac{3t_4^2}{8}\lambda_{040}^* + \frac{3t_5^2}{8}\lambda_{004}^* - \frac{t_4}{2}\lambda_{220}^* - \frac{t_5}{2}\lambda_{202}^* + \frac{t_4t_5}{2}\lambda_{022}^* \right]. \end{aligned} \quad (21)$$

After some simplifications, we obtain

$$Bias(\hat{S}_H^2) \cong \theta'' S_y^2 \left[\frac{3}{8}(t_4^2\lambda_{040}^* + t_5^2\lambda_{004}^*) - \frac{1}{2}(t_4\lambda_{220}^* + t_5\lambda_{202}^* - t_4t_5\lambda_{022}^*) \right], \quad (22)$$

where

$$\theta'' = \theta - \theta'.$$

We derived an MSE of $\hat{S}_H^2$ by squaring both sides of (20) and applying the expectation, which is expressed by the following equation, i.e.,

$$\begin{aligned} MSE(\hat{S}_H^2) \cong \theta S_y^4 \left[\lambda_{400}^* + \frac{t_4^2}{4}\lambda_{040}^* + \frac{t_5^2}{4}\lambda_{004}^* - t_4\lambda_{220}^* - t_5\lambda_{202}^* + \frac{t_4t_5}{2}\lambda_{022}^* \right] \\ - \theta' S_y^4 \left[\frac{t_4^2}{4}\lambda_{040}^* + \frac{t_5^2}{4}\lambda_{004}^* - t_4\lambda_{220}^* - t_5\lambda_{202}^* + \frac{t_4t_5}{2}\lambda_{022}^* \right]. \end{aligned}$$

After some simplifications, we obtain

$$MSE(\hat{S}_H^2) \cong S_y^4 \left[\theta\lambda_{400}^* + \frac{\theta''}{4}(t_4^2\lambda_{040}^* + t_5^2\lambda_{004}^* - 4t_4\lambda_{220}^* - 4t_5\lambda_{202}^* + 2t_4t_5\lambda_{022}^*) \right]. \quad (23)$$

3.2. Analysis of Continuity, Boundedness, and Partial Derivative Properties of the Proposed Estimator

To check the continuity, the boundedness, and the existence and continuity of the first- and second-order partial derivatives of the proposed estimator (18), we need to carefully analyze the properties of the estimator function.

Step 1: Restating the proposed estimator

The proposed estimator for the finite population variance S_y^2 is given below:

$$\hat{S}_H^2 = s_y^2 \exp \left[v_1 \left\{ \frac{d_1(\hat{s}_x^2 - s_x^2)}{d_1(\hat{s}_x^2 + s_x^2) + 2d_2} \right\} \right] \exp \left[v_2 \left\{ \frac{d_3(\hat{s}_r^2 - s_r^2)}{d_3(\hat{s}_r^2 + s_r^2) + 2d_4} \right\} \right],$$

where

- s_y^2, s_x^2 , and s_r^2 are sample variances from the second-phase sample.
- $\hat{s}_x^2, \hat{s}_r^2$ are sample variances from the first-phase sample.
- d_1, d_2, d_3, d_4 , are known constants, while v_1, v_2 are constants, either 1 or 2.

Step 2: Continuity of the function

Exponential function behavior: The exponential function $\exp(x)$ is well known for be-

ing continuous for all real values of x . This property holds true for the terms inside the exponents of the proposed estimator. Let us analyze the expression:

$$\exp \left[v_1 \left\{ \frac{d_1(\hat{s}_x^2 - s_x^2)}{d_1(\hat{s}_x^2 + s_x^2) + 2d_2} \right\} \right]$$

and

$$\exp \left[v_2 \left\{ \frac{d_3(\hat{s}_r^2 - s_r^2)}{d_3(\hat{s}_r^2 + s_r^2) + 2d_4} \right\} \right].$$

Since the exponential function is continuous for any real-valued input, we only need to ensure that the terms inside the exponents are continuous functions.

Continuity of the expressions Inside the exponentials: Each term inside the exponentials has the following form:

$$\frac{d_1(\hat{s}_x^2 - s_x^2)}{d_1(\hat{s}_x^2 + s_x^2) + 2d_2}, \quad \frac{d_3(\hat{s}_r^2 - s_r^2)}{d_3(\hat{s}_r^2 + s_r^2) + 2d_4},$$

these are rational functions (fractions), and rational functions are continuous wherever the denominator is non-zero. The denominators are $d_1(\hat{s}_x^2 + s_x^2) + 2d_2$ and $d_3(\hat{s}_r^2 + s_r^2) + 2d_4$, which are always positive (since s_x^2 and s_r^2 are variances and therefore non-negative), as long as d_2, d_4 are positive constants. Thus, the rational functions inside the exponentials are continuous. Therefore, the function $\hat{S}_H^2$ is continuous.

Step 3: Boundedness of the Function

For boundedness, we need to ensure that $\hat{S}_H^2$ does not approach infinity for any valid values of the sample variances s_x^2, s_y^2 , and s_r^2 .

1. Exponential terms: The exponential terms are bounded because of the following points:

- The ratios $\frac{d_1(\hat{s}_x^2 - s_x^2)}{d_1(\hat{s}_x^2 + s_x^2) + 2d_2}$ and $\frac{d_3(\hat{s}_r^2 - s_r^2)}{d_3(\hat{s}_r^2 + s_r^2) + 2d_4}$ are finite as long as s_x^2, s_r^2 are finite, which is guaranteed since they are sample variances.
- The exponential function of a real-valued number is always finite and bounded.

2. Multiplying by s_y^2 : Since s_y^2 is the sample variance, it is also finite and non-negative. Thus, the estimator $\hat{S}_H^2$ is bounded.

Step 4: Existence and continuity of partial derivatives

- **First-Order Partial Derivatives:** The first-order partial derivatives of $\hat{S}_H^2$ with respect to s_x^2, s_y^2 , and s_r^2 exist because the exponential function and the rational functions inside the exponents are smooth and differentiable. Thus, these partial derivatives are continuous.
- **Second-Order Partial Derivatives:** Similarly, the second-order partial derivatives exist and are continuous because the functions involved (exponentials and rational functions) are smooth and their derivatives are bounded.

Therefore, from the above observation, we can conclude that the function $\hat{S}_H^2$ is continuous and bounded. Moreover, the first- and second-order partial derivatives of $\hat{S}_H^2$ with respect to the sample variances exist, are continuous, and are bounded.

4. Mathematical Comparison

The suggested double exponential class of estimators $\hat{S}_H^2$ is compared and analyzed in this section with a number of other estimators, including T_1, T_2, T_3, T_4, T_5 , and $T_i (i = 6, 7, 8)$. Condition (1): By (1) and (23),

$$Var(T_1) > MSE(\hat{S}_H^2) \text{ if}$$

$$(\theta - \theta') \left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) < 0.$$

For $\theta - \theta' > 0$, that is, $\theta > \theta'$:

$$\left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) < 0. \quad (24)$$

Similarly, $\theta - \theta' < 0$, that is, $\theta < \theta'$:

$$\left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) > 0. \quad (25)$$

When (24) or (25) are satisfied, the estimator $\hat{S}_H^2$ performs well compared to $Var(T_1)$.
Condition (2): By (4) and (23),

$$MSE(T_2) > MSE(\hat{S}_H^2) \text{ if}$$

$$(\theta - \theta') \left[\lambda_{040}^* - 2\lambda_{022}^* - \frac{1}{4} \left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) \right] > 0.$$

For $\theta - \theta' < 0$, that is, $\theta < \theta'$:

$$\left[\lambda_{040}^* - 2\lambda_{022}^* - \frac{1}{4} \left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) \right] > 0. \quad (26)$$

Similarly $\theta - \theta' > 0$, that is, $\theta > \theta'$:

$$\left[\lambda_{040}^* - 2\lambda_{022}^* - \frac{1}{4} \left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) \right] < 0. \quad (27)$$

When (26) or (27) are satisfied, the estimator $\hat{S}_H^2$ performs well compared to $MSE(T_2)$.
Condition (3): By (6) and (23),

$$MSE(T_3) > MSE(\hat{S}_H^2) \text{ if}$$

$$(\theta - \theta') \left[\rho_{yx}^{*2} + \frac{1}{4} \left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) \right] < 0.$$

For $\theta - \theta' > 0$, that is, $\theta > \theta'$:

$$\left[\rho_{yx}^{*2} + \frac{1}{4} \left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) \right] < 0. \quad (28)$$

Similarly $\theta - \theta' < 0$, that is, $\theta < \theta'$:

$$\left[\rho_{yx}^{*2} + \frac{1}{4} \left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) \right] > 0. \quad (29)$$

When (28) or (29) are satisfied, the estimator $\hat{S}_H^2$ performs well compared to $MSE(T_3)$.
Condition (4): By (9) and (23),

$$MSE(T_4) > MSE(\hat{S}_H^2) \text{ if}$$

$$(\theta - \theta') \left[\left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) - (\lambda_{040}^* - 4\lambda_{220}^*) \right] < 0.$$

For $\theta - \theta' > 0$, that is, $\theta > \theta'$:

$$\left[\left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) - (\lambda_{040}^* - 4\lambda_{220}^*) \right] < 0. \quad (30)$$

Similarly $\theta - \theta' < 0$, that is, $\theta < \theta'$:

$$\left[\left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) - (\lambda_{040}^* - 4\lambda_{220}^*) \right] > 0. \quad (31)$$

When (30) or (31) are satisfied, the estimator $\hat{S}_H^2$ is performs well compared to $MSE(T_4)$.

Condition (5): By (12) and (23),

$$MSE(T_5) > MSE(\hat{S}_H^2) \text{ if}$$

$$(\theta - \theta') \left[\left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) - \left(4g^2 \lambda_{040}^* - 4g \lambda_{220}^* \right) \right] < 0.$$

For $\theta - \theta' > 0$, that is, $\theta > \theta'$:

$$\left[\left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) - \left(4g^2 \lambda_{040}^* - 4g \lambda_{220}^* \right) \right] < 0. \quad (32)$$

Similarly $\theta - \theta' < 0$, that is, $\theta < \theta'$:

$$\left[\left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) - \left(4g^2 \lambda_{040}^* - 4g \lambda_{220}^* \right) \right] > 0. \quad (33)$$

When (32) or (33) are satisfied, the estimator $\hat{S}_H^2$ performs well as compared to $MSE(T_5)$.

Condition (6): By (17) and (23),

$$MSE(T_i) > MSE(\hat{S}_H^2) \text{ if}$$

$$(\theta - \theta') \left[\left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) - \left(4t_i^2 \lambda_{040}^* - 4t_i \lambda_{220}^* \right) \right] < 0.$$

For $\theta - \theta' > 0$, that is, $\theta > \theta'$:

$$\left[\left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) - \left(t_i^2 \lambda_{040}^* - 4t_i \lambda_{220}^* \right) \right] < 0. \quad (34)$$

Similarly $\theta - \theta' < 0$, that is, $\theta < \theta'$:

$$\left[\left(t_4^2 \lambda_{040}^* + t_5^2 \lambda_{004}^* - 4t_4 \lambda_{220}^* - 4t_5 \lambda_{202}^* + 2t_4 t_5 \lambda_{022}^* \right) - \left(t_i^2 \lambda_{040}^* - 4t_i \lambda_{220}^* \right) \right] > 0. \quad (35)$$

When (34) or (35) are satisfied, the estimator $\hat{S}_H^2$ performs well compared to $MSE(T_i)$.

5. Numerical Analysis

In this section, we compare our proposed double exponential class of estimators to existing ones by utilizing the population obtained by simulation and three different real data sets in terms of the percent relative efficiency (PRE). This study uses three distinct real and simulated data sets. Our goal is to provide a comprehensive evaluation of the effectiveness as well as the reliability of the suggested class of estimators in various types of practical situations using the PRE criterion on these simulated and different data sets.

5.1. Simulation Study

We perform a simulation analysis to investigate the performance of the suggested double exponential estimators, which are based on the known largest and smallest values of

the auxiliary variable and its ranks in a two-phase sampling framework, in order to validate our theoretical findings introduced in Section 4. The auxiliary variable X is artificially produced in this simulation for six various populations, each of which correspond to a probability distribution that is discussed below.

- Data 1: $X \sim \text{Uniform}(3, 5)$.
- Data 2: $X \sim \text{Uniform}(6, 9)$.
- SData 3: $X \sim \text{Exponential}(5)$.
- Data 4: $X \sim \text{Exponential}(8)$.
- Data 5: $X \sim \text{Gamma}(2, 6)$.
- Data 6: $X \sim \text{Gamma}(7, 10)$.

In each distribution, the dependent variable Y is obtained by employing the following formula:

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

where X is selected from each population distribution,

$$r_{yx} = 0.70$$

is a correlation coefficient, and

$$e \sim N(0, 1)$$

denotes a random error term.

Using R-Software (latest v. 4.4.0), we developed the following procedures to calculate the mean squared error (MSE) as well as the percent relative efficiency (PRE) of the suggested class of estimators (Algorithm 1):

Algorithm 1 Procedure for Calculating the Mean Squared Error (MSE) and Percent Relative Efficiency (PRE)

Step 1: Using the given probability distributions, obtain a population Ψ of size $N = 1500$.

Step 2: To choose a first-phase sample S_1 of size m from Ψ by applying simple random sampling without replacement (SRSWOR).

Step 3: Using SRSWOR, choose a second-phase sample S_2 of size n from S_1 .

Step 4: From the above steps, find the optimal values for unknown constants in estimators by computing population total, X_M , X_m , R_M , and R_m .

Step 5: Apply a two-phase sampling design as described in **Step 2** and **Step 3**.

Step 6: For each sample size, compute PRE ($\hat{S}_i^2$) for each described estimator $\hat{S}_i^2$.

Step 7: Use the following formulas to calculate the percent relative efficiency values for each estimator after repeating **Steps 5** and **6** a total of 40,000 times. The findings for the simulated populations are shown in Table 3.

$$MSE(\hat{S}_i^2)_{\min} = \frac{\sum_{j=1}^{40000} (\hat{S}_i^2 - S_y^2)^2}{40000}$$

and

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

where $\hat{S}_i^2$ is one of $T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, H_k (k = 1, 2, \dots, 8)$.

Table 3. PRE values using different artificial populations.

Estimator	Uni(3,5)	Uni(6,9)	Exp(5)	Exp(8)	Gam(2,6)	Gam(7,10)
T_1	100	100	100	100	100	100
T_2	136.645	133.902	143.443	138.506	150.209	137.278
T_3	145.523	144.483	165.609	152.625	152.889	138.745
T_4	152.289	156.120	158.701	154.200	151.774	136.285
T_5	153.448	157.392	162.209	156.526	159.130	139.133
T_6	155.552	159.409	166.989	158.789	154.651	132.807
T_7	155.526	158.762	165.335	158.002	153.299	130.218
T_8	154.798	153.580	163.290	159.328	151.416	136.809
H_1	214.203	200.120	180.309	195.139	187.505	170.274
H_2	289.387	262.560	210.279	210.508	179.035	163.549
H_3	250.770	260.293	223.031	225.625	161.281	155.770
H_4	205.560	205.449	203.971	209.663	160.441	153.939
H_5	219.654	222.517	240.351	230.323	165.710	160.171
H_6	230.921	250.810	220.091	227.688	169.930	166.529
H_7	283.228	283.941	243.339	233.167	162.019	164.905
H_8	294.497	282.445	288.167	289.910	190.583	183.270

5.2. Numerical Examples

Using different data sets, we computed the percent relative efficiency all the estimators in order to evaluate their performance with our suggested double class of exponential type estimators. Below are detailed explanations of the data sets, with descriptive statistics given in Tables 4–6.

Data 1. (Source: [26], p. 226)

Y: The employment levels that each department records, which indicate the total number of employees;

X: Information on industrial activities, including the total number of factories in which these departments are formally registered;

R: A comparative perspective of industrial participation across departments is provided by the rankings given to each department based on the total number of factories they registered in 2012.

Data 2. (Source: [3], p. 24)

Y: The family's food bills, which have a direct connection to their job;

X: The entire weekly revenue received by the family, which represents their available funds during that period of time;

R: Ranking of families based on their weekly income.

Data 3. (Source: [26], p. 135)

Y: Total student enrollment at institutions in 2012;

X: Number of government institutions in 2012;

R: Ranking of government institutions for 2012.

Table 4. Descriptive statistics for data 1.

$N = 36$	$m = 15$	$n = 6$	$X_m = 39$
$X_M = 2379$	$R_m = 1$	$R_M = 36$	$S_y = 8512.266$
$S_x = 3410.428$	$S_r = 10.53$	$\bar{X} = 1659.58$	$\bar{Y} = 4015.22$
$\bar{R} = 18.50$	$C_y = 2.12$	$C_x = 2.06$	$C_r = 0.57$
$\rho_{yx} = 0.65$	$\rho_{yr} = 0.36$	$\rho_{xr} = 0.75$	$\lambda_{400} = 2051$
$\lambda_{040} = 2237$	$\lambda_{004} = 3297$	$\lambda_{220} = 1542$	$\lambda_{202} = 2698$
$\lambda_{022} = 3698$	$\theta = 0.14$	$\theta' = 0.04$	$\theta'' = 0.10$

Table 5. Descriptive statistics for data 2.

$N = 33$	$m = 15$	$n = 6$	$X_m = 58$
$X_M = 95$	$R_m = 1.5$	$R_M = 33$	$S_y = 10.13$
$S_r = 9.64$	$S_x = 10.58$	$\bar{X} = 72.55$	$\bar{Y} = 27.49$
$\bar{R} = 17$	$C_y = 0.37$	$C_x = 0.14$	$C_r = 0.57$
$\rho_{yx} = 0.25$	$\rho_{yr} = 0.20$	$\rho_{xr} = 0.98$	$\lambda_{400} = 5.55$
$\lambda_{040} = 2.08$	$\lambda_{004} = 1.10$	$\lambda_{220} = 2.22$	$\lambda_{202} = 0.94$
$\lambda_{022} = 1.54$	$\theta = 0.14$	$\theta' = 0.04$	$\theta'' = 0.10$

Table 6. Descriptive statistics for data 3.

$N = 36$	$m = 15$	$n = 6$	$X_m = 39$
$X_M = 2370$	$R_m = 1.5$	$R_M = 36$	$S_y = 32,601.14$
$S_r = 10.54$	$S_x = 2214.219$	$\bar{X} = 1054.39$	$\bar{Y} = 14,818.70$
$\bar{R} = 18.50$	$C_y = 22.07$	$C_x = 2.10$	$C_r = 0.56$
$\rho_{yx} = 0.69$	$\rho_{yr} = 0.39$	$\rho_{xr} = 0.84$	$\lambda_{400} = 236.50$
$\lambda_{040} = 209.80$	$\lambda_{004} = 329.70$	$\lambda_{220} = 397.50$	$\lambda_{202} = 365.80$
$\lambda_{022} = 469.80$	$\theta = 0.14$	$\theta' = 0.04$	$\theta'' = 0.10$

Finally, the following formula is considered to obtain the percent relative efficiency values for different estimators using the above data sets.

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

where $\hat{S}_k^2$ is one of $T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, H_i (i = 1, 2, \dots, 8)$.

6. Conclusions

In this research, we established an improved and effective double exponential estimator for population variance that use ranks of the auxiliary variable and information on the minimum and maximum values. A comparison with existing estimators is made straightforward by the theoretical assumptions that show how these proposed estimators are more efficient, as detailed in Section 4. To verify these theoretical assumptions, we conducted simulation studies and analyzed various types of real data sets. According to the results, which are shown in Table 3, the new estimators are consistently better than existing approaches in terms of PREs. These conclusions and the theoretical ideas from Section 4 are supported by the empirical results displayed in Table 7.

Table 7. PRE values using real data sets.

Estimator	Pop-1	Pop-2	Pop-3
T_1	100	100	100
T_2	141.830	143.664	121.469
T_3	143.225	104.673	125.532
T_4	152.034	128.010	115.634
T_5	141.858	143.509	121.380
T_6	141.839	143.655	121.460
T_7	141.839	143.660	121.460
T_8	141.849	142.044	121.469
H_1	167.220	146.790	165.289
H_2	170.456	150.281	172.520
H_3	169.723	147.00	168.254
H_4	167.489	146.520	166.768
H_5	164.056	146.526	165.879
H_6	158.309	146.531	160.446
H_7	166.615	146.923	170.236
H_8	172.869	151.402	177.216

We can conclude from the simulation and empirical data that the proposed estimators $\hat{S}_{H_k}^2$ ($k = 1, 2, \dots, 8$) are more precise and efficient compared to other estimators that were discussed in this paper. Among the proposed estimators, $\hat{S}_{H_8}^2$ is the most effective choice in the proposed classes of estimators and is therefore highly recommended.

We discussed the performance and properties of the new recommended double exponential family of estimators under a double-phase sampling method. Furthermore, the systematic sampling approach may be used to propose novel estimators, and our findings can help in the development of estimators with minimized MSE. This presents an intriguing direction for future research.

Author Contributions: Conceptualization, U.D., D.A., J.W. and W.E.; methodology, U.D., D.A., J.W. and W.E.; software, U.D., D.A. and J.W.; validation, U.D., D.A., J.W. and W.E.; formal analysis, U.D. and D.A.; investigation, U.D., D.A. and J.W.; resources, U.D., D.A. and J.W.; data curation, U.D., D.A., J.W. and W.E.; writing—original draft, U.D. and D.A.; writing—review & editing, U.D., D.A. and J.W.; visualization, U.D., D.A. and J.W.; supervision, U.D.; project administration, U.D., D.A. and W.E.; funding acquisition, W.E. All authors have read and agreed to the published version of the manuscript.

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

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Neyman, J. Contribution to the theory of sampling human population. *J. Am. Stat. Assoc.* **1938**, *33*, 101–116. [[CrossRef](#)]
2. Sukhatme, B.V. Some ratio-type estimators in two-phase sampling. *J. Am. Stat. Assoc.* **1962**, *57*, 628–632. [[CrossRef](#)]
3. Cochran, W.B. *Sampling Techniques*; John Wiley and Sons: Hoboken, NJ, USA, 1963.

4. Isaki, C.T. Variance estimation using auxiliary information. *J. Am. Stat. Assoc.* **1983**, *78*, 117–123. [\[CrossRef\]](#)
5. Rao, J.N.K. On double sampling for stratification and analytical surveys. *Biometrika* **1973**, *60*, 125–33. [\[CrossRef\]](#)
6. Vishwakarma, G.K.; Zeeshan, S.M. Generalized ratio-cum-product estimator for finite population mean under two-phase sampling scheme. *J. Mod. Appl. Stat. Methods* **2020**, *19*, 7. [\[CrossRef\]](#)
7. Zaman, T.; Kadilar, C. New class of exponential estimators for finite population mean in two-phase sampling. *Commun. Stat. Theory Methods* **2021**, *50*, 874–889. [\[CrossRef\]](#)
8. Bahl, S.; Tuteja, R. Ratio and product type exponential estimators. *J. Inf. Optim. Sci.* **1991**, *12*, 159–164. [\[CrossRef\]](#)
9. Upadhyaya, L.; Singh, H. An estimator for population variance that utilizes the kurtosis of an auxiliary variable in sample surveys. *Vikram Math. J.* **1999**, *19*, 14–17.
10. Dubey, V.; Sharma, H. On estimating population variance using auxiliary information. *Stat. Transit. New Ser.* **2008**, *9*, 7–18.
11. Kadilar, C.; Cingi, H. Ratio estimators for the population variance in simple and stratified random sampling. *Appl. Math. Comput.* **2006**, *173*, 1047–1059. [\[CrossRef\]](#)
12. Singh, H.; Chandra, P. An alternative to ratio estimator of the population variance in sample surveys. *J. Transp. Stat.* **2008**, *9*, 89–103.
13. Singh, H.P.; Solanki, R.S. A new procedure for variance estimation in simple random sampling using auxiliary information. *J. Stat. Pap.* **2013**, *54*, 479–497. [\[CrossRef\]](#)
14. Singh, H.P.; Pal, S.K. An efficient class of estimators of finite population variance using quartiles. *J. Appl. Stat.* **2016**, *43*, 1945–1958. [\[CrossRef\]](#)
15. Zaman, T.; Bulut, H. An efficient family of robust-type estimators for the population variance in simple and stratified random sampling. *Commun. Stat. Theory Methods* **2023**, *52*, 2610–2624. [\[CrossRef\]](#)
16. Mohanty, S.; Sahoo, J. A note on improving the ratio method of estimation through linear transformation using certain known population parameters. *Sankhyā Indian J. Stat. Ser. B* **1995**, *57*, 93–102.
17. Khan, M.; Shabbir, J. Some improved ratio, product, and regression estimators of finite population mean when using minimum and maximum values. *Sci. World J.* **2013**, *2013*, 431868. [\[CrossRef\]](#)
18. Daraz, U.; Shabbir, J.; Khan, H. Estimation of finite population mean by using minimum and maximum values in stratified random sampling. *J. Mod. Appl. Stat. Methods* **2018**, *17*, 20. [\[CrossRef\]](#)
19. Daraz, U.; Wu, J.; Albalawi, O. Double exponential ratio estimator of a finite population variance under extreme values in simple random sampling. *Mathematics* **2024**, *12*, 1737. [\[CrossRef\]](#)
20. Alomair, M.A.; Daraz, U. Dual transformation of auxiliary variables by using outliers in stratified random sampling. *Mathematics* **2024**, *12*, 2829. [\[CrossRef\]](#)
21. Daraz, U.; Alomair, M.A.; Albalawi, O.; Al Naim, A.S. New Techniques for Estimating Finite Population Variance Using Ranks of Auxiliary Variable in Two-Stage Sampling. *Mathematics* **2024**, *12*, 2741. [\[CrossRef\]](#)
22. Cekim, H.O.; Cingi, H. 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.* **2017**, *46*, 685–694.
23. Chatterjee, S.; Hadi, A.S. *Regression Analysis by Example*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
24. Walia, G.S.; Kaur, H.; Sharma, M. Ratio type estimator of population mean through efficient linear transformation. *Am. J. Math. Stat.* **2015**, *5*, 144–149.
25. Watson, D.J. The estimation of leaf area in field crops. *J. Agric. Sci.* **1937**, *27*, 474–483. [\[CrossRef\]](#)
26. Bureau of Statistics. *Punjab Development Statistics*; Government of the Punjab: Lahore, Pakistan, 2013.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.