*mathematics*

MDPI

*Article*

# A Comparative Study of New Ratio-Type Family of Estimators Under Stratified Two-Phase Sampling

**Abdulaziz S. Alghamdi [1] and Hleil Alrweili [2,*]**

[1] Department of Mathematics, College of Science & Arts, King Abdulaziz University, P.O. Box 344, Rabigh 21911, Saudi Arabia; ashalghamedi@kau.edu.sa
[2] Department of Mathematics, College of Science, Northern Border University, Arar 91431, Saudi Arabia
* Correspondence: hleil.alrweili@nbu.edu.sa

**Abstract:** Two-phase sampling is a useful technique for sample surveys, particularly when prior auxiliary data is not accessible. The ranks of the auxiliary variable often coincide with those of the research variable when two variables are correlated. By considering this relationship, we can significantly increase estimator accuracy. In this paper, we use the ranks of the auxiliary variable along with extreme values to estimate the population mean of the study variable. Up to a first-order approximation, we analyze the characteristics of the suggested class of estimators with an emphasis on biases and mean squared errors in stratified two-phase sampling. The theoretical results are verified using different datasets and a simulation study, which demonstrates that the proposed estimators outperform the existing ones in terms of percent relative efficiency.

**Keywords:** mean estimation; exponential estimators; stratified two-phase sampling; ranks; outliers; simulation; percent relative efficiency; mean absolute percentage error

**MSC:** 62D05

check for updates

Academic Editors: Yuhlong Lio and Heng Lian

Received: 15 November 2024
Revised: 5 January 2025
Accepted: 12 January 2025
Published: 21 January 2025

**Citation:** Alghamdi, A.S.; Alrweili, H. A Comparative Study of New Ratio-Type Family of Estimators Under Stratified Two-Phase Sampling. *Mathematics* **2025**, *13*, 327. https://doi.org/10.3390/math13030327

**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

To increase the accuracy of estimators, supplementary data are helpful in survey sampling. Estimators like the ratio, regression, and product-type depend on information from one or more auxiliary variables in addition to study variable data to increase their relative efficiency. Auxiliary variables such as the number of family members and total expenditure can be used, for example, to estimate total household income. Numerous studies have been carried out in an effort to develop new and enhanced estimators for population mean, total, median, and other overall characteristics. For additional information on these enhanced estimators and their characteristics, we refer to [1–5] and the references given therein.

Using auxiliary variables together with the study variable is a usual procedure in sampling theory. This helps to improve the design and the efficiency of estimators by focusing on the relationship between the two variables. However, before conducting the survey, the total population mean is often unknown in many practical applications. In such cases, it is preferable to use a two-phase sampling procedure. This method, also known as double sampling, involves selecting a sample from the population randomly in two distinct phases. It is commonly used in situations where conducting an initial phase prior to the main sampling phase is more efficient or cost effective. To collect data on auxiliary variables, two-phase sampling obtains a large initial sample from the population. The research variable and auxiliary data are observed in a second-phase sample that is either chosen from within the first-phase sample or drawn randomly from the population. This

method is a cost-effective sampling strategy, making it highly valuable in sample surveys, particularly when auxiliary information is not available in advance. Firstly, the idea of two-phase sampling was studied by [6]. Because of its ability to screen variables efficiently and at a lower cost, two-phase sampling has recently gained more interest. Studies on this method include works by [7–15].

Sampling survey data can sometimes contain outlier observations. When extreme values are included in the sample, the mean estimator may become skewed, leading to biased results. The two estimators based on a linear transformation utilizing the minimum and maximum information of known auxiliary variables were initially introduced by [16]. However, research on such designs was limited until [17] revisited the topic. By combining information on known extreme values, they suggested better ratio, product, and regression estimators for population means. In the presence of extreme values, [18] improved the estimation of the finite population mean within a two-phase sampling framework. Using the extreme values of auxiliary variables, [19] offered several types of modifications to estimate the finite population mean. A stratified random sampling strategy that takes extreme values into consideration was used by [20] to make great progress in estimating the finite population mean. A new family of estimators for population variance was also proposed by [21], which uses extreme values to minimize mean squared errors. In a more recent work, ref. [22] improved population variance estimation by transforming extreme values into many effective classes of estimators. For further information, see [23–29] and the references given therein.

It may seem appropriate to remove outlier observations from sample data, but it is important to understand that the presence of extreme values frequently results in a decline in the mean squared error *(MSE)* efficiency of classical estimators. This article uses these extreme values as auxiliary information rather than excluding them. We introduce a new improved class of estimators that use the lowest and maximum values of the auxiliary variable and ranks of the auxiliary variable to estimate the finite population mean in stratified two-phase sampling, improving on the technique of [22,27]. There are some advantages of our new proposed class of estimators in practical applications which are given below:

- By utilizing the extreme values and rankings of auxiliary variables, the new proposed class of estimators improves precision and efficiency when estimating the finite population mean. This new method to improve performance is especially valuable in real-world applications where survey data may contain outliers.
- 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.
- 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.

## 2. Practical Applications of the Proposed Estimators

The improved estimators have significant potential for application in various field studies and specific domains, such as economics and public health. These estimators can be directly applied to real-world scenarios where accurate predictions are essential.

- **Economics:** The estimators can greatly enhance economic forecasting. For example, by improving predictions of key economic indicators such as inflation, unemployment rates, and GDP growth, the proposed estimators can provide more reliable forecasts.

This, in turn, supports better-informed policy decisions and efficient resource allocation in both the public and private sectors.

- **Public Health:** In the domain of public health, the estimators can improve epidemiological modeling and health outcomes prediction. More accurate estimations can assist in forecasting the spread of diseases, evaluating healthcare demands, and guiding effective interventions. The ability to predict health trends with greater precision helps public health agencies allocate resources more efficiently and plan for future healthcare needs.
- **Environmental science:** The proposed estimators can also be applied in environmental science, particularly for climate modeling and pollution prediction. By providing more accurate estimates of environmental trends, the estimators can help predict long-term climate changes, pollution levels, and their potential impact on ecosystems. This can guide policy-making, disaster preparedness, and resource management.
- **Finance:** In finance, the estimators have applications in risk management and portfolio optimization. By improving the accuracy of financial models, the estimators can better forecast market trends, assess investment risks, and help in constructing more efficient portfolios. This, in turn, can enhance decision-making in asset management and financial planning.
- **Supply chain management:** The estimators can also be useful in supply chain management, where accurate demand forecasting is crucial. By applying the proposed estimators to predict future product demand or supply shortages, companies can better plan production schedules, optimize inventory levels, and improve logistics efficiency.

By applying these improved estimators across various fields, researchers and professionals can achieve more accurate models, leading to more effective solutions, better decision-making, and the more efficient use of resources.

The contents of this paper are set up as follows: The methods and notations used in this study are described completely in Section 3. An overview of existing estimators is given in Section 4. For the suggested class of estimators, we provide a detailed discussion in Section 5. A comprehensive mathematical comparison of these estimators is presented in Section 6. The simulation study described in Section 7 verifies the theoretical results defined in Section 6 by generating six different artificial populations using different probability distributions. Numerical examples are given as well in this section in order to highlight how our theoretical results are applied in practical situations. Finally, Section 9 discusses the main conclusions and proposes directions for future research.

## 3. Definitions and Symbols

Consider a finite population $\Omega = (\Omega_1, \Omega_2, \Omega_3, \ldots, \Omega_N)$ consisting of $N$ units. The population is segmented into $L$ strata, where each stratum has a size of $N_h$ for $h = 1, 2, \ldots, L$ such that the sum of all strata sizes equals the total population $N$, i.e., $\sum_{h=1}^{L} N_h = N$. Let $u_{hi}$ denote the value of the auxiliary variable $U$, $w_{hi}$ denote the rank of the auxiliary variable $W$, and $v_i$ denote the value of the study variable $V$ in the $hth$ stratum of the $ith (i = 1, 2, \ldots, N_h)$ unit, respectively.

In this paper, we suggested an improved class of estimators to estimate the finite population mean $\bar{V}$ of $V$ in the presence of auxiliary variable $U$. The definition of the two-phase sampling scheme is:

1. In the first stage, a sample of size $(m_h < N_h)$ is selected to estimate the population mean $\bar{U}$.
2. A sample size of $(n_h < m_h)$ is selected for the second stage in order to observe $v$ and $u$, respectively.

Consider the population means of the study variable ($V$), the auxiliary variable ($U$), and the auxiliary variable ranks ($W$), which are defined as:

$$\bar{V} = \sum_{h=1}^{L} \frac{Z_h \bar{V}_h}{N_h},$$

$$\bar{U} = \sum_{h=1}^{L} \frac{Z_h \bar{U}_h}{N_h},$$

$$\bar{W} = \sum_{h=1}^{L} \frac{Z_h \bar{W}_h}{N_h},$$

where

$$\bar{V}_h = \sum_{i=1}^{N_h} \frac{V_{hi}}{N_h},$$

$$\bar{U}_h = \sum_{i=1}^{N_h} \frac{U_{hi}}{N_h}$$

and

$$\bar{W}_h = \sum_{i=1}^{N_h} \frac{W_{hi}}{N_h},$$

denote the population means of the study variable, the auxiliary variable, and the auxiliary variable ranks in the *hth* stratum, while $Z_h$ is the stratum weight and is defined by:

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

Let the population variances (without replacement sampling) of $V, U$, and $W$ in the *hth* stratum be defined as:

$$\Psi_{V_h}^2 = \sum_{i=1}^{N_h} \frac{(V_{hi} - \bar{V}_h)^2}{N_h - 1},$$

$$\Psi_{U_h}^2 = \sum_{i=1}^{N_h} \frac{(U_{hi} - \bar{U}_h)^2}{N_h - 1}$$

and

$$\Psi_{W_h}^2 = \sum_{i=1}^{N_h} \frac{(W_{hi} - \bar{W}_h)^2}{N_h - 1}.$$

The population coefficients of variation for these variables in the *hth* stratum are defined as:

$$\phi_{v_h} = \frac{\Psi_{v_h}}{\bar{V}_h},$$

$$\phi_{u_h} = \frac{\Psi_{u_h}}{\bar{U}_h}$$

and

$$\phi_{w_h} = \frac{\Psi_{w_h}}{\bar{W}_h}.$$

Furthermore, the population correlation coefficients between $(V, U)$, $(V, W)$, and $(U, W)$ in the *hth* stratum are defined as:

$$\rho_{v_h u_h} = \frac{\Psi_{v_h u_h}}{\Psi_{v_h} \Psi_{u_h}},$$

$$\rho_{v_h w_h} = \frac{\Psi_{v_h w_h}}{\Psi_{v_h} \Psi_{w_h}}$$

and

$$\rho_{u_h w_h} = \frac{\Psi_{u_h w_h}}{\Psi_{u_h} \Psi_{w_h}}.$$

A sample of size $n_h$ from the *hth* stratum is chosen by simple random sample without replacement. Let the sample variances of the study variable, auxiliary variable, and ranks of the auxiliary variable in the *hth* stratum be defined as:

$$\hat{\Psi}^2_{v_h} = \sum_{i=1}^{n_h} \frac{(v_{hi} - \bar{v}_h)^2}{n_h - 1},$$

$$\hat{\Psi}^2_{u_h} = \sum_{i=1}^{n_h} \frac{(u_{hi} - \bar{u}_h)^2}{n_h - 1},$$

$$\hat{\Psi}^2_{w_h} = \sum_{i=1}^{n_h} \frac{(w_{hi} - \bar{w}_h)^2}{n_h - 1},$$

where

$$\bar{v}_h = \sum_{i=1}^{n_h} \frac{v_{hi}}{n_h},$$

$$\bar{u}_h = \sum_{i=1}^{n_h} \frac{u_{hi}}{n_h}$$

and

$$\bar{w}_h = \sum_{i=1}^{n_h} \frac{W_{hi}}{n_h},$$

represent the sample means of $V, U$, and $W$ in the *hth* stratum, respectively. In addition, let the sample coefficients of variation for variables $V, U$ in the *hth* stratum be defined as:

$$\hat{\phi}_{v_h} = \frac{\hat{\Psi}_{v_h}}{\bar{v}_h},$$

$$\hat{\phi}_{u_h} = \frac{\hat{\Psi}_{u_h}}{\bar{u}_h}$$

and

$$\hat{\phi}_{w_h} = \frac{\hat{\Psi}_{w_h}}{\bar{w}_h}.$$

Furthermore, let $\bar{u}'_h$ represent the sample mean calculated from the first phase sample with a size of $m_h$, while $\bar{v}_h$ and $\bar{u}_h$ denote the sample means derived from the second phase sample with a size of $n_h$.

## 4. Literature Review

This section examines the characteristics of existing available estimators, focusing on their biases and mean squared errors while utilizing a stratified two-phase sampling approach to estimate the finite population mean.

The usual unbiased estimator to estimate the population mean is given by:

$$\bar{v}^d_{st} = \sum_{h=1}^{L} Z_h \bar{v}_h. \tag{1}$$

The variance of $\bar{v}_{st}^{d}$, is given by:

$$V(\bar{v}_{st}^{d}) = \sum_{h=1}^{L} Z_h^2 \theta_h \bar{V}_h{}^2 \phi_{v_h}^2, \tag{2}$$

where

$$\theta_h = \left(\frac{1}{n_h} - \frac{1}{N_h}\right).$$

According to [6], a ratio estimator for $\bar{V}$ in stratified two-phase sampling is defined as

$$\bar{v}_{Rs}^{d} = \sum_{h=1}^{L} Z_h \bar{v}_h \left(\frac{\bar{u}}{\bar{u}_h'}\right). \tag{3}$$

The expressions in Equations (4) and (5) represent the bias and the $MSE$ of $\bar{v}_{Rs}^{d}$, are defined as:

$$Bias\left(\bar{v}_{Rs}^{d}\right) \cong \sum_{h=1}^{L} Z_h \theta_h'' \bar{V}_h \left(\phi_{u_h}^2 - \phi_{vu_h}\right) \tag{4}$$

and

$$MSE\left(\bar{v}_{Rs}^{d}\right) \cong \sum_{h=1}^{L} Z_h^2 \bar{V}_h^2 \left(\theta_h \phi_{v_h}^2 + \theta_h'' \phi_{u_h}^2 - 2\theta_h'' \phi_{vu_h}\right), \tag{5}$$

where

$$\theta_h' = \left(\frac{1}{m_h} - \frac{1}{N_h}\right)$$

and

$$\theta_h'' = \left(\frac{1}{n_h} - \frac{1}{m_h}\right),$$

respectively.

In stratified two-phase sampling, the product estimator of $\bar{V}$, is defined as:

$$\bar{v}_{Ps}^{d} = \sum_{h=1}^{L} Z_h \bar{v}_h \left(\frac{\bar{u_h}}{\bar{u}_h'}\right). \tag{6}$$

The expressions in Equations (7) and (8) represent the bias and $MSE$ of $\bar{v}_{Ps}^{d}$, which are defined as:

$$Bias\left(\bar{v}_{Ps}^{d}\right) \cong \sum_{h=1}^{L} Z_h \bar{V}_h \theta_h'' \phi_{vu_h} \tag{7}$$

and

$$MSE\left(\bar{v}_{Ps}^{d}\right) \cong \sum_{h=1}^{L} Z_h^2 \bar{V}_h^2 \left(\theta_h \phi_{v_h}^2 + \theta_h'' \phi_{u_h}^2 + 2\theta_h'' \phi_{vu_h}\right). \tag{8}$$

In stratified two-phase sampling, the classical regression estimator $\bar{v}_{lrs}^{d}$, is defined as:

$$\bar{v}_{lrs}^{d} = \sum_{h=1}^{L} Z_h \left[\bar{v}_h + b_{vu_h}(\bar{u}_h' - \bar{u}_h)\right], \tag{9}$$

where $b_{vu_h}$ is the sample regression coefficient.

The expressions in Equations (10) and (11) represent the bias and $MSE$ of $\bar{v}_{lrs}^{d}$, which are defined as:

$$Bias\left(\bar{v}_{lrs}^{d}\right) \cong -\sum_{h=1}^{L} Z_h \theta_h'' \beta_h \left(\frac{\mu_{12h}}{\phi_{vu_h}} - \frac{\mu_{03h}}{\phi_{u_h}^2}\right) \tag{10}$$

and

$$MSE\left(\bar{v}_{lrs}^{d}\right) \cong \sum_{h=1}^{L} Z_h^2 \bar{V}_h^2 \phi_{v_h}^2 \left(\theta_h - \theta_h'' \rho_{vu_h}^2\right), \tag{11}$$

where

$$\beta_h = \frac{\phi_{vu_h}}{\phi_{u_h}^2}$$

and

$$\mu_{rsh} = \frac{\sum_{i=1}^{N_h} (V_{ih} - \bar{V}_h)^r (X_{ih} - \bar{X}_h)^s}{N_h}.$$

According to [12], an exponential ratio and product estimators are expressed as:

$$\bar{v}_{Ts}^{d} = \sum_{h=1}^{L} Z_h \bar{v}_h \exp\left(\frac{\bar{u}_h' - \bar{u}_h}{\bar{u}_h' + \bar{u}_h}\right) \tag{12}$$

and

$$\bar{v}_{Qs}^{d} = \sum_{h=1}^{L} Z_h \bar{v}_h \exp\left(\frac{\bar{u_h} - \bar{u}_h'}{\bar{u_h} + \bar{u}_h'}\right). \tag{13}$$

The expressions in Equations (14) and (15) represent the $MSE$ of $\bar{v}_{Ts}^{d}$ and $\bar{v}_{Qs}^{d}$, which are defined as:

$$MSE\left(\bar{v}_{Ts}^{d}\right) \cong \sum_{h=1}^{L} Z_h^2 \bar{V}_h^2 \left[\theta_h \phi_{v_h}^2 + \theta_h'' \phi_{u_h}^2 \left(\frac{1}{4} - C_h\right)\right] \tag{14}$$

and

$$MSE\left(\bar{v}_{Qs}^{d}\right) \cong \sum_{h=1}^{L} Z_h^2 \bar{V}_h^2 \left[\theta_h \phi_{v_h}^2 + \theta_h'' \phi_{u_h}^2 \left(\frac{1}{4} + C_h\right)\right], \tag{15}$$

where

$$C_h = \left(\frac{\rho_{vu_h}}{\phi_{u_h}}\right) \phi_{v_h}.$$

The double sampling estimator in stratified two-phase sampling, proposed by [13], is expressed as:

$$\bar{v}_{RPs}^{d} \cong \sum_{h=1}^{L} \bar{v}_h \left\{k_h \frac{\bar{u}_h'}{\bar{u}_h} + (1 - k_h) \frac{\bar{u_h}}{\bar{u}_h'}\right\}, \tag{16}$$

where

$$k_h = \frac{1 + C_h}{2}.$$

The expressions in Equations (17) and (18) represent the bias and $MSE$ of $\bar{v}_{RPs}^{d}$, which are defined as:

$$Bias(\bar{v}_{RPs}^{d})_{\min} \cong \frac{1}{2} \sum_{h=1}^{L} Z_h \theta_h'' \bar{V}_h \phi_{u_h}^2 [1 + C_h(1 - 2C_h)] \tag{17}$$

and

$$MSE(\bar{y}_{RPs}^{d})_{\min} \cong \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \phi_{v_h}^2 \left[\theta_h (1 - \rho_{vu_h}^2) + \theta_h' \rho_{vu_h}^2\right]. \tag{18}$$

## 5. Generalized Proposed Class of Estimators

In this section, we provide a new improved family of estimators for population mean, motivated by the approaches discussed in [22,27]. Using a stratified two-phase sampling design, this improved family of estimators uses the known information of the auxiliary variable and the ranks of the auxiliary variable to increase estimation precision. The proposed estimator is expressed as:

$$\bar{v}_{est}^{d} = \sum_{h=1}^{L} Z_h \bar{v}_h \exp\left[k_{3h}\left\{\frac{t_{1h}(\bar{u}'_h - \bar{u}_h)}{t_{1h}(\bar{u}'_h - \bar{u}_h) + 2t_{2h}}\right\}\right] \exp\left[k_{4h}\left\{\frac{t_{3h}(\bar{w}'_h - \bar{w}_h)}{t_{3h}(\bar{w}'_h + \bar{w}_h) + 2t_{4h}}\right\}\right], \quad (19)$$

where $(t_{ih}, i = 1,2,3,4)$ are the parameters of the auxiliary variable, and $(k_{3h}, k_{4h})$ are known constant values of either 1 or 2. The notations $(U_{m_h}, U_{M_h})$ represent the smallest and largest values of the auxiliary variable, whereas the notations $(W_{m_h}, W_{M_h})$ denote the smallest and largest values of the auxiliary variable ranks. The known values of $(t_{ih}, i = 1,2,3,4)$ are given in Table 1, while Table 2 introduces the different estimators by using the (19).

**Table 1.** Different parameters of the auxiliary variables.

| $t_{1h}$ | $t_{2h}$ | $t_{3h}$ | $t_{4h}$ |
|---|---|---|---|
| $-\beta_{2(u_h)}$ | $U_{M_h} - U_{m_h}$ | $-\beta_{2(w_h)}$ | $W_{M_h} - W_{m_h}$ |
| $-\phi_{u_h}$ | $U_{M_h} - U_{m_h}$ | $-\phi_{w_h}$ | $W_{M_h} - W_{m_h}$ |
| $U_{M_h} - U_{m_h}$ | $-\phi_{u_h}$ | $W_{M_h} - W_{m_h}$ | $-\phi_{w_h}$ |
| $U_{M_h} - U_{m_h}$ | $\beta_{2(u_h)}$ | $W_{M_h} - W_{m_h}$ | $\beta_{2(w_h)}$ |
| $U_{M_h} - U_{m_h}$ | $-\beta_{2(u_h)}$ | $W_{M_h} - W_{m_h}$ | $-\beta_{2(w_h)}$ |
| $\beta_{2(u_h)}$ | $U_{M_h} - U_{m_h}$ | $\beta_{2(w_h)}$ | $W_{M_h} - W_{m_h}$ |
| $U_{M_h} - U_{m_h}$ | $\phi_{u_h}$ | $W_{M_h} - W_{m_h}$ | $\phi_{w_h}$ |
| $\phi_{u_h}$ | $U_{M_h} - U_{m_h}$ | $\phi_{w_h}$ | $W_{M_h} - W_{m_h}$ |

**Table 2.** Some classes of the proposed estimator in stratified two-phase sampling.

| **Subsets of $\bar{v}_{est}^{d}$** |
|---|
| $\bar{v}_{e_1}^{d} = \sum_{h=1}^{L} Z_h \bar{v}_h \exp\left[k_{3h}\left\{\frac{-\beta_{2(u_h)}(\bar{u}'_h - \bar{u}_h)}{-\beta_{2(u_h)}(\bar{u}'_h - \bar{u}_h) + 2(U_{M_h} - U_{m_h})}\right\}\right] \exp\left[k_{4h}\left\{\frac{-\beta_{2(w_h)}(\bar{w}'_h - \bar{w}_h)}{-\beta_{2(w_h)}(\bar{w}'_h + \bar{w}_h) + 2(W_{M_h} - W_{m_h})}\right\}\right]$ |
| $\bar{v}_{e_2}^{d} = \sum_{h=1}^{L} Z_h \bar{v}_h \exp\left[k_{3h}\left\{\frac{-\phi_{u_h}(\bar{u}'_h - \bar{u}_h)}{-\phi_{u_h}(\bar{u}'_h - \bar{u}_h) + 2(U_{M_h} - U_{m_h})}\right\}\right] \exp\left[k_{4h}\left\{\frac{-\phi_{w_h}(\bar{w}'_h - \bar{w}_h)}{-\phi_{w_h}(\bar{w}'_h + \bar{w}_h) + 2(W_{M_h} - W_{m_h})}\right\}\right]$ |
| $\bar{v}_{e_3}^{d} = \sum_{h=1}^{L} Z_h \bar{v}_h \exp\left[k_{3h}\left\{\frac{(U_{M_h} - U_{m_h})(\bar{u}'_h - \bar{u}_h)}{(U_{M_h} - U_{m_h})(\bar{u}'_h - \bar{u}_h) - 2\phi_{u_h}}\right\}\right] \exp\left[k_{4h}\left\{\frac{(W_{M_h} - W_{m_h})(\bar{w}'_h - \bar{w}_h)}{(W_{M_h} - W_{m_h})(\bar{w}'_h + \bar{w}_h) - 2\phi_{w_h}}\right\}\right]$ |
| $\bar{v}_{e_4}^{d} = \sum_{h=1}^{L} Z_h \bar{v}_h \exp\left[k_{3h}\left\{\frac{(U_{M_h} - U_{m_h})(\bar{u}'_h - \bar{u}_h)}{(U_{M_h} - U_{m_h})(\bar{u}'_h - \bar{u}_h) + 2\beta_{2(u_h)}}\right\}\right] \exp\left[k_{4h}\left\{\frac{(W_{M_h} - W_{m_h})(\bar{w}'_h - \bar{w}_h)}{(U_{M_h} - U_{m_h})(\bar{w}'_h + \bar{w}_h) + 2\phi_{w_h}}\right\}\right]$ |
| $\bar{v}_{e_5}^{d} = \sum_{h=1}^{L} Z_h \bar{v}_h \exp\left[k_{3h}\left\{\frac{(U_{M_h} - U_{m_h})(\bar{u}'_h - \bar{u}_h)}{(U_{M_h} - U_{m_h})(\bar{u}'_h - \bar{u}_h) - 2\beta_{2(u_h)}}\right\}\right] \exp\left[k_{4h}\left\{\frac{(W_{M_h} - W_{m_h})(\bar{w}'_h - \bar{w}_h)}{(U_{M_h} - U_{m_h})(\bar{w}'_h + \bar{w}_h) - 2\phi_{w_h}}\right\}\right]$ |
| $\bar{v}_{e_6}^{d} = \sum_{h=1}^{L} Z_h \bar{v}_h \exp\left[k_{3h}\left\{\frac{\beta_{2(u_h)}(\bar{u}'_h - \bar{u}_h)}{\beta_{2(u_h)}(\bar{u}'_h - \bar{u}_h) + 2(U_{M_h} - U_{m_h})}\right\}\right] \exp\left[k_{4h}\left\{\frac{\beta_{2(w_h)}(\bar{w}'_h - \bar{w}_h)}{\beta_{2(w_h)}(\bar{w}'_h + \bar{w}_h) + 2(W_{M_h} - W_{m_h})}\right\}\right]$ |
| $\bar{v}_{e_7}^{d} = \sum_{h=1}^{L} Z_h \bar{v}_h \exp\left[k_{3h}\left\{\frac{(U_{M_h} - U_{m_h})(\bar{u}'_h - \bar{u}_h)}{(U_{M_h} - U_{m_h})(\bar{u}'_h - \bar{u}_h) + 2\phi_{u_h}}\right\}\right] \exp\left[k_{4h}\left\{\frac{(W_{M_h} - W_{m_h})(\bar{w}'_h - \bar{w}_h)}{(W_{M_h} - W_{m_h})(\bar{w}'_h + \bar{w}_h) + 2\phi_{w_h}}\right\}\right]$ |
| $\bar{v}_{e_8}^{d} = \sum_{h=1}^{L} Z_h \bar{v}_h \exp\left[k_{3h}\left\{\frac{\phi_{u_h}(\bar{u}'_h - \bar{u}_h)}{\phi_{u_h}(\bar{u}'_h - \bar{u}_h) + 2(U_{M_h} - U_{m_h})}\right\}\right] \exp\left[k_{4h}\left\{\frac{\phi_{w_h}(\bar{w}'_h - \bar{w}_h)}{\phi_{w_h}(\bar{w}'_h + \bar{w}_h) + 2(W_{M_h} - W_{m_h})}\right\}\right]$ |

*Properties of the Proposed Class Estimators*

The properties of the proposed class of estimators can be obtained by defining the following error terms. Let

$$e_{0h} = \left(\frac{\bar{v}_h - \bar{V}_h}{\bar{V}_h}\right), e_{1h} = \left(\frac{\bar{u}_h - \bar{U}_h}{\bar{U}_h}\right), e_{2h} = \left(\frac{\bar{u}'_h - \bar{U}_h}{\bar{U}_h}\right), e_{3h} = \left(\frac{\bar{w}_h - \bar{W}_h}{\bar{W}_h}\right), e_{4h} = \left(\frac{\bar{w}'_h - \bar{W}_h}{\bar{W}_h}\right),$$

such that $E(e_{ih}) = 0$, $(i = 0, 1, 2, 3, 4)$.

Moreover,

$$E\left(e_{0h}^2\right) = \theta_h \phi_{v_h}^2, \quad E\left(e_{1h}^2\right) = \theta_h \phi_{u_h}^2, \quad E\left(e_{2h}^2\right) = \theta_h' \phi_{u_h}^2, \quad E\left(e_{3h}^2\right) = \theta_h \phi_{w_h}^2, \quad E\left(e_{4h}^2\right) = \theta_h' \phi_{w_h}^2,$$

$$E(e_{0h}e_{1h}) = \theta_h \phi_{vu_h}, \quad E(e_{0h}e_{2h}) = \theta_h' \phi_{vu_h}, \quad E(e_{0h}e_{3h}) = \theta_h \phi_{vw_h}, \quad E(e_{oh}e_{4h}) = \theta_h' \phi_{vw_h},$$
$$E(e_{1h}e_{2h}) = \theta_h' \phi_{u_h}^2, \quad E(e_{1h}e_{3h}) = \theta \phi_{uw_h}, E(e_{1h}e_{4h}) = \theta_h' \phi_{uw_h}, \quad E(e_{2h}e_{3h}) = \theta_h' \phi_{uw_h},$$
$$E(e_{2h}e_{4h}) = \theta_h' \phi_{uw_h}, \quad E(e_{3h}e_{4h}) = \theta_h' \phi_{w_h}^2.$$

In order to estimate the bias and mean squared error $(MSE)$ of the suggested estimator $\bar{v}_{est}^d$, we now rewrite Equation (19) to define it in terms of errors, i.e.,

$$\begin{aligned} \bar{v}_{est}^d = \sum_{h=1}^{L} Z_h \bar{V}_h (1 + e_{0h}) \exp\left[k_{3h}\left\{\tfrac{g_{2h}(e_{2h}-e_{1h})}{2}\left(1 + \tfrac{g_{2h}}{2}(e_{1h}+e_{2h})\right)^{-1}\right\}\right] \exp\left[k_{4h}\left\{\tfrac{g_{3h}(e_{4h}-e_{3h})}{2}\right.\right. \\ \left.\left.\left(1 + \tfrac{g_{3h}}{2}(e_{3h}+e_{4h})\right)^{-1}\right\}\right], \end{aligned} \tag{20}$$

where

$$g_{2h} = \frac{t_{1h}\bar{U}_h}{t_{1h}\bar{U}_h + t_{2h}}$$

and

$$g_{3h} = \frac{t_{3h}\bar{W}_h}{t_{3h}\bar{W}_h + t_{4h}}.$$

To drive the following expression by using the Taylor series to the first approximation order, we obtain:

$$\begin{aligned} \bar{v}_{est}^d - \bar{V} \cong \sum_{h=1}^{L} Z_h \bar{V}_h \Bigg[ & e_{0h} - \tfrac{k_{3h}g_{2h}}{2}(e_{1h} - e_{2h}) - \tfrac{k_{4h}g_{3h}}{2}(e_{3h} - e_{4h}) + \left(\tfrac{k_{3h}g_{2h}^2}{4} + \tfrac{k_{3h}^2 g_{2h}^2}{8}\right) e_{1h}^2 \\ & - \left(\tfrac{k_{3h}g_{2h}^2}{4} - \tfrac{k_{3h}^2 g_{2h}^2}{8}\right) e_{2h}^2 + \left(\tfrac{k_{4h}g_{3h}^2}{4} - \tfrac{k_{4h}^2 g_{3h}^2}{8}\right) e_{3h}^2 - \left(\tfrac{k_{4h}g_{3h}^2}{4} - \tfrac{k_{4h}^2 g_{3h}^2}{8}\right) e_{4h}^2 - \tfrac{k_{3h}g_{2h}}{2} e_{0h}e_{1h} \\ & + \tfrac{k_{3h}g_{2h}}{2} e_{0h}e_{2h} - \tfrac{k_4 g_{3h}}{2} e_{0h}e_{3h} + \tfrac{k_{4h}g_{3h}}{2} e_{0h}e_{4h} - \tfrac{k_{3h}^2 g_{2h}^2}{2} e_{1h}e_{2h} + \tfrac{k_{3h}k_{4h}g_{2h}g_{3h}}{4} e_{1h}e_{3h} \\ & - \tfrac{k_{3h}k_{4h}g_{2h}g_{3h}}{4} e_{1h}e_{4h} - \tfrac{k_{3h}k_{4h}g_{2h}g_{3h}}{4} e_{2h}e_{3h} + \tfrac{k_{3h}k_{4h}g_{2h}g_{3h}}{4} e_{2h}e_{4h} - \tfrac{k_{4h}^2 g_{3h}^2}{2} e_{3h}e_{4h} \Bigg]. \end{aligned} \tag{21}$$

Using (21), the bias of $\bar{v}_{est}^d$ is given by:

$$\begin{aligned} Bias\left(\bar{v}_{est}^d\right) \cong & \sum_{h=1}^{L} Z_h \theta_h \bar{V}_h \left[\left(\tfrac{k_{3h}^2 g_{2h}^2}{8} + \tfrac{k_{3h}g_{2h}^2}{4}\right)\phi_{u_h}^2 + \left(\tfrac{k_{4h}^2 g_{3h}^2}{8} + \tfrac{k_{4h}g_{3h}^2}{4}\right)\phi_{w_h}^2 - \tfrac{k_{3h}g_{2h}}{2}\phi_{vu_h} \right. \\ & \left. - \tfrac{k_{4h}g_{3h}}{2}\phi_{vw_h} + \tfrac{k_{3h}k_{4h}g_{2h}g_{3h}}{2}\phi_{uw_h}\right] - \sum_{h=1}^{L} Z_h \theta_h' \bar{V}_h \left[\left(\tfrac{k_{3h}^2 g_{2h}^2}{8} + \tfrac{k_{3h}g_{2h}^2}{4}\right)\phi_{u_h}^2 \right. \\ & \left. + \left(\tfrac{k_{4h}^2 g_{3h}^2}{8} + \tfrac{k_{4h}g_{3h}^2}{4}\right)\phi_{w_h}^2 - \tfrac{k_{3h}g_{2h}}{2}\phi_{vu_h} - \tfrac{k_{4h}g_{3h}}{2}\phi_{vw_h} + \tfrac{k_{3h}k_{4h}g_{2h}g_{3h}}{2}\phi_{uw_h}\right]. \end{aligned} \tag{22}$$

The first-order mean squared error $(MSE)$ is obtained by squaring both sides of (21) and applying the expected value, given below:

$$\begin{aligned} MSE\left(\bar{v}_{est}^d\right) \cong & \sum_{h=1}^{L} Z_h^2 \theta_h \bar{V}_h^2 \left[\phi_{v_h}^2 + \tfrac{k_{3h}^2 g_{2h}^2}{4}\phi_{u_h}^2 + \tfrac{k_{4h}^2 g_{3h}^2}{4}\phi_{w_h}^2 - k_{3h}g_{2h}\phi_{vu_h} - k_{4h}g_{3h}\phi_{vw_h} + \tfrac{k_{3h}k_{4h}g_{2h}g_{3h}}{2}\phi_{uw_h}\right] \\ & - \sum_{h=1}^{L} Z_h^2 \theta_h' \bar{V}_h^2 \left[\tfrac{k_{3h}^2 g_{2h}^2}{4}\phi_{u_h}^2 + \tfrac{k_{4h}^2 g_{3h}^2}{4}\phi_{w_h}^2 - k_{3h}g_{2h}\phi_{vu_h} - k_{4h}g_{3h}\phi_{vw_h} + \tfrac{k_{3h}k_{4h}g_{2h}g_{3h}}{2}\right]. \end{aligned} \tag{23}$$

We can rewrite the bias and $MSE$ for $\bar{v}_{est}^d$ by replacing the known constant values of $k_{3h}$ and $k_{4h}$ into (22) and (23). After a few simple simplifications, we get:

$$Bias\left(\bar{v}_{est}^d\right) \cong \sum_{h=1}^{L} Z_h \theta_h'' \bar{V}_h \left[\frac{3}{8}\left(g_{2h}^2\phi_{u_h}^2 + g_{3h}^2\phi_{w_h}^2\right) - \frac{1}{2}(g_{2h}\phi_{vu_h} + g_{3h}\phi_{vw_h} + g_{2h}g_{3h}\phi_{uw_h})\right] \tag{24}$$

and

$$MSE\left(\bar{v}_{est}^{d}\right) \cong \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \left[\theta_h \phi_{v_h}^2 + \frac{\theta_h''}{4}\left(g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h}\right)\right] \tag{25}$$

## 6. Analytical Comparison

This section presents a comparison between the new improved class of estimators $\bar{v}_{est}^{d}$ and the existing estimators, namely $\bar{v}^{d}$, $\bar{v}_{R_{st}}^{d}$, $\bar{v}_{lr_{st}}^{d}$, $\bar{v}_{T_{st}}^{d}$, $\bar{v}_{Q_{st}}^{d}$, and $\bar{v}_{RP_{st}}^{d}$.

Condition (i): By (2) and (25)

$$V(\bar{v}^{d}) > MSE\left(\bar{v}_{est}^{d}\right) \text{ if}$$

$$\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 (\theta_h - \theta_h')\left(g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h}\right) < 0.$$

For $\theta_h - \theta_h' > 0$, that is, $\theta_h > \theta_h'$

$$\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \left(g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h}\right) < 0. \tag{26}$$

Similarly, $\theta_h - \theta_h' < 0$, that is, $\theta_h < \theta_h'$

$$\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \left(g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h}\right) > 0. \tag{27}$$

When either (26) or (27) are satisfied, the estimator $\bar{v}_{est}^{d}$ demonstrates superior performance compared to $MSE(\bar{v}^{d})$.

Condition (ii): By (5) and (25)

$$MSE(\bar{v}_{Rst}^{d}) > MSE\left(\bar{v}_{est}^{d}\right) \text{ if}$$

$$4\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 (\theta_h - \theta_h')\left(\phi_{u_h}^2 - 2\phi_{vu_h}\right) > \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 (\theta_h - \theta_h')\left(g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h}\right).$$

For $\theta_h - \theta_h' < 0$, that is, $\theta_h < \theta_h'$

$$4\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \left(\phi_{u_h}^2 - 2\phi_{vu_h}\right) > \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \left(g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h}\right). \tag{28}$$

Similarly, $\theta_{;h} - \theta_h' > 0$, that is, $\theta_h > \theta_h'$

$$4\sum_{h=1}^{L} Z_h^2 \bar{V}_h^2 \left(\phi_{u_h}^2 - 2\phi_{vu_h}\right) < \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \left(g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h}\right). \tag{29}$$

When either (28) or (29) are satisfied, the estimator $\bar{v}_{est}^{d}$ demonstrates superior performance compared to $MSE(\bar{v}_{Rst}^{d})$.

Condition (iii): By (11) and (25)

$$MSE(\bar{v}_{lrst}^{d}) > MSE\left(\bar{v}_{est}^{d}\right) \text{ if}$$

$$4\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 (\theta_h - \theta_h')\rho_{vu_h}^{*2} < \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 (\theta_h - \theta_h')\left(g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h}\right).$$

For $\theta_h - \theta_h' > 0$, that is, $\theta_h > \theta_h'$

$$4\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \rho_{vu_h}^{*2} > \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \left(g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h}\right). \tag{30}$$

Similarly, $\theta_h - \theta'_h < 0$, that is, $\theta_h < \theta'_h$

$$4\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \rho_{vu_h}^{*2} < \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \left( g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h} \right). \quad (31)$$

When either (30) or (31) are satisfied, the estimator $\bar{y}_{est}^d$ demonstrates superior performance compared to $MSE(\bar{v}_{lrst}^d)$.

Condition (iv): By (14) and (25)

$$MSE(\bar{v}_{Tst}^d) > MSE\left(\bar{v}_{est}^d\right)$$

$\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 (\theta_h - \theta'_h) \phi_{u_h}^2 (1 - 4C_h) > \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 (\theta_h - \theta'_h) \left( g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h} \right)$.

For $\theta_h - \theta'_h > 0$, that is, $\theta_h > \theta'_h$

$$\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \phi_{u_h}^2 (1 - 4C_h) > \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \left( g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h} \right). \quad (32)$$

Similarly, $\theta_h - \theta'_h < 0$, that is, $\theta_h < \theta'_h$

$$\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \phi_{u_h}^2 (1 - 4C_h) < \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \left( g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h} \right). \quad (33)$$

When either (32) or (33) are satisfied, the estimator $\bar{v}_{est}^d$ demonstrates superior performance compared to $MSE(\bar{v}_{Tst}^d)$.

Condition (v): By (15) and (25)

$$MSE(\bar{v}_{Qst}^d) > MSE\left(\bar{v}_{est}^d\right) \text{ if}$$

$\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 (\theta_h - \theta'_h) \phi_{u_h}^2 (1 + 4C_h) > \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 (\theta_h - \theta'_h) (g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h})$.

For $\theta_h - \theta'_h > 0$, that is, $\theta_h > \theta'_h$

$$\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \phi_{u_h}^2 (1 + 4C_h) > \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \left( g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h} \right). \quad (34)$$

Similarly $\theta_h - \theta'_h < 0$, that is, $\theta_h < \theta'_h$

$$\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \phi_{u_h}^2 (1 + 4C_h) < \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \left( g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h} \right). \quad (35)$$

When either (34) or (35) are satisfied, the estimator $\bar{v}_{est}^d$ demonstrates superior performance compared to $MSE(\bar{y}_{Qst}^d)$.

Condition (vi): By (18) and (25)

$$MSE(\bar{v}_{RPst}^d)_{\min} > MSE\left(\bar{v}_{est}^d\right) \text{ if}$$

$4\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 (\theta_h - \theta'_h) \rho_{vu_h}^2 < \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 (\theta_h - \theta'_h) \left( g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h} \right)$.

For $\theta_h - \theta'_h > 0$, that is, $\theta_h > \theta'_h$

$$4\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \rho_{vu_h}^2 > \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \left( g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h} \right). \quad (36)$$

Similarly, $\theta_h - \theta'_h < 0$, that is, $\theta_h < \theta'_h$

$$4\sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \rho_{vu_h}^2 < \sum_{h=1}^{L} Z_h^2 \bar{V}_h{}^2 \left( g_{2h}^2 \phi_{u_h}^2 + g_{3h}^2 \phi_{w_h}^2 - g_{2h}\phi_{vu_h} - g_{3h}\phi_{vw_h} + 2g_{2h}g_{3h}\phi_{uw_h} \right). \quad (37)$$

When either (36) or (37) are satisfied, the estimator $\bar{v}_{est}^d$ demonstrates superior performance compared to $MSE(\bar{v}_{RPst}^d)$.

## 7. Numerical Comparison

This section presents a detailed comparison between our proposed class of estimators and existing estimators in terms of mean squared errors (*MSEs*), percent relative efficiencies (*PREs*), and mean absolute percentage errors (*MAPEs*). Three separate real datasets and simulated datasets are used in this study. Our goal is to provide a comprehensive evaluation of the effectiveness as well as the reliability of the suggested estimators in various types of practical situations using the *MSE*, *PRE*, and *MAPEs* on these simulated and different datasets.

### *7.1. Simulation Study*

We conduct a simulation analysis using the basic framework explained in reference [22] in order to confirm the theoretical results covered in Section 6. The auxiliary variable X is produced artificially in this simulation for six various populations, each of which corresponds to a probability distribution that is discussed below.

- **Population 1:** The variable $X$ follows a uniform distribution $(a_1 = 5, a_2 = 8)$;
- **Population 2:** The variable $X$ also follows a uniform distribution $(a_1 = 9, a_2 = 14)$;
- **Population 3:** The variable $X$ follows an exponential distribution $(\eta = 6)$;
- **Population 4:** The variable $X$ is also derived from an exponential distribution $(\eta = 9)$;
- **Population 5:** The variable $X$ has a gamma distribution $(\lambda_1 = 4, \lambda_2 = 6)$;
- **Population 6:** The variable $X$ is obtained from a gamma distribution $(\lambda_1 = 8, \lambda_2 = 10)$.

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,

$$r_{yx} = 0.75,$$

is a correlation coefficient, and $e \sim N(0, 1)$ denotes the random error term.

We used the following approaches in R-Software (latest v. 4.4.0) to calculate the *MSEs*, *PREs*, and *MAPEs* of the proposed enhanced estimators and other existing estimators:

**Step 1**: First, we start to generate a population of 1000 observations using the particular probability distributions defined above.
**Step 2**: From a population of $N_h$, choose a first stage sample of size $\acute{n_h}$ using simple random sampling without replacement (*SRSWOR)* to ensure equal selection chances without repetition.
**Step 3**: Utilizing the first sample obtained in the first phase as a starting point, apply the *SRSWOR* approach to determine the second phase $n_h$.
**Step 4:** Using the information from the previous phases, we estimate the population total, the auxiliary variable extreme, as well as the extreme values of the ranks of the auxiliary variable.
**Step 5:** We divide up the sample sizes into distinct strata within each population using the *(SRSWOR)* approach; the sample sizes are set at 25%, 35%, and 45%.
**Step 6:** Using all of the estimators described in this article, the mean squared errors and percent relative efficiencies values can be determined for each sample size. This step makes sure that each estimator's *MSEs* and *PREs* are examined for a set of sample sizes.
**Step 7:** After this, steps 5 and 6 are carried out 100,000 times to guarantee the reliability of the findings. A thorough study of the estimators' performance under simulated datasets can be found in Tables 3–5, which also displays the results for simulated populations.

**Table 3.** Mean squared error (*MSE*) values using the simulated data.

| Estimator | $Uni(5,8)$ | $Uni(9,14)$ | $Exp(6)$ | $Exp(9)$ | $Gam(4,6)$ | $Gam(8,10)$ |
|---|---|---|---|---|---|---|
| $\bar{v}^d$ | $2.02 \times 10^{-2}$ | $1.90 \times 10^{-2}$ | $7.80 \times 10^{-3}$ | $8.90 \times 10^{-2}$ | $8.06 \times 10^{-2}$ | $8.29 \times 10^{-2}$ |
| $v^d_{R_{st}}$ | $1.73 \times 10^{-2}$ | $2.09 \times 10^{-2}$ | $3.30 \times 10^{-3}$ | $8.09 \times 10^{-2}$ | $7.36 \times 10^{-2}$ | $8.03 \times 10^{-2}$ |
| $v^d_{lr_{st}}$ | $1.63 \times 10^{-2}$ | $1.71 \times 10^{-2}$ | $2.30 \times 10^{-3}$ | $7.58 \times 10^{-2}$ | $6.80 \times 10^{-2}$ | $8.13 \times 10^{-2}$ |
| $v^d_{T_{st}}$ | $1.68 \times 10^{-2}$ | $1.50 \times 10^{-2}$ | $2.08 \times 10^{-3}$ | $7.24 \times 10^{-2}$ | $6.33 \times 10^{-2}$ | $8.01 \times 10^{-2}$ |
| $v^d_{Q_{st}}$ | $1.98 \times 10^{-2}$ | $1.88 \times 10^{-2}$ | $1.80 \times 10^{-3}$ | $7.25 \times 10^{-2}$ | $8.05 \times 10^{-2}$ | $7.77 \times 10^{-2}$ |
| $v^d_{RP_{st}}$ | $5.88 \times 10^{-3}$ | $1.35 \times 10^{-2}$ | $1.90 \times 10^{-3}$ | $6.28 \times 10^{-2}$ | $5.90 \times 10^{-2}$ | $5.50 \times 10^{-2}$ |
| $\bar{v}^d_{e_1}$ | $4.28 \times 10^{-3}$ | $1.22 \times 10^{-2}$ | $1.71 \times 10^{-3}$ | $8.60 \times 10^{-3}$ | $4.41 \times 10^{-2}$ | $4.55 \times 10^{-2}$ |
| $\bar{v}^d_{e_2}$ | $2.30 \times 10^{-3}$ | $3.12 \times 10^{-3}$ | $4.54 \times 10^{-4}$ | $7.28 \times 10^{-3}$ | $2.53 \times 10^{-2}$ | $4.80 \times 10^{-2}$ |
| $\bar{v}^d_{e_3}$ | $2.05 \times 10^{-3}$ | $3.20 \times 10^{-3}$ | $4.56 \times 10^{-4}$ | $7.05 \times 10^{-3}$ | $1.94 \times 10^{-2}$ | $2.46 \times 10^{-2}$ |
| $\bar{v}^d_{e_4}$ | $2.26 \times 10^{-3}$ | $3.02 \times 10^{-3}$ | $4.33 \times 10^{-4}$ | $6.88 \times 10^{-3}$ | $3.64 \times 10^{-2}$ | $2.16 \times 10^{-2}$ |
| $\bar{v}^d_{e_5}$ | $2.61 \times 10^{-3}$ | $2.93 \times 10^{-3}$ | $4.90 \times 10^{-3}$ | $6.13 \times 10^{-3}$ | $3.08 \times 10^{-2}$ | $2.81 \times 10^{-2}$ |
| $\bar{v}^d_{e_6}$ | $1.70 \times 10^{-3}$ | $1.58 \times 10^{-3}$ | $1.60 \times 10^{-3}$ | $6.55 \times 10^{-3}$ | $2.15 \times 10^{-2}$ | $3.04 \times 10^{-2}$ |
| $\bar{v}^d_{e_7}$ | $5.90 \times 10^{-4}$ | $3.45 \times 10^{-3}$ | $1.65 \times 10^{-3}$ | $5.55 \times 10^{-3}$ | $2.03 \times 10^{-2}$ | $2.94 \times 10^{-2}$ |
| $\bar{v}^d_{e_8}$ | $4.02 \times 10^{-4}$ | $1.45 \times 10^{-3}$ | $1.45 \times 10^{-3}$ | $4.55 \times 10^{-3}$ | $1.67 \times 10^{-2}$ | $2.04 \times 10^{-2}$ |

**Table 4.** Percent relative efficiency (*PRE*) values using the simulated data.

| Estimator | $Uni(5,8)$ | $Uni(9,14)$ | $Exp(6)$ | $Exp(9)$ | $Gam(4,6)$ | $Gam(8,10)$ |
|---|---|---|---|---|---|---|
| $\bar{v}^d$ | 100 | 100 | 100 | 100 | 100 | 100 |
| $v^d_{R_{st}}$ | 104.663 | 109.827 | 236.364 | 110.012 | 109.511 | 103.238 |
| $v^d_{lr_{st}}$ | 107.447 | 116.564 | 339.130 | 117.412 | 118.520 | 101.968 |
| $v^d_{T_{st}}$ | 120.238 | 126.667 | 375.000 | 122.929 | 127.330 | 103.497 |
| $v^d_{Q_{st}}$ | 102.002 | 101.064 | 433.333 | 122.758 | 100.124 | 107.662 |
| $v^d_{RP_{st}}$ | 109.381 | 140.741 | 410.527 | 141.710 | 136.610 | 150.727 |
| $\bar{v}^d_{e_1}$ | 128.293 | 155.738 | 456.140 | 281.311 | 182.766 | 182.198 |
| $\bar{v}^d_{e_2}$ | 238.737 | 165.536 | 467.018 | 332.317 | 318.577 | 172.708 |
| $\bar{v}^d_{e_3}$ | 267.850 | 161.598 | 334.764 | 343.159 | 221.429 | 336.992 |
| $\bar{v}^d_{e_4}$ | 242.962 | 171.018 | 489.667 | 351.638 | 261.688 | 383.796 |
| $\bar{v}^d_{e_5}$ | 210.380 | 176.271 | 430.073 | 394.661 | 374.884 | 295.018 |
| $\bar{v}^d_{e_6}$ | 322.996 | 326.882 | 487.500 | 369.354 | 261.688 | 272.697 |
| $\bar{v}^d_{e_7}$ | 252.981 | 149.703 | 435.901 | 435.905 | 379.044 | 281.973 |
| $\bar{v}^d_{e_8}$ | 371.291 | 356.189 | 537.931 | 531.708 | 482.635 | 406.373 |

**Step 8:** In the final step, use the following formulae to compute the *MSEs*, *PREs*, and *MAPEs* for each estimator over all replications:

$$MSE(\bar{V}_f^d)_{\min} = \frac{\sum_{q=1}^{100000}\left(\bar{v}_{fq}^d - \bar{V}\right)^2}{100000},$$

$$PRE = \frac{Var(\bar{v}^d)}{MSE(\bar{v}_f^d)_{min}} \times 100$$

and

$$MAPE = \frac{1}{Q \times L}\sum_{q=1}^{Q}\sum_{h=1}^{L}\frac{|\bar{v}_{fq}^d - \bar{V}_h|}{\bar{V}_h} \times 100,$$

where $Q$ is the number of replications (100,000 in this case), $L$ is the number of strata, and $f$ is one of $Rst$, $lrst$, $Tst$, $Qst$, $RPst$, $e_1$, $e_2, \ldots, e_8$.

**Table 5.** Mean absolute percentage error ($MAPE$) values using the simulated data.

| Estimator | $Uni(5,8)$ | $Uni(9,14)$ | $Exp(6)$ | $Exp(9)$ | $Gam(4,6)$ | $Gam(8,10)$ |
|---|---|---|---|---|---|---|
| $\bar{v}^d$ | $1.33 \times 10^{-1}$ | $1.41 \times 10^{-1}$ | $1.53 \times 10^{-1}$ | $1.49 \times 10^{-1}$ | $1.62 \times 10^{-1}$ | $1.58 \times 10^{-1}$ |
| $v_{R_{st}}^d$ | $1.22 \times 10^{-1}$ | $1.31 \times 10^{-1}$ | $1.42 \times 10^{-1}$ | $1.37 \times 10^{-1}$ | $1.51 \times 10^{-1}$ | $1.46 \times 10^{-1}$ |
| $v_{lr_{st}}^d$ | $1.15 \times 10^{-1}$ | $1.23 \times 10^{-1}$ | $1.35 \times 10^{-1}$ | $1.29 \times 10^{-1}$ | $1.43 \times 10^{-1}$ | $1.38 \times 10^{-1}$ |
| $v_{T_{st}}^d$ | $1.22 \times 10^{-1}$ | $1.32 \times 10^{-1}$ | $1.43 \times 10^{-1}$ | $1.38 \times 10^{-1}$ | $1.52 \times 10^{-1}$ | $1.47 \times 10^{-1}$ |
| $v_{Q_{st}}^d$ | $1.23 \times 10^{-1}$ | $1.33 \times 10^{-1}$ | $1.45 \times 10^{-1}$ | $1.39 \times 10^{-1}$ | $1.53 \times 10^{-1}$ | $1.48 \times 10^{-1}$ |
| $v_{RP_{st}}^d$ | $1.16 \times 10^{-1}$ | $1.25 \times 10^{-1}$ | $1.36 \times 10^{-1}$ | $1.30 \times 10^{-1}$ | $1.45 \times 10^{-1}$ | $1.39 \times 10^{-1}$ |
| $\bar{v}_{e_1}^d$ | $8.10 \times 10^{-2}$ | $9.10 \times 10^{-2}$ | $1.01 \times 10^{-1}$ | $9.50 \times 10^{-2}$ | $1.12 \times 10^{-1}$ | $1.06 \times 10^{-1}$ |
| $\bar{v}_{e_2}^d$ | $7.90 \times 10^{-2}$ | $8.80 \times 10^{-2}$ | $9.70 \times 10^{-2}$ | $8.90 \times 10^{-2}$ | $1.05 \times 10^{-1}$ | $9.90 \times 10^{-2}$ |
| $\bar{v}_{e_3}^d$ | $7.70 \times 10^{-2}$ | $8.60 \times 10^{-2}$ | $9.50 \times 10^{-2}$ | $8.80 \times 10^{-2}$ | $1.03 \times 10^{-1}$ | $9.80 \times 10^{-2}$ |
| $\bar{v}_{e_4}^d$ | $7.80 \times 10^{-2}$ | $8.70 \times 10^{-2}$ | $9.60 \times 10^{-2}$ | $8.90 \times 10^{-2}$ | $1.05 \times 10^{-1}$ | $9.90 \times 10^{-2}$ |
| $\bar{v}_{e_5}^d$ | $8.20 \times 10^{-2}$ | $9.20 \times 10^{-2}$ | $1.02 \times 10^{-1}$ | $9.60 \times 10^{-2}$ | $1.13 \times 10^{-1}$ | $1.07 \times 10^{-1}$ |
| $\bar{v}_{e_6}^d$ | $7.60 \times 10^{-2}$ | $8.50 \times 10^{-2}$ | $9.30 \times 10^{-2}$ | $8.70 \times 10^{-2}$ | $1.02 \times 10^{-1}$ | $9.70 \times 10^{-2}$ |
| $\bar{v}_{e_7}^d$ | $7.70 \times 10^{-2}$ | $8.60 \times 10^{-2}$ | $9.50 \times 10^{-2}$ | $8.80 \times 10^{-2}$ | $1.03 \times 10^{-1}$ | $9.80 \times 10^{-2}$ |
| $\bar{v}_{e_8}^d$ | $6.90 \times 10^{-2}$ | $7.80 \times 10^{-2}$ | $8.70 \times 10^{-2}$ | $7.90 \times 10^{-2}$ | $9.50 \times 10^{-2}$ | $8.90 \times 10^{-2}$ |

### *7.2. Numerical Examples*

To assess the effectiveness of various estimators with our proposed improved class of estimators, we used some datasets to calculate the mean squared errors (MSEs), percent relative efficiencies (PREs), and mean absolute percentage error *(MAPE)*. The datasets are explained completely below, along with summary statistics:

**Data 1.** This population encompassing information from multiple divisions was taken from page 135 of the Bureau of Statistics [30], and gathered in Pakistan in 2012. It can be accessed for download on the Pakistan Bureau of Statistics website using the specified URL: https://www.pbs.gov.pk/content/microdata (accessed od 1 December 2024). The following are the explanations of the variables in the dataset:

$Y$: This variable shows how many students were enrolled in schools overall in 2012. It records the total number of students enrolled at all universities.

$X$: This refers to the number of public primary and secondary educational institutions in 2012. The availability of educational institutions in the public sector is measured by this measure.
$R$: This indicates the 2012 public elementary and secondary school ranking according to a set of criteria, which might include academic achievement, facilities, or other elements.

A summary for Dataset 1 is provided in Table 6. The data is categorized into two groups. The groups are set up as the following:

**Group 1**: This group consists of the divisions of Lahore, Sargodha, Gujranwala, and Rawalpindi; these divisions reflect areas with perhaps different student populations and educational facilities.

**Group 2:** This grou consists of the departments of Multan, Bahawalpur, Faisalabad, D.G. Khan, and Sahiwal. These divisions have distinct educational features from Group 1 and provide a varied viewpoint on the dynamics of regional education.

**Data 2.** This population consisting of information from different divisions was selected from page 226 [30] of the Bureau of Statistics, and was collected in Pakistan in 2012. It can be accessed for download on the Pakistan Bureau of Statistics website using the specified URL: https://www.pbs.gov.pk/content/microdata (accessed od 1 December 2024).

$Y$: The variable $Y$ represents the employment levels that were reported throughout the various departments in the year of 2012, and it serves as a measure of the distribution of the workforce.
$X$: This refers to the number of factories that were officially registered by the departments in the year of 2012, which is an indication of the level of industrial activity.
$R$: This represents the ranking of departments based on the number of factories that they registered in 2012, which reflects the relative growth or concentration of the industrial sector.

**Table 6.** Summary statistics for Dataset 1.

| | | | | |
|---|---|---|---|---|
| $N_1 = 18$ | $Z_1 = 0.500$ | $\bar{U_1} = 962$ | $\bar{V_1} = 162979$ | $\bar{W_1} = 9.500$ |
| $U_{M_1} = 1530$ | $U_{m_1} = 388$ | $W_{M_1} = 36$ | $W_{m_1} = 19$ | $\Psi_{u_1} = 308$ |
| $\Psi_{v_1} = 255887$ | $\Psi_{w_1} = 5.338$ | $\phi_{u_1} = 0.320$ | $\phi_{v_1} = 1.571$ | $\phi_{w_1} = 0.562$ |
| $\rho_{v_1u_1} = 0.145$ | $\rho_{v_1w_1} = 0.135$ | $\rho_{u_1w_1} = 0.802$ | $\theta_1' = 0.144$ | $\theta_1'' = 0.056$ |
| $N_2 = 18$ | $Z_2 = 0.500$ | $\bar{U_2} = 1146$ | $\bar{V_2} = 134458$ | $\bar{W_2} = 27.500$ |
| $U_{M_2} = 2370$ | $U_{m_2} = 58$ | $W_{m_2} = 19$ | $W_{M_2} = 36$ | $\Psi_{u_2} = 469.931$ |
| $\Psi_{v_2} = 50236$ | $\Psi_{w_2} = 5.338$ | $\phi_{u_2} = 0.409$ | $\phi_{v_2} = 0.374$ | $\phi_{w_2} = 0.194$ |
| $\rho_{v_2u_2} = 0.787$ | $\rho_{v_2w_2} = 0.657$ | $\rho_{u_2w_2} = 0.889$ | $\theta_2' = 0.144$ | $\theta_2'' = 0.056$ |

The data are divided into two distinct groups to simplify comparison analysis, with the summary statistics for dataset-2 are shown in Table 7.

**Group 1**: This group includes the divisions of Sargodha, Gujranwala, Rawalpindi, and Lahore, which represent some of the more industrialized and heavily inhabited districts.

**Group 2**: This group comprises the divisions of Multan, Bahawalpur, Faisalabad, D.G. Khan, and Sahiwal, including areas noted for agricultural activities as well as expanding industrial sectors.

**Data 3.** This dataset, obtained from page 24 [1], provides statistics on food costs and the weekly income of households.

$Y$: This variable indicates the family's food expenditure, which is connected to their working status and shows how food prices can change depending on their work environment.

$X$ : this represents the family's weekly income, which is an important reflection of their financial situation.
$R$ : This ranks the families according to their weekly income, enabling a comparison of the distribution of income.

For analysis, the dataset is divided into two groups; Table 8 presents the summary statistics for Dataset 3.

**Table 7.** Summary statistics for Dataset 2.

| | | | | |
|---|---|---|---|---|
| $N_1 = 18$ | $Z_1 = 0.500$ | $\bar{U_1} = 415$ | $\bar{V_1} = 85572$ | $\bar{W_1} = 9.500$ |
| $U_{M_1} = 2055$ | $U_{m1} = 24$ | $W_{M_1} = 18$ | $W_{m_1} = 1$ | $\Psi_{u_1} = 521.675$ |
| $\Psi_{v_1} = 248216$ | $\Psi_{w_1} = 5.338$ | $\phi_{u_1} = 1.258$ | $\phi_{v_1} = 2.901$ | $\phi_{w_1} = 0.562$ |
| $\rho_{v_1u_1} = 0.337$ | $\rho_{v_1w_1} = 0.304$ | $\rho_{u_1w_1} = 0.709$ | $\theta_1' = 0.144$ | $\theta_1'' = 0.056$ |
| $N_2 = 18$ | $Z_2 = 0.500$ | $\bar{U_2} = 257$ | $\bar{V_2} = 19294$ | $\bar{W_2} = 27.500$ |
| $U_{M_2} = 1674$ | $U_{m_2} = 52$ | $W_{m_2} = 19$ | $W_{M_2} = 36$ | $\Psi_{u_2} = 366$ |
| $\Psi_{v_2} = 37979$ | $\Psi_{w_2} = 5.338$ | $\phi_{u_2} = 1.423$ | $\phi_{v_2} = 1.969$ | $\phi_{w_2} = 0.194$ |
| $\rho_{v_2u_2} = 0.976$ | $\rho_{v_2w_2} = 0.565$ | $\rho_{u_2w_2} = 0.786$ | $\theta_2' = 0.144$ | $\theta_2'' = 0.056$ |

**Table 8.** Summary statistics for Dataset 3.

| | | | | |
|---|---|---|---|---|
| $N_1 = 18$ | $Z_1 = 0.500$ | $\bar{U_1} = 72.550$ | $\bar{V_1} = 27.490$ | $\bar{W_1} = 9.500$ |
| $U_{M_1} = 95$ | $U_{m1} = 28$ | $W_{M_1} = 18$ | $W_{m_1} = 1$ | $\Psi_{u_1} = 10.580$ |
| $\Psi_{v_1} = 10.130$ | $\Psi_{w_1} = 5.338$ | $\phi_{u_1} = 0.155$ | $\phi_{v_1} = 0.376$ | $\phi_{w_1} = 0.562$ |
| $\rho_{v_1u_1} = 0.337$ | $\rho_{v_1w_1} = 0.284$ | $\rho_{u_1w_1} = 0.557$ | $\theta_1' = 0.144$ | $\theta_1'' = 0.056$ |
| $N_2 = 18$ | $Z_2 = 0.500$ | $\bar{U_2} = 60.870$ | $\bar{V_2} = 20.820$ | $\bar{W_2} = 27.500$ |
| $U_{M_2} = 75$ | $U_{m_2} = 15$ | $W_{m_2} = 19$ | $W_{M_2} = 36$ | $\Psi_{u_2} = 8.980$ |
| $\Psi_{v_2} = 12.750$ | $\Psi_{w_2} = 5.338$ | $\phi_{u_2} = 0.142$ | $\phi_{v_2} = 0.269$ | $\phi_{w_2} = 0.194$ |
| $\rho_{v_2u_2} = 0.496$ | $\rho_{v_2w_2} = 0.297$ | $\rho_{u_2w_2} = 0.756$ | $\theta_2' = 0.144$ | $\theta_2'' = 0.056$ |

Finally, we apply the following formula to calculate the percent relative efficiency *(PREs)* for different datasets:

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

where $K$ is one of $Rst$, $lrst$, $Tst$, $Qst$, $RPst\, e_i (i = 1, 2, \ldots, 8)$.

## 8. Discussion

We carried out simulation studies and analyzed three real datasets to check the performance of the proposed improved class of estimators. The comparison between various estimators was based on the percent relative efficiency $(PRE)$ and the mean absolute percentage error $(MAPE)$ criteria. The results of the simulation analysis, including the MSE, PRE, and MAPE values for both the proposed class of estimators and the existing estimators, are summarized in Tables 3–5. Additionally, the results from the real datasets are displayed in Tables 9–11. Based on these analyses, we have reached the following key conclusions:

**1. Analysis of MSE, PRE, and MAPE**

- **MSE comparison:** In both the simulated data (Table 3) and real datasets (Table 9), the $MSE$ values of the newly proposed estimators are consistently lower than those of the classic estimators. For instance, in the simulated uniform distribution, the new estimators like $\bar{v}^d_{e_8}$ have a significantly smaller $MSE$ compared to $\bar{v}^d$. This pattern holds across different distributions and real-life datasets, as graphically presented in Figures 1 and 2, emphasizing the higher precision of the new estimators.
- **PRE evaluation:** Tables 4 and 10 reveal that the $PRE$ values of the new estimators are notably higher than those of the existing estimators. In the simulated data, the $PRE$ is much greater than that in various distributions, indicating the more efficient use of sample information. The same trend is observed in real datasets, validating the superiority of the new estimators in terms of efficiency.
- **MAPE analysis:** The $MAPE$ values in Tables 5 and 11 show that the new estimators have lower average absolute percentage errors. In both simulated and real data, the new estimators, such as $\bar{v}^d_{e_8}$, have smaller $MAPE$ values compared to traditional estimators, as also illustrated in Figures 3 and 4, further confirming their better performance.

**Table 9.** $MSE$ values using real populations.

| Estimator | Data 1 | Data 2 | Data 3 |
|---|---|---|---|
| $\bar{v}^d$ | 2458644474 | 2277471492 | 4.991 |
| $\bar{v}^d_{R_{st}}$ | 2325327219 | 2106243059 | 4.564 |
| $\bar{v}^d_{lr_{st}}$ | 2086419318 | 2005323090 | 4.879 |
| $\bar{v}^d_{T_{st}}$ | 2360073402 | 2123289511 | 4.628 |
| $\bar{v}^d_{Q_{st}}$ | 2621040437 | 2568789003 | 4.879 |
| $\bar{v}^d_{RP_{st}}$ | 2279370249 | 2091394778 | 4.540 |
| $\bar{v}^d_{e_1}$ | 1885405719 | 1891446231 | 4.090 |
| $\bar{v}^d_{e_2}$ | 1689232744 | 1679650342 | 3.836 |
| $\bar{v}^d_{e_3}$ | 1605592943 | 1578267448 | 3.714 |
| $\bar{v}^d_{e_4}$ | 1682954100 | 1651873834 | 3.953 |
| $\bar{v}^d_{e_5}$ | 1837716742 | 1860831352 | 4.2067 |
| $\bar{v}^d_{e_6}$ | 1443061255 | 1490706445 | 3.586 |
| $\bar{v}^d_{e_7}$ | 1753569322 | 1723803156 | 3.937 |
| $\bar{v}^d_{e_8}$ | 1363005851 | 1411256416 | 3.361 |

**Table 10.** $PRE$ values using real populations.

| Estimator | Data 1 | Data 2 | Data 3 |
|---|---|---|---|
| $\bar{v}^d$ | 100 | 100 | 100 |
| $\bar{v}^d_{R_{st}}$ | 105.733 | 108.120 | 109.356 |
| $\bar{v}^d_{lr_{st}}$ | 117.840 | 113.571 | 117.840 |
| $\bar{v}^d_{T_{st}}$ | 104.177 | 107.262 | 107.842 |
| $\bar{v}^d_{Q_{st}}$ | 103.804 | 105.659 | 102.293 |
| $\bar{v}^d_{RP_{st}}$ | 107.865 | 108.8973 | 109.696 |

**Table 10.** *Cont.*

| Estimator | Data 1 | Data 2 | Data 3 |
|---|---|---|---|
| $\bar{v}^d_{e_1}$ | 130.404 | 120.409 | 121.743 |
| $\bar{v}^d_{e_2}$ | 145.548 | 135.592 | 130.091 |
| $\bar{v}^d_{e_3}$ | 153.130 | 144.302 | 134.381 |
| $\bar{v}^d_{e_4}$ | 146.091 | 137.872 | 126.245 |
| $\bar{v}^d_{e_5}$ | 133.788 | 122.390 | 118.638 |
| $\bar{v}^d_{e_6}$ | 170.377 | 152.778 | 139.180 |
| $\bar{v}^d_{e_7}$ | 140.208 | 132.119 | 126.770 |
| $\bar{v}^d_{e_8}$ | 180.384 | 161.379 | 148.479 |

**Table 11.** *MAPE* values using real population.

| Estimator | Data 1 | Data 2 | Data 3 |
|---|---|---|---|
| $\bar{v}^d$ | $1.53 \times 10^{-1}$ | $1.41 \times 10^{-1}$ | $1.04 \times 10^{-1}$ |
| $\bar{v}^d_{R_{st}}$ | $1.35 \times 10^{-1}$ | $1.23 \times 10^{-1}$ | $9.23 \times 10^{-2}$ |
| $\bar{v}^d_{lr_{st}}$ | $1.27 \times 10^{-1}$ | $1.16 \times 10^{-1}$ | $8.45 \times 10^{-2}$ |
| $\bar{v}^d_{T_{st}}$ | $1.33 \times 10^{-1}$ | $1.22 \times 10^{-1}$ | $9.11 \times 10^{-2}$ |
| $\bar{v}^d_{Q_{st}}$ | $1.61 \times 10^{-1}$ | $1.51 \times 10^{-1}$ | $1.02 \times 10^{-1}$ |
| $\bar{v}^d_{RP_{st}}$ | $1.36 \times 10^{-1}$ | $1.23 \times 10^{-1}$ | $9.32 \times 10^{-2}$ |
| $\bar{v}^d_{e_1}$ | $1.05 \times 10^{-1}$ | $9.32 \times 10^{-2}$ | $7.18 \times 10^{-2}$ |
| $\bar{v}^d_{e_2}$ | $8.78 \times 10^{-2}$ | $7.65 \times 10^{-2}$ | $6.22 \times 10^{-2}$ |
| $\bar{v}^d_{e_3}$ | $7.89 \times 10^{-2}$ | $6.78 \times 10^{-2}$ | $5.36 \times 10^{-2}$ |
| $\bar{v}^d_{e_4}$ | $8.45 \times 10^{-2}$ | $7.33 \times 10^{-2}$ | $6.11 \times 10^{-2}$ |
| $\bar{v}^d_{e_5}$ | $1.01 \times 10^{-1}$ | $9.08 \times 10^{-2}$ | $7.05 \times 10^{-2}$ |
| $\bar{v}^d_{e_6}$ | $6.55 \times 10^{-2}$ | $5.44 \times 10^{-2}$ | $4.12 \times 10^{-2}$ |
| $\bar{v}^d_{e_7}$ | $8.22 \times 10^{-2}$ | $7.11 \times 10^{-2}$ | $6.03 \times 10^{-2}$ |
| $\bar{v}^d_{e_8}$ | $5.67 \times 10^{-2}$ | $4.55 \times 10^{-2}$ | $3.28 \times 10^{-2}$ |

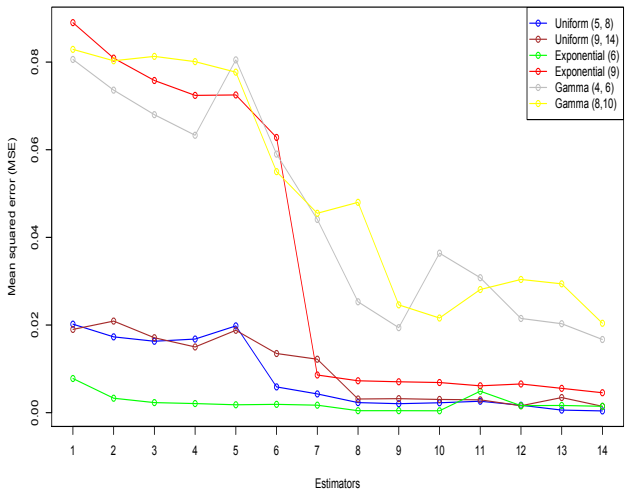



**Figure 1.** Graphical representation of the MSE values using artificial data.

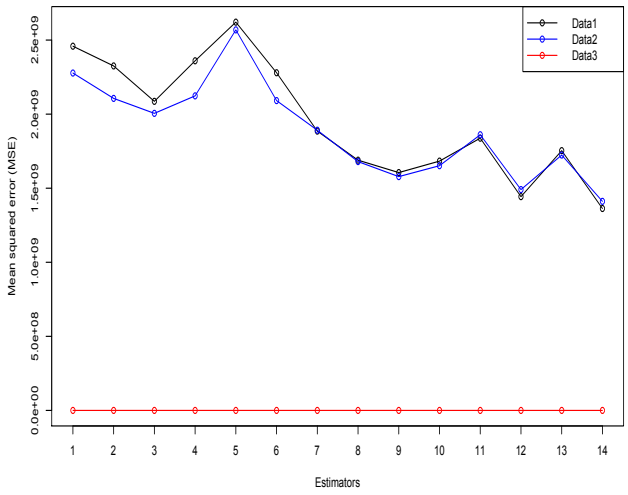



**Figure 2.** Graphical representation of the MSE values using different datasets.

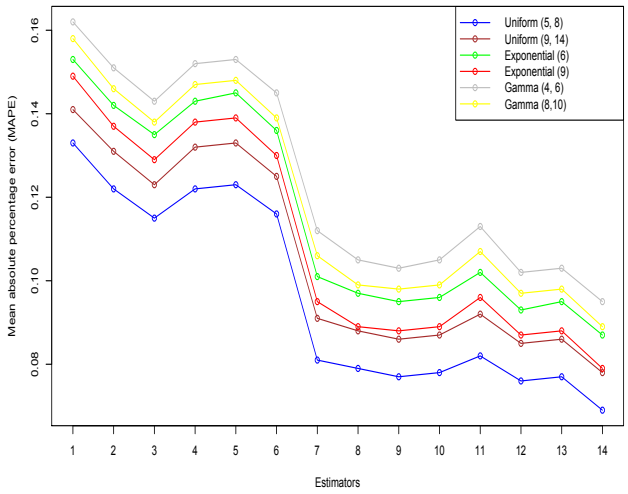



**Figure 3.** Graphical representation of the MAPE values using artificial data.

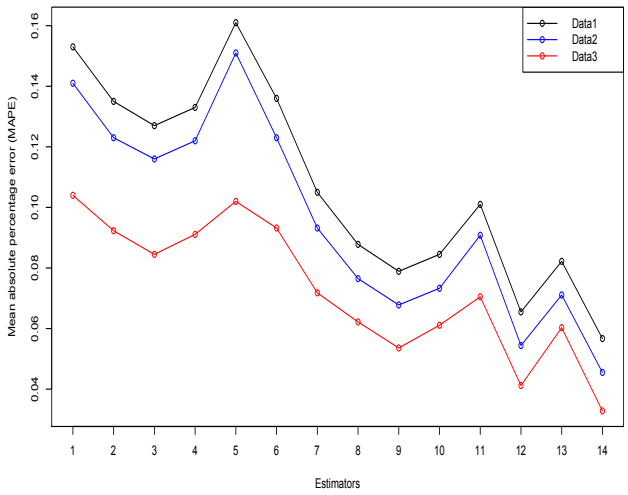



**Figure 4.** Graphical representation of the MAPE values using different datasets.

## 2. Significance and application of results

- **Field applications:** The proposed estimators have significant potential in multiple fields. In economics, they can enhance the accuracy of predicting inflation, unemployment rates, and GDP growth, leading to more informed policy decisions and efficient resource allocation. In public health, they can improve the prediction of disease spread and healthcare demand, enabling better planning and intervention. In environmental science, they can provide more accurate estimates of climate changes and pollution levels, guiding policy-making and disaster preparedness. In finance, they assist in risk management and portfolio optimization, and in supply chain management, they help in accurate demand forecasting for better production and logistics planning.
- **Methodological contribution:** The use of extreme values and ranks of the auxiliary variable in the new estimators not only improves accuracy but also provides a novel approach for handling data with outliers. This is especially valuable in real-world applications where such data are common. The results support the adoption of these estimators in practical situations and suggest further exploration in related research areas.

Therefore, the new estimators outperform existing ones in terms of accuracy and efficiency, as demonstrated by the comprehensive analysis of the $MSE$, $PRE$, and $MAPE$ values. Their practical applications across various fields highlight their importance and potential for improving decision-making and resource utilization.

## 9. Conclusions

In this work, we introduced a new family of effective estimators designed to compute the finite population mean. These estimators make use of the rank ordering and known minimum and maximum values of an auxiliary variable. The first degree of approximation was used to obtain the biases, mean squared errors $(MSEs)$, and minimum $MSE$ of the new family of estimators. Through extensive simulations and analyses of real datasets, it was conclusively demonstrated that these estimators outperform existing ones. The new estimators consistently show lower mean squared errors $(MSEs)$, higher percent relative efficiencies $(PREs)$, and lower mean absolute percentage errors $(MAPEs)$. This enhanced performance was observed across various data distributions and practical scenarios.

The proposed estimators $\bar{v}_i^d$ $(i = e_1, e_2, \ldots, e_8)$ are shown to be more efficient than the other estimators under evaluation by both empirical analysis and simulation findings. Among the proposed estimators, $\bar{v}_{e_8}^d$ is the most effective choice in the proposed classes of estimators and is therefore highly recommended.

Furthermore, we studied the characteristics of the suggested efficient estimators in the framework of a stratified two-phase sampling scheme. Future studies should explore the performance of the proposed estimators in additional contexts, such as big data analysis and predictive modeling in industrial applications. These areas present unique challenges, such as handling large-scale datasets and improving the accuracy of predictions, where the proposed estimators can provide valuable insights. Investigating their application in such fields is likely to further assess their robustness and utility in diverse, complex scenarios.

Moreover, we recommend the adoption of MAPE as a standard measure in similar studies to provide a more detailed understanding of estimator accuracy. Unlike other error measures, MAPE provides a percentage-based error measurement that is easier to interpret, particularly in real-world applications. Its inclusion in future studies can enhance comparability across different models and provide a clearer picture of their relative performance.

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

**Funding:** This project was funded by KAU Endowment (WAQF) at King Abdulaziz University, Jeddah. The authors, therefore, acknowledge with thanks WAQF and the Deanship of Scientific Research (DSR) for technical and financial support.

**Data Availability Statement:** Data are contained within the article.

**Acknowledgments:** The authors are grateful to KAU Endowment (WAQF) at King Abdulaziz University, Jeddah. The authors, therefore, acknowledge with thanks WAQF and the Deanship of Scientific Research (DSR) for technical and financial support.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Cochran, W.B. *Sampling Techniques*; John Wiley and Sons: Hoboken, NJ, USA, 1963.
2. Khoshnevisan, M.; Singh, R.; Chauhan, P.; Sawan, N. A general family of estimators for estimating population mean using known value of some population parameter(s). *Far East J. Theor. Stat.* **2007**, *22*, 181–191.
3. Rueda, M.M.; Arcos, A.; Martınez-Miranda, M.D.; Román, Y. Some improved estimators of finite population quantile using auxiliary information in sample surveys. *Comput. Stat. Data Anal.* **2004**, *45*, 825–848. [CrossRef]
4. Särndal, C.E. Sample survey theory vs. general statistical theory: Estimation of the population mean. *Int. Stat. Rev. Int. Stat.* **1972**, *40*, 1–12. [CrossRef]
5. Tarima, S.; Pavlov, D. Using auxiliary information in statistical function estimation. *ESAIM Probab. Stat.* **2006**, *10*, 11–23. [CrossRef]
6. Sukhatme, B.V. Some ratio-type estimators in two-phase sampling. *J. Am. Stat. Assoc.* **1962**, *57*, 628–632. [CrossRef]
7. Erinola, A.Y.; Singh, R.V.K.; Audu, A.; James, T. Modified class of estimator for finite population mean under two-phase sampling using regression estimation approach. *Asian J. Probab. Stat.* **2021**, *4*, 52–64. [CrossRef]
8. Garg, N.; Srivastava, M. A general class of estimators of a finite population mean using multi-auxiliary information under two stage sampling scheme. *J. Reliab. Stat. Stud.* **2009**, *2*, 103–118.
9. Guha, S.; Chandra, H. Improved estimation of finite population mean in two-phase sampling with subsampling of the nonrespondents. *Math. Popul. Stud.* **2021**, *28*, 24–44. [CrossRef]
10. Kamal, A.; Amir, N.; Dastagir, H. Some exponential type predictive estimators of finite population mean in two-phase sampling. *J. Stat. Comput. Interdiscip. Res.* **2020**, *2*, 51–57. [CrossRef]
11. Khare, B.B.; Khare, S. Generalized synthetic estimator for domain mean in two phase sampling using single auxiliary character. *J. Reliab. Stat. Stud.* **2019**, *12*, 139–151.
12. Singh, H.P.; Vishwakarma, G.K. Modified exponential ratio and product estimators for finite population mean in double sampling. *Austrian J. Stat.* **2007**, *36*, 217–225. [CrossRef]
13. Singh, H.P.; Espeio, M.R. Double sampling ratio-product estimator of a finite population mean in sample surveys. *J. Appl. Stat.* **2007**, *34*, 71–85. [CrossRef]
14. 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*, 1–16. [CrossRef]
15. 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]
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. Khan, M. Improvement in estimating the finite population mean under maximum and minimum values in double sampling scheme. *J. Stat. Appl. Probab. Lett.* **2015**, *2*, 115–121.
19. 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.

20. 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]
21. Daraz, U.; Khan, M. Estimation of variance of the difference-cum-ratio-type exponential estimator in simple random sampling. *Res. Math. Stat.* **2021**, *8*, 1899402. [CrossRef]
22. 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]
23. 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.
24. Chatterjee, S.; Hadi, A.S. *Regression Analysis by Example*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
25. Daraz, U.; Wu, J.; Alomair, M.A.; Aldoghan, L.A. New classes of difference cum-ratio-type exponential estimators for a finite population variance in stratified random sampling. *Heliyon* **2024**, *10*, e33402. [CrossRef]
26. Daraz, U.; Alomair, M.A.; Albalawi, O. Variance estimation under some transformation for both symmetric and asymmetric data. *Symmetry* **2024**, *16*, 957. [CrossRef]
27. Alomair, M.A.; Daraz, U. Dual transformation of auxiliary variables by using outliers in stratified random sampling. *Mathematics* **2024**, *12*, 2839. [CrossRef]
28. 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]
29. Daraz, U.; Wu, J.; Agustiana, D.; Emam, W. Finite population variance estimation using Monte Carlo simulation and real life application. *Symmetry* **2025**, *17*, 84. [CrossRef]
30. 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.