mathematics

MDPI

Article

Auxiliary-Driven Transformations for Efficient Mean Estimation Under Stratified Systematic Sampling

Hleil Alrweili [1] and Fatimah A. Almulhim [2,*]

[1] Department of Mathematics, College of Science, Northern Border University, Arar 73213, Saudi Arabia; hleil.alrweili@nbu.edu.sa
[2] Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
* Correspondence: faalmulhim@pnu.edu.sa

Abstract

This study proposes a flexible family of mean estimators for finite populations under stratified systematic sampling by combining power transformations with an exponential adjustment. The proposed family allows different estimator forms through suitable choices of transformation parameters and stratum-specific characteristics of the auxiliary variable. In particular, the exponential component is constructed using transformations based on the within-stratum standard deviation, mid-range, range, relative range, coefficient of variation, skewness, and excess kurtosis, together with selected combinations of these measures. The bias and mean squared error (MSE) of the proposed estimators are derived using a first-order Taylor approximation, and optimal values of the unknown constants are obtained by minimizing the MSE. Theoretical efficiency conditions are also established against several existing estimators. To assess the practical performance of the proposed family, extensive Monte Carlo simulations are conducted under different finite-population structures, followed by empirical evaluation using three real populations. The performance of the competing estimators is compared in terms of bias, MSE, root mean squared error (RMSE), and percent relative efficiency (PRE). A sensitivity analysis is also performed to examine the effects of the transformation and exponential adjustment parameters. The numerical results demonstrate the potential of the proposed family to achieve improved estimation efficiency across different population structures and provide a flexible framework for utilizing auxiliary information under stratified systematic sampling.

Keywords: stratified systematic sampling; auxiliary-driven estimation; transformation-based estimators; exponential adjustment; mean estimation; efficiency analysis; sensitivity analysis; finite populations

MSC: 62D05

Check for updates

Academic Editors: Wenbo Wu and Chenlu Ke

Received: 19 September 2026
Revised: 18 October 2026
Accepted: 21 October 2026
Published: 23 October 2026

Copyright: © 2026 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.

1. Literature Review

Stratified random sampling is one of the most effective probability sampling designs for populations consisting of distinct subgroups. In this design, the population is first divided into non-overlapping and relatively homogeneous strata, after which independent random samples are selected from each stratum. By accounting for variations among different subpopulations, stratified sampling generally improves the precision of population estimates compared with simple random sampling. Its flexibility also allows different sample allocation schemes to be adopted according to the characteristics of individual strata.

Nevertheless, the efficiency of this sampling design depends not only on the formation of appropriate strata but also on the selection of a suitable estimator. Since the distribution of auxiliary information and the characteristics of individual strata may vary considerably, constructing an efficient estimator is crucial to achieving the full potential of stratified random sampling.

The comparative efficiency of systematic and stratified random sampling has been extensively investigated, demonstrating that the relative performance of these designs depends largely on the underlying population structure [1]. Important theoretical developments have further clarified the principles, advantages, and limitations of systematic sampling under different practical settings [2]. The statistical properties and theoretical foundations of systematic sampling have also been studied in depth, providing a stronger basis for its application in survey sampling [3]. Additional theoretical investigations have identified several factors influencing the efficiency of systematic sampling and examined the conditions under which it performs well [4]. The development of optimal probability sampling strategies has further strengthened the theoretical understanding of systematic selection procedures [5]. Comprehensive discussions on the implementation of systematic sampling, together with its estimation procedures and practical applications in survey research, are available in the classical survey sampling literature [6,7]. A detailed summary of the theory, methodology, and practical applications of systematic sampling was later presented in one of the most influential texts on survey sampling [8].

Auxiliary information has long been recognized as an effective means of improving the precision of population mean estimation. The earliest ratio and product estimators were developed for simple random sampling [9], and their application was later extended to systematic sampling by several researchers [10–12]. These classical estimators perform well when a strong association exists between the study and auxiliary variables; however, their efficiency may decline when this relationship is weak or the underlying assumptions are not satisfied. To address these limitations, regression-based approaches were introduced to better exploit linear relationships between variables, although their performance can be affected by model misspecification. Subsequent methodological advancements led to the development of generalized and regression-type estimators, providing greater flexibility and improved estimation efficiency [13]. The literature was further enriched through the introduction of nonlinear estimator forms, including ratio-, product-, and difference-type transformations, which enhanced both robustness and adaptability in practical survey applications [14].

Continued research has resulted in numerous modifications of mean estimators under the systematic sampling framework. Almost unbiased ratio- and product-type estimators were proposed to reduce estimation bias and improve precision [15], followed by modified versions designed to make more effective use of auxiliary information [16]. More comprehensive classes of estimators were subsequently developed to achieve further gains in efficiency in population mean estimation under systematic sampling [17,18]. Additional contributions include the ratio-cum-product estimator, which integrates the advantages of both ratio and product methods [19], and the chain ratio estimator, which provides further refinement in estimation performance [20]. The potential benefits of weighted and transformed estimators have also been demonstrated, showing superior performance over conventional estimators under appropriate population conditions [21,22].

More recently, attention has shifted toward designing advanced estimators for finite population mean estimation across a variety of sampling schemes. Notable developments include a ratio-cum-product exponential class for adaptive cluster sampling [23], generalized robust regression estimators for rare and clustered populations [24], and memory-type estimators for systematic sampling in time-scaled surveys, all of which effectively utilize

auxiliary information to improve estimation accuracy [25,26]. A hybrid estimation strategy combining harmonic and geometric transformations of auxiliary information was proposed for mean estimation under systematic sampling [27]. The proposed estimators were examined through theoretical analysis, simulation experiments, and real-data applications. Extending the investigation of transformation-based methods, [28] proposed symmetric hybrid transformations for robust mean estimation under stratified systematic sampling, deriving optimal weights through MSE minimization. These studies illustrate different ways of incorporating auxiliary information into hybrid estimators and provide a foundation for further methodological development in stratified systematic sampling. Additional methodological developments and related discussions can be found in [29–34].

Research Gap in Stratified Systematic Sampling

Existing work on mean estimation under systematic sampling has produced a wide range of ratio, product, regression, exponential, and transformed estimators. These methods have shown that auxiliary information can substantially improve the estimation of a finite population mean when the auxiliary variable is informative. However, most of these approaches are constructed around a particular adjustment form, and their performance is therefore closely related to the population characteristics for which that adjustment is suitable. In stratified systematic sampling, this issue becomes more important because the relationship between the study and auxiliary variables, as well as the amount and type of variation in the auxiliary variable, may differ across strata.

A methodological issue that remains important in stratified systematic sampling is how to use different forms of auxiliary information within a single estimator structure. Existing ratio, product, regression, and exponential approaches use particular adjustment mechanisms, whereas changes in the transformation parameters can also alter how the auxiliary variable contributes to estimation. The present study investigates this flexibility by combining two power-transformation components with an exponential adjustment whose form depends on measurable characteristics of the auxiliary variable. This construction permits alternative subclasses to be examined under a common mathematical framework, with particular attention to their bias, mean squared error (MSE), and optimal constants. However, the relative performance of these alternative transformation choices and auxiliary-variable characteristics requires systematic theoretical and numerical investigation under stratified systematic sampling.

The present study addresses this gap by developing a unified class of mean estimators for stratified systematic sampling in which two power-transformation components are combined with an exponential adjustment. The construction permits positive, negative, and zero power choices and allows the exponential component to be defined through several measurable characteristics of the auxiliary variable. In addition, the study derives the corresponding bias and MSE expressions, obtains the optimal values of the unknown constants, and establishes theoretical conditions for comparing the proposed estimators with existing methods.

The proposed class is also examined theoretically against several established estimators, including the conventional systematic sample mean, ratio, product, regression, exponential ratio, and exponential product estimators. The resulting efficiency conditions indicate the situations in which the proposed estimators can achieve a lower MSE than the competing methods. Using stratum-specific measures of dispersion, range, relative variability, skewness, and tail behavior provides greater flexibility in selecting an auxiliary adjustment according to the structure of the population.

Finally, the proposed estimators are evaluated through simulation experiments and real-population applications. Their performance is assessed using bias, MSE, RMSE, and PRE,

allowing the theoretical findings to be examined under different population settings. The numerical results therefore provide an empirical basis for assessing the usefulness of the proposed subclasses in comparison with existing estimation methods.

2. Notation and Preliminary Definitions

Suppose the target population is represented by $\zeta = \{\zeta_1, \zeta_2, \ldots, \zeta_N\}$, where the N population units are partitioned into L mutually exclusive strata. For the hth stratum, let N_h represent its population size, with the stratum sizes satisfying

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

We denote the variable of primary interest by Y, whereas X represents an auxiliary variable whose values are known for every unit in the population.

For stratum h, the units are arranged according to a systematic sampling scheme, from which n_h observations are obtained. Assume that N_h is exactly divisible by n_h, giving

$$k_h = \frac{N_h}{n_h},$$

where k_h takes a positive integer value. Consequently, k_h different systematic samples can be formed, with each sample containing n_h units. A random start is used to determine the sample included in the survey. For the jth unit of sample i, the measurements on Y and X are represented by y_{hij} and x_{hij}, respectively. The indices are defined as

$$i = 1, 2, \ldots, k_h \quad \text{and} \quad j = 1, 2, \ldots, n_h.$$

For a selected systematic sample in stratum h, the sample averages of Y and X can be expressed as

$$\bar{y}_h = \frac{1}{n_h} \sum_{j=1}^{n_h} y_{hij}$$

and

$$\bar{x}_h = \frac{1}{n_h} \sum_{j=1}^{n_h} x_{hij},$$

respectively. The information obtained from all strata is combined through their respective population shares. This leads to the following stratified systematic estimates:

$$\bar{y}_{stsy} = \sum_{h=1}^{L} G_h \bar{y}_h, \qquad \bar{x}_{stsy} = \sum_{h=1}^{L} G_h \bar{x}_h,$$

where the contribution of stratum h is determined by

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

Hence, G_h represents the proportion of the entire population belonging to the hth stratum. Under proper randomization of systematic selection within each stratum, $\bar{y}_{stsy}$ and $\bar{x}_{stsy}$ provide unbiased estimates of the corresponding population means $\bar{Y}$ and $\bar{X}$.

The dispersion of Y and X within the hth stratum is characterized through their finite population variances. Their joint variation is described by the corresponding covariance. These quantities take the form

$$S_{y_h}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2,$$

$$S_{x_h}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)^2,$$

and

$$S_{y_h x_h} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)(x_{hi} - \bar{X}_h).$$

Here, $\bar{Y}_h$ and $\bar{X}_h$ represent the population means of Y and X in stratum h. The relative amount of variation in each variable can be expressed through the squared coefficients of variation as

$$C_{y_h}^2 = \frac{S_{y_h}^2}{\bar{Y}_h^2}$$

and

$$C_{x_h}^2 = \frac{S_{x_h}^2}{\bar{X}_h^2}.$$

The dependence among units generated by systematic selection is described through intra-class correlations. For the study variable and the auxiliary variable in stratum h, these measures are written as

$$\rho_{y_h} = \frac{E\left[\left(y_{hij} - \bar{Y}_h\right)\left(y_{hij'} - \bar{Y}_h\right)\right]}{E\left[\left(y_{hij} - \bar{Y}_h\right)^2\right]}$$

and

$$\rho_{x_h} = \frac{E\left[\left(x_{hij} - \bar{X}_h\right)\left(x_{hij'} - \bar{X}_h\right)\right]}{E\left[\left(x_{hij} - \bar{X}_h\right)^2\right]},$$

where j and j' refer to two different units of the selected systematic sample, so that $j \neq j'$.

The association between Y and X within stratum h is measured by the correlation coefficient

$$\rho_{y_h x_h} = \frac{E\left[\left(y_{hij} - \bar{Y}_h\right)\left(x_{hij} - \bar{X}_h\right)\right]}{\sqrt{E\left[\left(y_{hij} - \bar{Y}_h\right)^2\right] E\left[\left(x_{hij} - \bar{X}_h\right)^2\right]}}.$$

To facilitate the subsequent derivations, the sampling errors associated with the systematic sample means in stratum h are expressed in relative form as

$$\bar{y}_h = \bar{Y}_h(1 + F_{0h}), \qquad \bar{x}_h = \bar{X}_h(1 + F_{1h}),$$

where

$$F_{0h} = \frac{\bar{y}_h - \bar{Y}_h}{\bar{Y}_h}, \qquad F_{1h} = \frac{\bar{x}_h - \bar{X}_h}{\bar{X}_h},$$

such that

$$E(F_{0h}) = E(F_{1h}) = 0.$$

The second-order moments and cross-moments of the relative errors within each stratum are expressed as

$$E(F_{0h}^2) = \delta C_{yh}^2 \Omega_{yh},$$
$$E(F_{1h}^2) = \delta C_{xh}^2 \Omega_{xh},$$
$$E(F_{0h} F_{1h}) = \delta_h \rho_{yhxh} C_{yh} C_{xh} \sqrt{\Omega_{yhxh}},$$

where

$$\Omega_{y_h} = \left[\rho_{y_h}(n_h - 1) + 1\right],$$
$$\Omega_{x_h} = \left[\rho_{x_h}(n_h - 1) + 1\right],$$

$$\Omega_{y_h x_h} = \frac{\Omega_{y_h}}{\Omega_{x_h}} = \left[\frac{\rho_{y_h}(n_h - 1) + 1}{\rho_{x_h}(n_h - 1) + 1}\right],$$

$$K_{y_h x_h} = \rho_{y_h x_h}\left(\frac{C_{y_h}}{C_{x_h}}\right)$$

and

$$\delta_h = \frac{N_h - 1}{n_h N_h}.$$

Effect of Intra-Class Correlation on Estimation Efficiency

The precision of an estimator under systematic sampling depends not only on the sample size but also on the arrangement of population values before sample selection. When units are ordered according to a particular characteristic, observations separated by the sampling interval may exhibit a relationship that differs from the dependence between neighboring units. This feature distinguishes systematic sampling from designs in which units are selected independently. If the population contains a gradual trend, repeated pattern, or clusters of similar values, the selected units may represent the population with different levels of accuracy depending on the random starting point and the sampling interval. Consequently, the efficiency of an estimator cannot be assessed solely from the number of observations; the structure of the population must also be considered.

Intra-class correlation provides a useful way to describe the similarity among units belonging to the same group or sharing a common structure. When units within such groups have positively correlated study-variable values, the information contributed by additional observations may be less than would be expected under independence. This can reduce the effective information available for estimating the population mean. Conversely, a systematic arrangement that distributes the selected units across the full range of population variation may improve representation and lead to greater precision. The effect is not universally positive or negative, however, because it depends on the population ordering, the pattern of dependence, and the relationship between that pattern and the sampling interval. In particular, periodic structures can produce substantial differences in precision when the sampling interval aligns with, or cuts across, the underlying pattern.

These considerations are relevant to the proposed estimator family because its performance is determined by both the systematic sample and the use of auxiliary information. When the auxiliary variable X is strongly associated with the study variable Y, suitable transformations of X may help reduce the mean squared error (MSE) of the estimated population mean. Nevertheless, a strong association between X and Y does not, by itself, guarantee high efficiency under every systematic ordering. The effectiveness of an auxiliary adjustment also depends on how the variables vary across the selected units and on whether their relationship is adequately represented within each stratum. Therefore, the proposed estimators should be assessed in terms of their MSE and relative efficiency under different population structures, rather than assuming that a particular transformation will perform equally well in all circumstances.

3. Review of Existing Estimators

In this section, we discuss selected estimators introduced by different researchers in Section 1, focusing on systematic sampling techniques, their biases, and mean squared errors.

The overall sample mean ($V_1 = \bar{y}_{stsy}$) under stratified systematic sampling is given by

$$V_1 = \sum_{h=1}^{L} G_h \bar{y}_h. \tag{1}$$

In the absence of auxiliary information, the sampling variance (equivalently, the mean squared error) of V_1 can be expressed as

$$\mathrm{MSE}(V_1) \cong \sum_{h=1}^{L} G_h^2 \delta_h \bar{Y}_h^2 \Omega_{y_h} C_{y_h}^2. \tag{2}$$

The stratified ratio estimator ($V_2 = \bar{y}_{Rstsy}$) of [11] uses auxiliary information X within each stratum. The stratum-wise ratio estimates are then combined using the stratum weights to estimate the population mean $\bar{Y}$:

$$V_2 = \sum_{h=1}^{L} G_h \left(\frac{\bar{y}_h}{\bar{x}_h} \right) \bar{X}_h.$$

The bias and mean squared error of V_2 are given by:

$$\mathrm{Bias}(V_2) \cong \sum_{h=1}^{L} G_h \delta_h \bar{Y}_h \left[\Omega_{x_h} C_{x_h}^2 - \sqrt{\Omega_{y_h x_h}} K_{y_h x_h} C_{x_h}^2 \right] \tag{3}$$

and

$$\mathrm{MSE}(V_2) \cong \sum_{h=1}^{L} G_h^2 \delta_h \bar{Y}_h^2 \left[\Omega_{y_h} C_{y_h}^2 + \Omega_{x_h} C_{x_h}^2 \left(1 - 2K_{y_h x_h} \sqrt{\Omega_{y_h x_h}} \right) \right]. \tag{4}$$

The product-type estimator ($V_3 = \bar{y}_{Pstsy}$) proposed by [12] can be extended to stratified systematic sampling. The estimator is obtained separately within each stratum and then combined using the stratum weights:

$$V_3 = \sum_{h=1}^{L} G_h \left(\frac{\bar{y}_h}{\bar{X}_h} \right) \bar{x}_h.$$

The approximate bias and MSE of V_3 are given by:

$$\mathrm{Bias}(V_3) \cong \sum_{h=1}^{L} G_h \delta_h \bar{Y}_h K_{y_h x_h} C_{x_h}^2 \sqrt{\Omega_{y_h x_h}} \tag{5}$$

and

$$\mathrm{MSE}(V_3) \cong \sum_{h=1}^{L} G_h^2 \delta_h \bar{Y}_h^2 \left[\Omega_{y_h} C_{y_h}^2 + \Omega_{x_h} C_{x_h}^2 \left(1 + 2K_{y_h x_h} \sqrt{\Omega_{y_h x_h}} \right) \right]. \tag{6}$$

The classical regression estimator under stratified systematic sampling is expressed as:

$$V_4 = \sum_{h=1}^{L} G_h \left[\bar{y}_h + b_{y_h x_h} (\bar{X}_h - \bar{x}_h) \right],$$

here, $b_{y_h x_h}$ denotes the sample regression coefficient of y on x for stratum h. Using a Taylor series approximation, the MSE of V_4 is given by:

$$\mathrm{MSE}(V_4) \cong \sum_{h=1}^{L} G_h^2 \delta_h \bar{Y}_h^2 C_{y_h}^2 \Omega_{y_h} \left(1 - \rho_{y_h x_h}^2 \right). \tag{7}$$

Exponential ratio- and product-type estimators for the finite population mean were proposed by [16]. Under stratified systematic sampling, these estimators ($V_5 = \bar{y}_{Restsy}$, $V_6 = \bar{y}_{Pestsy}$) are calculated separately within each stratum and then combined using the corresponding stratum weights:

$$V_5 = \sum_{h=1}^{L} G_h \bar{y}_h \exp\left(\frac{\bar{X}_h - \bar{x}_h}{\bar{X}_h + \bar{x}_h} \right)$$

and

$$V_6 = \sum_{h=1}^{L} G_h \bar{y}_h \exp\left(\frac{\bar{x}_h - \bar{X}}{\bar{X}_h + \bar{x}_h} \right). \tag{8}$$

The respective bias and MSE of V_5 and V_6 are obtained as follows:

$$\text{Bias}(V_5) \cong \sum_{h=1}^{L} G_h \delta_h \bar{Y}_h \left[\frac{3}{8} \Omega_{x_h} C_{x_h}^2 - \frac{K_{y_h x_h}}{2} \sqrt{\Omega_{y_h x_h}} C_{x_h}^2 \right], \tag{9}$$

$$\text{Bias}(V_6) \cong \sum_{h=1}^{L} G_h \delta_h \bar{Y}_h \left[\frac{K_{y_h x_h}}{2} \sqrt{\Omega_{y_h x_h}} C_{x_h}^2 - \frac{1}{8} \Omega_{x_h} C_{x_h}^2 \right], \tag{10}$$

$$\text{MSE}(V_5) \cong \sum_{h=1}^{L} G_h^2 \delta_h \bar{Y}_h^2 \left[\Omega_{y_h} C_{y_h}^2 + \frac{1}{4} \Omega_{x_h} C_{x_h}^2 - K_{y_h x_h} \sqrt{\Omega_{y_h x_h}} C_{x_h}^2 \right] \tag{11}$$

and

$$\text{MSE}(V_6) \cong \sum_{h=1}^{L} G_h^2 \delta_h \bar{Y}_h^2 \left[\Omega_{y_h} C_{y_h}^2 + \frac{1}{4} \Omega_{x_h} C_{x_h}^2 + K_{y_h x_h} \sqrt{\Omega_{y_h x_h}} C_{x_h}^2 \right]. \tag{12}$$

Under systematic sampling, ref. [18] developed a class of exponential estimators for the population mean $\bar{Y}$. For the stratified case, the estimator ($V_7 = \bar{y}_{astsy}$) is formed by obtaining the estimate in each stratum and taking their weighted sum:

$$V_7 = \sum_{h=1}^{L} G_h \bar{y}_h \exp\left[\frac{t_h (\bar{x}_h - \bar{X}_h)}{\bar{x}_h + \bar{X}_h} \right],$$

where t_h is a unknown constant.

Using a Taylor series approximation, the bias and MSE of V_7 are obtained as follows:

$$\text{Bias}(V_7)_{opt} \cong -\frac{1}{2} \sum_{h=1}^{L} G_h \delta_h \bar{Y}_h \left[\frac{\Omega_{y_h x_h} C_{x_h}^2 K_{y_h x_h}^2}{D_{x_h}} \right] \tag{13}$$

and

$$\text{MSE}(V_7)_{opt} \cong \sum_{h=1}^{L} G_h^2 \delta_h \bar{Y}_h^2 \left[\Omega_{y_h} C_{y_h}^2 - \frac{K_{y_h x_h}^2 \Omega_{y_h x_h}}{\Omega_{x_h}} C_{x_h}^2 \right], \tag{14}$$

where

$$t_{h_{(\text{opt})}} = \frac{2 K_{y_h x_h} \sqrt{\Omega_{y_h x_h}}}{\Omega_{x_h}}.$$

4. Proposed Auxiliary-Driven Transformation Estimators

In this section, we introduce a general class of estimators for the finite population mean under stratified systematic sampling by using auxiliary information through power transformations and an exponential adjustment. The proposed form allows different combinations of ratio-type, inverse-ratio-type, and difference-type components by changing the transformation parameters. Different characteristics of the auxiliary variable are then used to construct specific subclasses of the proposed estimator.

$$\hat{\bar{Y}}_A = \sum_{h=1}^{L} G_h \left[f_{1h} \bar{y}_h \left(\frac{\bar{X}_h}{\bar{x}_h} \right)^{\ell_{1h}} + f_{2h} (\bar{X}_h - \bar{x}_h) \left(\frac{\bar{X}_h}{\bar{x}_h} \right)^{\ell_{2h}} \right] \exp\left[\frac{d_{1h} (\bar{X}_h - \bar{x}_h)}{d_{1h} (\bar{X}_h + \bar{x}_h) + 2 d_{2h}} \right], \tag{15}$$

where f_{1h} and f_{2h} are constants chosen to reduce bias and mean squared error (MSE), ℓ_{1h} and ℓ_{2h} are power transformation parameters, and d_{1h} and d_{2h} are transformation parameters. The different transformation values are presented in Table 1, while the final forms of the proposed estimators are given in Table 2.

Table 1. Values of transformations used in the proposed estimators.

Estimator	ℓ_{1h}	ℓ_{2h}	d_{1h}	d_{2h}
$\hat{\bar{Y}}_{A_1} = A_1$	0	0	1	S_{xh}
$\hat{\bar{Y}}_{A_2} = A_2$	1	0	1	M_h
$\hat{\bar{Y}}_{A_3} = A_3$	0	1	1	D_h
$\hat{\bar{Y}}_{A_4} = A_4$	1	1	1	$R_h\bar{X}_h$
$\hat{\bar{Y}}_{A_5} = A_5$	2	1	1	$(1 + C_{xh})S_{xh}$
$\hat{\bar{Y}}_{A_6} = A_6$	1	2	1	$(1 + R_h)S_{xh}$
$\hat{\bar{Y}}_{A_7} = A_7$	−1	0	1	$(1 + C_{xh} + R_h)S_{xh}$
$\hat{\bar{Y}}_{A_8} = A_8$	0	−1	1	$(1 + \lvert\gamma_{1h}\rvert)S_{xh}$
$\hat{\bar{Y}}_{A_9} = A_9$	−2	1	1	$(1 + \lvert\gamma_{2h}\rvert)S_{xh}$
$\hat{\bar{Y}}_{A_{10}} = A_{10}$	1	−2	1	$(1 + \lvert\gamma_{1h}\rvert + C_{xh})S_{xh}$
$\hat{\bar{Y}}_{A_{11}} = A_{11}$	2	−1	1	$(1 + \lvert\gamma_{2h}\rvert + R_h)S_{xh}$
$\hat{\bar{Y}}_{A_{12}} = A_{12}$	−1	2	1	$(1 + \lvert\gamma_{1h}\rvert + \lvert\gamma_{2h}\rvert)S_{xh}$

The parameter d_{2h} is specified using different stratum-specific characteristics of the auxiliary variable. Let S_{xh} denote the stratum-specific standard deviation of X, and define the coefficient of variation as

$$C_{xh} = \frac{S_{xh}}{\bar{X}_h}.$$

The stratum-specific mid-range and range are defined as

$$M_h = \frac{X_{h,\max} + X_{h,\min}}{2},$$

and

$$D_h = X_{h,\max} - X_{h,\min}.$$

The relative range is defined by

$$R_h = \frac{X_{h,\max} - X_{h,\min}}{X_{h,\max} + X_{h,\min}}.$$

Let γ_{1h} and γ_{2h} denote the stratum-specific skewness and excess kurtosis of the auxiliary variable X, respectively. Since C_{xh}, R_h, γ_{1h}, and γ_{2h} are dimensionless quantities, they are combined with S_{xh} or $\bar{X}_h$ whenever necessary so that d_{2h} has the same measurement unit as the auxiliary variable.

Based on these characteristics, twelve subclasses are generated by different combinations of the power parameters (ℓ_{1h}, ℓ_{2h}) and the auxiliary-variable transformations used in d_{2h}. The proposed subclasses are summarized in Table 1.

In Table 1, the first subclass corresponds to the basic difference-type structure of the proposed family. Setting $\ell_{1h} = 1$ produces the conventional ratio adjustment, whereas $\ell_{1h} = -1$ produces an inverse-ratio adjustment. Similarly, positive values of ℓ_{2h} generate ratio-type adjustments to the difference component, while negative values generate inverse-ratio forms. Values such as 2 and −2 provide higher-order ratio and inverse-ratio configurations. The final forms of the proposed estimators are given in Table 2.

Table 2. Transformed proposed estimators.

Estimator
$A_1 = \sum_{h=1}^{L} G_h\left[f_{1h}\bar{y}_h + f_{2h}(\bar{X}_h - \bar{x}_h)\right] \exp\left[\frac{d_{1h}(\bar{X}_h - \bar{x}_h)}{d_{1h}(\bar{X}_h + \bar{x}_h) + 2d_{2h}}\right]$
$A_2 = \sum_{h=1}^{L} G_h\left[f_{1h}\bar{y}_h\left(\frac{\bar{X}_h}{\bar{x}_h}\right) + f_{2h}(\bar{X}_h - \bar{x}_h)\right] \exp\left[\frac{d_{1h}(\bar{X}_h - \bar{x}_h)}{d_{1h}(\bar{X}_h + \bar{x}_h) + 2d_{2h}}\right]$
$A_3 = \sum_{h=1}^{L} G_h\left[f_{1h}\bar{y}_h + f_{2h}(\bar{X}_h - \bar{x}_h)\left(\frac{\bar{X}_h}{\bar{x}_h}\right)\right] \exp\left[\frac{d_{1h}(\bar{X}_h - \bar{x}_h)}{d_{1h}(\bar{X}_h + \bar{x}_h) + 2d_{2h}}\right]$
$A_4 = \sum_{h=1}^{L} G_h\left[f_{1h}\bar{y}_h\left(\frac{\bar{X}_h}{\bar{x}_h}\right) + f_{2h}(\bar{X}_h - \bar{x}_h)\left(\frac{\bar{X}_h}{\bar{x}_h}\right)\right] \exp\left[\frac{d_{1h}(\bar{X}_h - \bar{x}_h)}{d_{1h}(\bar{X}_h + \bar{x}_h) + 2d_{2h}}\right]$
$A_5 = \sum_{h=1}^{L} G_h\left[f_{1h}\bar{y}_h\left(\frac{\bar{X}_h}{\bar{x}_h}\right)^2 + f_{2h}(\bar{X}_h - \bar{x}_h)\left(\frac{\bar{X}_h}{\bar{x}_h}\right)\right] \exp\left[\frac{d_{1h}(\bar{X}_h - \bar{x}_h)}{d_{1h}(\bar{X}_h + \bar{x}_h) + 2d_{2h}}\right]$
$A_6 = \sum_{h=1}^{L} G_h\left[f_{1h}\bar{y}_h\left(\frac{\bar{X}_h}{\bar{x}_h}\right) + f_{2h}(\bar{X}_h - \bar{x}_h)\left(\frac{\bar{X}_h}{\bar{x}_h}\right)^2\right] \exp\left[\frac{d_{1h}(\bar{X}_h - \bar{x}_h)}{d_{1h}(\bar{X}_h + \bar{x}_h) + 2d_{2h}}\right]$
$A_7 = \sum_{h=1}^{L} G_h\left[f_{1h}\bar{y}_h\left(\frac{\bar{x}_h}{\bar{X}_h}\right) + f_{2h}(\bar{X}_h - \bar{x}_h)\right] \exp\left[\frac{d_{1h}(\bar{X}_h - \bar{x}_h)}{d_{1h}(\bar{X}_h + \bar{x}_h) + 2d_{2h}}\right]$
$A_8 = \sum_{h=1}^{L} G_h\left[f_{1h}\bar{y}_h + f_{2h}(\bar{X}_h - \bar{x}_h)\left(\frac{\bar{x}_h}{\bar{X}_h}\right)\right] \exp\left[\frac{d_{1h}(\bar{X}_h - \bar{x}_h)}{d_{1h}(\bar{X}_h + \bar{x}_h) + 2d_{2h}}\right]$
$A_9 = \sum_{h=1}^{L} G_h\left[f_{1h}\bar{y}_h\left(\frac{\bar{x}_h}{\bar{X}_h}\right)^2 + f_{2h}(\bar{X}_h - \bar{x}_h)\left(\frac{\bar{X}_h}{\bar{x}_h}\right)\right] \exp\left[\frac{d_{1h}(\bar{X}_h - \bar{x}_h)}{d_{1h}(\bar{X}_h + \bar{x}_h) + 2d_{2h}}\right]$
$A_{10} = \sum_{h=1}^{L} G_h\left[f_{1h}\bar{y}_h\left(\frac{\bar{X}_h}{\bar{x}_h}\right) + f_{2h}(\bar{X}_h - \bar{x}_h)\left(\frac{\bar{x}_h}{\bar{X}_h}\right)^2\right] \exp\left[\frac{d_{1h}(\bar{X}_h - \bar{x}_h)}{d_{1h}(\bar{X}_h + \bar{x}_h) + 2d_{2h}}\right]$
$A_{11} = \sum_{h=1}^{L} G_h\left[f_{1h}\bar{y}_h\left(\frac{\bar{X}_h}{\bar{x}_h}\right)^2 + f_{2h}(\bar{X}_h - \bar{x}_h)\left(\frac{\bar{x}_h}{\bar{X}_h}\right)\right] \exp\left[\frac{d_{1h}(\bar{X}_h - \bar{x}_h)}{d_{1h}(\bar{X}_h + \bar{x}_h) + 2d_{2h}}\right]$
$A_{12} = \sum_{h=1}^{L} G_h\left[f_{1h}\bar{y}_h\left(\frac{\bar{x}_h}{\bar{X}_h}\right) + f_{2h}(\bar{X}_h - \bar{x}_h)\left(\frac{\bar{X}_h}{\bar{x}_h}\right)^2\right] \exp\left[\frac{d_{1h}(\bar{X}_h - \bar{x}_h)}{d_{1h}(\bar{X}_h + \bar{x}_h) + 2d_{2h}}\right]$

The different specifications of d_{2h} introduce alternative exponential adjustments based on the characteristics of the auxiliary variable. In particular, $d_{2h} = S_{xh}$ uses the within-stratum standard deviation, $d_{2h} = M_h$ uses the mid-range, and $d_{2h} = D_h$ uses the observed range of the auxiliary variable. The specification $d_{2h} = R_h\bar{X}_h$ uses a scale-adjusted relative range. The terms $(1 + C_{xh})S_{xh}$ and $(1 + R_h)S_{xh}$ introduce relative dispersion and relative range information into the exponential component. Furthermore, $(1 + |\gamma_{1h}|)S_{xh}$ and $(1 + |\gamma_{2h}|)S_{xh}$ allow the exponential adjustment to respond to skewness and tail behavior, respectively. The remaining combined forms provide additional adjustments based on more than one characteristic of the auxiliary variable.

The proposed construction provides flexibility in adapting the exponential component to different distributional characteristics of the auxiliary variable across strata. In contrast to assigning arbitrary numerical constants to d_{1h} and d_{2h}, the proposed specifications are based on interpretable stratum-level characteristics of X. This provides a systematic basis for comparing alternative subclasses under different population structures.

The selected transformations are functions of the auxiliary variable and are evaluated separately within each stratum. Their use allows the proposed family to account for differences in dispersion, relative variability, observed range, skewness, and kurtosis across strata. The resulting subclasses can subsequently be evaluated using bias, mean squared error (MSE), and percent relative efficiency (PRE) under both simulated and real populations.

4.1. Optimal Bias and Mean Squared Error of the Proposed Class

To derive the bias and MSE, we first express the estimator in (15) through the corresponding error terms. This leads to the following expressions

$$\hat{\bar{Y}}_A = \sum_{h=1}^{L} G_h \left[f_{1h}\, \bar{Y}_h (1+F_{0h}) \left(\frac{1}{1+F_{1h}} \right)^{\ell_{1h}} - f_{2h}\, \bar{X}_h F_{1h} \left(\frac{1}{1+F_{1h}} \right)^{\ell_{2h}} \right] \exp\left[\frac{-z_h F_{1h}}{2} \left(1 + \frac{z_h F_{1h}}{2} \right)^{-1} \right],$$

where

$$z_h = \frac{d_{1h}\bar{X}_h}{d_{1h}\bar{X}_h + d_{2h}}.$$

Using a Taylor series expansion, the estimator can be expressed in the following approximate form

$$\begin{aligned} \hat{\bar{Y}}_A \cong \sum_{h=1}^{L} G_h \Bigg[& f_{1h}\bar{Y}_h(1+F_{0h}) \left\{ 1 - \ell_{1h}F_{1h} + \frac{\ell_{1h}(\ell_{1h}+1)}{2} F_{1h}^2 \right\} \\ & - f_{2h}\bar{X}_h F_{1h} \left\{ 1 - \ell_{2h}F_{1h} + \frac{\ell_{2h}(\ell_{2h}+1)}{2} F_{1h}^2 \right\} \Bigg] \\ & \times \exp\left[\frac{-z_h F_{1h}}{2} \left(1 - \frac{z_h F_{1h}}{2} + \frac{z_h^2 F_{1h}^2}{4} \right) \right]. \end{aligned}$$

$$\begin{aligned} \Rightarrow \hat{\bar{Y}}_A \cong \sum_{h=1}^{L} G_h \Bigg[& f_{1h}\bar{Y}_h(1+F_{0h}) \left\{ 1 - \ell_{1h}F_{1h} + \frac{\ell_{1h}(\ell_{1h}+1)}{2} F_{1h}^2 \right\} \\ & - f_{2h}\bar{X}_h F_{1h} \left\{ 1 - \ell_{2h}F_{1h} + \frac{\ell_{2h}(\ell_{2h}+1)}{2} F_{1h}^2 \right\} \Bigg] \\ & \times \exp\left(-\frac{z_h F_{1h}}{2} + \frac{z_h^2 F_{1h}^2}{4} \right). \end{aligned}$$

We know that

$$\exp\left(-\frac{z_h F_{1h}}{2} + \frac{z_h^2 F_{1h}^2}{4} \right) \cong 1 - \frac{z_h F_{1h}}{2} + \frac{3}{8} z_h^2 F_{1h}^2 + O(z_h^3).$$

Therefore

$$\begin{aligned} \hat{\bar{Y}}_A \cong \sum_{h=1}^{L} G_h \Bigg[& f_{1h}\bar{Y}_h \left\{ 1 + F_{0h} - \ell_{1h}F_{1h} + \frac{\ell_{1h}(\ell_{1h}+1)}{2} F_{1h}^2 - F_{0h}F_{1h}\ell_{1h} \right\} \\ & - f_{2h}\bar{X}_h F_{1h} + f_{2h}\ell_{2h}\bar{X}_h F_{1h}^2 \Bigg] \left[1 - \frac{z_h F_{1h}}{2} + \frac{3}{8} z_h^2 F_{1h}^2 \right]. \end{aligned}$$

After multiplication, terms involving relative errors of order higher than two are neglected. Subtracting $\sum_{h=1}^{L} G_h \bar{Y}_h$ from both sides then yields the following expression.

$$\begin{aligned} \hat{\bar{Y}}_A - \sum_{h=1}^{L} G_h \bar{Y}_h \cong & - \sum_{h=1}^{L} G_h \bar{Y}_h + \sum_{h=1}^{L} G_h \Bigg[f_{1h}\bar{Y}_h \Bigg\{ 1 + F_{0h} - \left(\ell_{1h} + \frac{z_h}{2} \right) F_{1h} \\ & - F_{0h} \left(\ell_{1h} + \frac{z_h}{2} \right) F_{1h} + \left\{ \frac{\ell_{1h}(\ell_{1h}+1)}{2} + \frac{\ell_{1h}z_h}{2} + \frac{3z_h^2}{8} \right\} F_{1h}^2 \Bigg\} \\ & + f_{2h}\bar{X}_h \left\{ -F_{1h} + \left(\ell_{2h} + \frac{z_h}{2} \right) F_{1h}^2 \right\} \Bigg]. \end{aligned} \tag{16}$$

Taking expectations on both sides of (16) and applying the relevant results yields

$$E(F_{0h}) = E(F_{1h}) = 0.$$

$$E(F_{0h}^2) = \delta C_{yh}^2 \Omega_{yh},$$

$$E(F_{1h}^2) = \delta C_{xh}^2 \Omega_{xh}$$

and

$$E(F_{0h}F_{1h}) = \delta_h \rho_{yhxh} C_{yh} C_{xh} \sqrt{\Omega_{yhxh}},$$

we obtain

$$\text{Bias}\left(\hat{\bar{Y}}_A\right) \cong \sum_{h=1}^{L} G_h\Bigg[(f_{1h}-1)\bar{Y}_h + f_{1h}\bar{Y}_h\Bigg\{-\left(\ell_{1h}+\frac{z_h}{2}\right)\delta_h\rho_{y_hx_h}C_{y_h}C_{x_h}\sqrt{\Omega_{y_hx_h}} + \left\{\frac{\ell_{1h}(\ell_{1h}+1)}{2}+\frac{\ell_{1h}z_h}{2}+\frac{3z_h^2}{8}\right\}\delta_h C_{x_h}^2\Omega_{x_h}\Bigg\} + f_{2h}\bar{X}_h\left(\ell_{2h}+\frac{z_h}{2}\right)\delta_h C_{x_h}^2\Omega_{x_h}\Bigg].$$

The above bias expression can be expressed as follows

$$\text{Bias}\left(\hat{\bar{Y}}_A\right) \cong -\sum_{h=1}^{L} G_h\Big[\bar{Y}_h - f_{1h}\bar{Y}_h K_{1h} - f_{2h}K_{2h}\Big], \tag{17}$$

where

$$K_{1h} = \left[1+\delta_h\left\{C_{x_h}^2\Omega_{x_h}\left(\frac{4\ell_{1h}(\ell_{1h}+1+z_h)+3z_h^2}{8}\right) - K_{y_hx_h}C_{x_h}^2\sqrt{\Omega_{y_hx_h}}\left(\frac{2\ell_{1h}+z_h}{2}\right)\right\}\right]$$

and

$$K_{2h} = \delta_h\bar{X}_h C_{x_h}^2\Omega_{x_h}\left(\frac{2\ell_{2h}+z_h}{2}\right).$$

Squaring (16), retaining terms up to the second order, and taking expectations on both sides gives the following mean squared error expression

$$\text{MSE}\left(\hat{\bar{Y}}_A\right) \cong \sum_{h=1}^{L} G_h^2\Big[\bar{Y}_h^2 + f_{1h}^2\bar{Y}_h^2K_{3h} + f_{2h}^2K_{4h} + 2f_{1h}f_{2h}\bar{Y}_hK_{5h} - 2f_{1h}\bar{Y}_h^2K_{1h} - 2f_{2h}\bar{Y}_hK_{2h}\Big], \tag{18}$$

where

$$K_{3h} = \left[1+\delta_h\left\{C_{y_h}^2\Omega_{y_h} + C_{x_h}^2\Omega_{x_h}(2\ell_{1h}^2+\ell_{1h}+2\ell_{1h}z_h+z_h^2) - 2K_{y_hx_h}C_{x_h}^2(2\ell_{1h}+z_h)\sqrt{\Omega_{y_hx_h}}\right\}\right],$$

$$K_{4h} = \delta_h\bar{X}_h^2C_{x_h}^2\Omega_{x_h}$$

and

$$K_{5h} = \delta_h\bar{X}_hC_{x_h}^2\Big[(\ell_{1h}+\ell_{2h}+z_h) - K_{y_hx_h}\sqrt{\Omega_{y_hx_h}}\Big].$$

To determine the optimal bias and MSE, we differentiate (18) with respect to f_{1h} and f_{2h}, the results are given below

$$f_{1h}^* = \frac{K_{1h}K_{4h}-K_{2h}K_{5h}}{K_{3h}K_{4h}-K_{5h}^2}$$

and

$$f_{2h}^* = \frac{\bar{Y}_h(K_{2h}K_{3h}-K_{1h}K_{5h})}{K_{3h}K_{4h}-K_{5h}^2}.$$

Using the optimal values f_{1h}^* and f_{1h}^* in (17) and (18), the resulting bias and MSE are obtained as

$$Bias\left(\hat{\bar{Y}}_A\right)_{min} \cong -\sum_{h=1}^{L} G_h\bar{Y}_h\left[1-\left(\frac{K_{2h}^2K_{3h}+K_{1h}^2K_{4h}-2K_{1h}K_{2h}K_{5h}}{K_{3h}K_{4h}-K_{5h}^2}\right)\right] \tag{19}$$

and

$$MSE\left(\hat{\bar{Y}}_A\right)_{min} \cong \sum_{h=1}^{L} G_h^2\bar{Y}_h^2\left[1-\left(\frac{K_{2h}^2K_{3h}+K_{1h}^2K_{4h}-2K_{1h}K_{2h}K_{5h}}{K_{3h}K_{4h}-K_{5h}^2}\right)\right]. \tag{20}$$

4.2. Practical Estimation of the Optimal Constants

The optimal constants derived in the preceding analysis depend on population characteristics that may not be available in practical survey applications. To implement the proposed estimator when these quantities are unknown, a plug-in approach can be adopted by replacing the unavailable population quantities with appropriate estimates obtained from the observed sample. For each stratum h, the sample mean $\bar{y}_h$ can be used as an estimate of $\bar{Y}_h$, while sample-based estimates of the coefficients of variation, correlation coefficients, and other required characteristics can be incorporated into the definitions of $K_{1h}, K_{2h}, K_{3h}, K_{4h}$, and K_{5h}. Characteristics of the auxiliary variable that cannot be estimated reliably from the available sample may instead be obtained from a pilot survey, previous survey information, or a population frame, when such information is available. Quantities already known from the sampling design or auxiliary population information need not be estimated again.

Let $\widehat{K}_{jh}, j = 1, \ldots, 5$, denote the corresponding estimates obtained by substituting the available sample-based quantities into the definitions of K_{jh}. The practical versions of the optimal constants can then be expressed as

$$\hat{f}_{1h} = \frac{\widehat{K}_{1h}\widehat{K}_{4h} - \widehat{K}_{2h}\widehat{K}_{5h}}{\widehat{K}_{3h}\widehat{K}_{4h} - \widehat{K}_{5h}^2},$$

and

$$\hat{f}_{2h} = \frac{\bar{y}_h\left(\widehat{K}_{2h}\widehat{K}_{3h} - \widehat{K}_{1h}\widehat{K}_{5h}\right)}{\widehat{K}_{3h}\widehat{K}_{4h} - \widehat{K}_{5h}^2}.$$

These expressions provide sample-based counterparts of the theoretical optimal constants in Section 4. The resulting feasible estimator is obtained by replacing f_{1h} and f_{2h} in (15) with $\hat{f}_{1h}$ and $\hat{f}_{2h}$, respectively, while retaining the specified transformation parameters for the selected subclass. This substitution is valid only when the estimated denominator is nonzero and the required quantities can be estimated adequately.

When the optimal constants are calculated from sample information, their estimated values may not coincide with those obtained from the population quantities. This difference can affect the bias and mean squared error of the resulting estimator, particularly when the sample sizes within individual strata are small or the estimated denominator is close to zero. Consequently, the minimum MSE derived in (20) describes the theoretical performance under the quantities used in the derivation and should not be interpreted as the guaranteed performance of the sample-based estimator. Further simulation studies could examine how the accuracy of the estimated constants affects the estimator across different sample sizes and population structures.

5. Theoretical Comparison Between MSEs

To test the efficiency of the proposed estimator family, their mean squared errors are compared with the mean squared errors of the existing estimators mentioned in Sections 3 and 4.

(i): Comparison between (2) and (20)

$$\mathrm{MSE}(V_1) > MSE\left(\hat{\bar{Y}}_A\right)_{min} \text{ if}$$

$$\sum_{h=1}^{L} G_h^2 \bar{Y}_h^2 \left[\delta_h \Omega_{y_h} C_{y_h}^2 - 1 + \frac{\beth_h}{K_{3h}K_{4h} - K_{5h}^2}\right] > 0,$$

where

$$\beth_h = K_{2h}^2 K_{3h} + K_{1h}^2 K_{4h} - 2K_{1h}K_{2h}K_{5h}.$$

(ii): Comparison between (4) and (20)

$$\mathrm{MSE}(V_2) > MSE\left(\hat{\bar{Y}}_A\right)_{min} \text{ if}$$

$$\sum_{h=1}^{L} G_h^2 \bar{Y}_h^2 \left[\delta_h \left\{ \Omega_{y_h} C_{y_h}^2 + \Omega_{x_h} C_{x_h}^2 \left(1 - 2K_{y_h x_h} \sqrt{\Omega_{y_h x_h}}\right) \right\} - 1 + \frac{\mathcal{J}_h}{K_{3h} K_{4h} - K_{5h}^2} \right] > 0.$$

(**iii**): Comparison between (6) and (20)

$$\mathrm{MSE}(V_3) > MSE\left(\hat{\bar{Y}}_A\right)_{min} \text{ if}$$

$$\sum_{h=1}^{L} G_h^2 \bar{Y}_h^2 \left[\delta_h \left\{ \Omega_{y_h} C_{y_h}^2 + \Omega_{x_h} C_{x_h}^2 \left(1 + 2K_{y_h x_h} \sqrt{\Omega_{y_h x_h}}\right) \right\} - 1 + \frac{\mathcal{J}_h}{K_{3h} K_{4h} - K_{5h}^2} \right] > 0.$$

(**iv**): Comparison between (7) and (20)

$$\mathrm{MSE}(V_4) > MSE\left(\hat{\bar{Y}}_A\right)_{min} \text{ if}$$

$$\sum_{h=1}^{L} G_h^2 \bar{Y}_h^2 \left[\delta_h C_{y_h}^2 \Omega_{y_h} \left(1 - \rho_{y_h x_h}^2\right) - 1 + \frac{\mathcal{J}_h}{K_{3h} K_{4h} - K_{5h}^2} \right] > 0.$$

(**v**): Comparison between (11) and (20)

$$\mathrm{MSE}(V_5) > MSE\left(\hat{\bar{Y}}_A\right)_{min} \text{ if}$$

$$\sum_{h=1}^{L} G_h^2 \bar{Y}_h^2 \left[\delta_h \left\{ \Omega_{y_h} C_{y_h}^2 + \frac{1}{4} \Omega_{x_h} C_{x_h}^2 - K_{y_h x_h} \sqrt{\Omega_{y_h x_h}} C_{x_h}^2 \right\} - 1 + \frac{\mathcal{J}_h}{K_{3h} K_{4h} - K_{5h}^2} \right] > 0.$$

(**vi**): Comparison between (12) and (20)

$$\mathrm{MSE}(V_6) > MSE\left(\hat{\bar{Y}}_A\right)_{min} \text{ if}$$

$$\sum_{h=1}^{L} G_h^2 \bar{Y}_h^2 \left[\delta_h \left\{ \Omega_{y_h} C_{y_h}^2 + \frac{1}{4} \Omega_{x_h} C_{x_h}^2 + K_{y_h x_h} \sqrt{\Omega_{y_h x_h}} C_{x_h}^2 \right\} - 1 + \frac{\mathcal{J}_h}{K_{3h} K_{4h} - K_{5h}^2} \right] > 0.$$

(**vii**): Comparison between (14) and (20)

$$\mathrm{MSE}(V_7) > MSE\left(\hat{\bar{Y}}_A\right)_{min} \text{ if}$$

$$\sum_{h=1}^{L} G_h^2 \bar{Y}_h^2 \left[\delta_h \left\{ \Omega_{y_h} C_{y_h}^2 - \frac{K_{y_h x_h}^2 \Omega_{y_h x_h}}{\Omega_{x_h}} C_{x_h}^2 \right\} - 1 + \frac{\mathcal{J}_h}{K_{3h} K_{4h} - K_{5h}^2} \right] > 0.$$

6. Analysis

6.1. Simulation Study

A Monte Carlo simulation study is conducted to examine the performance of the existing and proposed estimators under different finite-population structures. The simulation considers three population models representing linear, nonlinear and periodic relationships between the study variable and the auxiliary variable. The population is divided into four strata, and stratified systematic sampling is applied independently within each stratum. The MSE expressions derived in Sections 3 and 4 are evaluated for all competing estimators using the corresponding population quantities. The simulation considers three overall sample sizes and 25,000 independent repetitions for each population model and sample size.

6.1.1. Population Generation

The finite population contains $N = 2400$ units divided into $L = 4$ equal strata, with

$$N_h = 600, \qquad G_h = \frac{N_h}{N} = 0.25, \qquad h = 1, 2, 3, 4.$$

Within each stratum, the units are indexed by $i = 1, \ldots, N_h$. Three population models are generated to represent different relationships between the study variable Y and the auxiliary variable X.

Model I: Linear Stratified Population

The auxiliary variable is generated according to

$$X_{hi}^{(1)} = a_h + b_{hi} + \varepsilon_{x_{hi}},$$

where

$$(a_1, a_2, a_3, a_4) = (20, 35, 50, 65),$$

$$(b_1, b_2, b_3, b_4) = (0.025, 0.030, 0.035, 0.040),$$

and

$$\varepsilon_{x_{hi}} \sim N(0, 6^2).$$

The study variable is generated from

$$Y_{hi}^{(1)} = c_h + \beta_h X_{hi}^{(1)} + \varepsilon_{y_{hi}},$$

where

$$(c_1, c_2, c_3, c_4) = (10, 15, 20, 25),$$

$$(\beta_1, \beta_2, \beta_3, \beta_4) = (0.80, 0.90, 1.00, 1.10),$$

and

$$\varepsilon_{y_{hi}} \sim N(0, 12^2).$$

This model provides a stratified population with a positive and approximately linear association between Y and X, while allowing the relationship to differ across strata.

Model II: Nonlinear and Skewed Stratified Population

To represent a nonlinear population structure, the auxiliary variable is generated as

$$X_{hi}^{(2)} = a_h + b_{hi} + c_{hi}^2 + \varepsilon_{x_{hi}},$$

where

$$(a_1, a_2, a_3, a_4) = (15, 30, 45, 60),$$

$$(b_1, b_2, b_3, b_4) = (0.015, 0.020, 0.025, 0.030),$$

$$(c_1, c_2, c_3, c_4) = (1.5, 1.2, 0.9, 0.6) \times 10^{-5},$$

and

$$\varepsilon_{x,hi} \sim N(0, 8^2).$$

The corresponding study variable follows

$$Y_{hi}^{(2)} = c_h + \beta_h X_{hi}^{(2)} + \lambda_h \left(X_{hi}^{(2)}\right)^2 + \varepsilon_{y,hi},$$

where

$$(c_1, c_2, c_3, c_4) = (8, 12, 16, 20),$$

$$(\beta_1, \beta_2, \beta_3, \beta_4) = (0.70, 0.80, 0.90, 1.00),$$

$$(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0.0008, 0.0007, 0.0006, 0.0005),$$

and

$$\varepsilon_{y_{hi}} \sim N(0, 15^2).$$

This model allows the relationship between the study and auxiliary variables to depart from a linear form and provides greater variation in the auxiliary characteristics used by the proposed estimators.

Model III: Periodic Stratified Population

The third model introduces a periodic component into the auxiliary variable:

$$X_{hi}^{(3)} = a_h + b_h i + A_h \sin\left(\frac{2\pi i}{P_h}\right) + \varepsilon_{x_{hi}},$$

where

$$(a_1, a_2, a_3, a_4) = (25, 40, 55, 70),$$
$$(b_1, b_2, b_3, b_4) = (0.020, 0.025, 0.030, 0.035),$$
$$(A_1, A_2, A_3, A_4) = (6, 8, 10, 12),$$
$$(P_1, P_2, P_3, P_4) = (40, 50, 60, 70),$$

and

$$\varepsilon_{x_{hi}} \sim N(0, 7^2).$$

The study variable is generated as

$$Y_{hi}^{(3)} = c_h + \beta_h X_{hi}^{(3)} + \eta_h \cos\left(\frac{2\pi i}{P_h}\right) + \varepsilon_{y_{hi}},$$

where

$$(c_1, c_2, c_3, c_4) = (12, 18, 24, 30),$$
$$(\beta_1, \beta_2, \beta_3, \beta_4) = (0.85, 0.95, 1.05, 1.15),$$
$$(\eta_1, \eta_2, \eta_3, \eta_4) = (4, 5, 6, 7),$$

and

$$\varepsilon_{y_{hi}} \sim N(0, 14^2).$$

For every generated population, the required stratum-specific characteristics of the auxiliary variable are calculated before sampling. These quantities include S_{xh}, M_h, D_h, R_h, C_{xh}, γ_{1h}, and γ_{2h}, which are used according to the corresponding choices of d_{2h} in the proposed estimator family.

6.1.2. Simulation Procedure

The simulation procedure is carried out using 25,000 independent Monte Carlo repetitions for each population model and each overall sample size. A fixed random seed is used to ensure reproducibility. The overall sample sizes are selected as

$$n \in \{80, 120, 200\}.$$

Equal allocation is used across the four strata, giving

$$n_h = \frac{n}{4} \in \{20, 30, 50\}.$$

Since $N_h = 600$ for each stratum, the corresponding systematic sampling intervals are

$$k_h = \frac{N_h}{n_h} \in \{30, 20, 12\}.$$

The simulation proceeds as follows:

1. Generate a finite population of $N = 2400$ units and divide it into $L = 4$ equal-sized strata, with $N_h = 600$ units in each stratum. Generate the study variable Y and auxiliary variable X according to one of the three population models described in the previous Section 6.1.1.

2. For each stratum, calculate the population means, variances, covariances, coefficients of variation, correlations, and the auxiliary-variable characteristics required by the proposed estimator family.
3. For each overall sample size $n \in \{80, 120, 200\}$, allocate $n_h = n/4$ units to each stratum and calculate the corresponding systematic sampling interval

$$k_h = \frac{N_h}{n_h}.$$

A random starting position

$$d_h \in \{1, \ldots, k_h\}$$

is selected independently within each stratum, and the systematic sample from stratum h consists of

$$d_h,\ d_h + k_h,\ d_h + 2k_h, \ldots, d_h + (n_h - 1)k_h.$$

4. Using the population quantities obtained in Step 2, evaluate the MSE expressions of the existing estimators $V_1,\ V_2,\ \ldots, V_7$ derived in Section 3. For each proposed estimator in Section 4, evaluate its corresponding MSE expression using the same population quantities and the transformation values specified in Tables 1 and 2.
5. Repeat Steps 1–4 independently for 25,000 repetitions for each population model and each sample size. For every estimator, summarize the resulting MSE values. The corresponding RMSE is calculated as

$$\mathrm{RMSE}(\hat{\bar{Y}}_j) = \sqrt{\mathrm{MSE}(\hat{\bar{Y}}_j)},$$

where

$$\hat{\bar{Y}}_j = V_1,\ V_2, \ldots,\ V_7,\ A_1,\ A_2,\ \ldots,\ A_{12}.$$

6. The percent relative efficiency of each estimator is calculated with respect to the conventional stratified systematic estimator V_1 as

$$\mathrm{PRE}(\hat{\bar{Y}}_j) = \frac{\mathrm{MSE}(V_1)}{\mathrm{MSE}(\hat{\bar{Y}}_j)} \times 100.$$

The RMSE and PRE results are reported separately for the three population models and the three overall sample sizes. The simulation results are presented in Tables 3–5.

Table 3. RMSE and PRE of the estimators for $n = 80$ under the three simulation models.

Estimator	Model I		Model II		Model III	
	RMSE	**PRE**	**RMSE**	**PRE**	**RMSE**	**PRE**
V_1	5.083	100.000	6.207	100.000	5.503	100.000
V_2	4.706	116.685	5.760	116.111	5.091	116.874
V_3	5.359	89.958	6.618	87.952	5.880	87.578
V_4	4.430	131.664	5.464	128.996	4.782	132.415
V_5	4.620	121.047	5.628	121.634	4.952	123.485
V_6	5.210	95.179	6.434	93.078	5.711	92.862
V_7	4.234	144.115	5.220	141.379	4.602	143.053
A_1	3.850	174.392	4.870	162.440	4.226	169.638
A_2	3.594	200.070	4.635	179.207	4.024	187.034
A_3	3.704	188.345	4.761	169.860	4.133	177.394
A_4	3.436	218.829	4.494	190.796	3.916	197.468
A_5	3.539	206.333	4.617	180.714	3.997	189.635
A_6	3.254	244.025	4.318	206.666	3.753	215.039
A_7	3.454	216.533	4.550	186.086	3.963	192.797
A_8	3.630	196.060	4.731	172.096	4.093	180.786
A_9	3.364	228.282	4.453	194.182	3.888	200.291
A_{10}	3.492	211.823	4.653	177.963	4.043	185.346
A_{11}	3.328	233.250	4.403	198.737	3.847	204.604
A_{12}	3.562	203.719	4.591	182.770	4.012	188.146

Table 4. RMSE and PRE of the estimators for $n = 120$ under the three simulation models.

Estimator	Model I		Model II		Model III	
	RMSE	**PRE**	**RMSE**	**PRE**	**RMSE**	**PRE**
V_1	4.031	100.000	4.916	100.000	4.329	100.000
V_2	3.658	121.447	4.443	122.367	3.872	124.968
V_3	4.314	87.292	5.324	85.241	4.724	83.967
V_4	3.413	139.437	4.175	138.585	3.649	140.752
V_5	3.565	127.871	4.320	129.428	3.779	131.227
V_6	4.194	92.369	5.178	90.115	4.577	89.474
V_7	3.240	154.759	3.986	152.060	3.467	155.947
A_1	2.912	191.590	3.647	181.653	3.112	193.447
A_2	2.706	221.803	3.421	206.467	2.867	228.089
A_3	2.795	207.959	3.535	193.431	3.002	207.716
A_4	2.564	247.129	3.312	220.217	2.743	248.960
A_5	2.663	229.111	3.404	208.531	2.836	233.084
A_6	2.432	274.727	3.132	246.336	2.570	283.858
A_7	2.586	242.896	3.267	226.318	2.775	243.322
A_8	2.821	204.255	3.473	200.344	2.973	212.098
A_9	2.511	257.624	3.200	235.911	2.687	259.716
A_{10}	2.635	233.967	3.380	211.477	2.853	230.269
A_{11}	2.480	264.113	3.159	242.054	2.643	268.257
A_{12}	2.685	225.432	3.360	213.995	2.873	226.987

Table 5. RMSE and PRE of the estimators for $n = 200$ under the three simulation models.

Estimator	Model I		Model II		Model III	
	RMSE	**PRE**	**RMSE**	**PRE**	**RMSE**	**PRE**
V_1	3.070	100.000	3.686	100.000	3.238	100.000
V_2	2.727	126.773	3.243	129.168	2.846	129.467
V_3	3.337	84.665	4.052	82.740	3.581	81.774
V_4	2.521	148.235	3.022	148.747	2.707	143.057
V_5	2.643	134.994	3.149	136.950	2.782	135.475
V_6	3.243	89.638	3.948	87.154	3.439	88.632
V_7	2.379	166.597	2.871	164.820	2.493	168.723
A_1	2.112	211.257	2.610	199.347	2.261	205.040
A_2	1.954	246.754	2.438	228.456	2.069	244.863
A_3	2.027	229.610	2.513	215.116	2.166	223.491
A_4	1.817	285.583	2.319	252.586	1.977	268.182
A_5	1.917	256.423	2.387	238.435	2.053	248.532
A_6	1.727	315.951	2.174	287.226	1.799	323.759
A_7	1.844	277.163	2.290	258.943	1.999	262.312
A_8	2.034	227.762	2.467	223.289	2.175	221.419
A_9	1.802	290.504	2.236	271.785	1.901	290.011
A_{10}	1.931	252.843	2.369	241.951	2.090	240.005
A_{11}	1.771	300.541	2.202	280.000	1.867	300.688
A_{12}	1.955	246.493	2.400	235.718	2.028	254.936

6.2. Real Life Applications

The numerical study evaluates the proposed estimators using three real finite populations. For each population, the required population quantities are calculated directly from the observed data. The MSE expressions derived in Sections 3 and 4 are then evaluated using these population quantities for the conventional stratified systematic estimator, the existing estimators $V_1, V_2 \ldots, V_7$, and the proposed estimator sub-classes. The PRE of each estimator is calculated relative to the conventional stratified systematic estimator. The results are

presented in terms of RMSE and PRE to provide a direct comparison between the existing and proposed estimators. The results are presented in Table 6.

Population 1: The first population comes from the education sector of the (Punjab Development Statistics 2014) [35], using Tables 89 and 90 on pages 135–136. The data correspond to the 2012–2013 period and provide information on government primary schools for boys and girls together with student enrolment and staff figures across the divisions and districts of Punjab. For the present study, the information on the number of schools and student enrolment serves as the two variables of interest. The population is divided into two strata, with Y_1 and Y_2 denoting the student-enrolment values and X_1 and X_2 representing the corresponding numbers of schools in the two strata, respectively.

$N_1 = 36, N_2 = 36, n_1 = 6, n_2 = 6, k_1 = 6, k_2 = 6,\ \bar{Y}_1 = 74690.890,\ \bar{Y}_2 = 47948.420,\ \bar{X}_1 = 1054.389,$
$\bar{X}_2 = 223.0556, X_{1_{min}} = 388, X_{2_{min}} = 84, X_{1_{max}} = 2370, X_{2_{max}} = 478, C_{y_1} = 0.812,$
$C_{y_2} = 0.461, C_{x_1} = 0.382, C_{x_2} = 0.409, S_{y_1} = 60681.580, S_{y_2} = 22127.630, S_{x_1} = 402.609,$
$S_{x_2} = 91.274, S_{y_1x_1} = 2244236, S_{y_2x_2} = 780056.200, \rho_{y_1} = 0.029, \rho_{y_2} = 0.006, \rho_{x_1} = -0.345,$
$\rho_{x_2} = -0.207, \rho_{y_1x_1} = 0.091, \rho_{y_2x_2} = 0.386, \gamma_{11} = 1.098, \gamma_{12} = 0.661, \gamma_{21} = 1.911, \gamma_{22} = 0.315,$
$\Omega_{y_1} = 0.458, \Omega_{y_2} = 1.330, \Omega_{x_1} = 0.463, \Omega_{x_2} = 0.503, \Omega_{y_1x_1} = 0.991,\ \Omega_{y_2x_2} = 2.646,$
$K_{y_1x_1} = 0.195, K_{y_2x_2} = 0.436,\ \delta_1 = 0.162037, \delta_2 = 0.162037.$

Population 2: The data come from the industrial sector of the Punjab Development Statistics 2014, using the table "Number of Registered Factories, Employment Level by Division and District" on page 226 [35]. The data cover 2012 and 2013 and provide information on registered factories and employment across the divisions and districts of Punjab. In this study, the employment level serves as the study variable, while the number of registered factories serves as the auxiliary variable. The population is divided into two strata, with Y_1 and Y_2 representing the employment values and X_1 and X_2 denoting the corresponding numbers of registered factories in the two strata, respectively.

$N_1 = 36, N_2 = 36, n_1 = 6, n_2 = 6, k_1 = 6, k_2 = 6, \bar{Y}_1 = 25431.470, \bar{Y}_2 = 25791.36, \bar{X}_1 = 328.806,$
$\bar{X}_2 = 333.000, X_{1_{min}} = 24, X_{2_{min}} = 24, X_{1_{max}} = 1986, X_{2_{max}} = 2055, C_{y_1} = 1.630, C_{y_2} = 1.635,$
$C_{x_1} = 1.334, C_{x_2} = 1.359, S_{y_1} = 41459.890, S_{y_2} = 42163.640, S_{x_1} = 438.519, S_{x_2} = 452.713,$
$S_{y_1x_1} = 16579414, S_{y_2x_2} = 9915213, \rho_{y_1} = 0.194, \rho_{y_2} = 0.226, \rho_{x_1} = 0.146, \rho_{x_2} = 0.136, \rho_{y_1x_1} = 0.912,$
$\rho_{y_2x_2} = 0.519, \gamma_{11} = 2.554, \gamma_{12} = 2.575, \gamma_{21} = 6.714, \gamma_{22} = 6.861, \Omega_{y_1} = 1.970, \Omega_{y_2} = 2.128,$
$\Omega_{x_1} = 1.729, \Omega_{x_2} = 1.682, \Omega_{y_1x_1} = 1.139,\ \Omega_{y_2x_2} = 1.265, K_{y_1x_1} = 1.115, K_{y_2x_2} = 0.624,$
$\delta_1 = 0.162037, \delta_2 = 0.162037.$

Population 3: This population represents the medical education sector of Punjab and is organized into two strata. The study variables y_1 and y_2 denote the numbers of medical institutions in the two strata, while x_1 and x_2 are the corresponding auxiliary variables representing the teaching staff in medical institutes. The auxiliary information is used to provide additional information about the structure of the medical education system and to improve the precision of the estimators under stratified systematic sampling. The data are taken from the Education-Professional/Medical Institutes section of the Punjab Development Statistics, covering pages 129–132 in Table 86 [35].

$N_1 = 60, N_2 = 60, n_1 = 10, n_2 = 10, k_1 = 6, k_2 = 6, \bar{Y}_1 = 536.583, \bar{Y}_2 = 391.700, \bar{X}_1 = 6570.467,$
$\bar{X}_2 = 6459.800, X_{1_{min}} = 208, X_{2_{min}} = 92, X_{1_{max}} = 85231, X_{2_{max}} = 27545, C_{y_1} = 1.073, C_{y_2} = 1.406,$
$C_{x_1} = 1.781, C_{x_2} = 1.337, S_{y_1} = 575.789, S_{y_2} = 550.898, S_{x_1} = 11707.320, S_{x_2} = 8635.260,$
$S_{y_1x_1} = 1145236, S_{y_2x_2} = 4624843, \gamma_{11} = 5.40, \gamma_{12} = 1.134, \gamma_{21} = 35.272, \gamma_{22} = -0.135, \rho_{y_1} = 0.781,$
$\rho_{y_2} = 0.900, \rho_{x_1} = 0.212, \rho_{x_2} = 0.907, \rho_{y_1x_1} = 0.169, \rho_{y_2x_2} = 0.972, \Omega_{y_1} = 8.029, \Omega_{y_2} = 9.10,$
$\Omega_{x_1} = 2.905, \Omega_{x_2} = 9.164, \Omega_{y_1x_1} = 2.764,\ \Omega_{y_2x_2} = 0.994, K_{y_1x_1} = 0.102, K_{y_2x_2} = 1.023,$
$\delta_1 = 0.09833333, \delta_2 = 0.09833333.$

Table 6. RMSE and PRE of the estimators for three real data-sets.

Estimator	Model I		Model II		Model III	
	RMSE	**PRE**	**RMSE**	**PRE**	**RMSE**	**PRE**
V_1	9746.597	100.000	17,041.710	100.000	365.187	100.000
V_2	10,054.967	93.960	12,054.273	199.869	335.058	118.793
V_3	11,631.160	70.210	27,644.258	38.003	639.184	32.642
V_4	9512.200	104.989	11,618.191	215.153	259.372	198.237
V_5	9130.526	113.950	15,526.099	120.476	392.975	86.358
V_6	10,873.099	80.352	20,541.377	68.828	419.306	75.852
V_7	8768.568	123.55	14,136.670	145.322	304.549	143.785
A_1	7737.109	158.689	11,038.531	238.344	234.549	242.417
A_2	7291.303	178.688	10,765.684	250.578	227.082	258.622
A_3	7899.086	152.241	11,405.617	223.249	197.525	341.811
A_4	7434.519	171.870	10,953.365	242.065	182.920	398.573
A_5	7252.074	180.627	10,698.206	253.749	199.201	336.084
A_6	7028.432	192.304	10,377.145	269.694	195.089	350.40
A_7	7007.290	193.467	11,529.771	218.467	188.1367	376.776
A_8	7579.282	165.367	11,117.853	234.955	202.870	324.037
A_9	7681.056	161.014	11,182.750	232.236	208.343	307.237
A_{10}	7814.700	155.554	10,995.944	240.194	222.329	269.798
A_{11}	8014.839	147.882	10,163.164	281.160	239.760	231.994
A_{12}	8091.340	145.099	10,895.427	244.645	227.867	256.843

6.3. Interpretation of the Results

The results presented in Tables 3–5 show that the proposed estimators provide lower RMSE values than the conventional stratified systematic estimator under all three simulation models. The PRE values of the proposed estimators remain above 100, indicating improvements over the reference estimator. The results also show that increasing the sample size from $n = 80$ to $n = 120$ and $n = 200$ reduces the RMSE for both the existing and proposed estimators. Figure 1 also shows that the proposed estimators generally have higher PRE values than the conventional stratified systematic estimator and most of the existing estimators. This pattern is observed under Models I, II, and III, indicating that the proposed estimators maintain their performance across different population structures. Among the existing estimators, V_4 and V_7 generally produce lower RMSE values than the conventional estimator. In contrast, V_3 and V_6 have PRE values below 100 under the reported simulation settings. The proposed estimators show a different pattern, with considerably higher PRE values across the three models. For example, at $n = 200$, the PRE of A_6 reaches 315.951, 287.226, and 323.759 for Models I, II, and III, respectively. The results indicate that the choice of transformation can have an important effect on estimator efficiency.

The real-data analysis provides further evidence of the differences among the estimators. In Population-1, the proposed estimators have smaller RMSE values than the reference estimator, with A_7 giving the lowest RMSE among the proposed estimators. In Population-2, A_{11} gives the lowest RMSE among the proposed estimators, whereas A_4 gives the lowest RMSE in Population-3. Their corresponding PRE values are 193.467, 281.160, and 398.573, respectively. The differences among the three populations are also visible in Figure 2. Although the proposed estimators perform better than the conventional estimator in the reported cases, the estimator producing the smallest RMSE varies from one population to another. This variation is reasonable because the three populations have different characteristics and different relationships between the study and auxiliary variables. It indicates that the choice of transformation can influence estimator performance in practical applications.

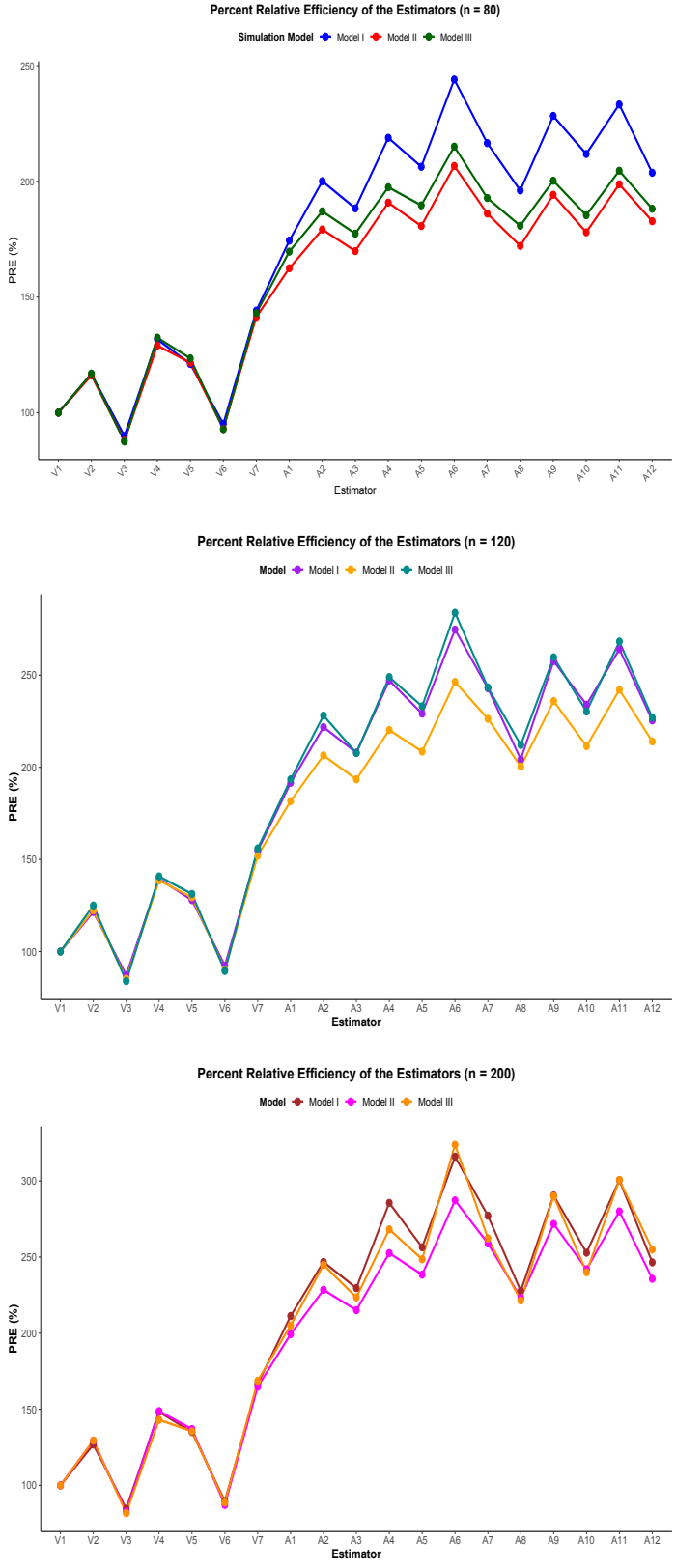


Figure 1. Simulation results for the three population models, comparing existing estimators (V_1–V_7) with the proposed estimators (A_1–A_{12}).

Overall, the results in Tables 3–6, together with Figures 1 and 2, support the theoretical findings developed in Section 5. The proposed family provides several transformation-based alternatives that can reduce estimation error when informative auxiliary information is available. The numerical results also indicate that the choice of a particular subclass should

be based on the characteristics of the population and the auxiliary variable rather than on the assumption that one transformation will perform identically in every setting.

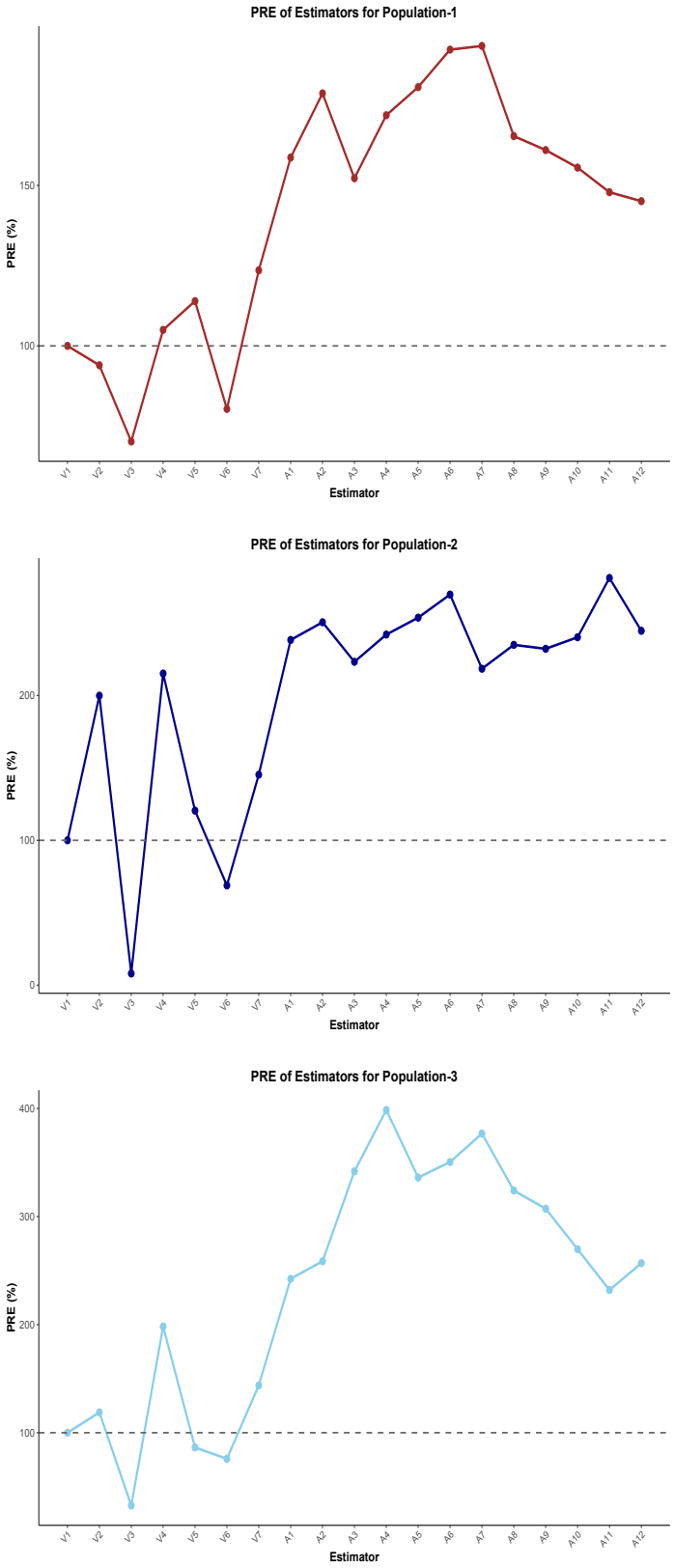


Figure 2. Comparative results for the three real-data populations, illustrating the performance of the existing estimators (V_1–V_7) and the proposed estimators (A_1–A_{12}).

7. Conclusions

This study proposed a family of estimators for estimating the finite population mean under stratified systematic sampling. The proposed estimators use different power transformations of the auxiliary variable together with auxiliary characteristics such as the standard deviation, range, relative range, coefficient of variation, skewness, and kurtosis. The theoretical expressions for bias and MSE were derived using first-order approximations, and the optimal constants were obtained through MSE minimization.

The simulation study considered three population models representing linear, nonlinear, and periodic relationships. The results showed a clear reduction in RMSE as the sample size increased. The proposed estimators also produced PRE values greater than 100 across the reported simulation settings, showing lower RMSE than the conventional stratified systematic estimator. These findings provide numerical support for the theoretical comparisons developed in the study. The analysis of the three real populations produced similar overall findings, although the individual subclasses behaved differently from one population to another. The proposed estimators generally provided lower RMSE values than the conventional estimator, while the subclass with the smallest RMSE changed across the populations. This result indicates that the proposed family provides several transformation choices that can be considered according to the characteristics of the available auxiliary information.

There are some limitations to the present study. The theoretical results are based on first-order approximations, and the numerical analysis uses a limited number of simulated and real populations. In addition, the theoretical MSE calculations rely on population quantities that may not always be known in practical surveys. Therefore, the results should be interpreted within the settings considered in this study.

Future Research Directions

Further research may examine the performance of the proposed estimators when their optimal constants are obtained from sample-based estimates rather than known population quantities. Particular attention could be given to the effects of stratum-specific sample sizes, variation in population characteristics, and the accuracy of estimated auxiliary measures on bias, mean squared error, and numerical stability. Such an investigation would help identify the conditions under which the resulting estimators perform reliably in practical survey applications.

The proposed framework may also be extended to settings involving multiple auxiliary variables, unequal allocation among strata, non-response, missing observations, and measurement errors. Further studies could investigate its behaviour under alternative systematic sampling designs and population structures with pronounced periodic or nonlinear patterns. In addition, data-driven methods could be explored to construct auxiliary information for populations in which the relationship between the study variable and auxiliary variables is more complex. These extensions would help establish the scope and practical usefulness of the proposed estimation approach across a wider range of survey settings.

Overall, the findings indicate that the proposed estimator family provides a useful alternative for finite population mean estimation under stratified systematic sampling, particularly when relevant auxiliary information is available.

Author Contributions: Conceptualization, H.A.; Methodology, H.A.; Software, H.A. and F.A.A.; Validation, H.A. and F.A.A.; Formal analysis, H.A. and F.A.A.; Investigation, H.A. and F.A.A.; Resources, H.A. and F.A.A.; Data curation, H.A. and F.A.A.; Writing—original draft, H.A.; Writing—review & editing, F.A.A.; Visualization, F.A.A.; Supervision, F.A.A.; Project administration, H.A. and F.A.A.; Funding acquisition, F.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to the Deanship of Scientific Research and Libraries in Princess Nourah bint Abdulrahman University for funding this research work through the Suporting Publication in Top-impact journals Initiative (SPTIF-2026).

Data Availability Statement: The real data are secondary, and their source is given in the data section, while the simulated data have been generated using R software (latest v. 4.4.0).

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

References

1. Cochran, W.G. Relative efficiency of systematic and stratified random samples for a certain class of population. *Ann. Math. Stat.* **1946**, *17*, 164–177. [CrossRef]
2. Yates, F. Systematic sampling. *Philos. Trans. R. Soc. A* **1948**, *241*, 345–377. [CrossRef]
3. Madow, W.G. On the theory of systematic sampling. *Ann. Math. Stat.* **1948**, *19*, 333–354. [CrossRef]
4. Gautschi, W. Some remarks on systematic sampling. *Ann. Math. Stat.* **1957**, *28*, 385–394. [CrossRef]
5. Hajeck, J. Optimum strategy and other problems in probability sampling. *Čas. Pěst. Mat.* **1959**, *84*, 387–423. [CrossRef]
6. Kish, L. *Survey Sampling*; John Wiley & Sons: New York, NY, USA, 1965.
7. Murthy, M.N. *Sampling Theory and Methods*, 2nd ed.; Statistical Publishing Society: Calcutta, India, 1967.
8. Cochran, W.G. *Sampling Techniques*, 3rd ed.; John Wiley and Sons: Hoboken, NJ, USA, 1977.
9. Hansen, M.H.; Hurwitz, W.N. On the theory of sampling from finite populations. *Ann. Math. Stat.* **1943**, *14*, 333–362. [CrossRef]
10. Lahiri, D.B. On the question of bias of systematic sampling. *Proc. World Pet. Congr.* **1954**, *6*, 349–362.
11. Swain, A.K.P.C. The use of systematic sampling in ratio estimate. *J. Indian Stat. Assoc.* **1964**, *2*, 160–164.
12. Shukla, N.D. Systematic sampling and product method of estimation. In *Proceedings of All India Seminar on Demography and Statistics*; BHU: Varanasi, India, 1971.
13. Kushwaha, K.S.; Singh, H.P. Class of almost unbiased ratio and product estimators in systematic sampling. *J. Indian Soc. Agric. Stat.* **1989**, *41*, 193–205.
14. Kushwaha, S.N.S.; Kushwaha, K.S. A class of ratio, product and difference (RPD) estimators in systematic sampling. *Microelectron. Reliab.* **1993**, *33*, 455–45 [CrossRef]
15. Singh, R.; Singh, H.P. Almost unbiased ratio and product-type estimators in systematic sampling. *Qüestiió Quadr. Estad. Investig. Oper.* **1998**, *22*, 403–416.
16. Singh, H.P.; Tailor, R.; Jatwa, N.K. Modified ratio and product estimators for population mean in systematic sampling. *J. Mod. Appl. Stat. Methods* **2011**, *10*, 424–435. [CrossRef]
17. Singh, H.P.; Solanki, R.S. An efficient class of estimators for the population mean using auxiliary information in systematic sampling. *J. Stat. Theory Pract.* **2012**, *6*, 274–285. [CrossRef]
18. Singh, H.P.; Jatwa, N.K. A class of exponential type estimators in systematic sampling. *Econ. Qual. Control* **2013**, *27*, 195–208. [CrossRef]
19. Tailor, T.; Jatwa, N.K.; Singh, H.P. A ratio-cum-product estimator of finite population mean in systematic sampling. *Stat. Transit.* **2013**, *14*, 391–398. [CrossRef]
20. Khan, M.; Singh, R. Estimation of population mean in chain ratio-type estimator under systematic sampling. *J. Probab. Stat.* **2015**, *2015*, 248374. [CrossRef]
21. Noor-ul-Amin, M.; Javaid, A.; Hanif, M. Estimation of population mean in systematic random sampling using auxiliary information. *J. Stat. Manag. Syst.* **2017**, *20*, 1095–1106. [CrossRef]
22. Javaid, A.; Noor-ul-Amin, M.; Hanif, M. Modified ratio estimator in systematic random sampling under non-response. *Proc. Natl. Acad. Sci. India Sect. A Phys. Sci.* **2019**, *89*, 817–825. [CrossRef]
23. Mishra, R.; Singh, R.; Adichwal, N.K. A novel ratio cum product type exponential class of estimators of finite population mean in adaptive cluster sampling. *Braz. J. Biom.* **2025**, *43*, e-43745. [CrossRef]
24. Subza, M.; Alqurashi, T.; Chandawat, D.; Tamboli, S.; Raja, T.A.; Attri, A.K.; Wani, S.A. Generalized robust regression techniques and adaptive cluster sampling for efficient estimation of population mean in case of rare and clustered populations. *Sci. Rep.* **2025**, *15*, 2069. [CrossRef] [PubMed]
25. Qureshi, M.N.; Khalil, S.; Hanif, M. Generalized semi exponential type estimator under systematic sampling. *J. Stat. Theory Appl.* **2018**, *17*, 283–290. [CrossRef]
26. Nagy, M.; Qureshi, M.N.; Shaheen, N.; Hanif, M. Mean estimation using memory-type estimators in systematic sampling for time-scaled surveys. *Mathematics* **2026**, *14*, 1180. [CrossRef]
27. Almulhim, F.A.; Aljohani, H.M.; Daraz, U. A unified hybrid estimation strategy using multiple auxiliary transformations in systematic sampling with simulation and real-life applications. *Axioms* **2026**, *15*, 590. [CrossRef]

28. Almulhim, F.A.; Daraz, U.; Aljohani, H.M.; Shabbir, J. Symmetric hybrid transformations for robust mean estimation under stratified systematic sampling. *Aims Math.* **2026**, *11*, 32046–32080. [CrossRef]
29. Khan, M.; Shabbir, J. Some improved ratio, product, and regression estimators of finite population mean using minimum and maximum values. *Sci. World J.* **2013**, *2013*, 431868. [CrossRef] [PubMed]
30. Iftikhar, A.; Shi, H.; Hussain, S.; Qayyum, A.; El-Morshedy, M.; Al-Marzouki, S. Estimation of finite population mean in presence of maximum and minimum values under systematic sampling scheme. *Aims Math.* **2022**, *7*, 9825–9834. [CrossRef]
31. El-Morshedy, M.; Hussain, S.; Ullah, K.; Khalil, A.; Shabbir, J.; Mansoor, W. Finite population mean estimation under systematic sampling scheme in presence of maximum and minimum values using two auxiliary variables. *Math. Probl. Eng.* **2022**, *2022*, 2703178. [CrossRef]
32. Koçyiğit, E.G. New memory-type estimators for systematic sampling. *J. Adv. Res. Nat. Appl. Sci.* **2025**, *11*, 224–236. [CrossRef]
33. Karim, A.; Khan, H.; Mahmood, Y.; Riaz, M.; Ahmad, S. On estimation and monitoring of population mean using systematic sampling under an exponentially weighted moving average scheme. *Pak. J. Stat. Oper. Res.* **2024**, *20*, 517–531 [CrossRef]
34. Pal, S.K.; Mahmud, S.A.; Singh, H.P. An efficient estimation of finite population mean through difference estimator in systematic sampling. *Afr. Mat.* **2025**, *36*, 14. [CrossRef]
35. Bureau of Statistics. *Punjab Development Statistics Government of the Punjab, Lahore, Pakistan*; Bureau of Statistics: Islamabad, Pakistan, 2014.

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.