symmetry

MDPI

Article

Auxiliary-Information Estimation Under Stratified Double-Phase Systematic Sampling: Theoretical Properties and Performance Under Symmetric and Asymmetric Error Models

Umer Daraz [1] and Fatimah A. Almulhim [2,*]

[1] Department of Management Sciences, College of Business Administration, Hunan University, Changsha 410082, China; umerdaraz099@gmail.com
[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

Estimating a finite population mean under stratified double-phase systematic sampling requires methods that can use auxiliary information effectively while accounting for the structure of the sampling design. This study introduces two estimators based on harmonic and geometric combinations of auxiliary-variable sample means collected at the first and second sampling phases. Approximate bias and mean squared error (MSE) expressions are developed, and the unknown coefficients are optimized by minimizing the corresponding MSE functions. The theoretical properties of the proposed estimators are investigated, and their performance is compared with six existing estimators through a Monte Carlo simulation involving three artificial finite populations. These populations represent linear association, nonlinear association with asymmetric disturbances, and periodic patterns in the ordering of population units. Three nested sample-size configurations are considered to examine how estimation accuracy changes as the number of sampled units increases. Empirical MSE and percentage relative efficiency are used as performance measures, while graphical representations provide a visual comparison of the competing estimators across population models and sampling configurations. In addition, three real finite populations are considered to assess their performance using observed data. The proposed harmonic- and geometric-based approaches provide alternative ways to incorporate auxiliary information under stratified double-phase systematic sampling, with their relative advantages assessed through theoretical analysis, numerical comparisons, and graphical evaluation.

Keywords: stratified double-phase systematic sampling; finite population mean; auxiliary information; Monte Carlo simulation; symmetric and asymmetric error distributions; percentage relative efficiency

MSC: 62D05

Check for updates

Academic Editor: Calogero Vetro

Received: 25 October 2026
Revised: 8 November 2026
Accepted: 16 November 2026
Published: 17 November 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. Background

In many survey studies, obtaining reliable information about a finite population requires a sampling design that balances statistical precision with practical simplicity. Systematic random sampling is widely used in finite population studies because it offers a straightforward way to select representative units from an ordered sampling frame. The method involves choosing an initial unit at random and then selecting subsequent

units at fixed intervals until the required sample size is obtained. Unlike simple random sampling, this procedure distributes the selected units throughout the population through a systematic selection mechanism, making it useful in agricultural surveys, environmental monitoring, industrial quality assessment, economic studies, and social investigations. However, the effectiveness of systematic sampling may depend on how the population units are arranged, particularly when the sampling frame contains trends, recurring patterns, or other forms of structure. Such features can influence the precision of population estimates and have encouraged researchers to investigate alternative estimation techniques for this design. Early work by [1,2] established important foundations for the study of systematic sampling. Subsequent research considered the use of auxiliary information to improve estimation, including ratio-based methods introduced by [3] and product-based approaches investigated by [4]. These developments have contributed to the methodological basis for constructing and assessing more efficient estimators under systematic sampling.

Auxiliary information can play an important role in improving the precision of population mean estimation when the study variable is related to an additional variable whose population characteristics are known. Rather than relying exclusively on observations collected for the study variable, an estimator can use this supplementary information to obtain more efficient estimates. The way in which auxiliary information is incorporated depends on the nature of the relationship between the variables and the structure of the sampling design. Within systematic sampling, several estimation strategies have been developed to make better use of this relationship. For instance, ref. [5] examined ratio and product estimators with an emphasis on reducing bias, followed by further developments presented by [6]. The search for improved estimation procedures continued with the family of estimators proposed by [7] and the ratio-cum-product approach investigated by [8]. Another line of research explored exponential transformations as a means of incorporating auxiliary information into estimation, as reported by [9,10]. Extending this line of investigation to settings where information is collected over successive occasions, ref. [11] developed memory-type estimators for population mean estimation under systematic sampling. Their approach incorporates information retained from earlier occasions, providing an additional way to utilize auxiliary data in surveys where the information available at the current occasion is linked to previous observations.

The development of finite population mean estimators has involved various strategies for utilizing auxiliary information under different sampling designs. A broader class using ratio, product, and difference estimation was subsequently considered [12]. The use of auxiliary information for estimating the population mean was further explored under systematic random sampling [13], while modified ratio estimation was investigated to address non-response [14]. For rare and clustered populations, adaptive cluster sampling has been combined with exponential ratio-cum-product estimation [15] and robust regression techniques [16]. In addition, an exponentially weighted moving average scheme has been considered for estimating and monitoring the population mean under systematic sampling [17]. More recently, difference estimation has also been examined as an approach to improving finite population mean estimation under systematic sampling [18]. In addition, recent studies have proposed hybrid estimation strategies involving multiple auxiliary transformations and symmetric hybrid transformations to enhance mean estimation under systematic and stratified systematic sampling [19,20].

Motivation and Research Objectives

Existing research has established several approaches for estimating the finite population mean under systematic sampling, yet opportunities remain to investigate alternative ways of utilizing auxiliary information collected at different sampling phases. In particular,

limited attention has been given to constructing estimators that combine first- and second-phase auxiliary information through harmonic and geometric mean functions within a stratified double-phase systematic sampling framework. This provides the motivation for developing alternative estimation procedures that account for the structure of the sampling design while making effective use of auxiliary information available at both phases.

To address this issue, the present study proposes two estimators based on harmonic and geometric combinations of auxiliary-variable sample means obtained from the two sampling phases. Their approximate bias and mean squared error (MSE) expressions are derived, and the estimator coefficients are optimized by minimizing the corresponding MSE expressions. The performance of the two proposed estimators is assessed against six established methods using Monte Carlo simulations and three real-population datasets. The comparison considers how effectively the estimators perform under varying population structures and sampling configurations, providing insight into their practical usefulness and relative efficiency.

The proposed framework is intended for situations in which auxiliary information is collected from a first-phase sample and more detailed information on the study and auxiliary variables is obtained from a second-phase subsample within the strata. This sampling approach is particularly useful when collecting information on the main study variable requires considerable time, effort, or expense. Information gathered during the two phases is used to develop harmonic- and geometric-based estimators, which are then compared with existing methods. Their performance is examined across different population structures and sample sizes to understand how well they work under varying sampling conditions and where they may be most useful in practice.

2. Basic Notations

Consider a finite population of N units partitioned into L non-overlapping strata. The population in stratum h is denoted by

$$T_h = \{T_{h1}, T_{h2}, \ldots, T_{hN_h}\}, \qquad h = 1, 2, \ldots, L,$$

where N_h denotes the number of units in the hth stratum and

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

Let λ_Y denote the study variable and λ_X the auxiliary variable. The objective is to estimate the finite population mean of the study variable by using auxiliary information collected through a stratified double-phase systematic sampling design.

The sampling procedure is carried out independently within each stratum at both phases, with the second-phase sample selected as a systematic subsample of the corresponding first-phase sample. Consequently, the sampling units selected at the second phase are nested within those selected at the first phase. The procedure is described below.

First-phase systematic sampling: At the first phase, a systematic sample of size n_{1h} is drawn from stratum h, where

$$N_h = n_{1h} k_{1h},$$

and k_{1h} is a positive integer representing the systematic sampling interval. The first-phase random start, denoted by R_{1h}, is selected uniformly from

$$R_{1h} \in \{1, 2, \ldots, k_{1h}\}, \qquad P(R_{1h} = r_h) = \frac{1}{k_{1h}}.$$

For a selected starting position $R_{1h} = r_h$, the selected first-phase sample is

$$s_{1h}(r_h) = \left\{ T_{h,r_h}, T_{h,r_h+k_{1h}}, \ldots, T_{h,r_h+(n_{1h}-1)k_{1h}} \right\}.$$

The auxiliary variable λ_X is observed for every unit included in this sample. Its first-phase sample mean in stratum h is

$$\bar{\lambda}_{x_{1h}} = \frac{1}{n_{1h}} \sum_{i \in s_{1h}} \lambda_{x_{hi}}.$$

The first-phase sample therefore provides auxiliary information from a relatively large number of units in each stratum before the study variable is measured at the subsequent phase.
Second-phase systematic subsampling: At the second phase, a systematic subsample of size n_{2h} is selected from the first-phase sample in stratum h, with

$$n_{2h} < n_{1h} < N_h, \qquad n_{1h} = n_{2h} k_{2h},$$

where k_{2h} is the second-phase systematic sampling interval. The second-phase random start R_{2h} satisfies

$$R_{2h} \in \{1, 2, \ldots, k_{2h}\},$$

with

$$P(R_{2h} = q_h \mid R_{1h} = r_h) = \frac{1}{k_{2h}}.$$

Accordingly, conditional on the selected first-phase sample, the second-phase sample associated with the starting position q_h comprises the units located at the following positions.

$$q_h,\ q_h + k_{2h}, \ldots, q_h + (n_{2h} - 1)k_{2h}$$

within that first-phase sample. Denote this subsample by $s_{2h}(r_h, q_h)$. By construction,

$$s_{2h}(r_h, q_h) \subset s_{1h}(r_h).$$

At the second phase, both λ_Y and λ_X are observed for the selected units. The corresponding stratum-specific sample means are

$$\bar{\lambda}_{y_{2h}} = \frac{1}{n_{2h}} \sum_{j \in s_{2h}} \lambda_{y_{hj}},$$

$$\bar{\lambda}_{x_{2h}} = \frac{1}{n_{2h}} \sum_{j \in s_{2h}} \lambda_{x_{hj}}.$$

The overall population mean of the study variable is estimated using the stratum-weighted second-phase sample mean,

$$\bar{\lambda}_{y_{2st}} = \sum_{h=1}^{L} W_h \bar{\lambda}_{y_{2h}},$$

where

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

$$\sum_{h=1}^{L} W_h = 1.$$

Similarly, the combined auxiliary means obtained at the first and second phases are

$$\bar{\lambda}_{x_{1st}} = \sum_{h=1}^{L} W_h \bar{\lambda}_{x_{1h}},$$

$$\bar{\lambda}_{x_{2st}} = \sum_{h=1}^{L} W_h \bar{\lambda}_{x_{2h}}.$$

The first-phase auxiliary mean uses information from all first-phase stratum samples, whereas the second-phase means incorporate observations from the smaller subsamples. These quantities provide the basis for developing estimators that combine the study variable with auxiliary information under the stratified double-phase systematic sampling design.

The population means of the study and auxiliary variables within stratum h are defined as

$$\bar{\lambda}_{Y_h} = \frac{1}{N_h} \sum_{i=1}^{N_h} \lambda_{y_{hi}},$$

$$\bar{\lambda}_{X_h} = \frac{1}{N_h} \sum_{i=1}^{N_h} \lambda_{x_{hi}}.$$

The corresponding overall population means are

$$\bar{\lambda}_Y = \sum_{h=1}^{L} W_h \bar{\lambda}_{Y_h},$$

$$\bar{\lambda}_X = \sum_{h=1}^{L} W_h \bar{\lambda}_{X_h}.$$

Within stratum h, the finite population variances of the two variables are

$$S^2_{\lambda_{y_h}} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (\lambda_{y_{hi}} - \bar{\lambda}_{Y_h})^2$$

and

$$S^2_{\lambda_{x_h}} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (\lambda_{x_{hi}} - \bar{\lambda}_{X_h})^2.$$

Their covariance is given by

$$S_{\lambda_{y_h x_h}} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (\lambda_{y_{hi}} - \bar{\lambda}_{Y_h})(\lambda_{x_{hi}} - \bar{\lambda}_{X_h}).$$

The stratum-specific coefficients of variation are

$$C_{\lambda_{y_h}} = \frac{S_{\lambda_{y_h}}}{\bar{\lambda}_{Y_h}},$$

$$C_{\lambda_{x_h}} = \frac{S_{\lambda_{x_h}}}{\bar{\lambda}_{X_h}},$$

provided that the corresponding stratum means are nonzero. The finite population correlation coefficient between the study and auxiliary variables within stratum h is

$$\rho_{\lambda_{y_h x_h}} = \frac{S_{\lambda_{y_h x_h}}}{S_{\lambda_{y_h}} S_{\lambda_{x_h}}},$$

provided that both standard deviations are positive.

Systematic sampling may produce dependence between observations selected at different positions in the same stratum. To describe this feature, let $\rho_{\lambda_{x_{1h}}}$ denote the correlation between distinct first-phase auxiliary observations at specified sample positions in stratum h. Similarly, let $\rho_{\lambda_{y_{2h}}}$ and $\rho_{\lambda_{x_{2h}}}$ denote the corresponding correlations for the study and auxiliary variables observed at the second phase.

Using design-based covariance and variance operators, these correlations can be expressed as

$$\rho_{\lambda_{x_{1h}}} = \frac{\mathrm{Cov}(\lambda_{x_{hi}}, \lambda_{x_{hi'}})}{\sqrt{\mathrm{Var}(\lambda_{x_{hi}})\,\mathrm{Var}(\lambda_{x_{hi'}})}}, \qquad i \neq i',$$

and

$$\rho_{\lambda_{y_{2h}}} = \frac{\mathrm{Cov}(\lambda_{y_{hj}}, \lambda_{y_{hj'}})}{\sqrt{\mathrm{Var}(\lambda_{y_{hj}})\,\mathrm{Var}(\lambda_{y_{hj'}})}}, \qquad j \neq j',$$

together with

$$\rho_{\lambda_{x_{2h}}} = \frac{\mathrm{Cov}(\lambda_{x_{hj}}, \lambda_{x_{hj'}})}{\sqrt{\mathrm{Var}(\lambda_{x_{hj}})\,\mathrm{Var}(\lambda_{x_{hj'}})}}, \qquad j \neq j'.$$

The association between the study and auxiliary variables measured on the same second-phase units in stratum h is represented by

$$\rho_{\lambda_{y_{2h}x_{2h}}} = \frac{\mathrm{Cov}(\lambda_{y_{hj}}, \lambda_{x_{hj}})}{\sqrt{\mathrm{Var}(\lambda_{y_{hj}})\,\mathrm{Var}(\lambda_{x_{hj}})}}.$$

This correlation concerns paired observations on the second-phase sample units and should not be confused with either the within-stratum population correlation $\rho_{\lambda_{y_h x_h}}$ or the overall population correlation $\rho_{\lambda_{yx}}$.

The preceding correlations may vary across strata and across systematic sample positions. Consequently, a common intraclass correlation should be used only when the population structure and sampling design provide a reasonable justification. When the correlations are position-dependent, their definitions should specify the relevant pairs or an explicitly defined average over those pairs.

For the development of estimators of $\bar{\lambda}_Y$, define the relative errors as

$$\tau_{0h} = \frac{\bar{\lambda}_{y_{2h}} - \bar{\lambda}_{Y_h}}{\bar{\lambda}_{Y_h}},$$

$$\tau_{1h} = \frac{\bar{\lambda}_{x_{1h}} - \bar{\lambda}_{X_h}}{\bar{\lambda}_{X_h}},$$

and

$$\tau_{2h} = \frac{\bar{\lambda}_{x_{2h}} - \bar{\lambda}_{X_h}}{\bar{\lambda}_{X_h}}.$$

These definitions require nonzero values of $\bar{\lambda}_{Y_h}$ and $\bar{\lambda}_{X_h}$.

Under the specified nested random-start design, with independent random starts across strata and uniform randomization within each stratum, the combined sample means are design-unbiased for their respective population means. It follows that

$$E(\tau_{0h}) = E(\tau_{1h}) = E(\tau_{2h}) = 0.$$

Moreover

$$E(\tau_{0h}^2) = \zeta_2 C_{\lambda_{y_h}}^2 \Pi_{\lambda_{y_h}},$$

$$E(\tau_{1h}^2) = \zeta_{1h} C_{\lambda_{x_h}}^2 \Pi_{\lambda_{x_h}},$$

$$E(\tau_{2_h}^2) = \zeta_{2h} C_{\lambda_{x_h}}^2 \Pi_{\lambda_{x_h}},$$

$$E(\tau_{0_h}\tau_{1h}) = \zeta_{1h} T_{\lambda_{y_h x_h}} C_{\lambda_{x_h}}^2 \sqrt{\Pi_{\lambda_{y_h x_h}}},$$

$$E(\tau_{0h}\tau_{2h}) = \zeta_{2h} T_{\lambda_{y_h x_h}} C_{\lambda_{x_h}}^2 \sqrt{\Pi_{\lambda_{y_h x_h}}},$$

$$E(\tau_{1h}\tau_{2h}) = \zeta_{1h} C_{\lambda_{x_h}}^2 \Pi_{\lambda_{x_h}},$$

where

$$\Pi_{\lambda_{x_h}} = 1 + (n_{2h} - 1)\rho_{\lambda_{x_h}},$$

$$\Pi_{\lambda_{y_h}} = 1 + (n_{2h} - 1)\rho_{\lambda_{y_h}},$$

$$\Pi_{\lambda_{y_h x_h}} = \frac{\Pi_{\lambda_{y_h}}}{\Pi_{\lambda_{x_h}}} = \frac{1 + (n_{2h} - 1)\rho_{\lambda_{y_h}}}{1 + (n_{2h} - 1)\rho_{\lambda_{x_h}}},$$

$$T_{\lambda_{y_h x_h}} = \rho_{\lambda_{y_h x_h}} \left(\frac{C_{\lambda_{y_h}}}{C_{\lambda_{x_h}}} \right),$$

$$\zeta_{1h} = \frac{N_h - 1}{n_{1h} N_h}, \qquad \zeta_{2h} = \frac{N_h - 1}{n_{2h} N_h}$$

and

$$\zeta_{3h} = \zeta_{2h} - \zeta_{1h}.$$

The bias and mean squared error expressions developed in this study are based on Taylor-series approximations of the proposed estimators with respect to the relative error terms. It is important to distinguish this approximation procedure from the sampling design used to obtain the observations. Although the algebraic expansion is general, the resulting bias and MSE depend on the design-specific moments of the relative errors. In the present study, these moments are considered under the stratified double-phase systematic sampling framework, which involves random starts at both sampling phases and a nested selection of the second-phase sample within the first-phase sample. The correlation terms are introduced to represent the dependence among observations selected at systematic positions. Accordingly, the applicability of the derived expressions depends on the validity of the adopted correlation structure and the corresponding design-based moment expressions.

For ease of reference, the main statistical symbols and their meanings are summarized in Table 1. The table brings together the notation used to describe the finite population, the two-phase sampling procedure, the population characteristics, and the estimator performance measures. It is intended to help readers follow the theoretical development and the subsequent comparisons without repeatedly searching for individual definitions.

Table 1. Summary of Notations.

Symbol	Definition	Symbol	Definition
N	Population size	L	Number of strata
N_h	Size of stratum h	W_h	Stratum weight, N_h/N
n_{1h}	First-phase sample size	n_{2h}	Second-phase sample size
k_{1h}	First-phase sampling interval	k_{2h}	Second-phase sampling interval

Table 1. *Cont.*

Symbol	Definition	Symbol	Definition
R_{1h}	First-phase random start	R_{2h}	Second-phase random start
s_{1h}	First-phase sample	s_{2h}	Second-phase subsample
$\bar{\lambda}_{Y_h}$	Population mean of study variable	$\bar{\lambda}_{X_h}$	Population mean of auxiliary variable
$\bar{\lambda}_Y$	Overall population mean of study variable	$\bar{\lambda}_X$	Overall population mean of auxiliary variable
$\bar{\lambda}_{y_{2h}}$	Second-phase sample mean of study variable	$\bar{\lambda}_{x_{1h}}$	First-phase sample mean of auxiliary variable
$\bar{\lambda}_{x_{2h}}$	Second-phase sample mean of auxiliary variable	$\bar{\lambda}_{y_{2st}}$	Stratified sample mean
$S^2_{\lambda_{y_h}}$	Study-variable variance	$S^2_{\lambda_{x_h}}$	Auxiliary-variable variance
$S_{\lambda_{y_h x_h}}$	Study-auxiliary covariance	$\rho_{\lambda_{y_h x_h}}$	Population correlation coefficient
$C_{\lambda_{y_h}}$	Study-variable coefficient of variation	$C_{\lambda_{x_h}}$	Auxiliary-variable coefficient of variation
$\rho_{\lambda_{x_{1h}}}$	First-phase systematic correlation	$\rho_{\lambda_{x_{2h}}}$	Second-phase systematic correlation
$\rho_{\lambda_{y_{2h}}}$	Study-variable systematic correlation	$\rho_{\lambda_{y_{2h} x_{2h}}}$	Phase-two cross-correlation
τ_{0h}	Relative error for study variable	τ_{1h}	Relative error for first-phase auxiliary variable
τ_{2h}	Relative error for second-phase auxiliary variable	$\zeta_{1h}, \zeta_{2h}, \zeta_{3h}$	Correction terms
$\Pi_{\lambda_{x_h}}$	Auxiliary design factor	$\Pi_{\lambda_{y_h}}$	Study-variable design factor
$\Pi_{\lambda_{y_h x_h}}$	Cross-variable design factor	$T_{\lambda_{y_h x_h}}$	Association term
Bias$(\cdot)$	Bias of estimator	MSE$(\cdot)$	Mean squared error
PRE	Percent relative efficiency	D	Number of Monte Carlo replications
$\mathbb{E}(\cdot)$	Expectation	Var$(\cdot)$	Variance
Cov$(\cdot,\cdot)$	Covariance	h	Stratum index

3. Review of Estimation Methods

The estimation of the finite population mean under stratified double-phase systematic sampling can be approached through several established procedures. This section outlines selected existing estimators that provide a basis for comparison with the methods proposed in this study. Their theoretical properties and numerical performance are considered to evaluate the potential advantages of the proposed approaches.

In the absence of auxiliary information, the population mean $\bar{\lambda}_Y$ is estimated directly by the second-phase sample mean, denoted by $\bar{\lambda}_{y_{2st}} = \hat{\bar{\Gamma}}_{y_{st}}$, defined as

$$\hat{\bar{\Gamma}}_{y_{st}} = \sum_{h=1}^{L} W_h \bar{\lambda}_{y_{2h}}. \tag{1}$$

Under the sampling assumptions adopted in this study, this estimator is unbiased for the population mean. Consequently, its bias is zero, and its mean squared error (MSE) is equal to its sampling variance, which is expressed as follows:

$$\text{MSE}(\hat{\bar{\Gamma}}_{y_{st}}) = \sum_{h=1}^{L} W_h^2 \zeta_{2h} \bar{\lambda}^2_{Y_h} \Pi_{\lambda_{y_h}} C^2_{\lambda_{y_h}}. \tag{2}$$

The ratio-based approach introduced in [3] is extended to the stratified double-phase systematic sampling framework, yielding the following estimator of the finite population mean:

$$\hat{\bar{\Gamma}}_{y_{Rst}} = \sum_{h=1}^{L} W_h \bar{\lambda}_{y_{2h}} \left(\frac{\bar{\lambda}_{x_{1h}}}{\bar{\lambda}_{x_{2h}}} \right). \tag{3}$$

The corresponding first-order bias is

$$\text{Bias}(\hat{\bar{\Gamma}}_{y_{Rst}}) \cong \sum_{h=1}^{L} W_h \zeta_{3h} \bar{\lambda}_{Y_h} C^2_{\lambda_{x_h}} \left[\Pi_{\lambda_{x_h}} - T_{\lambda_{y_h x_h}} \sqrt{\Pi_{\lambda_{y_h x_h}}} \right], \tag{4}$$

whereas its mean squared error is given by

$$\text{MSE}(\hat{\bar{\Gamma}}_{y_{Rst}}) \cong \sum_{h=1}^{L} W_h^2 \bar{\lambda}^2_{Y_h} \left[\zeta_{2h} C^2_{\lambda_{y_h}} \Pi_{\lambda_{y_h}} + \zeta_{3h} C^2_{\lambda_{x_h}} \left(\Pi_{\lambda_{x_h}} - 2T_{\lambda_{y_h x_h}} \sqrt{\Pi_{\lambda_{y_h x_h}}} \right) \right], \tag{5}$$

where

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

and

$$\zeta_{3h} = \zeta_{2h} - \zeta_{1h} = \frac{(N_h - 1)(n_{1h} - n_{2h})}{n_{1h} n_{2h} N_h}.$$

The product-based method developed in [4] provides an alternative means of using auxiliary information. Its corresponding form under stratified double-phase systematic sampling is given by

$$\hat{\bar{\Gamma}}_{y_{Pst}} = \sum_{h=1}^{L} W_h \bar{\lambda}_{y_{2h}} \left(\frac{\bar{\lambda}_{x_{2h}}}{\bar{\lambda}_{x_{1h}}} \right). \tag{6}$$

Using a Taylor approximation, the bias of this estimator is

$$\text{Bias}(\hat{\bar{\Gamma}}_{y_{Pst}}) \cong \sum_{h=1}^{L} W_h \zeta_{3h} \bar{\lambda}_{Y_h} T_{\lambda_{y_h x_h}} C^2_{\lambda_{x_h}} \sqrt{\Pi_{\lambda_{y_h x_h}}} \tag{7}$$

and its MSE is

$$\text{MSE}(\hat{\bar{\Gamma}}_{y_{Pst}}) \cong \sum_{h=1}^{L} W_h^2 \bar{\lambda}^2_{Y_h} \left[\zeta_{2h} C^2_{\lambda_{y_h}} \Pi_{\lambda_{y_h}} + \zeta_{3h} C^2_{\lambda_{x_h}} \left(\Pi_{\lambda_{x_h}} + 2T_{\lambda_{y_h x_h}} \sqrt{\Pi_{\lambda_{y_h x_h}}} \right) \right]. \tag{8}$$

Exponential transformations of ratio and product estimators have also been considered in the literature. The estimators presented in [6] are adapted to stratified double-phase systematic sampling and are expressed as follows:

$$\hat{\bar{\Gamma}}_{y_{Rest}} = \sum_{h=1}^{L} W_h \bar{\lambda}_{y_{2h}} \exp\left(\frac{\bar{\lambda}_{x_{1h}} - \bar{\lambda}_{x_{2h}}}{\bar{\lambda}_{x_{1h}} + \bar{\lambda}_{x_{2h}}} \right) \tag{9}$$

and

$$\hat{\bar{\Gamma}}_{y_{Pest}} = \sum_{h=1}^{L} W_h \bar{\lambda}_{y_{2h}} \exp\left(\frac{\bar{\lambda}_{x_{2h}} - \bar{\lambda}_{x_{1h}}}{\bar{\lambda}_{x_{1h}} + \bar{\lambda}_{x_{2h}}} \right). \tag{10}$$

The first-order bias expressions for these estimators are

$$\text{Bias}(\hat{\bar{\Gamma}}_{y_{Rest}}) \cong \frac{1}{2} \sum_{h=1}^{L} W_h \zeta_{3h} \bar{\lambda}_{Y_h} C^2_{\lambda_{x_h}} \left[\frac{3}{4} \Pi_{\lambda_{x_h}} - T_{\lambda_{y_h x_h}} \sqrt{\Pi_{\lambda_{y_h x_h}}} \right], \tag{11}$$

and

$$\text{Bias}(\hat{\bar{\Gamma}}_{y_{Pest}}) \cong \frac{1}{2} \sum_{h=1}^{L} W_h \zeta_{3h} \bar{\lambda}_{Y_h} C^2_{\lambda_{x_h}} \left[T_{\lambda_{y_h x_h}} \sqrt{\Pi_{\lambda_{y_h x_h}}} - \frac{3}{4} \Pi_{\lambda_{x_h}} \right]. \tag{12}$$

Their corresponding first-order MSE expressions are

$$\mathrm{MSE}(\hat{\bar{\Gamma}}_{y_{Rest}}) \cong \sum_{h=1}^{L} W_h^2 \bar{\lambda}_{Y_h}^2 \left[\zeta_{2h} C_{\lambda_{y_h}}^2 \Pi_{\lambda_{y_h}} + \zeta_{3_h} C_{\lambda_{x_h}}^2 \left(\frac{\Pi_{\lambda_{x_h}}}{4} - T_{\lambda_{y_h x_h}} \sqrt{\Pi_{\lambda_{y_h x_h}}} \right) \right] \tag{13}$$

and

$$\mathrm{MSE}(\hat{\bar{\Gamma}}_{y_{Pest}}) \cong \sum_{h=1}^{L} W_h^2 \bar{\lambda}_{Y_h}^2 \left[\zeta_{2h} C_{\lambda_{y_h}}^2 \Pi_{\lambda_{y_h}} + \zeta_{3_h} C_{\lambda_{x_h}}^2 \left(\frac{\Pi_{\lambda_{x_h}}}{4} + T_{\lambda_{y_h x_h}} \sqrt{\Pi_{\lambda_{y_h x_h}}} \right) \right]. \tag{14}$$

An additional exponential-type estimator, originally introduced in [9], is reformulated for stratified double-phase systematic sampling. The resulting estimator is defined as follows:

$$\hat{\bar{\Gamma}}_{y_{est}} = \sum_{h=1}^{L} W_h \bar{\lambda}_{y_{2h}} \exp \left[\frac{z_h (\bar{\lambda}_{x_{2h}} - \bar{\lambda}_{x_{1h}})}{\bar{\lambda}_{x_{1h}} + \bar{\lambda}_{x_{2h}}} \right], \tag{15}$$

where z_h is an unknown constant.

After determining the optimum value of z_h, the first-order bias of the estimator becomes

$$\mathrm{Bias}(\hat{\bar{\Gamma}}_{y_{est}})_{\mathrm{opt}} \cong -\frac{1}{2} \sum_{h=1}^{L} W_h \bar{\lambda}_{Y_h} \zeta_{3_h} \sqrt{\Pi_{\lambda_{y_h x_h}}} \left[\frac{\rho_{\lambda_{y_h x_h}}^2 C_{\lambda_{y_h}}^2 \sqrt{\Pi_{\lambda_{y_h x_h}}}}{\Pi_{\lambda_{x_h}}} - T_{\lambda_{y_h x_h}} C_{\lambda_{x_h}}^2 \right] \tag{16}$$

and the corresponding minimum MSE is

$$\mathrm{MSE}(\hat{\bar{\Gamma}}_{y_{est}})_{\mathrm{opt}} \cong \sum_{h=1}^{L} W_h^2 \bar{\lambda}_{Y_h}^2 C_{\lambda_{y_h}}^2 \left[\zeta_{2h} \Pi_{\lambda_{y_h}} - \zeta_{3h} \left(\frac{\rho_{\lambda_{y_h x_h}}^2 \Pi_{\lambda_{y_h x_h}}}{\Pi_{\lambda_{x_h}}} \right) \right]. \tag{17}$$

The value of z_h that minimizes the first-order MSE is

$$z_{h_{(\mathrm{opt})}} = -\frac{2\rho_{\lambda_{y_h x_h}} C_{\lambda_{y_h}} \sqrt{\Pi_{\lambda_{y_h x_h}}}}{C_{\lambda_{x_h}} \Pi_{\lambda_{x_h}}}. \tag{18}$$

4. Proposed Estimators Under Stratified Double-Phase Systematic Sampling

This section introduces two new estimators for the finite population mean within the framework of stratified double-phase systematic sampling. Both formulations incorporate the second-phase sample mean of the study variable together with auxiliary information collected at the two sampling phases. The distinction between the proposed estimators lies in how the auxiliary-variable means are combined: one employs a harmonic-mean function, whereas the other uses a geometric-mean function. The unknown coefficients are determined by minimizing the MSE, and the resulting properties are examined to assess the efficiency of the proposed methods.

4.1. First Proposed Estimator

The estimator is constructed by using a harmonic-mean function of the auxiliary-variable sample means from the first and second phases into the estimation of the study-variable population mean. The coefficients b_{1h} and b_{2h} control the contributions of the study-variable component and the auxiliary-information component, respectively. Their optimal values are obtained by minimizing the MSE under the stratified double-phase systematic

sampling design. This formulation allows information from both sampling phases to contribute to the estimation process. The proposed estimator is expressed as follows:

$$\hat{\bar{\Gamma}}_{D_{1st}} = \sum_{h=1}^{L} W_h \left[b_{1h}\bar{\lambda}_{y_{2h}} + b_{2h}\left(\frac{2\bar{\lambda}_{x_{1h}}\bar{\lambda}_{x_{2h}}}{\bar{\lambda}_{x_{1h}} + \bar{\lambda}_{x_{2h}}} \right) \right], \tag{19}$$

here, b_{1h} and b_{2h} represent the unknown coefficients assigned to the study-variable and auxiliary-variable components, respectively. The terms $\bar{\lambda}_{y_{2h}}$, $\bar{\lambda}_{x_{1h}}$, and $\bar{\lambda}_{x_{2h}}$ denote the relevant sample means calculated at the second phase for the study variable and at the first and second phases for the auxiliary variable, respectively, under stratified double-phase systematic sampling. The coefficients are determined by minimizing the MSE of the proposed estimator.

Theorem 1. *For the proposed harmonic-mean-based estimator $\hat{\bar{\Gamma}}_{D_{1st}}$ under stratified double-phase systematic sampling, the expressions for its minimum bias and minimum MSE are given by the following results.*

$$\begin{aligned} \text{Bias}(\hat{\bar{\Gamma}}_{D_{1st}})_{\text{opt}} \cong \sum_{h=1}^{L} W_h \left(\frac{\bar{\lambda}_{Y_h}}{Q_h} \right) \Bigg[& \left(4 - \zeta_{3h} C^2_{\lambda_{x_h}} \Pi_{\lambda_{x_h}} \right) \\ & \left\{ \zeta_{2_h} C^2_{\lambda_{y_h}} \Pi_{\lambda_{y_h}} \left(4 - \zeta_{3h} C^2_{\lambda_{x_h}} \Pi_{\lambda_{x_h}} \right) - 4(\zeta_{1h} + \zeta_{2h}) U_{\lambda_{y_h x_h}} \right\} \\ & + C^2_{\lambda_{x_h}} \Pi_{\lambda_{x_h}} \left\{ 4(3\zeta_{1h} + \zeta_{2h}) - \zeta^2_{3h} C^2_{\lambda_{xh}} \Pi_{\lambda_{x_h}} \right\} - Q_h \Bigg] \end{aligned} \tag{20}$$

and

$$\begin{aligned} \text{MSE}(\hat{\bar{\Gamma}}_{D_{1st}})_{\min} \cong \sum_{h=1}^{L} W_h^2 \left(\frac{\bar{\lambda}^2_{Y_h}}{Q_h^2} \right) \Bigg[& (F_h - Q_h + 4G_h)^2 + \zeta_{2h} C^2_{\lambda_{y_h}} \Pi_{\lambda_{y_h}} F_h^2 \\ & + 4G_h C^2_{\lambda_{x_h}} \Bigg\{ (3\zeta_{1h} + \zeta_{2h}) \Pi_{\lambda_{xh}} G_h \\ & + (\zeta_{1h} + \zeta_{2h}) T_{\lambda_{y_h x_h}} \sqrt{\Pi_{\lambda_{y_h x_h}}} F_h - \frac{\zeta_{3h}}{2} \Pi_{\lambda_{xh}} (F_h - Q_h + 4G_h) \Bigg\} \Bigg], \end{aligned} \tag{21}$$

where

$$Q_h = 4\left(1 + \zeta_{2h} C^2_{\lambda_{yh}} \Pi_{\lambda_{yh}}\right) \left\{ 4 + (5\zeta_{1h} - \zeta_{2h}) C^2_{\lambda_{xh}} \Pi_{\lambda_{xh}} \right\} - \left[4 + 2(\zeta_{1h} + \zeta_{2h}) U_{\lambda_{y_h x_h}} - \zeta_{3h} C^2_{\lambda_{x_h}} \Pi_{\lambda_{x_h}} \right]^2,$$

$$T_{\lambda_{y_h x_h}} = \rho_{\lambda_{y_h x_h}} \left(\frac{C_{\lambda_{y_h}}}{C_{\lambda_{x_h}}} \right),$$

$$U_{\lambda_{y_h x_h}} = T_{\lambda_{y_h x_h}} C^2_{\lambda_{x_h}} \sqrt{\Pi_{\lambda_{y_h x_h}}},$$

$$F_h = \left[C^2_{\lambda_{x_h}} \Pi_{\lambda_{x_h}} \left\{ 4(3\zeta_{1h} + \zeta_{2h}) - \zeta^2_{3h} C^2_{\lambda_{x_h}} \Pi_{\lambda_{x_h}} \right\} - 2(\zeta_{1h} + \zeta_{2h}) \left(4 - \zeta_{3h} C^2_{\lambda_{x_h}} \Pi_{\lambda_{x_h}} \right) U_{\lambda_{y_h x_h}} \right],$$

$$G_h = \zeta_{2h} C^2_{\lambda_{y_h}} \Pi_{\lambda_{y_h}} \left(4 - \zeta_{3h} C^2_{\lambda_{x_h}} \Pi_{\lambda_{x_h}} \right) - 2(\zeta_{1h} + \zeta_{2h}) U_{\lambda_{y_h x_h}}.$$

Proof. The detailed proof of Theorem 1 is presented in Appendix A. □

4.2. Second Proposed Estimator

The second proposed estimator uses the geometric mean of the auxiliary-variable sample means from the two phases of stratified double-phase systematic sampling. The

coefficient b_{3h} controls the contribution of the study-variable sample mean, while b_{4h} determines the contribution of the geometric-mean component. These coefficients are optimized using the bias and MSE criteria. The estimator is expressed as follows:

$$\hat{\bar{\Gamma}}_{D_{2st}} = \sum_{h=1}^{L} W_h \left(b_{3h} \bar{\lambda}_{y_{2h}} + b_{4h} \sqrt{\bar{\lambda}_{x_{1h}} \bar{\lambda}_{x_{2h}}} \right), \tag{22}$$

where b_{3h} and b_{4h} are unknown coefficients for the study-variable and auxiliary-variable components, respectively. The quantities $\bar{\lambda}_{y_{2h}}$, $\bar{\lambda}_{x_{1h}}$, and $\bar{\lambda}_{x_{2h}}$ represent the corresponding sample means from the two phases of stratified double-phase systematic sampling.

The applicability of the proposed estimators depends on the values of the auxiliary-variable means. The theoretical development requires a nonzero population mean of the auxiliary variable, as the coefficient of variation and relative-error terms are defined using this mean. For the harmonic-mean-based estimator, the sum of the first- and second-phase auxiliary sample means must be nonzero to avoid an undefined expression. For the geometric-mean-based estimator, the product of these two sample means must be nonnegative for the square-root term to be real. Nevertheless, positive auxiliary-variable values are preferable for the conventional interpretation of the geometric-mean component. Thus, the proposed estimators are not directly applicable when the auxiliary-variable population mean is zero under the present formulation.

Theorem 2. *Under stratified double-phase systematic sampling, the minimum bias and minimum mean squared error of the second proposed estimator, $\hat{\bar{\Gamma}}_{D_{2st}}$, are given by the following expressions*

$$\begin{aligned} \mathrm{Bias}(\hat{\bar{\Gamma}}_{D_{2st}})_{\min} \cong 4 \sum_{h=1}^{L} W_h \left(\frac{\bar{\lambda}_{Y_h} C^2_{\lambda_{x_h}}}{K_h} \right) \Bigg[& (\zeta_{1h} + \zeta_{2h}) U_{\lambda_{y_h x_h}} \left\{ 4(\zeta_{1h} + \zeta_{2h}) T_{\lambda_{y_h x_h}} \sqrt{\Pi_{\lambda_{y_h x_h}}} - \zeta_3 \Pi_{\lambda_{x_h}} \right\} \\ & - 2\zeta_{2h}(8\zeta_{1h} + \zeta_{3h}) C^2_{\lambda_{y_h}} \Pi_{\lambda_{y_h}} \Pi_{\lambda_{x_h}} \Bigg] \end{aligned} \tag{23}$$

and

$$\begin{aligned} \mathrm{MSE}(\hat{\bar{\Gamma}}_{D_{2st}})_{\min} \cong \sum_{h=1}^{L} W_h^2 \left(\frac{\bar{\lambda}^2_{Y_h} C^2_{\lambda_{x_h}}}{K_h} \right) \Bigg[& 64 \zeta_{1h} \zeta_{2h} \Pi_{\lambda_{x_h}} C^2_{\lambda_{y_h}} \Pi_{\lambda_{y_h}} \\ & - C^2_{\lambda_{x_h}} \left\{ 4(\zeta_{1h} + \zeta_{2h}) T_{\lambda_{y_h x_h}} \sqrt{\Pi_{\lambda_{y_h x_h}}} - \zeta_{3h} \Pi_{\lambda_{x_h}} \right\}^2 \Bigg], \end{aligned} \tag{24}$$

where

$$K_h = 64 \left(1 + \zeta_{2h} C^2_{\lambda_{y_h}} \Pi_{\lambda_{y_h}} \right) \left(1 + \zeta_{1h} C^2_{\lambda_{x_h}} \Pi_{\lambda_{x_h}} \right) - \left\{ 8 + 4(\zeta_{1h} + \zeta_{2h}) U_{\lambda_{y_h x_h}} - \zeta_{3h} C^2_{\lambda_{x_h}} \Pi_{\lambda_{x_h}} \right\}^2.$$

Proof. The detailed proof of Theorem 2 is Appendix B. □

4.3. Practical Estimation of the Optimal Coefficients

In practical survey applications, the population quantities involved in the derivation of the optimal coefficients are rarely known in advance. Therefore, the theoretical coefficients of the proposed estimators $\hat{\bar{\Gamma}}_{D_{1st}}$ and $\hat{\bar{\Gamma}}_{D_{2st}}$ cannot generally be used directly. A practical alternative is to estimate the required population characteristics from the observations collected during the two sampling phases. Specifically, the unknown means, variances, and covariance-related quantities can be replaced by their corresponding sample-based estimates, provided that these estimates are obtained in accordance with the adopted

stratified double-phase systematic sampling design. Substituting these estimates into the expressions for the optimal coefficients yields their estimated values, which can then be used to calculate $\hat{\hat{\Gamma}}_{D_{1st}}$ and $\hat{\hat{\Gamma}}_{D_{2st}}$. When suitable historical information or pilot-survey results are available, they may also be considered in estimating the required characteristics. This procedure makes the proposed estimators applicable when the population parameters are unknown. Nevertheless, the coefficients obtained in this manner are estimates of their theoretical optimal values, and their sampling variability may affect the resulting estimation performance. A detailed investigation of the stability and efficiency of the sample-based coefficients is therefore an important direction for further research.

5. Simulation Study and Empirical Evaluation

This section investigates the performance of the existing and proposed estimators under stratified double-phase systematic sampling. The assessment is conducted using artificially generated finite populations, with the sampling design applied separately within each stratum. The study considers how estimation accuracy changes with the population structure, the relationship between the study and auxiliary variables, and the number of units observed at each phase. Particular attention is given to the two proposed estimators, which use auxiliary information through harmonic and geometric mean functions.

The simulation results are to be evaluated using the empirical mean squared error (MSE) and percentage relative efficiency (PRE). The conventional stratum-weighted second-phase sample mean serves as the reference estimator. In addition to the simulation experiments, real population are used to examine the performance of the competing procedures under observed population conditions. Together, these assessments provide a framework for examining the accuracy and practical relevance of the proposed estimators.

5.1. Simulation Design

A finite-population Monte Carlo experiment is designed to compare the existing estimators with the two proposed estimators under different population configurations. Unlike a design in which observations are sampled directly from the entire population, the present experiment follows the stratified double-phase structure defined in Section 5. Auxiliary-variable observations are collected from the first-phase systematic samples, whereas both the study and auxiliary variables are observed in the second-phase subsamples. This arrangement preserves the nesting of the two samples and allows the estimators to be assessed under the sampling mechanism for which they were developed.

Construction of artificial populations

Three artificial finite populations are constructed to represent different forms of association between the study and auxiliary variables and different patterns in the ordering of population units. Each population contains $N = 2160$ units divided into $L = 3$ non-overlapping strata of equal size:

$$N_h = 720, \qquad h = 1, 2, 3, \qquad N = \sum_{h=1}^{3} N_h = 2160.$$

The stratum weights are therefore

$$W_h = \frac{N_h}{N} = \frac{1}{3}, \qquad h = 1, 2, 3.$$

Let $i = 1, \ldots, 720$ denote the ordered unit position within stratum h. Define the scaled position

$$u_i = \frac{i - \frac{1}{2}}{720}.$$

For each model, the generated values are retained as a fixed finite population throughout the subsequent sampling repetitions. The population characteristics used in the analysis are calculated from the complete generated populations rather than from an individual sample.

Model I: Stratum-specific linear association

The first population represents a setting in which the auxiliary variable changes gradually across the ordered units and the study variable is approximately linearly related to it. The auxiliary variable is generated as

$$\lambda_{X_{hi}} = 7 + 2h + 4u_i + 0.6\sin(2\pi u_i) + e_{x_{hi}},$$

where

$$e_{x_{hi}} \sim N(0, 0.16).$$

The corresponding study-variable values are defined by

$$\lambda_{Y_{hi}} = 3 + 1.15\lambda_{X_{hi}} + e_{y_{hi}},$$

where

$$e_{y_{hi}} \sim N(0, \sigma_{y_h}^2), \qquad \sigma_{y_h} = 0.7 + 0.15h.$$

The error terms are generated independently across units and strata, and the two sets of disturbances are mutually independent. This construction introduces a gradual ordering effect in the auxiliary variable while allowing the variability of the study variable to differ among strata. It consequently provides a setting in which auxiliary information may be useful for improving estimation precision.

Model II: Nonlinear association with asymmetric disturbances

The second population represents a curved relationship between the two variables. The auxiliary-variable values are generated from

$$\lambda_{X_{hi}} = 5 + h + 3u_i^2 + 0.25e_{xhi},$$

where

$$e_{x_{hi}} \sim N(0, 1).$$

The study variable is then defined as

$$\lambda_{Y_{hi}} = 4 + 0.30\left(\lambda_{X_{hi}}\right)^2 + \left(0.4 + 0.06\lambda_{X_{hi}}\right)e_{y_{hi}},$$

where

$$e_{y_{hi}} = \exp(\lambda_{Z_{hi}}) - \exp\left(\frac{\sigma^2}{2}\right), \qquad \lambda_{Z_{hi}} \sim N(0, \sigma^2), \qquad \sigma^2 = 0.36.$$

All random quantities used in generating this population are independent across units, and the auxiliary-variable disturbances are independent of the study-variable disturbances. The subtraction of $\exp(\sigma^2/2)$ gives the transformed error zero expectation. Its asymmetric distribution and the quadratic relationship introduce features that differ from those of the first model.

Model III: Periodic ordering with stratum-specific phases

The third population is constructed to examine the effect of recurring patterns in the ordered population. Its auxiliary variable is given by

$$\lambda_{X_{hi}} = 11 + h + 1.8 \sin\left(\frac{2\pi i}{53} + \phi_h\right) + e_{x_{hi}},$$

where

$$\phi_h = \frac{\pi(h-1)}{4}, \qquad e_{x_{hi}} \sim N(0, 0.25).$$

The study variable is generated according to

$$\lambda_{Y_{hi}} = 14 + 0.55\lambda_{X_{hi}} + 2.2 \cos\left(\frac{2\pi i}{53} + \phi_h\right) + e_{y_{hi}},$$

with

$$e_{y_{hi}} \sim N(0, 1.44).$$

The error terms are independent across units and strata and independent of the auxiliary-variable disturbances. The stratum-specific phase shifts produce different alignments of the periodic patterns across the three strata. This model is intended to assess estimator performance when the ordering of the population may affect the systematic samples and the relationship between the two variables is not adequately described by a linear trend alone.

For all three models, the random disturbances are generated once when the finite populations are constructed. The resulting population values are then held fixed during the sampling experiment. This distinction is important because the subsequent Monte Carlo repetitions represent variation arising from the sampling design, rather than repeated regeneration of the underlying populations. The finite population means, variances, covariances, coefficients of variation, and correlations are calculated separately for each stratum and model. These quantities are also used when evaluating the population-based optimal coefficients appearing in the theoretical expressions for the proposed estimators.

Sampling configurations

Three combinations of first- and second-phase sample sizes are considered. The sample sizes are specified separately within each stratum as follows:

$$(n_{1h}, n_{2h}) \in \{(72, 18), (120, 30), (180, 45)\}, \qquad h = 1, 2, 3.$$

Thus, the corresponding overall sample sizes are

$$n_1 \in \{216, 360, 540\}, \qquad n_2 \in \{54, 90, 135\}.$$

For each configuration, the first-phase sample size is four times the second-phase sample size. Since $N_h = 720$, the first-phase systematic intervals are

$$k_{1h} = \frac{N_h}{n_{1h}} \in \{10, 6, 4\},$$

and the second-phase interval is fixed at

$$k_{2h} = \frac{n_{1h}}{n_{2h}} = 4.$$

All intervals are positive integers, so each configuration satisfies the divisibility requirements of the specified systematic sampling design. The three configurations allow the effect

of increasing the first- and second-phase sample sizes to be examined while retaining the same phase-two subsampling interval.

The sample size ratio between the two sampling phases was fixed at 4:1 throughout the simulation study. This configuration was adopted to maintain a consistent relationship between the phase-one and phase-two sample sizes across the considered simulation settings. However, the choice of this ratio should not be interpreted as an optimal allocation for all survey situations. Different ratios may change the relative amount of auxiliary and study-variable information available at the two phases, potentially influencing the efficiency of the proposed estimators. Therefore, the performance of $\hat{\bar{\Gamma}}_{D_{1st}}$ and $\hat{\bar{\Gamma}}_{D_{2st}}$ may vary under alternative sample size ratios. Investigating the effect of different phase allocation ratios is an important extension of the present study.

Simulation procedure

For each artificial population and sample-size configuration, the following steps are carried out.

1. Generate the three strata according to the selected population model and retain the resulting finite population without modification throughout the Monte Carlo experiment.
2. Within each stratum h, select a first-phase random start R_{1h} uniformly from $\{1, \ldots, k_{1h}\}$. Construct the first-phase systematic sample

$$s_{1h} = \{R_{1h} + jk_{1h} : j = 0, \ldots, n_{1h} - 1\}.$$

   Observe the auxiliary variable λ_X for every selected unit and calculate its first-phase sample mean $\bar{\lambda}_{x_{1h}}$.
3. Conditional on the first-phase sample, independently select a second-phase random start R_{2h} uniformly from $\{1, 2, 3, 4\}$. Let $s_{1h}(j)$ denote the unit at position j in the ordered first-phase sample. The second-phase subsample is

$$s_{2h} = \{s_{1h}(R_{2h} + 4j) : j = 0, \ldots, n_{2h} - 1\}.$$

   Observe both λ_X and λ_Y for these units and calculate $\bar{\lambda}_{x_{2h}}$ and $\bar{\lambda}_{y_{2h}}$. By construction, $s_{2h} \subset s_{1h}$.
4. Calculate the conventional estimator

$$\bar{\lambda}_{y_{2st}} = \sum_{h=1}^{3} W_h \bar{\lambda}_{y_{2h}},$$

   and all existing and proposed estimators defined in Sections 3 and 4. For the theoretical benchmark, the optimal coefficients are calculated from the complete finite-population characteristics and kept fixed across repetitions. This benchmark represents performance when the population characteristics required by the optimal coefficient formulas are available. Its interpretation should be distinguished from that of a feasible implementation using estimated coefficients.
5. Repeat the nested two-phase sampling procedure $D = 8000$ times for each combination of population model and sample-size configuration. The random starts are redrawn at each repetition, independently across strata, and the second-phase start is selected according to the specified conditional design. The finite populations remain unchanged.

6. For estimator $\hat{\hat{\Gamma}}_j$, calculate the empirical design MSE as

$$\mathrm{MSE}\left(\hat{\hat{\Gamma}}_j\right) = \frac{1}{D}\sum_{d=1}^{D}\left(\hat{\hat{\Gamma}}_{jd} - \bar{\lambda}_Y\right)^2,$$

where $\hat{\hat{\Gamma}}_{jd}$ is the estimate obtained in repetition d and

$$\bar{\lambda}_Y = \sum_{h=1}^{3} W_h \bar{\lambda}_{Y_h}$$

is the known mean of the complete finite population. The same definition is used for the conventional estimator and every competing estimator.

7. Compute the percentage relative efficiency of estimator $\hat{\hat{\Gamma}}_j$ with respect to the conventional second-phase stratified sample mean:

$$\mathrm{PRE} = \frac{\mathrm{MSE}\left(\hat{\hat{\Gamma}}_{y_{st}}\right)}{\mathrm{MSE}(\hat{\hat{\Gamma}}_{jd})} \times 100,$$

where $\hat{\hat{\Gamma}}_{jd} = \hat{\hat{\Gamma}}_{y_{st}}, \hat{\hat{\Gamma}}_{y_{Rst}}, \hat{\hat{\Gamma}}_{y_{Pst}}, \hat{\hat{\Gamma}}_{y_{Rest}}, \hat{\hat{\Gamma}}_{y_{Pest}}, \hat{\hat{\Gamma}}_{y_{est}}, \hat{\hat{\Gamma}}_{D_{1st}}, \hat{\hat{\Gamma}}_{D_{2st}}$.

In this study, 8000 Monte Carlo replications were considered to assess the performance of the proposed and existing estimators. This number was chosen to obtain reliable MSE and PRE estimates while keeping the computational burden manageable. Although a larger number of replications can further reduce simulation variability, 8000 replications do not necessarily ensure complete stability for all simulation settings. The precision of the results also depends on the variability of the estimators. Therefore, the reported MSE and PRE values are based on the chosen simulation size, and further investigation using a larger number of replications could be considered to examine their stability more closely.

Assessment of estimator performance

The empirical MSE and PRE values are summarized separately for the three population models and the three sample-size configurations. Comparisons focus on whether the harmonic- and geometric-mean components improve estimation accuracy relative to the conventional estimator and the existing alternatives. The results are also examined across increasing sample sizes to determine whether the relative performance of the estimators changes as more information becomes available at both phases. Any claims concerning the superiority of a proposed estimator are based on the computed results for the relevant population and sampling configuration rather than assumed in advance.

5.2. Empirical Evaluation

The empirical performance of the proposed estimators is assessed using three real finite populations based on the data reported in [21]. For each population, the necessary summary statistics are computed from the observed data and substituted into the MSE expressions presented in Sections 3 and 4, under the stratified double-phase systematic sampling design. The evaluation considers the conventional estimator $\hat{\hat{\Gamma}}_{y_{st}}$, the existing estimators $\hat{\hat{\Gamma}}_{y_{Rst}}$, $\hat{\hat{\Gamma}}_{y_{Pst}}$, $\hat{\hat{\Gamma}}_{y_{Rest}}$, $\hat{\hat{\Gamma}}_{y_{Pest}}$, and $\hat{\hat{\Gamma}}_{y_{est}}$, together with the two proposed estimators, $\hat{\hat{\Gamma}}_{D_{1st}}$ and $\hat{\hat{\Gamma}}_{D_{2st}}$. The percentage relative efficiency (PRE) of each method is calculated with respect to the conventional estimator, which serves as the reference. This comparison provides a basis for assessing the efficiency of the competing estimators across the three

populations. The corresponding population summary statistics are reported in Table 5, and PRE is computed using the following formula:

$$PRE = \frac{MSE\left(\hat{\hat{\Gamma}}_{y_{st}}\right)}{MSE\left(\hat{\hat{\Gamma}}_{t}\right)} \times 100,$$

where $t = y_{st}, y_{Rst}, y_{Pst}, y_{Rest}, y_{Pest}, y_{est}, D_{1st}$, and D_{2st}. The corresponding results are reported in Table 6.

5.3. Discussion

This study examines the performance of two proposed estimators based on harmonic and geometric combinations of auxiliary information under stratified double-phase systematic sampling. Their performance is evaluated through Monte Carlo simulations involving linear association, nonlinear association with asymmetric disturbances, and periodic population structures. The results show that both proposed estimators can improve estimation efficiency compared with the conventional stratified second-phase sample mean. Tables 2–4 present the simulation results for the three population models under different first- and second-phase sample-size configurations. Across these tables, the proposed estimators generally achieve higher PRE values than the conventional estimator. The geometric-based estimator, $\hat{\hat{\Gamma}}_{D_{2st}}$, performs particularly well. For example, in Table 2, its PRE values are 383.43, 413.30, and 276.28 for Populations I, II, and III, respectively. In Table 4, these values increase to 233.30, 559.57, and 247.74. These findings indicate that the proposed methods can reduce estimation error, although their relative performance varies with the population structure and sampling configuration. Among the two methods, the geometric-based estimator, $\hat{\hat{\Gamma}}_{D_{2st}}$, demonstrates particularly strong performance across the numerical comparisons.

Figure 1 provides a graphical comparison of the eight estimators across the three simulated populations. The plots complement Tables 2–4 by illustrating the differences in PRE among the existing and proposed methods. The geometric-based estimator shows particularly strong relative efficiency in the first two population models, while both proposed estimators perform favourably in the periodic population. The results also show that increasing the sample sizes does not necessarily produce the same change in PRE for every estimator. Therefore, estimator performance should be assessed using both the absolute MSE values and their relative efficiencies rather than relying on sample size alone.

Table 2. MSE and PRE of competing estimators for three population models at $(n_{1h}, n_{2h}) = (72, 18)$.

Estimator	Population 1		Population 2		Population 3	
	MSE	**PRE**	**MSE**	**PRE**	**MSE**	**PRE**
$\hat{\hat{\Gamma}}_{y_{st}}$	0.02871883	100.00	0.04384512	100.00	0.03945757	100.00
$\hat{\hat{\Gamma}}_{y_{Rst}}$	0.02266443	126.71	0.01918332	228.56	0.04966747	79.44
$\hat{\hat{\Gamma}}_{y_{Pst}}$	0.04671020	61.48	0.08792821	49.86	0.06763093	58.34
$\hat{\hat{\Gamma}}_{y_{Rest}}$	0.02419764	118.68	0.02909169	150.71	0.03974875	99.27
$\hat{\hat{\Gamma}}_{y_{Pest}}$	0.03622392	79.28	0.06345344	69.10	0.04875964	80.92
$\hat{\hat{\Gamma}}_{y_{est}}$	0.02261780	126.97	0.01296839	338.09	0.03830230	103.02
$\hat{\hat{\Gamma}}_{D_{1st}}$	0.00859028	334.32	0.01160199	377.91	0.0182765	215.89
$\hat{\hat{\Gamma}}_{D_{2st}}$	0.00748989	383.43	0.01060831	413.30	0.01428169	276.28

Table 3. MSE and PRE of competing estimators for three population models at $(n_{1h}, n_{2h}) = (120, 30)$.

Estimator	Population 1		Population 2		Population 3	
	MSE	PRE	MSE	PRE	MSE	PRE
$\hat{\bar{\Gamma}}_{y_{st}}$	0.01457481	100.00	0.02473893	100.00	0.02091200	100.00
$\hat{\bar{\Gamma}}_{y_{Rst}}$	0.01137530	128.13	0.01076094	229.90	0.02291047	91.28
$\hat{\bar{\Gamma}}_{y_{Pst}}$	0.02390656	60.97	0.05227299	47.33	0.03077383	67.95
$\hat{\bar{\Gamma}}_{y_{Rest}}$	0.01220892	119.38	0.01606158	154.03	0.02042409	102.39
$\hat{\bar{\Gamma}}_{y_{Pest}}$	0.01847367	78.90	0.03680502	67.22	0.02436472	85.83
$\hat{\bar{\Gamma}}_{y_{est}}$	0.01048144	139.05	0.00848530	291.55	0.01928149	108.46
$\hat{\bar{\Gamma}}_{D_{1st}}$	0.00938118	155.36	0.00766295	322.84	0.01211742	172.58
$\hat{\bar{\Gamma}}_{D_{2st}}$	0.00778070	187.32	0.00696356	355.26	0.01011696	206.70

Table 4. MSE and PRE of competing estimators for three population models at $(n_{1h}, n_{2h}) = (180, 45)$.

Estimator	Population 1		Population 2		Population 3	
	MSE	PRE	MSE	PRE	MSE	PRE
$\hat{\bar{\Gamma}}_{y_{st}}$	0.00894303	100.00	0.01658506	100.00	0.01534467	100.00
$\hat{\bar{\Gamma}}_{y_{Rst}}$	0.00744754	120.08	0.00555084	298.78	0.01825003	84.08
$\hat{\bar{\Gamma}}_{y_{Pst}}$	0.01483351	60.29	0.03761576	44.09	0.02087095	73.52
$\hat{\bar{\Gamma}}_{y_{Rest}}$	0.00764691	116.95	0.00981813	168.92	0.01574527	97.46
$\hat{\bar{\Gamma}}_{y_{Pest}}$	0.01133785	78.88	0.02585099	64.16	0.01705185	89.99
$\hat{\bar{\Gamma}}_{y_{est}}$	0.00732335	122.12	0.00348901	475.35	0.01409864	108.84
$\hat{\bar{\Gamma}}_{D_{1st}}$	0.00458173	195.19	0.00316404	524.17	0.00673948	227.61
$\hat{\bar{\Gamma}}_{D_{2st}}$	0.00383324	233.30	0.00296387	559.57	0.00619383	247.74

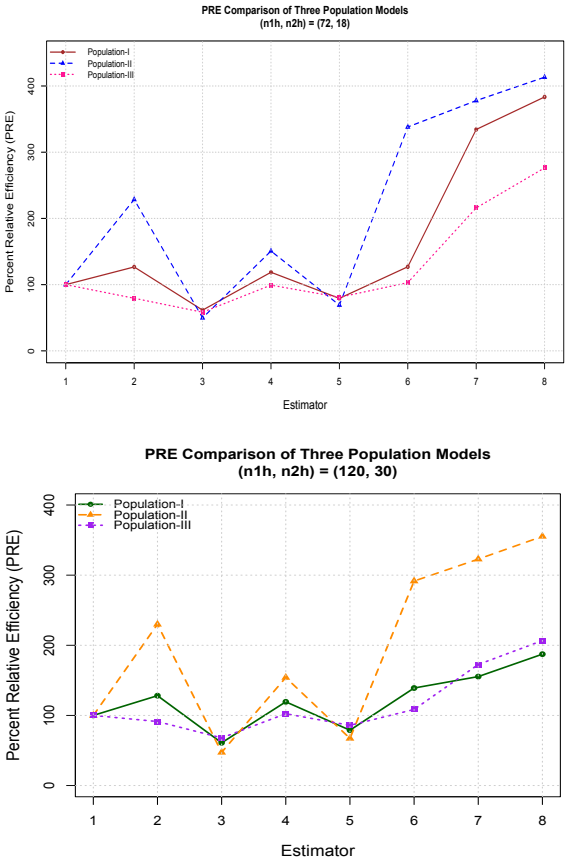


Figure 1. *Cont.*

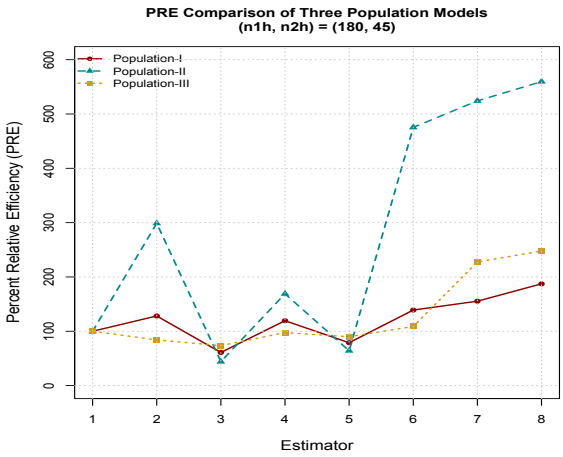


Figure 1. The comparative analysis examines the performance of eight estimators across three simulated populations, including six established methods (Estimators 1–6) and two newly developed approaches (Estimators 7 and 8).

The improvements in estimation efficiency achieved by the proposed estimators differ across the three population models because the underlying relationships between the auxiliary and study variables are not identical. In Model I, the approximately linear association provides a relatively direct structure for utilizing auxiliary information. Model II involves a nonlinear relationship accompanied by asymmetric disturbances, which may affect the extent to which the auxiliary variable improves estimation. In Model III, the periodic patterns and stratum-specific phase shifts introduce additional effects related to the ordering of population units under systematic sampling. The correlation between the auxiliary and study variables may partly explain the observed differences in efficiency; however, it should not be considered the sole determining factor. The form of the association, the variability of the errors, and the population ordering may also influence the performance of the proposed estimators. Thus, the variation in PRE across the models reflects the combined influence of these population characteristics rather than correlation alone.

The empirical results in Table 6 provide further evidence of the performance of the proposed methods. The harmonic-based estimator, $\hat{\hat{\Gamma}}_{D_{1st}}$, obtains PRE values of 632.46, 420.41, and 334.19 for Populations 1, 2, and 3, respectively. The corresponding values for the geometric-based estimator, $\hat{\hat{\Gamma}}_{D_{2st}}$, are 1026.44, 687.92, and 613.81. These results indicate that both methods reduce the empirical MSE relative to the conventional estimator in the populations examined, with the geometric-based approach providing the larger improvement. Nevertheless, the magnitude of the gains varies between populations, suggesting that the effectiveness of each estimator depends partly on the characteristics of the study and auxiliary variables. Figure 2 presents the same comparisons graphically and highlights the substantial efficiency gains achieved by the proposed estimators.

A closer examination of the real-population results reveals that the use of auxiliary information does not necessarily lead to improved estimation precision. The performance of conventional ratio, product, and exponential estimators depends on how effectively their specific functional forms represent the relationship between the study and auxiliary variables. When this relationship does not adequately support the assumptions underlying a particular estimator, the use of auxiliary information may increase rather than reduce the estimation error. In contrast, the proposed harmonic- and geometric-based estimators provide a more flexible way of combining auxiliary information obtained from the first and second sampling phases. Their coefficients are determined by minimizing the corresponding approximate MSE expressions, allowing the contribution of the available information to

be adjusted according to the population characteristics. This optimization-based construction offers a possible explanation for their consistently lower MSE values in the three real populations considered. The geometric-based estimator, in particular, exhibits the strongest performance, with PRE values exceeding those of the harmonic-based estimator in all three cases. These findings suggest that the improvements in efficiency are associated not merely with the inclusion of auxiliary information, but also with the way in which that information is combined and weighted within the proposed estimation framework. Nevertheless, the observed superiority should be interpreted in the context of the populations examined, since the optimized coefficients depend on population characteristics that may not be fully available in practical applications.

Table 5. Population characteristics for the three populations.

	Population 1		**Population 2**		**Population 3**	
Parameter	**Stratum 1**	**Stratum 2**	**Stratum 1**	**Stratum 2**	**Stratum 1**	**Stratum 2**
N_h	36	36	36	36	60	60
n_{1h}	12	12	18	18	20	20
n_{2h}	4	4	6	6	5	5
k_{1h}	3	3	2	2	3	3
k_{2h}	3	3	3	3	4	4
$\bar{\lambda}_{X_h}$	1054.3889	223.05556	328.8056	333.0000	6570.4667	6459.8000
$\bar{\lambda}_{Y_h}$	74,690.8889	47,948.4167	25,431.4720	25,791.3611	536.5833	391.7000
$S_{\lambda_{x_h}}$	396.9786	89.9972	438.5196	452.7130	11,707.3181	8635.2600
$S_{\lambda_{y_h}}$	59,832.8464	21,818.1399	41,459.8925	42,163.6438	575.7893	550.8980
$C_{\lambda_{x_h}}$	0.3765	0.4035	1.3337	1.3595	1.7818	1.3368
$C_{\lambda_{y_h}}$	0.8011	0.4550	1.6303	1.6347	1.0731	1.4064
$S_{\lambda_{y_h x_h}}$	2,181,896.4877	758,387.9491	16,579,414.3516	9,915,213.0000	1,145,236.2486	4,624,843.1254
$\rho_{\lambda_{y_h x_h}}$	0.09186	0.38622	0.91191	0.51945	0.16989	0.9722
$\rho_{\lambda_{x_h}}$	0.34495	−0.20648	0.14499	0.13571	0.21118	0.89902
$\rho_{\lambda_{y_h}}$	0.02971	0.00576	0.19323	0.22450	0.77800	0.89379
$\Pi_{\lambda_{x_h}}$	2.03486	0.3806	1.72495	1.67855	1.84474	4.59606
$\Pi_{\lambda_{y_h}}$	1.08914	1.01727	1.96617	2.12252	4.11201	4.57518
$\Pi_{\lambda_{y_h x_h}}$	0.535163	2.673142	1.13984	1.26450	2.22905	0.99545
ζ_{1h}	0.08102	0.08102	0.04861	0.04861	0.04917	0.04917
ζ_{2h}	0.24306	0.24306	0.16204	0.16204	0.19667	0.19667
ζ_{3h}	0.16204	0.16204	0.11343	0.11343	0.14750	0.14750

Table 6. MSE and PRE values of the estimators for the three populations.

Estimator	**Population 1**		**Population 2**		**Population 3**	
	MSE	**PRE**	**MSE**	**PRE**	**MSE**	**PRE**
$\hat{\bar{\Gamma}}_{y_{st}}$	266,365,363.83	100.00	289,758,839.41	100.00	135,299.80	100.00
$\hat{\bar{\Gamma}}_{y_{Rst}}$	306,568,181.13	86.89	277,928,985.95	104.26	213,015.94	63.52
$\hat{\bar{\Gamma}}_{y_{Pst}}$	368,079,260.74	72.37	531,176,955.06	54.55	274,882.94	49.22
$\hat{\bar{\Gamma}}_{y_{Rest}}$	268,727,183.20	99.12	255,145,379.90	113.57	146,995.46	92.04
$\hat{\bar{\Gamma}}_{y_{Pest}}$	299,482,723.00	88.94	381,769,364.46	75.90	177,928.96	76.04
$\hat{\bar{\Gamma}}_{y_{est}}$	245,843,266.28	108.35	252,726,012.60	114.65	132,582.53	102.05
$\hat{\bar{\Gamma}}_{D_{1st}}$	42,115,770.83	632.46	68,923,349.50	420.41	40,485.82	334.19
$\hat{\bar{\Gamma}}_{D_{2st}}$	25,950,452.18	1026.44	42,121,049.31	687.92	22,042.67	613.81

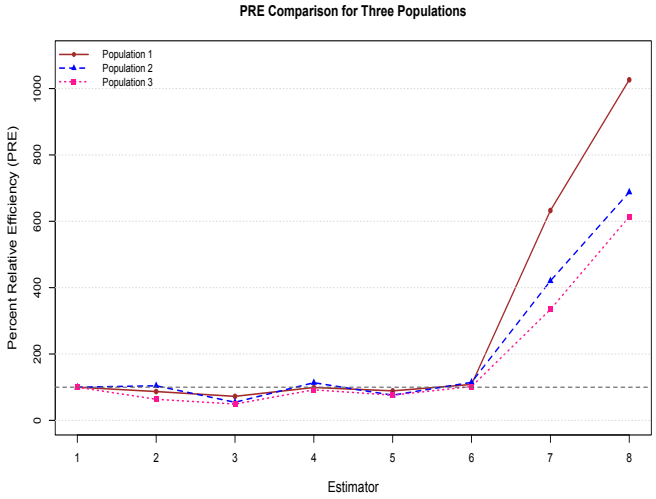


Figure 2. Eight estimators, comprising six existing methods and two proposed approaches, are compared using three real-data populations.

Overall, the simulation and real-population results suggest that both proposed estimators are useful alternatives to the existing methods considered in this study, with the geometric-based estimator performing best in the reported comparisons. However, the optimal coefficients are derived using population characteristics that may not be fully known in practical applications. The reported results should therefore be interpreted as a theoretical performance benchmark. Further work is needed to evaluate feasible versions of the estimators using estimated coefficients and to investigate their performance under a broader range of population structures.

6. Conclusions and Future Research

This study proposed two estimators for finite population mean estimation under stratified double-phase systematic sampling. The methods incorporate auxiliary information through harmonic- and geometric-based formulations, with their approximate bias and MSE expressions developed and their coefficients optimized using the MSE criterion. Their performance was assessed through Monte Carlo simulations and three real-population applications.

The numerical findings indicate that both proposed estimators can outperform the conventional estimator in the settings considered. In particular, the geometric-based estimator achieves the highest PRE values in the three real-population comparisons. These results suggest that combining auxiliary information from both sampling phases through suitable transformations can provide an effective approach to improving estimation precision.

Future research could examine the proposed estimators under a wider range of population structures, including stronger skewness, outliers, weaker auxiliary-variable associations, and different periodic patterns. Another important direction is to develop feasible procedures for estimating the optimal coefficients when the required population characteristics are unknown. Further extensions could investigate additional auxiliary variables, alternative transformations, non-response, and other sampling designs. These developments would help establish the practical usefulness and limitations of the proposed methods across a broader range of survey applications.

Author Contributions: Conceptualization, U.D.; Methodology, U.D. and F.A.A.; Software, U.D. and F.A.A.; Validation, U.D. and F.A.A.; Formal analysis, U.D. and F.A.A.; Investigation, U.D. and F.A.A.; Resources, U.D.; Data curation, U.D. and F.A.A.; Writing—original draft, U.D.; Writing—review and editing, U.D. and F.A.A.; Visualization, U.D.; Supervision, F.A.A.; Project administration, U.D. and

F.A.A.; Funding acquisition, F.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R515), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

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

Acknowledgments: The authors would like to thank Princess Nourah bint Abdulrahman University for supporting this work.

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

Appendix A

Proof. The proposed estimator $\hat{\bar{\Gamma}}_{D_{1st}}$ can be reformulated using the corresponding relative error terms as follows:

$$\hat{\bar{\Gamma}}_{D_{1st}} = \sum_{h=1}^{L} W_h \left[b_{1h} \bar{\lambda}_{Y_h} (1 + \tau_{0h}) + b_{2h} \bar{\lambda}_{X_h} \left\{ \frac{2(1+\tau_{1h})(1+\tau_{2h})}{2 + \tau_{1h} + \tau_{2h}} \right\} \right].$$

$$\Rightarrow \hat{\bar{\Gamma}}_{D_{1st}} = \sum_{h=1}^{L} W_h \left[b_{1h} \bar{\lambda}_{Y_h} (1 + \tau_{0h}) + b_{2h} \bar{\lambda}_{X_h} (1 + \tau_{1h} + \tau_{2h} + \tau_{1h}\tau_{2h}) \left(1 + \frac{\tau_{1h} + \tau_{2h}}{2} \right)^{-1} \right].$$

Using a Taylor expansion, we have

$$(1 + \tau_{1h} + \tau_{2h} + \tau_{1h}\tau_{2h}) \left(1 + \frac{\tau_{1h} + \tau_{2h}}{2} \right)^{-1} \approx 1 + \frac{\tau_{1h} + \tau_{2h}}{2} - \frac{(\tau_{1h} - \tau_{2h})^2}{4}.$$

As a result,

$$\hat{\bar{\Gamma}}_{D_{1st}} \cong \sum_{h=1}^{L} W_h \left[b_{1h} \bar{\lambda}_{Y_h} (1 + \tau_{0h}) + b_{2h} \bar{\lambda}_{X_h} \left(1 + \frac{\tau_{1h} + \tau_{2h}}{2} - \frac{(\tau_{1h} - \tau_{2h})^2}{4} \right) \right].$$

Subtracting $\sum_{h=1}^{L} W_h \bar{\lambda}_{Y_h}$ from both sides gives

$$\begin{aligned} \hat{\bar{\Gamma}}_{D_{1st}} - \sum_{h=1}^{L} W_h \lambda_{Y_h} \cong \sum_{h=1}^{L} W_h \Bigg[& - \bar{\lambda}_{Y_h} + b_{1h} \bar{\lambda}_{Y_h} + b_{2h} \bar{\lambda}_{X_h} + b_{1h} \bar{\lambda}_{Y_h} \tau_{0h} \\ & + \frac{b_{2h} \bar{\lambda}_{X_h}}{2} (\tau_{1h} + \tau_{2h}) - \frac{b_{2h} \bar{\lambda}_{X_h}}{4} (\tau_{1h} - \tau_{2h})^2 \Bigg]. \end{aligned} \tag{A1}$$

Taking the expectation of (A1) yields the approximate bias as follows:

$$\text{Bias}(\hat{\bar{\Gamma}}_{D_{1st}}) \cong \sum_{h=1}^{L} \left[(b_{1h} - 1) \bar{\lambda}_{Y_h} + \frac{b_{2h} \bar{\lambda}_{X_h}}{4} \left(4 - \zeta_{3h} C^2_{\lambda_{x_h}} \Pi_{\lambda_{x_h}} \right) \right]. \tag{A2}$$

The MSE is derived by squaring (A1) and retaining terms up to the second order in the relative error components:

$$\begin{aligned}\left(\hat{\bar{\Gamma}}_{D_{1st}} - \sum_{h=1}^{L} W_h \bar{\lambda}_{Y_h}\right)^2 \cong \sum_{h=1}^{L} W_h^2 \Bigg[& \{(b_{1h}-1)\bar{\lambda}_{Y_h} + b_{2h}\bar{\lambda}_{X_h}\}^2 \\ & + 2b_{1h}\bar{\lambda}_{Y_h}\{(b_{1h}-1)\bar{\lambda}_{Y_h} + b_{2h}\bar{\lambda}_{X_h}\}\tau_{0h} \\ & + b_{2h}\bar{\lambda}_{X_h}\{(b_{1h}-1)\bar{\lambda}_{Y_h} + b_{2h}\bar{\lambda}_{X_h}\}(\tau_{1h}+\tau_{2h}) \\ & + b_{1h}^2\bar{\lambda}_{Y_h}^2\tau_{0h}^2 + b_{1h}b_{2h}\bar{\lambda}_{Y_h}\bar{\lambda}_{X_h}\tau_{0h}(\tau_{1h}+\tau_{2h}) \\ & + \frac{b_{2h}^2\bar{\lambda}_{X_h}^2}{4}(\tau_{1h}+\tau_{2h})^2 - \frac{b_{2h}\bar{\lambda}_{X_h}}{2} \\ & \times \{(b_{1h}-1)\bar{\lambda}_{Y_h} + b_{2h}\bar{\lambda}_{X_h}\}(\tau_{1h}-\tau_{2h})^2 \Bigg].\end{aligned} \tag{A3}$$

By taking expectations and substituting the corresponding moment expressions, the approximate MSE is obtained as follows:

$$\begin{aligned}\mathrm{MSE}(\hat{\bar{\Gamma}}_{D_{1st}}) \cong \sum_{h=1}^{L} W_h^2 \Bigg[& \{(b_{1h}-1)\bar{\lambda}_{Y_h} + b_{2h}\bar{\lambda}_{X_h}\}^2 + b_{1h}^2\bar{\lambda}_{Y_h}^2\zeta_{2h}C_{\lambda_{y_h}}^2\Pi_{\lambda_{y_h}} \\ & + \frac{b_{2h}^2\bar{\lambda}_{X_h}^2}{4}(3\zeta_{1h}+\zeta_{2h})C_{\lambda_{x_h}}^2\Pi_{\lambda_{x_h}} \\ & + b_{1h}b_{2h}\bar{\lambda}_{Y_h}\bar{\lambda}_{X_h}(\zeta_{1h}+\zeta_{2h})T_{\lambda_{y_hx_h}}C_{\lambda_{x_h}}^2\sqrt{\Pi_{\lambda_{y_hx_h}}} \\ & - \frac{b_{2h}\bar{\lambda}_{X_h}}{2}\{(b_{1h}-1)\bar{\lambda}_{Y_h} + b_{2h}\bar{\lambda}_{X_h}\}\zeta_{3h}C_{\lambda_{x_h}}^2\Pi_{\lambda_{x_h}} \Bigg].\end{aligned} \tag{A4}$$

where

$$\rho_{\lambda_{y_hx_h}}C_{\lambda_{y_h}} = T_{\lambda_{y_hx_h}}C_{\lambda_{x_h}}.$$

Under stratified double-phase systematic sampling, the optimal values of b_{1h} and b_{2h} are obtained by differentiating the MSE in (A4) with respect to each constant and equating the derivatives to zero. Solving the resulting equations gives the optimal constants b_{1h}^* and b_{2h}^*, as follows:

$$b_{1h}^* = \frac{C_{\lambda_{x_h}}^2\Pi_{\lambda_{x_h}}\left\{4(3\zeta_{1h}+\zeta_{2h}) - \zeta_{3h}^2C_{\lambda_{x_h}}^2\Pi_{\lambda_{x_h}}\right\} - 2(\zeta_{1h}+\zeta_{2h})\left(4 - \zeta_{3h}C_{\lambda_{x_h}}^2\Pi_{\lambda_{x_h}}\right)U_{\lambda_{y_hx_h}}}{Q_h}$$

and

$$b_{2h}^* = \frac{4\bar{\lambda}_{Y_h}}{\bar{\lambda}_{X_h}}\left[\frac{\zeta_{2h}C_{\lambda_{y_h}}^2\Pi_{\lambda_{y_h}}\left(4 - \zeta_{3h}C_{\lambda_{x_h}}^2\Pi_{\lambda_{x_h}}\right) - 2(\zeta_{1h}+\zeta_{2h})U_{\lambda_{y_hx_h}}}{Q_h}\right],$$

where

$$Q_h = 4\left(1+\zeta_{2h}C_{\lambda_{yh}}^2\Pi_{\lambda_{yh}}\right)\left\{4+(5\zeta_{1h}-\zeta_{2h})C_{\lambda_{xh}}^2\Pi_{\lambda_{xh}}\right\} - \left[4+2(\zeta_{1h}+\zeta_{2h})U_{\lambda_{y_hx_h}} - \zeta_{3h}C_{\lambda_{x_h}}^2\Pi_{\lambda_{x_h}}\right]^2,$$

$$U_{\lambda_{y_hx_h}} = T_{\lambda_{y_hx_h}}C_{\lambda_{x_h}}^2\sqrt{\Pi_{\lambda_{y_hx_h}}}.$$

Using the optimal coefficients b_{1h}^{*} and b_{2h}^{*}, the minimum bias and MSE of the proposed estimator under stratified double-phase systematic sampling are derived from (A2) and (A4), respectively, as follows:

$$\begin{aligned}\mathrm{Bias}(\hat{\bar{\Gamma}}_{D_{1st}})_{\mathrm{opt}} \cong \sum_{h=1}^{L} W_h \left(\frac{\bar{\lambda}_{Y_h}}{Q_h}\right)\Bigg[&\left(4-\zeta_{3h}C_{\lambda_{x_h}}^2 \Pi_{\lambda_{x_h}}\right)\\ &\times\left\{\zeta_{2h}C_{\lambda_{y_h}}^2 \Pi_{\lambda_{y_h}}\left(4-\zeta_{3h}C_{\lambda_{x_h}}^2 \Pi_{\lambda_{x_h}}\right)-4(\zeta_{1h}+\zeta_{2h})U_{\lambda_{y_h x_h}}\right\}\\ &+C_{\lambda_{x_h}}^2 \Pi_{\lambda_{x_h}}\left\{4(3\zeta_{1h}+\zeta_{2h})-\zeta_{3h}^2 C_{\lambda_{x_h}}^2 \Pi_{\lambda_{x_h}}\right\}-Q_h\Bigg]\end{aligned}$$

and

$$\begin{aligned}\mathrm{MSE}(\hat{\bar{\Gamma}}_{D_{1st}})_{\min} \cong \sum_{h=1}^{L} W_h^2 \left(\frac{\bar{\lambda}_{Y_h}^2}{Q_h^2}\right)\Bigg[&(F_h-Q_h+4G_h)^2+\zeta_{2h}C_{\lambda_{y_h}}^2 \Pi_{\lambda_{y_h}} F_h^2\\ &+4G_h C_{\lambda_{x_h}}^2\Bigg\{(3\zeta_{1h}+\zeta_{2h})\Pi_{\lambda_{xh}} G_h\\ &+(\zeta_{1h}+\zeta_{2h})T_{\lambda_{y_h x_h}}\sqrt{\Pi_{\lambda_{y_h x_h}}}\,F_h-\frac{\zeta_{3h}}{2}\Pi_{\lambda_{xh}}(F_h-Q_h+4G_h)\Bigg\}\Bigg],\end{aligned}$$

where

$$F_h=\left[C_{\lambda_{x_h}}^2 \Pi_{\lambda_{x_h}}\left\{4(3\zeta_{1h}+\zeta_{2h})-\zeta_{3h}^2 C_{\lambda_{x_h}}^2 \Pi_{\lambda_{x_h}}\right\}-2(\zeta_{1h}+\zeta_{2h})\left(4-\zeta_{3h}C_{\lambda_{x_h}}^2 \Pi_{\lambda_{x_h}}\right)U_{\lambda_{y_h x_h}}\right],$$

$$G_h=\zeta_{2h}C_{\lambda_{y_h}}^2 \Pi_{\lambda_{y_h}}\left(4-\zeta_{3h}C_{\lambda_{x_h}}^2 \Pi_{\lambda_{x_h}}\right)-2(\zeta_{1h}+\zeta_{2h})U_{\lambda_{y_h x_h}}.$$

□

Appendix B

Proof. For the approximate bias and MSE derivations, we first represent the proposed estimator $\hat{\bar{\Gamma}}_{D_{2st}}$ using the corresponding relative error terms as follows:

$$\hat{\bar{\Gamma}}_{D_{2st}}=\sum_{h=1}^{L} W_h\left[b_{3h}\bar{\lambda}_{Y_h}(1+\tau_{0h})+b_{4h}\bar{\lambda}_{X_h}\sqrt{(1+\tau_{1h})(1+\tau_{2h})}\right].$$

$$\Rightarrow \hat{\bar{\Gamma}}_{D_{2st}}=\sum_{h=1}^{L} W_h\left[b_{3h}\bar{\lambda}_{Y_h}(1+\tau_{0h})+b_{4h}\bar{\lambda}_{X_h}(1+\tau_{1h}+\tau_{2h}+\tau_{1h}\tau_{2h})^{1/2}\right]. \tag{A5}$$

Applying the binomial expansion to the square-root component and retaining terms up to the second order, the higher-order terms are neglected, yielding the following expression:

$$(1+\tau_{1h}+\tau_{2h}+\tau_{1h}\tau_{2h})^{1/2}\cong 1+\frac{\tau_{1h}+\tau_{2h}+\tau_{1h}\tau_{2h}}{2}-\frac{(\tau_{1h}^2+\tau_{2h}^2+2\tau_{1h}\tau_{2h})}{8}.$$

Therefore

$$\hat{\bar{\Gamma}}_{D_{2st}}\cong\sum_{h=1}^{L} W_h\left[b_{3h}\bar{\lambda}_{Y_h}(1+\tau_{0h})+b_{4h}\bar{\lambda}_{X_h}\left\{1+\frac{\tau_{1h}+\tau_{2h}+\tau_{1h}\tau_{2h}}{2}-\frac{(\tau_{1h}^2+\tau_{2h}^2+2\tau_{1h}\tau_{2h})}{8}\right\}\right].$$

Subtracting $\sum_{h=1}^{L} W_h \bar{\lambda}_{Y_h}$ from both sides, we obtain

$$\hat{\bar{\Gamma}}_{D_{2st}} - \sum_{h=1}^{L} W_h \bar{\lambda}_{Y_h} \cong \sum_{h=1}^{L} W_h \left[-\bar{\lambda}_{Y_h} + b_{3h}\bar{\lambda}_{Y_h}(1+\tau_{0h}) + b_{4h}\bar{\lambda}_{X_h}\left\{1 + \frac{\tau_{1h}+\tau_{2h}+\tau_{1h}\tau_{2h}}{2} - \frac{(\tau_{1h}^2+\tau_{2h}^2+2\tau_{1h}\tau_{2h})}{8}\right\}\right]. \tag{A6}$$

The approximate bias is obtained by applying the expectation operator to the expansion in (A6), giving the following result:

$$\text{Bias}(\hat{\bar{\Gamma}}_{D_{2st}}) \cong \sum_{h=1}^{L} W_h \left[(b_{3h}-1)\bar{\lambda}_{Y_h} + b_{4h}\bar{\lambda}_{X_h}\left\{1 - \frac{\zeta_{3h}C_{\lambda_{x_h}}^2 \Pi_{\lambda_{x_h}}}{8}\right\}\right]. \tag{A7}$$

Squaring the expansion in (A6), retaining second-order relative error terms, and taking expectations gives the approximate MSE as follows:

$$\begin{aligned}\text{MSE}(\hat{\bar{\Gamma}}_{D_{2st}}) \cong \sum_{h=1}^{L} W_h^2 \Bigg[& \bar{\lambda}_{Y_h}^2 + b_{3h}^2\bar{\lambda}_{Y_h}^2\left(1+\zeta_{2h}C_{\lambda_{y_h}}^2\Pi_{\lambda_{y_h}}\right) + b_{4h}^2\bar{\lambda}_{X_h}^2\left(1+\zeta_{1h}C_{\lambda_{x_h}}^2\Pi_{\lambda_{x_h}}\right) \\ & + b_{3h}b_{4h}\bar{\lambda}_{Y_h}\bar{\lambda}_{X_h}\left\{2 + (\zeta_{1h}+\zeta_{2h})U_{\lambda_{y_hx_h}} - \frac{\zeta_{3h}}{4}C_{\lambda_{x_h}}^2\Pi_{\lambda_{x_h}}\right\} \\ & - 2b_{3h}\bar{\lambda}_{Y_h}^2 - 2b_{4h}\bar{\lambda}_{Y_h}\bar{\lambda}_{X_h}\Bigg].\end{aligned} \tag{A8}$$

The optimal values of b_{3h} and b_{4h} are determined by differentiating the MSE expression in (A8) with respect to each constant.

$$b_{3h}^* = 8\left[\frac{(8\zeta_{1h}+\zeta_{3h})C_{\lambda_{x_h}}^2\Pi_{\lambda_{x_h}} - 4(\zeta_{1h}+\zeta_{2h})U_{\lambda_{y_hx_h}}}{K_h}\right]$$

and

$$b_{4h}^* = \frac{8\bar{\lambda}_{Y_h}}{\bar{\lambda}_{X_h}}\left[\frac{8\zeta_{2h}C_{\lambda_{y_h}}^2\Pi_{\lambda_{y_h}} - 4(\zeta_{1h}+\zeta_{2h})U_{\lambda_{y_hx_h}} + \zeta_{xh}C_{\lambda_{x_h}}^2\Pi_{\lambda_{x_h}}}{K_h}\right],$$

where

$$K_h = 64\left(1+\zeta_{2h}C_{\lambda_{y_h}}^2\Pi_{\lambda_{y_h}}\right)\left(1+\zeta_{1h}C_{\lambda_{x_h}}^2\Pi_{\lambda_{x_h}}\right) - \left\{8 + 4(\zeta_{1h}+\zeta_{2h})U_{\lambda_{y_hx_h}} - \zeta_{3h}C_{\lambda_{x_h}}^2\Pi_{\lambda_{x_h}}\right\}^2.$$

Substituting the optimal coefficients b_{3h}^* and b_{4h}^* into (A7) and (A8) yields the resulting minimum bias and MSE expressions as follows:

$$\begin{aligned}\text{Bias}(\hat{\bar{\Gamma}}_{D_{2st}})_{\min} \cong 4\sum_{h=1}^{L} W_h\left(\frac{\bar{\lambda}_{Y_h}C_{\lambda_{x_h}}^2}{K_h}\right)\Bigg[& (\zeta_{1h}+\zeta_{2h})U_{\lambda_{y_hx_h}}\left\{4(\zeta_{1h}+\zeta_{2h})T_{\lambda_{y_hx_h}}\sqrt{\Pi_{\lambda_{y_hx_h}}} - \zeta_3\Pi_{\lambda_{x_h}}\right\} \\ & - 2\zeta_{2h}(8\zeta_{1h}+\zeta_{3h})\Pi_{\lambda_{x_h}}C_{\lambda_{y_h}}^2\Pi_{\lambda_{y_h}}\Bigg]\end{aligned}$$

and

$$\mathrm{MSE}(\hat{\bar{\Gamma}}_{D_{2st}})_{\min} \cong \sum_{h=1}^{L} W_h^2 \left(\frac{\bar{\lambda}_{Y_h}^2 C_{\lambda_{x_h}}^2}{K_h} \right) \left[64\zeta_{1h}\zeta_{2h} C_{\lambda_{y_h}}^2 \Pi_{\lambda_{y_h}} \Pi_{\lambda_{x_h}} - C_{\lambda_{x_h}}^2 \left\{ 4(\zeta_{1h} + \zeta_{2h}) T_{\lambda_{y_h x_h}} \sqrt{\Pi_{\lambda_{y_h x_h}}} - \zeta_{3h} \Pi_{\lambda_{x_h}} \right\}^2 \right].$$

□

References

1. Yates, F. Systematic sampling. *Philos. Trans. R. Soc. Lond. Ser. A* **1948**, *241*, 345–377. [CrossRef]
2. Madow, W.G. On the theory of systematic sampling. *Ann. Math. Stat.* **1948**, *19*, 333–354. [CrossRef]
3. Swain, A.K.P.C. The use of systematic sampling in ratio estimate. *J. Ind. Stat. Assoc.* **1964**, *2*, 160–164.
4. Shukla, N.D. Systematic sampling and product method of estimation. In *Proceedings of the All India Seminar on Demography and Statistics;* BHU: Varanasi, India, 1971.
5. 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.
6. Singh, H.P.; Tailor, R.; Jatwa, N.K. Modified ratio and product estimators for population mean in systematic sampling. *Mod. Appl. Stat. Methods* **2011**, *10*, 424–435. [CrossRef]
7. 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]
8. 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]
9. Singh, H.P.; Jatwa, N.K. A class of exponential type estimators in systematic sampling. *Econ. Qual. Control* **2013**, *27*, 195–208. [CrossRef]
10. Qureshi, M.N.; Khalil, S.; Hanif, M. Generalized semi exponential type estimator under systematic sampling. *J. Stat. Theory Appl.* **2018**, *17*, 283–290. [CrossRef]
11. 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]
12. Kushwaha, S.N.S.; Kushwaha, K.S. A class of ratio, product and difference (RPD) estimators in systematic sampling. *Microelectron. Reliab.* **1993**, *33*, 455–457. [CrossRef]
13. 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]
14. 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]
15. 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*, e43745. [CrossRef]
16. Subzar, 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. Correction in *Sci. Rep.* **2025**, *15*, 8034. [CrossRef] [PubMed]
17. 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]
18. 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]
19. 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]
20. 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]
21. 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.