Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2026 (44) 12** : 1–18.
ISSN-0037-8712
doi:10.5269/bspm.83507

Improved Framework for Mean Estimation under Missingness with Forestry Data Application

Swapnil V. Shinkar, Anup Kumar Sharma and Awadhesh K. Pandey

ABSTRACT: Reliable assessment of environmental conditions relies on statistical analysis of key indicators such as forest data, soil quality, rainfall, and air and water quality. To improve estimation of population parameters, various methods have been developed that effectively incorporate auxiliary information. The current paper constructs effective and versatile families of estimators of the finite-population mean of a study variable y in two-phase sampling, including random non-response in the second phase in the framework of a missing-at-random process and based on two auxiliary variables. Broad class of estimators are constructed as one-to-one functions that satisfy regularity conditions. Large-sample error expansions are used to derive first-order biases and mean-squared errors for both sampling scenarios. Optimality conditions are derived which results in explicit minimum-mse expressions that quantify efficiency gains under various correlations. Analytical efficiency comparisons demonstrate that suggested estimation methods outperform the conventional responding-sample mean estimator under mild non-negativity conditions. Empirical evaluation on four natural populations (including forestry survey data) shows relative efficiencies exceeding 100% across non-response rates p = 0.02–0.10, with τ_m consistently outperforming τ_u, $\tau_m^\star$ outperforming $\tau_u^\star$, and Case II often slightly more efficient than Case I. These findings indicate that incorporating two auxiliary variables within two-phase designs can substantially reduce non-response-induced inefficiency and enable more cost-effective, precise mean estimation in large-scale surveys.

Keywords: Population mean, double sampling, random non-response, mean square error (mse), efficiency.

Contents

2020 *Mathematics Subject Classification:* 62D05, 94A20.
Submitted April 16, 2026. Published June 09, 2026.

1. Introduction

Simple random sampling (SRS) is commonly employed for data collection when the population is homogeneous. However, substantial heterogeneity in real-world populations often renders SRS inefficient for estimating population parameters. In such contexts, two-phase sampling provides a more effective alternative. The survey sampling literature identifies two-phase sampling as an efficient method for improving estimation when auxiliary information is only partially available or costly to obtain. Cochran (1977)[4] demonstrated that two-phase sampling is particularly advantageous when population parameters of auxiliary variables are unknown and must be estimated from a preliminary large sample. This methodology enables the effective use of auxiliary information while maintaining control over survey costs.

Murthy (1967) [15] further emphasized that the two-phase approach offers flexibility in practical survey situations where complete information cannot be collected in a single phase. By dividing the sampling process into two stages, researchers can initially gather inexpensive auxiliary data and subsequently collect detailed information from a subsample. This strategy enhances efficiency without a proportional increase in cost.

In survey sampling, the use of auxiliary variables correlated with the study variable is a standard practice to enhance the efficiency and reliability of estimators. Classical estimation techniques, including ratio, regression and product estimators, utilize auxiliary information to achieve greater precision than estimators that disregard such data. In practical applications, auxiliary information frequently takes the form of attributes. For example, crop yield may be associated with a specific seed variety, milk production with cattle breed and agricultural output with particular crop types. Incorporating these auxiliary substantially improves estimation accuracy.

Despite these advantages, complete information on auxiliary variables is often unavailable for the entire population. In such situations, two-phase sampling provides a practical and cost-effective solution, as formalized by Cochran (1977) [4]. In this approach, a large initial sample is selected to collect auxiliary information, followed by a second-phase subsample for observing the study variable. This method is particularly effective when population parameters of auxiliary variables are unknown or when collecting auxiliary information is costly. In two-phase sampling, auxiliary characteristics are estimated during the first phase, while detailed observations of the study variable are obtained in the second utilizing auxiliary information in two-phase sampling significantly enhances estimator precision, which is a primary benefit of this approach. Sukhatme (1962) [29] demonstrated that ratio and regression estimators in two-phase sampling yield lower mean square errors (MSEs) than the conventional estimators, especially when the auxiliary variable is highly correlated with the study variable. This efficiency gain serves as a major motivation for adopting two-phase designs.

A substantial body of literature has developed estimators for two-phase sampling frameworks, building on these foundational findings. Notable contributions include works by Sahoo et al. (1993) [17], Samiuddin and Hanif (2007) [18], Chaudhari and Singh (2012) [3], Hamad et al. (2013) [7], Malik and Singh (2015) [13], Zaman and Kadilar (2021) [32], and Liu and Arslan (2021) [12], among others.

The use of auxiliary characteristics to estimate population parameters in sample surveys is well established for achieving precise estimates. However, information on auxiliary characteristics is not always available for all population units. Two-phase sampling is recognized as an effective method for generating valid estimates of auxiliary variable population parameters from the first-phase sample. For instance, in a household survey, household size may serve as an auxiliary characteristic for estimating family expenditure. Several pioneering studies in context of two-phase sampling include those by Chand (1975) [2], Kiregyera (1980) [11] and Singh and Sharma (2014, 2015) [22,23], among others.

Despite these methodological advances, most existing studies assume complete response from sampled units, a condition that is seldom met in practice. Non-response frequently arises in survey sampling, resulting in incomplete data and potential biased estimates if not properly addressed. Rubin (1976) [16] introduced the concept of missing data mechanisms, including Missing Completely at Random (MCAR), Missing at Random (MAR) and Missing Not at Random (MNAR). The MAR mechanisms is particularly significant, as it permits valid statistical inference when the probability of missingness depends solely on observed data. In real-world survey and observational studies, complete data collection is seldom achievable due to non-response, measurement limitations, and logistical constraints. For instance, in healthcare surveys, patients may fail to report sensitive information such as income or lifestyle habits, while in longitudinal studies, attrition leads to missing follow-up observations. Similarly, in agricultural and forestry surveys, data may be incomplete due to inaccessible locations, adverse environmental conditions, or equipment limitations. Household surveys frequently encounter partial responses, particularly for sensitive variables like expenditure or income. Such incompleteness, if ignored, can introduce significant bias and reduce the efficiency of estimators. Earlier contributions by Hansen and Hurwitz (1946) [9], as well as subsequent studies by Tracy and Osahan (1994) [31], Heitjan and Basu (1996) [10], Singh and Joarder (1998) [26], Singh et al. (2000) [27], and Singh and Tracy (2001) [25], have examined the impact of non-response on the estimation of population parameters.

To address the effect of missing data, a range of imputation techniques has been proposed, including mean, ratio, regression, hot-deck and cold-deck imputation and nearest neighbor methods. These approaches seek to restore the structural completeness of datasets and facilitate valid statistical analysis. Researchers have consistently developed improved imputation based estimators to enhance the efficiency of population mean estimation in scenarios with incomplete data. The practical relevance of the proposed estimators can be clearly understood through real-life survey scenarios. In healthcare studies, variables such as blood glucose (study variable) are often missing for a subset of individuals, while auxiliary variables like age and body mass index are readily available. Similarly, in agricultural and forestry surveys, primary variables such as crop yield or timber volume are expensive and prone to non-response, whereas auxiliary variables like land area, tree height, or crown diameter are easily obtainable. These situations naturally conform to a two-phase sampling framework, where auxiliary information is collected in the first phase and detailed observations are obtained from a subsample in the second phase. Under the Missing at Random (MAR) assumption, the probability of non-response depends only on observed auxiliary variables, enabling valid adjustment through the proposed estimators.

Auxiliary information plays a crucial role not only in estimation but also in addressing non-response. When complete auxiliary information is available, traditional techniques such as ratio and regression estimation can be effectively implemented. However, when auxiliary population parameters are unknown, two-phase sampling becomes essential. Several authors, including Chand (1975) [2], Kiregyera (1980) [11], Chaudhary and Singh (2012) [3], Mehta and tailor (2020) [14], Basit et al. (2023) [1], and Grover and Sharma (2023) [6], have made significant contributions to the development of two-phase sampling methodologies that incorporate auxiliary information and multiple auxiliary variables.

Motivated by these considerations and utilizing two auxiliary variables, this study proposes efficient classes of estimators for the population mean in the presence of random non-response under the missing at random mechanism in two-phase sampling. The advantages of the proposed class of estimators are demonstrated through empirical studies, with recommendations provided for survey practitioners. The subsequent sections of the paper is structured as follows. Section 2 describes the two-phase sampling design, notation, and assumptions, including the treatment of non-response under the Missing at Random framework. Section 3 develops the proposed estimators and derives their bias and mean square error expressions. Section 4 and 5 discusses optimality conditions and theoretical efficiency comparisons. Section 6 provides empirical validation using real datasets. Section 7 concludes with key findings and future research directions. Finally, Section 8 concludes the work with some closing remarks.

2. Problem Formulation

2.1. Sampling Structure and Notations

A finite population of N units considered, with sampling conducted in two distinct phases. The study variable of interest is denoted by y, while x and z represent auxiliary variables. The aim is to be estimate the population mean $\bar{Y}$ of y using information from the auxiliary variables. In this context, the population mean $\bar{X}$ of x is unknown, whereas complete population level information on z is available. To estimate $\bar{X}$, a first phase sample $S^{'}$ of size n is selected from the population U using simple random sampling without replacement (SRSWOR). Data on both auxiliary variables x and z are collected from this initial sample. Subsequently, a second phase sample S of size m, where $(m < n)$, is drawn using SRSWOR from the first-phase sample under specified conditions. In this phase, observations on the study variable y and the auxiliary variable x are recorded.

Case-I: Second phase sample S is drawn as a subsample of the first phase sample $(i.e. S \subset S^{'})$. **Case-II**:

Second phase sample S is drawn independently of the first phase sample $S^{'}$ $(i.e. S \cap S^{'})$. Hence onwards, we use the following notations:
$\bar{Y}$: The population mean of study variable y.
$\bar{X}, \bar{Z}$: The population means of auxiliary variables x and z respectively.
$\bar{y}_m, \bar{x}_m, \bar{x}_n, \bar{z}_n$: The sample means of the respective variables based on the sample sizes shown in the suffices.
$\bar{y}^* = \frac{1}{(m-2)} \sum_i^{m-2} y_i, \bar{x}^* = \frac{1}{(m-2)} \sum_i^{m-2} x_i$: The sample means of the respective variables based on the responding part of the second phase sample S.
In the initial phase sample $S^{'}$ is assumed to exhibit a complete response scenario. In contrast, in the second-phase, sample S may experience non-response for both variables y and x, or for only one variable. This random non-response is modeled using a discrete probability distribution, as detailed in the following:

In cases of random non-response within the second-phase sample S, which consists of m units, r $\{r = 0, 1, 2, ..., (m-2)\}$ denotes the number of units for which data could not be collected due to non-response. Data for variables affected by non-response are obtained from the remaining (m - 2) units that did respond in the second phase sample. According to Singh and Joarder (1998) [26], r follows a discrete probability distribution, with assumption that $0 \leq r \leq (m-2)$ and p represents the probability of non-response among the (m - 2) possible non-response values. See Singh et al. (2017) [24] and Sharma (2017) [19].

$$p(r) = \frac{m-r}{mq+2p} \begin{pmatrix} m-2 \\ r \end{pmatrix} p^r q^{m-2-r}; r = 0, 1, 2, ...(m-2) \tag{2.1}$$

where q = (1-p)

2.2. Suggested Strategy

To achieve a more accurate estimation, a ratio type estimator for the population mean $\bar{Y}$ should be constructed within a two-phase sampling framework. This approach utilizes auxiliary information from variable x, for which the population mean $\bar{X}$ is unknown, in accordance with the Sukhatme (1962) [29],

$$\bar{y}_{rd} = \bar{y}_m \frac{\bar{x}_n}{\bar{x}_m} \tag{2.2}$$

Following Sharma and Singh (2018) [20], can develop a wide range of estimators for the population mean $\bar{Y}$ in a two-phase sampling framework when $\bar{X}$ is unknown but information on $\bar{Z}$ is available.

$$\tau_{\mathfrak{g}} = \bar{y}\mathfrak{g}(u, v) \tag{2.3}$$

here $u = \frac{\bar{x}_m}{\bar{X}}, v = \frac{S^2_{xm}}{S^2_x}$. The parametric functions $\mathfrak{g} = (u, v)$ models how sample and population statistics relate, enabling adaptable estimation. Regulatory conditions, like those of Shrivastava and Jhajj (1981) [28], apply $\mathfrak{g}$ is defined such that,

$$\mathfrak{g}(1, 1) = 1$$

In respond to the aforementioned suggestions, two general estimators for the population mean $\bar{Y}$ were introduced to enhance estimation and flexibility. These estimators are implemented within a two-phase sampling framework under two distinct scenarios.
Situation I: The suggested general classes of estimators indicated by τ_u and τ_m for estimating the population mean $\bar{Y}$ under a two phase sampling framework are described as follows where random non-response only influences the study variable y.

$$\tau_u = \mathfrak{f}(\bar{y}^\star_m, \bar{x}_m, \bar{x}_n) \tag{2.4}$$

$$\tau_m = \mathfrak{g}(\bar{y}^\star_m, \bar{x}_m, \bar{x}_n, \bar{z}_n) \tag{2.5}$$

we consider the function $\mathfrak{f}(\bar{y}^\star_m, \bar{x}_m, \bar{x}_n)$ and $\mathfrak{g}(\bar{y}^\star_m, \bar{x}_m, \bar{x}_n, \bar{z}_n)$ as one to one function of $(\bar{y}^\star_m, \bar{x}_m, \bar{x}_n)$ and $(\bar{y}^\star_m, \bar{x}_m, \bar{x}_n, \bar{z}_n)$ such that

$$\mathcal{F}(\bar{Y},\bar{X},\bar{X})=\bar{Y}\Rightarrow \left.\frac{\partial\mathcal{F}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n)}{\partial\bar{y}_m^\star}\right|_{(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n)}=1 \tag{2.6}$$

and

$$\mathcal{G}(\bar{Y},\bar{X},\bar{X},\bar{Z})=\bar{Y}\Rightarrow \left.\frac{\partial\mathcal{G}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n,\bar{z}_n)}{\partial\bar{y}_m^\star}\right|_{(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n,\bar{z}_n)}=1 \tag{2.7}$$

with $(\bar{Y},\bar{X},\bar{X})$, $(\bar{Y},\bar{X},\bar{X},\bar{Z})$, $\mathcal{F}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n)$ and $\mathcal{G}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n,\bar{z}_n)$ satisfy the following conditions:
(i) Whatever be the chosen samples, $(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n)$ and $(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n,\bar{z}_n)$ assume values in a closed convex subspace R^3 and R^4, of the three and four dimensional real space containing the point $(\bar{Y},\bar{X},\bar{X})$ and $(\bar{Y},\bar{X},\bar{X},\bar{Z})$.
(ii) The function $\mathcal{F}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n)$ and $\mathcal{G}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n,\bar{z}_n)$ is continuous and bounded in R^3 and R^4.
(iii) The first and second partial derivative of $\mathcal{F}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n)$ and $\mathcal{G}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n,\bar{z}_n)$ exist and are continuous and bounded in R^3 and R^3.
The class of estimators τ_u and τ_m is very comprehensive, as shown by equation (2.4) and (2.5). For any parametric functions $\mathcal{F}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n)$ and $\mathcal{G}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n,\bar{z}_n)$ that satisfy the above regularity conditions, $\mathcal{F}(\bar{Y},\bar{X},\bar{X})=\bar{Y}$ and $\mathcal{G}(\bar{Y},\bar{X},\bar{X},\bar{Z})=\bar{Y}$ may produce estimators of $\bar{Y}$. The class of estimators τ_u includes, for instances, the following ratio, product, regression and exponential type estimators of $\bar{Y}$.

$$\tau_1=\bar{y}_m^\star\frac{\bar{x}_n}{\bar{x}_m},\qquad \tau_2=\bar{y}_m^\star\frac{\bar{x}_m}{\bar{x}_n},\qquad \tau_3=\bar{y}_m^\star+\alpha_1(\bar{x}_n-\bar{x}_m),\qquad \tau_4=\bar{y}_m^\star-\alpha_1(\bar{x}_n-\bar{x}_m),$$

$$\tau_5=\bar{y}_m^\star exp\left(\frac{\bar{x}_n-\bar{x}_m}{\bar{x}_n+\bar{x}_m}\right),\quad \tau_6=\bar{y}_m^\star\left(\frac{\bar{x}_m}{\bar{x}_n}\right)^{\alpha_2},\quad \tau_7=\bar{y}_m^\star exp\left(\frac{\bar{x}_m-\bar{x}_n}{\bar{x}_m+\bar{x}_n}\right),\quad \tau_8=\bar{y}_m^\star\left\{2-\left(\frac{\bar{x}_m}{\bar{x}_n}\right)^{\alpha_3}\right\},$$

$$\tau_9=\bar{y}_m^\star\left(\frac{\bar{x}_n}{\bar{x}_n+\alpha_4(\bar{x}_m-\bar{x}_n)}\right),\quad \tau_{10}=\bar{y}_m^\star\left\{\alpha_5\frac{\bar{x}_n}{\bar{x}_m}+(1-\alpha_5)\left(\frac{\bar{x}_n}{\bar{x}_m}\right)\right\},\quad \tau_{11}=\bar{y}_m^\star\left\{\alpha_6\frac{\bar{x}_m}{\bar{x}_n}+(1-\alpha_6)\left(\frac{\bar{x}_m}{\bar{x}_n}\right)\right\},$$

$$\tau_{12}=\bar{y}_m^\star\left\{\alpha_7+(1-\alpha_7)\left(\frac{\bar{x}_m}{\bar{x}_n}\right)\right\},\qquad \tau_{13}=\bar{y}_m^\star\left\{\alpha_8+(1-\alpha_8)\left(\frac{\bar{x}_n}{\bar{x}_m}\right)\right\},$$

$$\tau_{14}=\bar{y}_m^\star\left\{\frac{\alpha_9\bar{x}_n+(1-\alpha_9)\bar{x}_m}{\alpha_9\bar{x}_m+(1-\alpha_9)\bar{x}_n}\right\},\quad \tau_{15}=\bar{y}_m^\star\left\{\alpha_{10}exp\left(\frac{\bar{x}_n-\bar{x}_m}{\bar{x}_n+\bar{x}_m}\right)+(1-\alpha_{10})exp\left(\frac{\bar{x}_m-\bar{x}_n}{\bar{x}_m+\bar{x}_n}\right)\right\},etc...$$

Similarly, the following estimators of $\bar{Y}$ are the members of the class estimator τ_m as

$$\delta_1=\bar{y}_m^\star\frac{\bar{x}_n}{\bar{x}_m}\frac{\bar{Z}}{\bar{z}_n},\qquad \delta_2=\bar{y}_m^\star\frac{\bar{x}_m}{\bar{x}_n}\frac{\bar{z}_n}{\bar{Z}},\qquad \delta_3=\frac{\bar{y}_m^\star}{\bar{x}_m}\{\bar{x}_n+b_1(\bar{Z}-\bar{z}_n)\},\qquad \delta_4=\bar{y}_m^\star+b_2\left(\frac{\bar{x}_n}{\bar{z}_n}\bar{Z}-\bar{x}_m\right),$$

$$\delta_5=\bar{y}_m^\star+b_3\{\bar{x}_n+b_4(\bar{Z}-\bar{z}_n)-\bar{x}_m\},\quad \delta_6=\bar{y}_m^\star exp\left(\frac{\frac{\bar{x}_n}{\bar{z}_n}\bar{Z}-\bar{x}_m}{\frac{\bar{x}_n}{\bar{z}_n}\bar{Z}+\bar{x}_m}\right),\quad \delta_7=\bar{y}_m^\star exp\left(\frac{\bar{x}_m-\frac{\bar{x}_n}{\bar{z}_n}\bar{Z}}{\bar{x}_m+\frac{\bar{x}_n}{\bar{z}_n}\bar{Z}}\right),$$

$$\delta_8=\bar{y}_m^\star exp\left(\frac{\bar{x}_n-\bar{x}_m}{\bar{x}_n+\bar{x}_m}\right)\frac{\bar{Z}}{\bar{z}_n},\quad \delta_9=\bar{y}_m^\star exp\left(\frac{\bar{x}_n-\bar{x}_m}{\bar{x}_n+\bar{x}_m}\right)exp\left(\frac{\bar{Z}-\bar{z}_n}{\bar{Z}+\bar{z}_n}\right),\quad \delta_{10}=\bar{y}_m^\star\left\{\frac{\bar{x}_m}{\bar{x}_n+b_1(\bar{Z}-\bar{z}_n)}\right\},$$

$$\delta_{11}=\{\bar{y}_m^\star+b_5(\bar{x}_n-\bar{x}_m)\}exp\left(\frac{\bar{Z}-\bar{z}_n}{\bar{Z}+\bar{z}_n}\right),$$

$$\delta_{12}=\bar{y}_m^\star\left\{b_6exp\left(\frac{\frac{\bar{x}_n}{\bar{z}_n}\bar{Z}-\bar{x}_m}{\frac{\bar{x}_n}{\bar{z}_n}\bar{Z}+\bar{x}_m}\right)+(1-b_6)exp\left(\frac{\bar{x}_m-\frac{\bar{x}_n}{\bar{z}_n}\bar{Z}}{\bar{x}_m+\frac{\bar{x}_n}{\bar{z}_n}\bar{Z}}\right)\right\},etc...$$

Situation II: Both the study variable y and the auxiliary variable x are expected to exhibit random non-response in this context. Therefore the following formulation presents the proposed generic class of estimators for the population mean $\bar{Y}$ within a two-phase sampling framework.

$$\tau_u^\star = \varphi(\bar{y}_m^\star, \bar{x}_m^\star, \bar{x}_n) \tag{2.8}$$

$$\tau_m^\star = \Psi(\bar{y}_m^\star, \bar{x}_m^\star, \bar{x}_n, \bar{z}_n) \tag{2.9}$$

we consider the function $\varphi(\bar{y}_m^\star, \bar{x}_m^\star, \bar{x}_n)$ and $\Psi(\bar{y}_m^\star, \bar{x}_m^\star, \bar{x}_n, \bar{z}_n)$ as one to one function of $(\bar{y}_m^\star, \bar{x}_m, \bar{x}_n)$ and $(\bar{y}_m^\star, \bar{x}_m^\star, \bar{x}_n, \bar{z}_n)$ such that

$$\varphi(\bar{Y}, \bar{X}, \bar{X}) = \bar{Y} \Rightarrow \left.\frac{\partial\varphi(\bar{y}_m^\star, \bar{x}_m, \bar{x}_n, \bar{z}_n)}{\partial\bar{y}_m^\star}\right|_{(\bar{y}_m^\star, \bar{x}_m, \bar{x}_n, \bar{z}_n)} = 1 \tag{2.10}$$

and

$$\Psi(\bar{Y}, \bar{X}, \bar{X}, \bar{X}) = \bar{Y} \Rightarrow \left.\frac{\partial\Psi(\bar{y}_m^\star, \bar{x}_m^\star, \bar{x}_n, \bar{z}_n)}{\partial\bar{y}_m^\star}\right|_{(\bar{y}_m^\star, \bar{x}_m, \bar{x}_n, \bar{z}_n)} = 1 \tag{2.11}$$

with $(\bar{Y}, \bar{X}, \bar{X})$, $(\bar{Y}, \bar{X}, \bar{X}, \bar{Z})$, $\varphi(\bar{y}_m^\star, \bar{x}_m, \bar{x}_n)$ and $\Psi(\bar{y}_m^\star, \bar{x}_m^\star, \bar{x}_n, \bar{z}_n)$ satisfy the following conditions similar to those given for class of estimators τ_u and τ_m above.
Following a similar line of development it can be shown that the estimator classes $\tau_u^\star$ and $\tau_m^\star$ are quite broad in scope and several specific estimators can be obtained as special members of these classes.
Estimators belonging to the class estimators $\tau_u^\star$:

$$\rho_1 = \bar{y}_m^\star \frac{\bar{x}_n}{\bar{x}_m^\star}, \qquad \rho_2 = \bar{y}_m^\star \frac{\bar{x}_m^\star}{\bar{x}_n}, \qquad \rho_3 = \bar{y}_m^\star + \alpha_1(\bar{x}_n - \bar{x}_m^\star), \qquad \rho_4 = \bar{y}_m^\star - \alpha_1(\bar{x}_n - \bar{x}_m^\star),$$

$$\rho_5 = \bar{y}_m^\star exp\left(\frac{\bar{x}_n - \bar{x}_m^\star}{\bar{x}_n + \bar{x}_m^\star}\right), \quad \rho_6 = \bar{y}_m^\star \left(\frac{\bar{x}_m^\star}{\bar{x}_n}\right)^{\alpha_2}, \quad \rho_7 = \bar{y}_m^\star exp\left(\frac{\bar{x}_m^\star - \bar{x}_n}{\bar{x}_m^\star + \bar{x}_n}\right), \quad \rho_8 = \bar{y}_m^\star\left\{2 - \left(\frac{\bar{x}_m^\star}{\bar{x}_n}\right)^{\alpha_3}\right\},$$

$$\rho_9 = \bar{y}_m^\star\left(\frac{\bar{x}_n}{\bar{x}_n + \alpha_4(\bar{x}_m^\star - \bar{x}_n)}\right), \quad \rho_{10} = \bar{y}_m^\star\left\{\alpha_5\frac{\bar{x}_n}{\bar{x}_m^\star} + (1-\alpha_5)\left(\frac{\bar{x}_n}{\bar{x}_m^\star}\right)\right\}, \quad \rho_{11} = \bar{y}_m^\star\left\{\alpha_6\frac{\bar{x}_m^\star}{\bar{x}_n} + (1-\alpha_6)\left(\frac{\bar{x}_m^\star}{\bar{x}_n}\right)\right\},$$

$$\rho_{12} = \bar{y}_m^\star\left\{\alpha_7 + (1-\alpha_7)\left(\frac{\bar{x}_m^\star}{\bar{x}_n}\right)\right\}, \qquad \rho_{13} = \bar{y}_m^\star\left\{\alpha_8 + (1-\alpha_8)\left(\frac{\bar{x}_n}{\bar{x}_m^\star}\right)\right\},$$

$$\rho_{14} = \bar{y}_m^\star\left\{\frac{\alpha_9\bar{x}_n + (1-\alpha_9)\bar{x}_m^\star}{\alpha_9\bar{x}_m^\star + (1-\alpha_9)\bar{x}_n}\right\}, \quad \rho_{15} = \bar{y}_m^\star\left\{\alpha_{10}exp\left(\frac{\bar{x}_n - \bar{x}_m^\star}{\bar{x}_n + \bar{x}_m^\star}\right) + (1-\alpha_{10})exp\left(\frac{\bar{x}_m^\star - \bar{x}_n}{\bar{x}_m^\star + \bar{x}_n}\right)\right\}, etc...$$

Estimators belonging to the class estimators $\tau_m^\star$:

$$\mathcal{R}_1 = \bar{y}_m^\star \frac{\bar{x}_n^{'}}{\bar{x}_m}, \qquad \mathcal{R}_2 = \bar{y}_m^\star \frac{\bar{x}_m}{\bar{x}_n^{'}}, \qquad \mathcal{R}_3 = \frac{\bar{y}_m^\star}{\bar{x}_m^\star}\{\bar{x}_n + b_1(\bar{Z} - \bar{z}_n)\}, \qquad \mathcal{R}_4 = \bar{y}_m^\star + b_2(\bar{x}_n^{'} - \bar{x}_m),$$

$$\mathcal{R}_5 = \bar{y}_m^\star + b_3\{\bar{x}_n + b_4(\bar{Z} - \bar{z}_n) - \bar{x}_m^\star)\}, \qquad \mathcal{R}_6 = \bar{y}_m^\star exp\left(\frac{\bar{x}_n^{'} - \bar{x}_m^\star}{\bar{x}_n^{'} + \bar{x}_m^\star}\right), etc...$$

where, $\bar{x}_n^{'} = \frac{\bar{x}_n}{\bar{z}_n}\bar{Z}$

3. Properties of τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$

Under the assumptions of large samples, the mean square errors and bias of the suggested estimator classes τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$ are obtain to the first order of approximations using the following transformations:

$$\bar{y}_m^\star = \bar{Y}(1+\varrho_0), \quad \bar{x}_m = \bar{X}(1+\varrho_1), \quad \bar{x}_n = \bar{X}(1+\varrho_2), \quad \bar{z}_n = \bar{Z}(1+\varrho_3), \qquad \bar{x}_m^\star = \bar{X}(1+\varrho_4)$$

For every $i=(0,1,2,3,4)$, it is assumed that $E(\varrho_i)=0$ and $|\varrho_i|\leq 1$. For scenarios I and II of the two phase sampling layout discussed in section 2, the bias and mean square error of the suggested estimator classes τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$ have been independently derived under these conditions. The resulting expressions are presented below.

3.1. Bias and mean square errors of the Proposed Classes of Estimators under case I

In this section, the second-phase sample S, with size m is selected as a subsample from the first-phase sample $S^{'}$, which has size n. The outcomes resulting from this sampling strategy are presented below.

$$E(\varrho_0^2) = f_1^\star C_y^2, \quad E(\varrho_1^2) = f_1 C_x^2, \quad E(\varrho_2^2) = f_2 C_x^2, \quad E(\varrho_3^2) = f_2 C_z^2, \qquad E(\varrho_4^2) = f_1^\star C_x^2,$$

$$E(\varrho_0\varrho_1) = f_1\rho_{yx}C_yC_x, \quad E(\varrho_0\varrho_2) = f_2\rho_{yx}C_yC_x, \quad E(\varrho_0\varrho_3) = f_2\rho_{yz}C_yC_z, \quad E(\varrho_0\varrho_4) = f_1^\star\rho_{yx}C_yC_x,$$

$$E(\varrho_1\varrho_2) = f_2C_x^2, \quad E(\varrho_1\varrho_3) = f_2\rho_{xz}C_xC_z, \quad E(\varrho_1\varrho_4) = f_1C_x^2, \quad E(\varrho_2\varrho_3) = f_2\rho_{xz}C_xC_z,$$

$$E(\varrho_2\varrho_4) = f_2C_x^2, \qquad E(\varrho_3\varrho_4) = f_2\rho_{xz}C_xC_z$$

where,$f_1^\star = \left(\frac{1}{mq+2p}-\frac{1}{N}\right)$, $f_1=\left(\frac{1}{m}-\frac{1}{N}\right)$, $f_2=\left(\frac{1}{n}-\frac{1}{N}\right)$, $f_3=(f_1-f_2)=\left(\frac{1}{m}-\frac{1}{n}\right)$ and $C_y=\frac{S_y}{\bar{Y}}$, $C_x=\frac{S_x}{\bar{X}}$, $C_z=\frac{S_z}{\bar{Z}}$, ρ_{yx}, ρ_{yz} and ρ_{xz} are population coefficient of variation and population correlation coefficient between the variable y, x and z respectively.
Under the stated transformations, the estimators τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$ can be expressed in the following forms:
To represents the estimator class τ_u in terms of the error components e's, the function $\mathcal{F}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n)$ is expanded around the point $(\bar{Y},\bar{X},\bar{X})$ using a third Taylor's series expansion. This yields the following expression:

$$\begin{aligned}
\mathcal{F}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n) &= \mathcal{F}(\bar{Y},\bar{X},\bar{X}) + (\bar{y}_m^\star-\bar{Y})H_1 + (\bar{x}_m-\bar{X})H_2 + (\bar{x}_n-\bar{X})H_3 \\
&+\frac{1}{2}\left\{(\bar{y}_m^\star-\bar{Y})^2H_{11} + (\bar{x}_m-\bar{X})^2H_{22} + (\bar{x}_n^\star-\bar{X})^2H_{33}\right. \\
&\left.+2(\bar{y}_m^\star-\bar{Y})(\bar{x}_m-\bar{X})H_{12} + 2(\bar{y}_m^\star-\bar{Y})(\bar{x}_n-\bar{X})H_{13} + 2(\bar{x}_m-\bar{Y})(\bar{x}_n-\bar{X})H_{23}\right\} \\
&+\frac{1}{6}\left\{(\bar{y}_m^\star-\bar{Y})\frac{\partial}{\partial\bar{y}_m^\star} + (\bar{x}_m-\bar{X})\frac{\partial}{\partial\bar{x}_m} + (\bar{x}_n-\bar{X})\frac{\partial}{\partial\bar{x}_n}\right\}^3 \mathcal{F}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n) + ...
\end{aligned}$$

where, $H_1 = \left.\frac{\partial\mathcal{F}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n)}{\partial\bar{y}_m^\star}\right|_{(\bar{Y},\bar{X},\bar{X})}$, $H_2 = \left.\frac{\partial\mathcal{F}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n)}{\partial\bar{x}_m}\right|_{(\bar{Y},\bar{X},\bar{X})}$,$H_3 = \left.\frac{\partial\mathcal{F}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n)}{\partial\bar{x}_n}\right|_{(\bar{Y},\bar{X},\bar{X})}$, and $(H_{11},H_{22},H_{33},H_{12},H_{13},H_{23})$ are the second order partial derivative of $\mathcal{F}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n)$ at the point $(\bar{Y},\bar{X},\bar{X})$ and $\bar{y}_m^\star=\bar{Y}+\theta(\bar{y}_m^\star-\bar{Y})$, $\bar{x}_m=\bar{X}+\theta(\bar{x}_m-\bar{X})$, $\bar{x}_n=\bar{X}+\theta(\bar{x}_n-\bar{X})$ for $0<\theta<1$.
It can be note some terms under the mentioned function of $\mathcal{F}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n)$ in the equation (2.6) and (2.7) as follows

$$\mathcal{F}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n)=\bar{Y}\Rightarrow H_1=1, \qquad and \qquad H_{11}=\left.\frac{\partial^2}{\partial\bar{y}_m^{\star 2}}\mathcal{F}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n)\right|_{(\bar{y}_m,\bar{x}_m,\bar{x}_n)}=0$$

$$\mathcal{G}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n,\bar{z}_n)=\bar{Y}\Rightarrow J_1=1, \qquad \textit{and} \qquad J_{11}=\frac{\partial^2}{\partial \bar{y}_m^{\star 2}}\mathcal{G}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n,\bar{z}_n)\Big|_{(\bar{y}_m,\bar{x}_m,\bar{x}_n,\bar{z}_n)}=0$$

Since the population mean $\bar{X}$ of the auxiliary variable x is not known, it becomes necessary to impose the following constraint as
$H_2=-H_3$
Thus, the expression of the estimator τ_u up to first degree approximation is given by

$$\begin{aligned}(\tau_u-\bar{Y})=\mathcal{F}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n) &= \bar{Y}\varrho_0+\bar{X}(\varrho_1-\varrho_2)H_2\\ &+\frac{1}{2}\{\bar{X}^2(\varrho_1^2H_{22}+\varrho_2^2H_{33}+2\varrho_1\varrho_2H_{23})+2\bar{Y}\bar{X}(\varrho_0\varrho_1H_{12}+\varrho_0\varrho_2H_{13})\}\end{aligned} \tag{3.1}$$

Similarly, we have the expression of τ_m, $\tau_u^\star$ and $\tau_m^\star$ up to the first degree of approximation as

$$\begin{aligned}(\tau_m-\bar{Y})=\mathcal{G}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n,\bar{z}_n) &= \bar{Y}\varrho_0+\bar{X}(\varrho_1-\varrho_2)J_2+\bar{z}\varrho_3J_4\\ &+\frac{1}{2}\{\bar{X}^2(\varrho_1^2J_{22}+\varrho_2^2J_{33}+2\varrho_1\varrho_2J_{23})+\bar{z}^2\varrho_3^2J_{44} \qquad (3.2)\\ &+2\bar{Y}\bar{X}(\varrho_0\varrho_1J_{12}+\varrho_0\varrho_2J_{13})\\ &+2\bar{Y}\bar{Z}\varrho_0\varrho_3J_{14}+2\bar{X}\bar{Z}(\varrho_1\varrho_3J_{24}+\varrho_2\varrho_3J_{34})\}\end{aligned} \tag{3.3}$$

$$\begin{aligned}(\tau_u^\star-\bar{Y}) &= \mathcal{F}(\bar{y}_m^\star,\bar{x}_m^\star,\bar{x}_n)=\bar{Y}\varrho_0+\bar{X}(\varrho_4-\varrho_2)K_2\\ &+\frac{1}{2}\{\bar{X}^2(\varrho_4^2K_{22}+\varrho_2^2K_{33}+2\varrho_2\varrho_4K_{23})+2\bar{Y}\bar{X}(\varrho_0\varrho_4K_{12}+\varrho_0\varrho_{2K_{13}})\}\end{aligned} \tag{3.4}$$

$$\begin{aligned}(\tau_m^\star-\bar{Y}) &= \mathcal{G}(\bar{y}_m^\star,\bar{x}_m^\star,\bar{x}_n,\bar{z}_n)=\bar{Y}\varrho_0+\bar{X}(\varrho_4-\varrho_2)S_2+\bar{z}\varrho_3S_4+\frac{1}{2}\{\bar{X}^2(\varrho_4^2S_{22}+\varrho_2^2S_{33}+2\varrho_4\varrho_2S_{23})\\ &+\bar{z}^2\varrho_3^2S_{44}+2\bar{Y}\bar{X}(\varrho_0\varrho_4S_{12}+\varrho_0\varrho_2S_{13})+2\bar{Y}\bar{Z}\varrho_0\varrho_3S_{14}+2\bar{X}\bar{Z}(\varrho_4\varrho_3S_{24}+\varrho_2\varrho_3S_{34})\}\end{aligned} \tag{3.5}$$

where, $J_1=\frac{\partial\mathcal{G}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n,\bar{z}_n)}{\partial\bar{y}_m^\star}\Big|_{(\bar{Y},\bar{X},\bar{X},\bar{Z})}$, $J_2=\frac{\partial\mathcal{G}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n,\bar{z}_n)}{\partial\bar{x}_m}\Big|_{(\bar{Y},\bar{X},\bar{X},\bar{Z})}$, $J_3=\frac{\partial\mathcal{G}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n,\bar{z}_n)}{\partial\bar{x}_n}\Big|_{(\bar{Y},\bar{X},\bar{X},\bar{Z})}$,
and $J_4=\frac{\partial\mathcal{G}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n,\bar{z}_n)}{\partial\bar{z}_n}\Big|_{(\bar{Y},\bar{X},\bar{X},\bar{Z})}$, if $\bar{X}$ is unknown
$K_1=\frac{\partial\mathcal{F}(\bar{y}_m^\star,\bar{x}_m^\star,\bar{x}_n)}{\partial\bar{y}_m^\star}\Big|_{(\bar{Y},\bar{X},\bar{X})}$, $K_2=\frac{\partial\mathcal{F}(\bar{y}_m^\star,\bar{x}_m^\star,\bar{x}_n)}{\partial\bar{x}_m^\star}\Big|_{(\bar{Y},\bar{X},\bar{X})}$, $K_3=\frac{\partial\mathcal{F}(\bar{y}_m^\star,\bar{x}_m^\star,\bar{x}_n)}{\partial\bar{x}_n}\Big|_{(\bar{Y},\bar{X},\bar{X})}$ if $\bar{X}$ is unknown and
$S_1=\frac{\partial\mathcal{G}(\bar{y}_m^\star,\bar{x}_m^\star,\bar{x}_n,\bar{z}_n)}{\partial\bar{y}_m^\star}\Big|_{(\bar{Y},\bar{X},\bar{X},\bar{Z})}$, $S_2=\frac{\partial\mathcal{G}(\bar{y}_m^\star,\bar{x}_m^\star,\bar{x}_n,\bar{z}_n)}{\partial\bar{x}_m^\star}\Big|_{(\bar{Y},\bar{X},\bar{X},\bar{Z})}$, $S_3=\frac{\partial\mathcal{G}(\bar{y}_m^\star,\bar{x}_m^\star,\bar{x}_n,\bar{z}_n)}{\partial\bar{x}_n}\Big|_{(\bar{Y},\bar{X},\bar{X},\bar{Z})}$,
$S_4=\frac{\partial\mathcal{G}(\bar{y}_m^\star,\bar{x}_m^\star,\bar{x}_n,\bar{z}_n)}{\partial\bar{z}_n}\Big|_{(\bar{Y},\bar{X},\bar{X},\bar{Z})}$ if $\bar{X}$ is unknown
$J_2=-J_3$, $K_2=-K_3$ and $S_2=-S_3$ (under the assumed imposed constraints) and J_{11},J_{22},J_{33} $J_{44},J_{12},J_{13},J_{14},J_{23},J_{24},J_{34}$ are the second order partial derivative of $\mathcal{G}(\bar{y}_m^\star,\bar{x}_m,\bar{x}_n,\bar{z}_n)$ at the point $(\bar{Y},\bar{X},\bar{X},\bar{Z})$, $K_{11},K_{22},K_{33},K_{12},K_{13},K_{23}$ are the second order partial derivative of $\varphi(\bar{y}_m^\star,\bar{x}_m^\star,\bar{x}_n)$ at the point $(\bar{Y},\bar{X},\bar{X})$ and $S_{11},S_{22},S_{33},S_{44},S_{12},S_{13},S_{14},S_{23},S_{24},S_{34}$ are the second order partial derivative of $\Psi(\bar{y}_m^\star,\bar{x}_m^\star,\bar{x}_n,\bar{z}_n)$ at the point $(\bar{Y},\bar{X},\bar{X},\bar{Z})$.
By considering the expectation on both sides of equations (3.1) - (3.4) and substituting the expected values of ϱ_i for $(i=0,1,2,3,4)$, the bias B(.) and the mean square error M(.) of the estimator classes up to the first order of approximations as follows.

$$B(\tau_u)=\frac{1}{2}\bar{X}^2C_x^2\{f_1H_{22}+f_2(H_{33}+2H_{23})\}+\bar{Y}\bar{X}\rho_{yx}C_yC_x(f_1H_{12}+f_2H_{13}) \tag{3.6}$$

$$\begin{aligned}B(\tau_m) &= \frac{1}{2}\bar{X}^2C_x^2\{f_1J_{22}+f_2(J_{33}+2J_{23})\}+\frac{1}{2}\bar{Z}^2f_2C_z^2J_{44}+\bar{Y}\bar{X}\rho_{yx}C_yC_x(f_1J_{12}+f_2J_{13})\\ &+f_2\bar{Y}\bar{Z}\rho_{yz}C_yC_zJ_{14}+f_2\bar{X}\bar{Z}\rho_{xz}C_xC_z(J_{24}+J_{34})\end{aligned} \tag{3.7}$$

$$B(\tau_u^\star) = \frac{1}{2}\bar{X}^2 C_x^2\{f_1^\star K_{22} + f_2(K_{33} + 2K_{23})\} + \bar{Y}\bar{X}\rho_{yx}C_yC_x(f_1^\star K_{12} + f_2K_{13}) \quad (3.8)$$

$$\begin{aligned} B(\tau_m^\star) &= \frac{1}{2}\bar{X}^2 C_x^2\{f_1^\star S_{22} + f_2(S_{33} + 2S_{23})\} + \bar{Y}\bar{X}(f_1^\star S_{12} + f_2S_{13})\rho_{yx}C_yC_x + f_2\bar{Y}\bar{Z}\rho_{yz}C_yC_zS_{14} \\ &\quad + f_2\bar{X}\bar{Z}(S_{24} + S_{34})\rho_{xz}C_xC_z + \frac{1}{2}\bar{Z}^2 f_2C_z^2S_{44} \end{aligned} \quad (3.9)$$

$$M(\tau_u) = \bar{Y}^2 f_1^\star C_y^2 + \bar{X}^2 f_3C_x^2H_2^2 + 2\bar{Y}\bar{X}f_3\rho_{yx}C_yC_xH_2 \quad (3.10)$$

$$M(\tau_m) = \bar{Y}^2 f_1^\star C_y^2 + \bar{X}^2 f_3C_x^2J_2^2 + \bar{Z}^2 f_2C_z^2J_4^2 + 2\bar{Y}\bar{X}f_3\rho_{yx}C_yC_xJ_2 + 2\bar{Y}\bar{Z}f_2\rho_{yz}C_yC_zJ_4 \quad (3.11)$$

$$M(\tau_u^\star) = \bar{Y}^2 f_1^\star C_y^2 + \bar{X}^2 C_x^2(f_1^\star - f_2)K_2^2 + 2\bar{Y}\bar{X}(f_1^\star - f_2)\rho_{yx}C_yC_xK_2 \quad (3.12)$$

$$\begin{aligned} M(\tau_m^\star) &= \bar{Y}^2 f_1^\star C_y^2 + \bar{X}^2 C_x^2(f_1^\star - f_2)S_2^2 + 2\bar{Y}\bar{X}(f_1^\star - f_2)\rho_{yx}C_xC_yS_2 \\ &\quad + \bar{Z}^2 f_2C_z^2S_4^2 + 2\bar{Y}\bar{Z}f_2\rho_{yz}C_yC_zS_4 \end{aligned} \quad (3.13)$$

3.2. Bias and MSE of Estimators under case II

$$E(\varrho_0^2) = f_1^\star C_y^2, \quad E(\varrho_1^2) = f_1C_x^2, \quad E(\varrho_2^2) = f_2C_x^2, \quad E(\varrho_3^2) = f_2C_z^2, \quad E(\varrho_4^2) = f_1^\star C_x^2,$$

$$E(\varrho_0\varrho_1) = f_1\rho_{yx}C_yC_x, \quad E(\varrho_0\varrho_4) = f_1^\star\rho_{yx}C_yC_x, \quad E(\varrho_1\varrho_4) = f_1C_x^2, \quad E(\varrho_2\varrho_3) = f_2\rho_{xz}C_xC_z,$$

$$E(\varrho_0\varrho_2) = 0, \quad E(\varrho_0\varrho_3) = 0, \quad E(\varrho_1\varrho_2) = 0, \quad E(\varrho_1\varrho_3) = 0, \quad E(\varrho_2\varrho_4) = 0, \quad E(\varrho_3\varrho_4) = 0$$

By applying the approach from this section 3.1 and the results in section 3.2, we obtain the bias B(.) and mean square error M(.) expressions for the proposed estimator classes τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$, accurate up to the first order of approximations as presented below. Under Case II, where the first-phase and second-phase samples are drawn independently, the estimators $\bar{x}_n$ and $\bar{y}_m^\star$ are based on independent random samples. Hence, their associated error terms are uncorrelated, implying $E(\varrho_0\varrho_2) = Cov(\varrho_0, \varrho_2) = 0$. Similar arguments apply to other cross-product expectations involving variables from different phases. This result follows directly from the independence of sampling designs and is consistent with standard two-phase sampling theory. These derivations highlight the improved performance and accuracy offered by our estimator classes compared to existing methods,

$$B(\tau_u) = \frac{1}{2}\bar{X}^2 C_x^2(f_1H_{22} + f_2H_{33}) + f_1\bar{Y}\bar{X}\rho_{yx}C_yC_xH_{12} \quad (3.14)$$

$$B(\tau_m) = \frac{1}{2}\bar{X}^2 C_x^2(f_1J_{22} + f_2J_{33}) + \frac{1}{2}\bar{Z}^2 f_2C_z^2J_{44} + \bar{X}C_x(f_1\bar{Y}\rho_{yx}C_yJ_{12} + f_2\bar{Z}\rho_{xz}C_zJ_{34}) \quad (3.15)$$

$$B(\tau_u^\star) = \frac{1}{2}\bar{X}^2 C_x^2(f_1^\star K_{22} + f_2K_{33}) + f_1^\star\bar{Y}\bar{X}\rho_{yx}C_yC_xK_{12} \quad (3.16)$$

$$B(\tau_m^\star) = \tfrac{1}{2}\bar{X}^2 C_x^2(f_1^\star S_{22} + f_2S_{33}) + \bar{X}C_x(f_1\bar{Y}\rho_{yx}C_yS_{12} + f_2\bar{Z}\rho_{xz}C_zS_{34}) + \tfrac{1}{2}\bar{Z}^2 f_2C_z^2S_{44} \quad (3.17)$$

$$M(\tau_u) = \bar{Y}^2 f_1^\star C_y^2 + \bar{X}^2 C_x^2(f_1 + f_2)H_2^2 + 2\bar{Y}\bar{X}f_1\rho_{yx}C_yC_xH_2 \quad (3.18)$$

$$M(\tau_m) = \bar{Y}^2 f_1^\star C_y^2 + \bar{X}^2 C_x^2 (f_1 + f_2) J_2^2 + \bar{Z}^2 f_2 C_z^2 J_4^2 + 2\bar{Y}\bar{X} f_1 \rho_{yx} C_y C_x J_2 - 2\bar{X}\bar{Z} f_2 \rho_{xz} C_x C_z J_4 J_2 \quad (3.19)$$

$$M(\tau_u^\star) = \bar{Y}^2 f_1^\star C_y^2 + \bar{X}^2 C_x^2 (f_1^\star + f_2) K_2^2 + 2\bar{Y}\bar{X} f_1^\star \rho_{yx} C_y C_x K_2 \quad (3.20)$$

$$\begin{aligned} M(\tau_m^\star) &= \bar{Y}^2 f_1^\star C_y^2 + \bar{X}^2 C_x^2 (f_1^\star + f_2) S_2^2 + 2\bar{Y}\bar{X} f_1^\star \rho_{yx} C_x C_y S_2 \\ &\quad + \bar{Z}^2 f_2 C_z^2 S_4^2 - 2\bar{X}\bar{Z} f_2 \rho_{xz} C_x C_z S_2 S_4 \end{aligned} \quad (3.21)$$

Remark 3.1:(i)The bias and mean square errors of the various estimators described in Section 2.3, which are members of the proposed classes, can be derived by substituting the relevant derivative values. Use equations (3.5) - (3.12) and (3.13) - (3.20).
(ii) When p = 0, indicating the absence of non-response, the expected values of sample statistics affected by random non-response are equivalent to the standard expressions derived under complete response conditions.

4. Minimum MSE of Estimators τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$

Equations (3.9)–(3.12) and (3.17)–(3.20) and remarks 3.1 indicate that the mean square errors of the proposed estimator classes depend on the derivative terms H_2, J_2, J_4, K_2, S_2 and S_4. The aim is to determine the optimal values of these derivatives to minimize the mean square errors of the estimator classes τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$ for each of the two distinct cases within the two-phase sampling framework considered in this study.

4.1. Minimum MSE under case I

When the second phase sample S of size m is selected as a subsample from of the first phase sample S' of size n, the conditions that ensure minimum mean square error for the proposed classes of estimators are derived as follows,

$$H_2 = K_2 = -\rho_{yx}\frac{\bar{Y}}{\bar{X}}\frac{C_y}{C_x}, \qquad J_2 = S_2 = -\rho_{yx}\frac{\bar{Y}}{\bar{X}}\frac{C_y}{C_x}, \qquad J_4 = S_4 = -\rho_{yx}\frac{\bar{Y}}{\bar{Z}}\frac{C_y}{C_z} \quad (4.1)$$

Substituting the derivatives of equation (4.1) in equations (3.9)-(3.12), which yield the minimum mean square errors of the proposed estimator classes τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$ as

$$min.M(\tau_u) = \bar{Y}^2 C_y^2 (f_1^\star - f_3 \rho_{yx}^2) \quad (4.2)$$

$$min.M(\tau_m) = \bar{Y}^2 C_y^2 (f_1^\star - f_2 \rho_{yz}^2 - f_3 \rho_{yx}^2) \quad (4.3)$$

$$min.M(\tau_u^\star) = \bar{Y}^2 C_y^2 \{f_1^\star - (f_1^\star - f_2)\rho_{yx}^2\} \quad (4.4)$$

and

$$min.M(\tau_m^\star) = \bar{Y}^2 C_y^2 \{f_1^\star - (f_1^\star - f_2)\rho_{yx}^2 - f_2 \rho_{yz}^2\} \quad (4.5)$$

Remark : Before obtaining the estimators of mean suqare errors that shown in equation (4.2) - (4.5) we need the following lemma.
Lemma 1: The maximum likelihood estimator of the non-response probability 'p' and the response probability 'q' are given by

$$p = \frac{(n + r - 1) - \sqrt{(n + r - 1)^2 - \frac{4nr(n-3)}{(n-2)}}}{2(n-3)} \quad (4.6)$$

and q = (1-p)
The lemma is established by equating $\frac{\partial log_{p_r}}{\partial p}$ to zero. Solving the resulting quadratic equation in p yields a single admissible solution within the interval $0 \leq p \leq 1$ as presented in equation (4.6).

An estimator of the equations (4.2)-(4.5) is given by

$$min.M(\tau_u) = \bar{Y}^2 c_y^{\star 2}(f_1^\star - f_3\rho_{yx}^2) \tag{4.7}$$

$$min.M(\tau_m) = \bar{Y}^2 c_y^{\star 2}(f_1^\star - f_2\rho_{yz}^2 - f_3\rho_{yx}^2) \tag{4.8}$$

$$min.M(\tau_u^\star) = \bar{Y}^2 c_y^{\star 2}\{f_1^\star - (f_1^\star - f_2)\rho_{yx}^2\} \tag{4.9}$$

and

$$min.M(\tau_m^\star) = \bar{Y}^2 c_y^{\star 2}\{f_1^\star - (f_1^\star - f_2)\rho_{yx}^2 - f_2\rho_{yz}^2\} \tag{4.10}$$

4.2. Minimum MSE of Proposed strategies under case II

When the second phase sample S is chosen independently using SRSWOR from the first phase sample $S^{'}$, the conditions require to achieve minimum mean squared error for the proposed estimator classes are derived as follows,

$$\begin{aligned} H_2 &= -\rho_{yx}\frac{f_1}{(f_1+f_2)}\frac{\bar{Y}}{\bar{X}}\frac{C_y}{C_x}, \quad J_2 = \frac{f_1\bar{Y}\rho_{yx}C_y}{f_1\bar{X}C_x + f_2\bar{X}C_x(1-\rho_{xz}^2)}, \quad J_4 = -f_1\rho_{yx}\rho_{xz}\frac{\bar{Y}}{\bar{Z}}\frac{C_y}{C_z}\frac{1}{f_1+f_2(1-\rho_{xz}^2)}, \\ K_2 &= -\rho_{yx}\frac{f_1^\star}{(f_1+f_2)}\frac{\bar{Y}}{\bar{X}}\frac{C_y}{C_x}, \quad S_2 = \frac{f_1^\star\bar{Y}\rho_{yx}C_y}{f_1^\star\bar{X}C_x + f_2\bar{X}C_x(1-\rho_{xz}^2)}, \quad S_4 = -f_1\rho_{yx}\rho_{xz}\frac{\bar{Y}}{\bar{Z}}\frac{C_y}{C_z}\frac{1}{f_1^\star+f_2(1-\rho_{xz}^2)} \end{aligned} \tag{4.11}$$

Substituting these optimum values of the derivative in equation (4.11) in equations (3.17)-(3.20), we have the expression of minimum mean square errors of the proposed classes of estimators as.

$$min.M(\tau_u) = \bar{Y}^2 C_y^2\left(f_1^\star - \frac{f_1^2}{(f_1+f_2)}\rho_{yx}^2\right) \tag{4.12}$$

$$min.M(\tau_m) = \bar{Y}^2 C_y^2\left[f_1^\star + \frac{f_1^2\rho_{yx}^2}{f_1+f_2(1-\rho_{xz}^2)}\left(\frac{(f_1+f_2)}{f_1+f_2(1-\rho_{xz}^2)} - \frac{f_2\rho_{xz}^2}{f_1+f_2(1-\rho_{xz}^2)} - 2\right)\right] \tag{4.13}$$

$$min.M(\tau_u^\star) = \bar{Y}^2 C_y^2\left(f_1^\star - \frac{f_1^{\star 2}}{(f_1^\star+f_2)}\rho_{yx}^2\right) \tag{4.14}$$

and

$$min.M(\tau_m^\star) = \bar{Y}^2 C_y^2\left[f_1^\star + \frac{f_1^{\star 2}\rho_{yx}^2}{f_1^\star+f_2(1-\rho_{xz}^2)}\left(\frac{(f_1^\star+f_2)}{f_1^\star+f_2(1-\rho_{xz}^2)} - \frac{f_2\rho_{xz}^2}{f_1^\star+f_2(1-\rho_{xz}^2)} - 2\right)\right] \tag{4.15}$$

An estimator of the equations (4.12)-(4.15) is given by

$$min.M(\tau_u) = \bar{Y}^2 c_y^{\star 2}\left(f_1^\star - \frac{f_1^2}{(f_1+f_2)}\rho_{yx}^2\right) \tag{4.16}$$

$$min.M(\tau_m) = \bar{Y}^2 c_y^{\star 2}\left[f_1^\star + \frac{f_1^2\rho_{yx}^2}{f_1+f_2(1-\rho_{xz}^2)}\left(\frac{(f_1+f_2)}{f_1+f_2(1-\rho_{xz}^2)} - \frac{f_2\rho_{xz}^2}{f_1+f_2(1-\rho_{xz}^2)} - 2\right)\right] \tag{4.17}$$

$$min.M(\tau_u^\star) = \bar{Y}^2 c_y^{\star 2}\left(f_1^\star - \frac{f_1^{\star 2}}{(f_1^\star + f_2)}\rho_{yx}^2\right) \tag{4.18}$$

and

$$min.M(\tau_m^\star) = \bar{Y}^2 c_y^{\star 2}\left[f_1^\star + \frac{f_1^{\star 2}\rho_{yx}^2}{f_1^\star + f_2(1-\rho_{xz}^2)}\left(\frac{(f_1^\star + f_2)}{f_1^\star + f_2(1-\rho_{xz}^2)} - \frac{f_2\rho_{xz}^2}{f_1^\star + f_2(1-\rho_{xz}^2)} - 2\right)\right] \tag{4.19}$$

where $f_1^\star = \left(\frac{1}{mq+2p} - \frac{1}{N}\right)$, $c_y^\star = \frac{s_y^\star}{\bar{y}^\star}$, $c_x^\star = \frac{s_x^\star}{\bar{x}^\star}$, $c_z^\star = \frac{s_z^\star}{\bar{z}^\star}$ are sample coefficient of variation of y, x, and z respectively and $\rho_{yx}^\star$, $\rho_{yz}^\star$ and $\rho_{xz}^\star$ are correlation coefficients between the variables y and x, y and z, x and z respectively.

5. Efficiency Comparisons of the Proposed Methods τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$

To assess the performance of the proposed estimator classes τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$ under the two distinct cases of the two-phase sampling framework considered in this study their efficiencies are compared with that of conventional estimators of the population mean including $\bar{y}^\star$, the sample mean estimator in the presence of random non-response.
The mean squared error of this estimator has been derived up to the first order approximations for cases I and II of the two phase sampling design and the corresponding expressions are provided in the equations below.

$$V(\bar{y}^\star) = f_1^\star \bar{Y}^2 C_y^2 \tag{5.1}$$

and

$$V(\bar{y}^\star) = f_1^\star \bar{Y}^2 C_y^2 \tag{5.2}$$

5.1. Efficiency Comparisons of the Estimators τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$

In this section compares the efficiencies of the proposed estimator classes τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$ under their respective optimality criteria to the the existing estimator $\bar{y}^\star$. In the second phase sample of size n, the comparison is done for cases where non-response occurs either on the study variable alone or on both the study variable y and the auxiliary variable x.
Case I
(a) Comparisons of effectiveness of estimaters τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$
It can be observed from the equation that (4.2) - (4.5), and (5.1) that
(i) τ_u is more effective than $\bar{y}^\star$, given that $V(\bar{y}^\star) \geq min.M(\tau_u)$

$$f_3\rho_{yx}^2 \geq 0 \tag{5.3}$$

It can be inferred from equation (4.2) that the estimator class τ_u consistently outperforms $\bar{y}^\star$ since $(f_1, f_2, f_3 > 0)$.
Similarly,
(ii) τ_m is more accurate than $\bar{y}^\star$, given by $V(\bar{y}^\star) \geq min.M(\tau_m)$

$$\frac{f_2}{f_3} \geq -\frac{\rho_{yx}^2}{\rho_{yz}^2} \tag{5.4}$$

(iii) $\tau_u^\star$ is more effective than $\bar{y}^\star$, given that $V(\bar{y}^\star) \geq min.M(\tau_u^\star)$

$$(f_1^\star - f_2)\rho_{yx}^2 \geq 0 \tag{5.5}$$

(iv) $\tau_m^\star$ is more accurate than $\bar{y}^\star$, given by $V(\bar{y}^\star) \geq min.M(\tau_m^\star)$

$$\frac{f_1^\star}{f_2} \geq \frac{\rho_{yx}^2 - \rho_{yz}^2}{\rho_{yz}^2} \tag{5.6}$$

Case II
(a) Comparisons of effectiveness of estimators τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$
It could be concluded from the equations that (4.12) - (4.15), and (5.2) that
(i) τ_u is more effective than $\bar{y}^\star$, given that $V(\bar{y}^\star) \geq min.M(\tau_u)$

$$\frac{f_1^2}{(f_1 + f_2)}\rho_{yx}^2 \geq 0 \tag{5.7}$$

(ii) τ_m is more accurate than $\bar{y}^\star$, given by $V(\bar{y}^\star) \geq min.M(\tau_m)$

$$\frac{f_1^2\rho_{yx}^2}{\mathcal{A}}\left(2 + \frac{f_2\rho_{xz}^2}{\mathcal{A}} - \frac{(f_1 + f_2)}{\mathcal{A}}\right) \geq 0 \tag{5.8}$$

(iii) $\tau_u^\star$ is more effective than $\bar{y}^\star$, given that $V(\bar{y}^\star) \geq min.M(\tau_u^\star)$

$$\frac{f_1^\star}{(f_1^\star + f_2)}\rho_{yx}^2 \geq 0 \tag{5.9}$$

(iv) $\tau_m^\star$ is more accurate than $\bar{y}^\star$, given by $V(\bar{y}^\star) \geq min.M(\tau_m^\star)$

$$\frac{f_1^{\star 2}\rho_{yx}^2}{\mathcal{B}}\left(2 + \frac{f_2\rho_{xz}^2}{\mathcal{B}} - \frac{(f_1^\star + f_2)}{\mathcal{B}}\right) \geq 0 \tag{5.10}$$

where $\mathcal{A} = \frac{1}{f_1+f_2(1-\rho_{xz}^2)}$, $\mathcal{B} = \frac{1}{f_1^\star+f_2(1-\rho_{xz}^2)}$

6. Numerical Illustration

To demonstrates the effectiveness of the proposed estimator classes τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$, four natural population datasets have been selected. The sources of these populations, the descriptions of the variables y, x, and z along with the corresponding parameter values are provided below:
Population I: Source: Shukla (1966) [21]
Y: fibber
X: height
Z: base diameter
$N = 50$, $n = 25$, $m = 20$, $\bar{Y} = 2.5840$, $C_y^2 = 0.0866$, $C_x^2 = 0.1163$, $C_z^2 = 0.0170$, $\rho_{yx} = 0.4800$, $\rho_{yz} = 0.3700$, $\rho_{xz} = 7300$

Population II- Source: Fisher (1936) [5]
Y: sepal width
X: sepal length
Z: petal length
$N = 34$, $n = 15$, $m = 10$, $\bar{Y} = 2.770$, $C_y^2 = 0.012566$, $C_x^2 = 0.007343$, $C_z^2 = 0.011924$, $\rho_{yx} = 0.5605$, $\rho_{yz} = 0.5259$, $\rho_{xz} = 0.7540$

Population III - Source: Handique et al. (2011) [8]
Y: forest timber volume in cubic meter (Cum) in 0.1 ha sample plot
X: average tree height in the sample plot in meter (m)
Z: average crown diameter in the sample plot in meter (m)
$N = 2500$, $n = 200$, $m = 25$, $\bar{Y} = 4.63$, $\bar{X} = 21.09$, $\bar{Z} = 13.55$, $C_y^2 = 0.9025$, $C_x^2 = 0.9604$, $C_z^2 = 0.4096$, $\rho_{yx} = 0.79$, $\rho_{yz} = 0.72$, $\rho_{xz} = 0.66$

Population IV - Source: Sukhatme and Chand (1977) [30]
Y: Apple trees age in 1964.
X: Bushels of apples harvested in 1964.
Z: Bushels of apples harvested in 1959.

$N = 200$, $n = 30$, $m = 20$, $\bar{Y} = 0.103182 \times 10^4$, $\bar{X} = 0.293458 \times 10^4$, $\bar{Z} = 0.365149 \times 10^4$, $C_y^2 = 2.55280$, $C_x^2 = 4.02504$, $C_z^2 = 2.09379$, $\rho_{yx} = 0.93$, $\rho_{yz} = 0.77$, $\rho_{xz} = 0.84$

To obtain a clear understanding of the performance of the proposed estimator classes τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$ the percent relative efficiencies (PRE's) of these estimators evaluated under their respective optimality conditions have been computed for various levels of the non-response rate p.
The results are presented in tables 1 and 2, where the PREs of the proposed estimator classes are calculated relative to the sample mean estimator $\bar{y}^\star$.

$$PRE = \frac{V(\bar{y}^\star)}{M(\tau)} \tag{6.1}$$

Table 1: **Percent relative efficiencies of the class of estimators τ_u and τ_m with respect to estimator $\bar{y}^\star$ when non-response situations occur only on the study variable y at the second phase sample.**

Estimators	Population I			
	Case I		Case II	
	τ_u	τ_m	τ_u	τ_m
P=0.02	108.0524	119.4864	115.4924	120.5519
P=0.04	107.7935	118.7957	114.9610	119.8173
P=0.06	107.5417	118.1288	114.4467	119.1082
P=0.08	107.2968	117.4843	113.9486	118.4235
P=0.10	107.0586	116.8612	113.4659	117.7618
	Population II			
Estimators	Case I		Case II	
	τ_u	τ_m	τ_u	τ_m
P=0.02	116.9608	140.3892	125.1568	133.3557
P=0.04	116.5116	139.1119	124.4451	132.3522
P=0.06	116.0717	137.8740	123.7506	131.3764
P=0.08	115.6406	136.6735	123.0727	130.4274
P=0.10	115.2183	135.5089	122.4109	129.5039
	Population III			
Estimators	Case I		Case II	
	τ_u	τ_m	τ_u	τ_m
P=0.02	218.0328	250.2832	221.6106	235.1927
P=0.04	213.2683	243.3609	216.6245	229.3247
P=0.06	208.7093	236.8136	211.8595	223.7444
P=0.08	204.3426	230.6117	207.3013	218.4312
P=0.10	200.1565	224.7284	202.9367	213.3663
	Population IV			
Estimators	Case I		Case II	
	τ_u	τ_m	τ_u	τ_m
P=0.02	145.7597	312.3005	208.3944	350.9832
P=0.04	144.4191	299.4063	203.9107	333.9793
P=0.06	143.1082	287.5799	199.6327	318.6053
P=0.08	141.8259	276.6940	195.5464	304.6377
P=0.10	140.5713	266.6407	191.6392	291.8919

Table 2: Percent relative efficiencies of the class of estimators $\tau_u^\star$ and $\tau_m^\star$ with respect to estimator $\bar{y}^\star$ when non-response situations occur only on the study variable y as well as auxiliary variable x at the second phase sample.

	Population I			
Estimators	Case I		Case II	
	$\tau_u^\star$	$\tau_m^\star$	$\tau_u^\star$	$\tau_m^\star$
P=0.02	108.8558	120.4695	116.2654	121.4972
P=0.04	109.3853	120.7320	116.4901	121.6795
P=0.06	109.9075	120.9895	116.7156	121.8607
P=0.08	110.4225	121.2421	116.9420	122.0406
P=0.10	110.9306	121.4901	117.1693	122.2194
	Population II			
Estimators	Case I		Case II	
	$\tau_u^\star$	$\tau_m^\star$	$\tau_u^\star$	$\tau_m^\star$
P=0.02	117.9366	141.7973	126.1422	134.5824
P=0.04	118.4513	141.8860	126.3999	134.7761
P=0.06	118.9637	141.9737	126.6592	134.9694
P=0.08	119.4738	142.0603	126.9201	135.1623
P=0.10	119.9816	142.1459	127.1828	135.3549
	Population III			
Estimators	Case I		Case II	
	$\tau_u^\star$	$\tau_m^\star$	$\tau_u^\star$	$\tau_m^\star$
P=0.02	223.6890	257.7651	227.3926	241.7625
P=0.04	224.3648	257.9167	227.9561	242.1558
P=0.06	225.0445	258.0685	228.5245	242.5510
P=0.08	225.7281	258.2203	229.0977	242.9484
P=0.10	226.4156	258.3723	229.6759	243.3478
	Population IV			
Estimators	Case I		Case II	
	$\tau_u^\star$	$\tau_m^\star$	$\tau_u^\star$	$\tau_m^\star$
P=0.02	149.5222	330.0974	214.9885	373.0432
P=0.04	151.9824	333.8495	216.9345	376.2570
P=0.06	154.5147	337.6724	218.9394	379.5343
P=0.08	157.5147	341.5683	221.0058	382.8767
P=0.10	159.8084	345.5390	223.1367	386.2863

7. Interpretation from Generalized Empirical Study

1) Table 1 presents the percent relative efficiencies (PRE) of the two estimator classes, τ_u and τ_m, in comparison with the conventional estimator $\bar{y}^*$ under non-response conditions for the study variable y during second-phase sampling. The results encompass four populations and two non-response scenarios (Case I and Case II), with non-response rates p ranging from 0.02 to 0.10.

a) For all populations and in both cases, the relative efficiency of all estimators decreases slightly as the non-response rate p increases. This is expected. Increasing non-responses generally reduces the accuracy of the estimators.

b) Estimator τ_m consistently outperforms τ_u in terms of efficiency across all populations, cases, and non-response levels, indicating that τ_m is more robust and efficient in handling non-responses to the study variable. Across most populations, Case II yielded slightly higher efficiency than Case I.

2) Table 2 shows the per cent relative efficiencies (PRE) of two estimator classes, $\tau_u^\star$ and $\tau_m^\star$, compared to the conventional estimator $\bar{y}^\star$. The results span four populations and two non-response scenarios (Case

I and II). They show how the PREs for both classes change with second-phase sampling non-response for y and with non-response rates p from 0.02 to 0.10. Both $\tau_u^\star$ and $\tau_m^\star$ are more efficient than $\bar{y}^\star$ under non-response, demonstrating their clear advantage in maintaining higher efficiency when non-response occurs.

a) For all populations and in both cases, the relative efficiency of all estimators increased slightly as the non-response rate p increased.

b) The estimator $\tau_m^\star$ consistently outperforms $\tau_u^\star$ in terms of efficiency across all populations, cases, and non-response levels, indicating that $\tau_m^\star$ is more robust and efficient in handling non-responses to the study variable. Across most populations, Case II results in slightly higher efficiencies compared to Case I.

8. Conclusions

This study develops broad, flexible classes of estimators for the population mean under a two-phase sampling framework, with random non-response governed by a missing-at-random (MAR) mechanism. Unlike conventional approaches that assume complete response, the proposed methodology explicitly accounts for non-response in the second phase and utilises information from two auxiliary variables to improve estimation efficiency.

General classes of estimators τ_u, τ_m, $\tau_u^\star$ and $\tau_m^\star$ were formulated to address two practical situations: (i) non-response affecting only the study variable y and (ii) non-response affecting both the study and auxiliary variables. Theoretical expressions for bias and mean square error were derived up to the first-order approximation under both sub-sampling and independent second-phase sampling schemes. Optimality conditions were obtained by minimising the mean-squared errors with respect to the involved derivative parameters, yielding efficient estimators within each class.

The proposed classes are sufficiently general to include ratio, product, regression, and exponential-type estimators as special cases. Analytical results demonstrate that the incorporation of auxiliary information substantially reduces mean square error compared to the usual sample mean estimator in the presence of non-response.

Acknowledgment

The authors sincerely thank the reviewers for their valuable comments and constructive suggestions, which have significantly improved the quality and clarity of the manuscript.

References

1. Basit, Z., Masood, S., & Bhatti, I. (2023). A class of estimators for estimation of population mean under random non-response in two phase successive sampling. Journal of Statistical Theory and Applications, 22(4), 309-338.

2. Chand, L. (1975), Some ratio-type estimators based on two or more auxiliary variables. Ph.D., dissertation, Iowa State University, Ames, Iwoa.

3. Choudhury, S., & Singh, B. K. (2012). A class of chain ratio–product type estimators with two auxiliary variables under double sampling scheme. Journal of the Korean Statistical Society, 41(2), 247-256.

4. Cochran, W. G. (1977). Sampling techniques. john wiley & sons.

5. Fisher, R. A. (1936), The use of multiple measurements in taxonomic problems, Annals of Eugenics, 7,179-188.

6. Grover, L. K., & Sharma, A. (2023). Predictive estimation of finite population mean in case of missing data under two-phase sampling. Journal of Statistical Theory and Applications, 22(4), 283-308.

7. Hamad, N., Hanif, M., Haider, H. (2013), A regression type estimator with two auxiliary variables for two-phase sampling. Open J. Stat. 3, 74–78

8. Handique, B. K., Das, G., Kalita, M. C. (2011), Comparative studies on forest sampling techniques with satellite remote sensing inputs, Project Report, NESAC, 12–40.

9. Hansen, M.H., Hurwitz, W.N. (1946), The problem of non-response in sample surveys. J. Am. Stat. Assoc. 41(236), 517–529

10. Heitzan, D.F., Basu, S. (1996): Distinguish 'missing at random' and 'missing completely at random. Amer. Stat. 50:207–217.

11. Kiregyera, B. (1980). A chain ratio-type estimator in finite population double sampling using two auxiliary variables, Metrika, 27, (217-223).

12. Liu, X., & Arslan, M. (2021). A general class of estimators on estimating population mean using the auxiliary proportions under simple and two phase sampling. AIMS Mathematics, 6(12), 13592-13607.

13. Malik, S., Singh, R. (2015), Estimation of population mean using information on auxiliary attribute in two-phase sampling. Appl. Math. Comput. 261, 114–118

14. Mehta, P., Tailor, R. (2020). Chain ratio type estimators using known parameters of auxiliary variates in double sampling. J. Reliab. Stat. Stud. 13(2–4), 243–252.

15. Murthy, M.N. (1967), Sampling theory and methods. Statistical Publishing Society, Calcutta, India

16. Rubin, D. B. (1976), Inference and missing data, Biometrika, 63(3), (581-592).

17. Sahoo, J., Sahoo, L. N., & Mohanty, S. (1993). A regression approach to estimation in two-phase sampling using two auxiliary variables. Current Science, 73-75.

18. Samiuddin, M., & Hanif, M. (2007). Estimation of population mean in single and two phase sampling with or without additional information. PAKISTAN JOURNAL OF STATISTICS-ALL SERIES-, 23(2), 99.

19. Sharma, A. K. (2017), An effective imputation method to minimize the effect of random non-response in estimation of population variance on successive occasions, Kuwait Journal of Science, 44(4),91-97.

20. Sharma, A. K. and Singh, G. N.(2018), Efficient Class of Estimators of Population Variance Under Random Non-response in Two-Phase Successive Sampling, Proceedings of the National Academy of Sciences, India Section A: Physical Sciences, 88(2),171-180.

21. Shukla, G. K. (1966), An alternative multivariate ratio estimate for finite population, Calcutta Statistical Association Bulletin, 15,127-134.

22. Singh, G. N. and Sharma, A. K. (2014), An improved estimation procedure of population mean in two phase sampling, Journal of International Academy of Physical Sciences, 18(1), (1-10).

23. Singh, G. N. and Sharma, A. K. (2015), Some improved estimators for population mean in double sampling, International Journal of Mathematics & Computation, 26(3), (36-45).

24. Singh, G. N., Bandyopadhyay, A. and Sharma, A. K. (2017), Effectual variance estimation strategy in two-occasion successive sampling in presence of random non response, Communications in Statistics-Theory and Methods, 46(14), (7201-7224).

25. Singh, H. P. and Tracy, D. S. (2001), Estimation of population mean in presence of random non-response in sample surveys, Statistica, LXI, (231-248).

26. Singh, S. and Joarder, A. H. (1998), Estimation of finite population variance using random non-response in survey sampling, Metrika, 47, (241-249).

27. Singh, S., Joarder, A. H. and Tracy, D. S. (2000), Regression type estimators for random non-response in survey sampling, Statistica, LX(1), (39-43).

28. Srivastava, S. K. and Jhajj, H. S. (1981), A class of estimators of population mean in survey sampling using auxiliary information, Biometrika, 68, (341-343).

29. Sukhatme, B.V. Some ratio-type estimators in two-phase sampling. J. Am. Stat. Assoc. 1962, 57, 628–632.

30. Sukhatme, B. V. and Chand, L. (1977), Multivariate ratio-type estimators, Proceedings of American Statistical Association, Social Statistics Section, 927–931.

31. Tracy, D.S. and Osahan, S. S. (1994), Random non-response on study variable versus on study as well as auxiliary variables, Statistica, 54, (163-168).

32. Zaman, T., Kadilar, C. (2021), Exponential ratio and product type estimators of the mean in stratified two phase sampling. AIMS Math. 6(5), 4265–4279

Swapnil V Shinkar,
Department of Mathematics,
National Institute of Technology, Raipur, India.
E-mail address: `shinkars2008@gmail.com`

and

Anup Kumar Sharma, Department of Mathematics
Department of Mathematics,
National Institute of Technology, Raipur, India.

E-mail address: aksharma.ism@gmail.com

and

Awadhesh K Pandey,
Jindal School of Banking and Finanace,
O. P. Jindal Global University, Sonipat, India.
E-mail address: avipbhu@gmail.com