A Generalized Approach for Estimation of a Finite Population Mean in Two-Phase Sampling

Abstract

The present paper deals with the formulation of a generalized class of ratio-cum-product estimators for the estimation of a finite population mean in two-phase sampling. The mathematical expressions for the mean square errors (MSEs) of the preliminary and proposed estimators are derived. It has been established that the proposed class encompasses a wide range of estimators for specific choices of the scalars. The relative performance of the proposed class, as compared to the preliminary estimators, is evaluated using the MSE criterion. In addition, optimum sample sizes of the first-phase and second-phase samples are obtained using a specified cost function. The theoretical findings are empirically assessed by computing MSEs and percent relative efficiencies (PREs) of various estimators based on real population datasets. The findings of the study reveal that the proposed class of estimators are superior as compared to the preliminary estimators for the estimation of mean.

Keywords: *Study variable; auxiliary variable; two-phase sampling; cost function analysis; mean square error; percent relative efficiency.*

2022 Mathematics Subject Classification: 62D05

1 Introduction

The problem of estimation of population mean of a target variable (i.e., study variable) is of utmost importance in sample survey dealing with almost every diversified fields of study, for instance, crop estimation surveys, clinical studies, demographic studies, economic studies and much more. The fundamental procedures for estimation of a population quantity (for instance, population mean and/or population variance) are the ratio, product and regression methods. In these methods, the prior information on auxiliary variable(s) is utilized at the estimation stage. However, in the absence of such prior information, the two-phase sampling design is utilized for the estimation of concerned population quantity.

"The term two-phase sampling was first introduced by Neyman [1]. The two-phase sampling method for the estimation of population mean ($\bar{Y}$) of a study variable *Y*, on utilizing information on single auxiliary variable *X*", consists of the following steps:

(i) A preliminary larger sample of size n' is selected from the population consisting of *N* units, and the sample mean ($\bar{x}'$) of the auxiliary variable *X* is measured.

(ii) A sub-sample of size n is selected from the preliminary sample, and the sample means ($\bar{y}$, $\bar{x}$) of the variables Y and X, respectively, are measured.

At each phase, simple random sampling without replacement (SRSWOR) scheme is utilized for the selection of samples. Also, the population mean ($\bar{X}$) of the auxiliary variable X is assumed to be unknown. The notations involved are as follows:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \; , \; \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \; , \; \bar{x}' = \frac{1}{n'}\sum_{i=1}^{n'} x_i \, .$$

The sample obtained above in step (i) is regarded as the first-phase sample, whereas the sample obtained in step (ii) is regarded as the second-phase sample.

Considering the utility of two-phase sampling, various scientists and researchers have made their valuable contributions by formulating estimators for the population mean on utilizing the ratio, product and regression methods (for instance, Sukhatme [2], Srivastava [3], Sisodia and Dwivedi [4], Diana and Tommasi [5], Singh and Ruiz Espejo [6], Handique [7], Malik and Tailor [8], Vishwakarma and Kumar [9], Kumar and Vishwakarma [10], Kumar and Tiwari [11], Erinola et al. [12], Kumar and Tiwari [13] and Oyeyemi et al. [14].

Some recent noteworthy contributions towards the estimation of mean in survey sampling have been made by Kumar and Vishwakarma [15], Zeeshan et al. [16], Rather et al. [17], Rather and Kadilar [18], and Rather et al. [19].

Due to the wide applicability of survey sampling for the estimation of parameters (i.e., mean and variance) of a study variable, various scientists and researchers have developed estimators, which are sometimes not precise as desired. Considering the given fact, an attempt is made in the present paper by formulating a generalized class of ratio-cum-product estimators for the estimation of population mean of the study variable. The proposed class incorporates a wide range of preliminary estimators as the specific members for suitable choices of the scalars used in the study. Some of the members are observed to be the best in terms of precision, and hence are regarded as the asymptotic optimum estimators (AOEs) of the proposed class. The findings of the study exhibit the superiority of the proposed class of estimators as compared to the preliminary conventional estimators.

2 Some Preliminary Estimators of the Population Mean

The ratio estimator developed by Sukhatme [2] for the population mean $\bar{Y}$ of the study variable Y in two-phase sampling is given by:

$$\bar{y}_{Rd} = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right), \qquad (1)$$

where $\bar{x}' = \sum_{i=1}^{n'} X_i / n'$ denotes the first-phase sample mean of the auxiliary variable *X*. Also, $\bar{y} = \sum_{i=1}^{n} Y_i / n$ and $\bar{x} = \sum_{i=1}^{n} X_i / n$ denote the second-phase sample means of the variables *Y* and *X*, respectively.

The usual product estimator for the population mean $\bar{Y}$ under two-phase sampling is given by:

$$\bar{y}_{Pd} = \bar{y}\left(\frac{\bar{x}}{\bar{x}'}\right). \qquad (2)$$

Kwathekar and Ajagonkar [20] suggested the following ratio and product estimators for $\bar{Y}$ in two-phase sampling by utilizing information on coefficient of variation (C_x) of the variable X :

$$\bar{y}_{KA}^{R} = \bar{y}\left(\frac{\bar{x}' + C_x}{\bar{x} + C_x}\right), \qquad (3)$$

and

$$\bar{y}_{KA}^{P} = \bar{y}\left(\frac{\bar{x} + C_x}{\bar{x}' + C_x}\right). \qquad (4)$$

Singh and Ruiz Espejo [6] suggested the following ratio-product type estimator for $\bar{Y}$ under two-phase sampling:

$$\bar{y}_{RP}^{d} = \bar{y}\left[k\frac{\bar{x}'}{\bar{x}} + (1-k)\frac{\bar{x}}{\bar{x}'}\right]. \qquad (5)$$

Singh et al. [21] suggested a class of ratio-cum-product estimator for $\bar{Y}$ as follows:

$$\bar{y}_{SEA} = \bar{y}\left\{k\left(\frac{\bar{x}' + C_x}{\bar{x} + C_x}\right) + (1-k)\left(\frac{\bar{x} + C_x}{\bar{x}' + C_x}\right)\right\}. \qquad (6)$$

Malik and Tailor [8] suggested the following ratio estimator for $\bar{Y}$ in two-phase sampling by utilizing the information on correlation coefficient (ρ_{YX}) of the variable X :

$$T_{MT} = \bar{y}\left(\frac{\bar{x}' + \rho_{YX}}{\bar{x} + \rho_{YX}}\right). \qquad (7)$$

To the first order of approximation, the mean square errors (MSEs) of various preliminary estimators described above are as follows:

$$MSE(\bar{y}_{Rd}) = \bar{Y}^2 \left\{ f_1 C_Y^{\ 2} + f_3 \left(C_X^{\ 2} - 2\rho_{YX} C_Y C_X \right) \right\}, \qquad (8)$$

$$MSE(\bar{y}_{Pd}) = \bar{Y}^2 \left\{ f_1 C_Y^{\ 2} + f_3 \left(C_X^{\ 2} + 2\rho_{YX} C_Y C_X \right) \right\}, \qquad (9)$$

$$MSE(\bar{y}_{KA}^{R}) = \bar{Y}^2 \left\{ f_1 C_Y^{\ 2} + f_3 \delta \left(\delta C_X^{\ 2} - 2\rho_{YX} C_Y C_X \right) \right\}, \qquad (10)$$

$$MSE(\bar{y}_{KA}^{P}) = \bar{Y}^2 \left\{ f_1 C_Y^{\ 2} + f_3 \delta \left(\delta C_X^{\ 2} + 2\rho_{YX} C_Y C_X \right) \right\}, \qquad (11)$$

$$MSE(\bar{y}_{RP}^{d}) = \bar{Y}^2 \left[f_1 C_Y^2 + (2k-1) f_3 \left\{ (2k-1) C_X^2 - 2\rho_{YX} C_Y C_X \right\} \right], \qquad (12)$$

$$MSE(\bar{y}_{SEA}) = \bar{Y}^2 \left[f_1 C_Y^2 + (2k-1) f_3 \delta \left\{ (2k-1) \delta C_X^2 - 2\rho_{YX} C_Y C_X \right\} \right], \qquad (13)$$

and

$$MSE(\bar{y}_{MT}) = \bar{Y}^2 \left\{ f_1 C_Y^{\ 2} + f_3 \omega \left(\omega C_X^{\ 2} - 2\rho_{YX} C_Y C_X \right) \right\}. \qquad (14)$$

Furthermore, the minimum attainable MSEs of the estimators $\bar{y}_{RP}^{d}$ and $\bar{y}_{SEA}$ are given, respectively, by

$$MSE(\bar{y}_{RP}^{d})_{\min} = \bar{Y}^2 C_Y^{\ 2} \left(f_1 - f_3 \rho_{YX}^2 \right), \qquad (15)$$

and

$$MSE(\bar{y}_{SEA})_{\min} = \bar{Y}^2 C_Y^{\ 2} (f_1 - f_3 \rho_{YX}^2). \qquad (16)$$

It is well known that the sample mean ($\bar{y}$) is an unbiased estimator of the population mean ($\bar{Y}$), and its variance under simple random sampling without replacement (SRSWOR) scheme is given by

$$Var(\bar{y}) = f_1 \bar{Y}^2 C_y^2. \qquad (17)$$

The notations used above are as follows:

$$f_1 = \left(\frac{1}{n} - \frac{1}{N} \right), \quad f_2 = \left(\frac{1}{n'} - \frac{1}{N} \right), \quad f_3 = f_1 - f_2 = \left(\frac{1}{n} - \frac{1}{n'} \right),$$

$$\delta = \frac{\bar{X}}{\bar{X} + C_X}, \ \omega = \frac{\bar{X}}{\bar{X} + \rho_{YX}}, \ C_Y = \frac{S_Y}{\bar{Y}}, \ C_X = \frac{S_X}{\bar{X}}, \ \rho_{YX} = \frac{S_{YX}}{S_Y S_X},$$

$$S_Y^{\ 2} = \frac{1}{(N-1)} \sum_{i=1}^{N} \left(y_i - \bar{Y} \right)^2, \quad S_X^{\ 2} = \frac{1}{(N-1)} \sum_{i=1}^{N} \left(x_i - \bar{X} \right)^2, \text{ and } S_{YX} = \frac{1}{(N-1)} \sum_{i=1}^{N} \left(y_i - \bar{Y} \right)\left(x_i - \bar{X} \right).$$

3 Proposed Class of Estimators

Motivated by the work of Singh et al. [21], we develop the following generalized class of ratio-cum-product estimators of the population mean $\left(\bar{Y}\right)$ under two-phase sampling:

$$T=\bar{y}\left[k\left\{\frac{\alpha\bar{x}'+\gamma}{\alpha\bar{x}+\gamma}\right\}+\left(1-k\right)\left\{\frac{\alpha\bar{x}+\gamma}{\alpha\bar{x}'+\gamma}\right\}\right], \qquad (18)$$

where k,α,γ are the scalar quantities. The optimum values of these scalars are obtained on minimizing the MSE of the proposed class T. Moreover, it is worth mentioning that for suitable choices of the scalars k,α,γ, the proposed class T encompasses a wide range of members, which are depicted in Table 1.

Table 1: Members of the class of estimators T

Authors	**Estimators**	k	α	γ
Classical mean estimator	$\bar{y}$	1	0	*
Sukhatme [2]	$\bar{y}_{Rd}=\bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)$	1	*	0
	$\bar{y}_{Pd}=\bar{y}\left(\frac{\bar{x}}{\bar{x}'}\right)$	0	*	0
Kwathekar and Ajagonkar [20]	$\bar{y}_{KA}^{R}=\bar{y}\left(\frac{\bar{x}'+C_x}{\bar{x}+C_x}\right)$	1	1	C_X
	$\bar{y}_{KA}^{P}=\bar{y}\left(\frac{\bar{x}+C_x}{\bar{x}'+C_x}\right)$	0	1	C_X
Singh and Ruiz Espejo [6]	$\bar{y}_{RP}^{d}=\bar{y}\left[k\frac{\bar{x}'}{\bar{x}}+(1-k)\frac{\bar{x}}{\bar{x}'}\right]$	k	1	0
Singh et al. [21]	$\bar{y}_{SEA}=\bar{y}\left\{k\left(\frac{\bar{x}'+C_x}{\bar{x}+C_x}\right)+\left(1-k\right)\left(\frac{\bar{x}+C_x}{\bar{x}'+C_x}\right)\right\}$	k	1	C_X
Malik and Tailor [8]	$T_{MT}=\bar{y}\left(\frac{\bar{x}'+\rho_{YX}}{\bar{x}+\rho_{YX}}\right)$	1	1	ρ_{YX}

*Any value can be assigned to the scalar.

4 MSE of the Proposed Class

To obtain the MSE of the proposed class T, we consider

$\bar{y}=\bar{Y}\left(1+e_0\right)$, $\bar{x}=\bar{X}\left(1+e_1\right)$, and $\bar{x}'=\bar{X}\left(1+e_1'\right)$.

Then, we have

$$\left.\begin{aligned}
&E(e_0)=E(e_1)=E(e_1^{'})=0,\\
&E(e_0^2)=f_1C_Y^{\,2},\ E(e_1^2)=f_1C_X^{\,2},\ E(e_1^{'2})=f_2C_X^{\,2},\\
&E(e_0e_1)=f_1\rho_{YX}C_YC_X,\ E(e_0e_1^{'})=f_2\rho_{YX}C_YC_X,\\
&\text{and } E(e_1e_1^{'})=f_2C_X^{\,2}.
\end{aligned}\right\} \qquad (19)$$

Now, expressing T in terms of e_0, e_1 and $e_1^{'}$,we have

$$T=\bar{Y}(1+e_0)\left[k\left(1+\psi e_1^{'}\right)\left(1+\psi e_1\right)^{-1}+\left(1-k\right)\left(1+\psi e_1\right)\left(1+\psi e_1^{'}\right)^{-1}\right], \qquad (20)$$

where $\psi=\dfrac{\alpha\bar{X}}{\alpha\bar{X}+\gamma}$.

Expanding the R.H.S. of (20), multiplying out and retaining the first order error terms, we have

$$T=\bar{Y}\left[k\left(1-\psi e_1+\psi e_1^{'}+e_0\right)+\left(1-k\right)\left(1-\psi e_1^{'}+\psi e_1+e_0\right)\right], \qquad (21)$$

$$\text{or } T=\bar{Y}\left[k\left\{2\psi e_1^{'}-2\psi e_1\right\}+\left\{\left(1+e_0\right)+\psi\left(e_1-e_1^{'}\right)\right\}\right], \qquad (22)$$

$$\text{i.e., } T-\bar{Y}=\bar{Y}\left[2\psi k\left\{e_1^{'}-e_1\right\}+\left\{e_0+\psi\left(e_1-e_1^{'}\right)\right\}\right]. \qquad (23)$$

Squaring both sides of (23), taking the expectation, and using the results of (19), we obtain the MSE of proposed class T to the first order of approximation as:

$$MSE(T)=\bar{Y}^2\left[f_1C_Y^{\,2}+(2k-1)f_3\psi\left\{(2k-1)\psi C_X^{\,2}-2\rho_{YX}C_YC_X\right\}\right]. \qquad (24)$$

The optimum value of ψ is obtained on using the following condition:

$$\frac{\partial}{\partial\psi}MSE(T)=0. \qquad (25)$$

On solving (25), the optimum value of ψ, which minimizes the MSE of T, is obtained as:

$$\psi_{opt}=\frac{1}{(2k-1)}\frac{\rho_{YX}C_Y}{C_X}. \qquad (26)$$

On substituting the optimum value of ψ from (26) in (24), the minimum attainable MSE of T is obtained as:

$$MSE(T)_{\min}=\bar{Y}^2C_Y^{\,2}\left(f_1-f_3\rho_{YX}^2\right). \qquad (27)$$

5 Efficiency Comparisons

In order to evaluate the relative performances of the proposed class T as compared to the preliminary estimators, the MSE criterion is used on utilizing the equations (8) to (14), (17) and (24):

(i) $MSE(T) < Var(\bar{y})$ if

$$(2k-1)\psi < \frac{2\rho_{YX}C_Y}{C_X},$$

(ii) $MSE(T) < Var(\bar{y}_{Rd})$ if

$$(2k-1)\psi < \frac{2\rho_{YX}C_Y}{C_X} - 1,$$

(iii) $MSE(T) < Var(\bar{y}_{Pd})$ if

$$(2k-1)\psi < \frac{2\rho_{YX}C_Y}{C_X} + 1,$$

(iv) $MSE(T) < Var(\bar{y}_{KA}^{R})$ if

$$(2k-1)\psi < \frac{2\rho_{YX}C_Y}{C_X} - \delta,$$

(v) $MSE(T) < Var(\bar{y}_{KA}^{p})$ if

$$(2k-1)\psi < \frac{2\rho_{YX}C_Y}{C_X} + \delta,$$

(vi) $MSE(T) < Var(\bar{y}_{RP}^{d})$ if

$$(2k-1)(\psi+1) < \frac{2\rho_{YX}C_Y}{C_X},$$

(vii) $MSE(T) < Var(\bar{y}_{SEA})$ if

$$(2k-1)(\psi+\delta) < \frac{2\rho_{YX}C_Y}{C_X},$$

(viii) $MSE(T) < Var(\bar{y}_{MT})$ if

$$(2k-1)\psi < \frac{2\rho_{YX}C_Y}{C_X} - \omega.$$

6 Cost Function Analysis

In this section, the optimum sample sizes of the first-phase and second-phase samples have been obtained by utilizing a cost function of the form:

$$c = c_1 n' + c_2 n, \qquad (28)$$

where c denotes the total sampling cost involved in the survey. Moreover, c_1 and c_2 denote, respectively, the costs per unit associated with the first-phase and second-phase samples.

The optimum values of n' and n that minimize the MSE of T for a specified cost $c \le c_0$, are obtained on using the Lagrangian function L as follows:

$$L = MSE(T) + \lambda\left(c_1 n' + c_2 n - c_0\right), \qquad (29)$$

where λ is the Lagrange's multiplier.

Also, from (24), we have

$$MSE(T) = \bar{Y}^2\left[f_1 C_Y^{\ 2} + (2k-1) f_3 \psi \left\{(2k-1)\psi C_X^{\ 2} - 2\rho_{YX} C_Y C_X\right\}\right],$$

$$\text{or } MSE(T) = f_1 S_Y^{\ 2} - f_3\left\{2(2k-1)\psi R S_{YX} - (2k-1)^2 \psi^2 R^2 S_X^{\ 2}\right\}, \qquad (30)$$

where the notations used are as follows:

$C_Y = \dfrac{S_Y}{\bar{Y}}$, $C_X = \dfrac{S_X}{\bar{X}}$, $R = \dfrac{\bar{Y}}{\bar{X}}$ and $S_{YX} = \rho_{YX} S_Y S_X$.

Hence, on simplifying (30), we have

$$MSE(T) = \left(\frac{1}{n} - \frac{1}{N}\right) S_Y^{\ 2} - \left(\frac{1}{n} - \frac{1}{n'}\right)\xi, \qquad (31)$$

where $\xi = 2(2k-1)\psi R S_{YX} - (2k-1)^2 \psi^2 R^2 S_X^{\ 2}$.

Substituting (31) in (29), we get

$$L = \left(\frac{1}{n} - \frac{1}{N}\right) S_Y^{\ 2} - \left(\frac{1}{n} - \frac{1}{n'}\right)\xi + \lambda(c_1 n' + c_2 n - c_0). \qquad (32)$$

Now, differentiating (32) with respect to n' and n, equating the results to zero, and then using (28), we obtain the optimum values of n' and n as follows:

$$n' = \frac{c}{c_1 + c_2\sqrt{A}}, \qquad (33)$$

and

$$n = n'\sqrt{A} = \frac{c\sqrt{A}}{c_1 + c_2\sqrt{A}}, \qquad (34)$$

where $A = \dfrac{c_1\left(S_Y{}^2 - \xi\right)}{c_2\xi}$.

Hence, substituting the values of n' and n, from (33) and (34) in (31), the optimum *MSE* of the proposed class T is obtained as follows:

$$MSE(T)_{opt} = \frac{\left(c_1 + c_2\sqrt{A}\right)}{c}\left[\frac{1}{\sqrt{A}}S_Y^2 - \frac{1}{\sqrt{A}}\xi + \xi\right] - \frac{1}{N}S_Y^2. \qquad (35)$$

In a similar manner, the optimum values of n' and n, along with the optimum MSEs, for the members of the proposed class T are computed and presented in Table 2.

Table 2: Optimum n' and n, along with the optimum MSEs of various members of the proposed class T

Estimator	n'	n	**Optimum *MSEs***
$\bar{y}$	---	$\frac{c}{c_2}$	$\left(\frac{c_2}{c} - \frac{1}{N}\right)S_Y^2$
$\bar{y}_{Rd}$	$\frac{c}{c_1 + c_2\sqrt{A_1}}$	$\frac{c\sqrt{A_1}}{c_1 + c_2\sqrt{A_1}}$	$\frac{\left(c_1 + c_2\sqrt{A_1}\right)}{c}\left[\frac{1}{\sqrt{A_1}}S_Y^2 - \frac{1}{\sqrt{A_1}}\xi_1 + \xi_1\right] - \frac{1}{N}S_Y^2$
$\bar{y}_{Pd}$	$\frac{c}{c_1 + c_2\sqrt{A_2}}$	$\frac{c\sqrt{A_2}}{c_1 + c_2\sqrt{A_2}}$	$\frac{\left(c_1 + c_2\sqrt{A_2}\right)}{c}\left[\frac{1}{\sqrt{A_2}}S_Y^2 - \frac{1}{\sqrt{A_2}}\xi_2 + \xi_2\right] - \frac{1}{N}S_Y^2$
$\bar{y}_{KA}^{R}$	$\frac{c}{c_1 + c_2\sqrt{A_3}}$	$\frac{c\sqrt{A_3}}{c_1 + c_2\sqrt{A_3}}$	$\frac{\left(c_1 + c_2\sqrt{A_3}\right)}{c}\left[\frac{1}{\sqrt{A_3}}S_Y^2 - \frac{1}{\sqrt{A_3}}\xi_3 + \xi_3\right] - \frac{1}{N}S_Y^2$
$\bar{y}_{KA}^{P}$	$\frac{c}{c_1 + c_2\sqrt{A_4}}$	$\frac{c\sqrt{A_4}}{c_1 + c_2\sqrt{A_4}}$	$\frac{\left(c_1 + c_2\sqrt{A_4}\right)}{c}\left[\frac{1}{\sqrt{A_4}}S_Y^2 - \frac{1}{\sqrt{A_4}}\xi_4 + \xi_4\right] - \frac{1}{N}S_Y^2$
$\bar{y}_{RP}^{d}$	$\frac{c}{c_1 + c_2\sqrt{A_5}}$	$\frac{c\sqrt{A_5}}{c_1 + c_2\sqrt{A_5}}$	$\frac{\left(c_1 + c_2\sqrt{A_5}\right)}{c}\left[\frac{1}{\sqrt{A_5}}S_Y^2 - \frac{1}{\sqrt{A_5}}\xi_5 + \xi_5\right] - \frac{1}{N}S_Y^2$

$\bar{y}_{SEA}$	$\frac{c}{c_1+c_2\sqrt{A_6}}$	$\frac{c\sqrt{A_6}}{c_1+c_2\sqrt{A_6}}$	$\frac{\left(c_1+c_2\sqrt{A_6}\right)}{c}\left[\frac{1}{\sqrt{A_6}}S_Y^2-\frac{1}{\sqrt{A_6}}\xi_6+\xi_6\right]-\frac{1}{N}S_Y^2$
$\bar{y}_{MT}$	$\frac{c}{c_1+c_2\sqrt{A_7}}$	$\frac{c\sqrt{A_7}}{c_1+c_2\sqrt{A_7}}$	$\frac{\left(c_1+c_2\sqrt{A_7}\right)}{c}\left[\frac{1}{\sqrt{A_7}}S_Y^2-\frac{1}{\sqrt{A_7}}\xi_7+\xi_7\right]-\frac{1}{N}S_Y^2$

The notations used in Table 2 are as follows:

$$A_1=\frac{c_1\left(S_Y{}^2-\xi_1\right)}{c_2\xi_1},\ \xi_1=2RS_{YX}-R^2S_Y{}^2,$$

$$A_2=\frac{c_1\left(S_Y{}^2-\xi_2\right)}{c_2\xi_2},\ \xi_2=-2RS_{YX}-R^2S_Y{}^2,$$

$$A_3=\frac{c_1\left(S_Y{}^2-\xi_3\right)}{c_2\xi_3},\ \xi_3=2\delta RS_{YX}-\delta^2R^2S_Y{}^2,$$

$$A_4=\frac{c_1\left(S_Y{}^2-\xi_4\right)}{c_2\xi_4},\ \xi_4=-2\delta RS_{YX}-\delta^2R^2S_Y{}^2,$$

$$A_5=\frac{c_1\left(S_Y{}^2-\xi_5\right)}{c_2\xi_5},\ \xi_5=2\left(2k-1\right)RS_{YX}-\left(2k-1\right)^2R^2S_X{}^2,$$

$$A_6=\frac{c_1\left(S_Y{}^2-\xi_6\right)}{c_2\xi_6},\ \xi_6=2\left(2k-1\right)\delta RS_{YX}-\left(2k-1\right)^2\delta^2R^2S_X{}^2,$$

and

$$A_7=\frac{c_1\left(S_Y{}^2-\xi_7\right)}{c_2\xi_7},\ \xi_7=2\omega RS_{YX}-\omega^2R^2S_Y{}^2.$$

7 Empirical Study

The relative performance of the proposed class T, as compared to the preliminary estimators, is evaluated empirically by utilizing four real population data sets. The populations considered in the analysis are described below:

Population I- Source: Sahoo and Swain [22]

Y : Yield of rice per plant,

X : Number of tillers.

Population II- Source: Das [23]

Y : Number of agricultural labourers for 1971,

X : Number of agricultural labourers for 1961.

Population III- Source: Handique [7]

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).

Population IV- Source: Maddala [24]

Y : Consumption per capita,

X : Deflated prices of veal.

Table 3: Summary of population values

Population	N	n'	n	$\bar{Y}$	$\bar{X}$	ρ_{YX}	C_Y	C_X
I	50	30	15	12.842	9.04	0.7133	0.3957	0.2627
II	278	50	25	39.0680	25.1110	0.7213	1.4451	1.6198
III	2500	200	25	4.63	21.09	0.79	0.95	0.98
IV	16	8	4	7.6375	75.4313	-0.6823	0.2278	0.0986

The values of scalars ω and δ are obtained for the above considered populations, and are summarized in Table 4.

Table 4: Values of scalars ω and δ for various populations

Population	**I**	**II**	**III**	**IV**

$\delta = \frac{\bar{X}}{\bar{X} + C_X}$	0.9718	0.9394	0.9556	0.9987
$\omega = \frac{\bar{X}}{\bar{X} + \rho_{YX}}$	0.9269	0.9721	0.9639	1.0091

The MSEs and percent relative efficiencies (PREs) of various estimators of $\bar{Y}$ have been computed and the concerned values are depicted in Tables 5 and 6, respectively. The PREs are obtained by using the formula:

$$PRE(\phi, \bar{y}) = \frac{Var(\bar{y})}{MSE(\phi)} \times 100,$$

where $\phi = \bar{y}$, $\bar{y}_{Rd}$, $\bar{y}_{Pd}$, $\bar{y}_{KA}^{R}$, $\bar{y}_{KA}^{P}$, $\bar{y}_{RP}^{d}$, $\bar{y}_{SEA}$, $\bar{y}_{MT}$, T.

Table 5: MSEs of various estimators of $\bar{Y}$

Estimator	Population			
	I	II	III	IV
$\bar{y}$	1.21	116.03	0.766	0.567
$\bar{y}_{Rd}$	0.769	93.04	0.383	*
$\bar{y}_{Pd}$	*	*	*	0.414
$\bar{y}_{KA}^{R}$	0.771	89.88	0.369	*
$\bar{y}_{KA}^{P}$	*	*	*	0.415
$\bar{y}_{RP}^{d}$	0.767	82.86	0.343	0.391
$\bar{y}_{SEA}$	0.767	82.86	0.343	0.391
$\bar{y}_{MT}$	0.775	91.51	0.372	*
T	0.767	82.86	0.343	0.391

*Data is not applicable.

Table 6: PREs of various estimators of $\bar{Y}$ with respect to the sample mean $\bar{y}$

Estimator	Population			
	I	II	III	IV
$\bar{y}$	100	100	100	100

$\bar{y}_{Rd}$	156.66	124.71	200.01	*
$\bar{y}_{Pd}$	*	*	*	136.77
$\bar{y}_{KA}^{R}$	156.28	129.10	207.35	*
$\bar{y}_{KA}^{P}$	*	*	*	136.74
$\bar{y}_{RP}^{d}$	157.09	140.03	223.02	145.00
$\bar{y}_{SEA}$	157.09	140.03	223.02	145.00
$\bar{y}_{MT}$	155.42	126.79	206.06	*
T	**157.09**	**140.03**	**223.02**	**145.00**

Bold values indicate the maximum PRE

*Data is not applicable

The following results are obtained from Table 5:

(1) In all the four populations, the proposed class T attains the minimum MSEs. Moreover, among the members of the proposed class T, the estimators $\bar{y}_{RP}^{d}$ and $\bar{y}_{SEA}$ can be regarded as the asymptotic optimum estimators (AOEs) as these estimators achieve the minimum MSEs as compared to the other members of class T.

(2) In populations II and III, the estimator $\bar{y}_{KA}^{R}$ is more efficient as compared to the estimators $\bar{y}_{Rd}$ and $\bar{y}_{MT}$.

(3) In population III, the MSEs of the estimators $\bar{y}_{Pd}$ and $\bar{y}_{KA}^{P}$ are nearly the same, and hence for the concerned population, both the estimators are equally efficient for the estimation of population mean ($\bar{Y}$).

In a similar manner, the following results are obtained from Table 6:

(1) In all the four populations, the proposed class T attains the maximum PREs. Also, among the members of class T, the estimators $\bar{y}_{RP}^{d}$ and $\bar{y}_{SEA}$ are superior as compared to the other members, i.e., $\bar{y}$, $\bar{y}_{Rd}$, $\bar{y}_{Pd}$, $\bar{y}_{KA}^{R}$, $\bar{y}_{KA}^{P}$ and $\bar{y}_{MT}$.

(2) In populations II and III, the PREs of estimator $\bar{y}_{KA}^{R}$ are more as compared to the estimators $\bar{y}_{Rd}$ and $\bar{y}_{MT}$.

(3) In population III, the PREs of the estimators $\bar{y}_{Pd}$ and $\bar{y}_{KA}^{P}$ are almost the same.

8 Conclusion

In this paper, a generalized class of ratio-cum-product estimators has been developed for estimating the finite population mean ($\bar{Y}$) of the study variable Y in two-phase sampling. The mathematical expressions for the MSEs of the preliminary estimators and the proposed class of estimators (T) are derived under large sample approximation. It has been established that the proposed class (T) encompasses a wide range of estimators for specific choices of the scalars.

The optimum sample sizes of the first-phase and second-phase samples, under a specified cost function, have been obtained for the proposed class as well as for the members of the proposed class. Moreover, the optimum MSEs have been computed for the specified cost, and the findings are demonstrated in Table 2.

The efficiency comparisons of the proposed class with the conventional preliminary estimators have been validated empirically using the MSE and PRE criteria on using real population data sets, and it has been revealed that the proposed class T is superior as compared to the preliminary estimators for the estimation of population mean $\bar{Y}$ of the study variable Y. Moreover, it has been established that, among the members of the proposed class T, the estimators $\bar{y}_{RP}^{d}$ and $\bar{y}_{SEA}$ can be regarded as the asymptotic optimum estimators (AOEs) of the proposed class T.

On the basis of theoretical and empirical results, we conclude that the proposed class T has greater scope and relevance for the estimation of finite population mean. Hence, the present paper provides significant contributions towards the theory of estimation of mean in two-phase sampling.

Acknowledgment

The authors are highly thankful to the Chief Editor and the anonymous reviewers for their valuable suggestions that led to the improvement of the contents of manuscript in the present form.

References

[1] Neyman J. Contribution to the theory of sampling human populations. Journal of American Statistical Association. 1938;33: 101-116.

[2] Sukhatme BV. Some ratio-type estimators in two-phase sampling. Journal of the American Statistical Association. 1962;57:628–632.

[3] Srivastava SK. A two phase sampling estimator in sample surveys. Australian Journal of Statistics. 1970;12(1):23-27.

[4] Sisodia BVS, Dwivedi VK, A class of ratio product type estimator in double sampling. Biometrical Journal. 1982;24(4):419-424.

[5] Diana G, Tommasi C. Optimal estimation for finite population mean in two-phase sampling. Statistical Methods and Applications. 2003;12(1):41-48.

[6] Singh HP, Ruiz Espejo M. Double sampling ratio-product estimator of a finite population mean in sample surveys. Journal of Applied Statistics. 2007;34(1): 71-85.

[7] Handique BK. A class of regression-cum-ratio estimators in two-phase sampling for utilizing information from high resolution satellite data. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences.2012;I(4):71-76.

[8] Malik KA, Tailor R. Ratio type estimator of population mean in double sampling. International Journal of Advanced Mathematics and Statistics. 2013;1(1): 34-39.

[9] Vishwakarma GK, Kumar M. An efficient class of estimators for the mean of a finite population in two-phase sampling using multi-auxiliary variates. Communications in Mathematics and Statistics. 2015;3(4):477–489.

[10] Kumar M, Vishwakarma GK. Estimation of mean in double sampling using exponential technique on multi-auxiliary variates. Communications in Mathematics and Statistics. 2017;5(4):429–445.

[11] Kumar M, Tiwari A. A generalized class of estimators of population mean in two-phase sampling using two auxiliary variables. International Journal of Mathematics and Statistics. 2021;22(3): 18-27.

[12] Erinola AY, Singh RVK, Audu A, James T. Modified class of estimator for finite population mean under two-phase sampling using regression estimation approach. Asian Journal of Probability and Statistics. 2003;14(4):52-64.

[13] Kumar M, Tiwari A. A composite class of ratio estimators for a finite population mean in two-phase sampling. International Journal of Statistics and Reliability Engineering. 2022;9(3):364-370.

[14] Oyeyemi GM, Muhammad I, Kareem AO. Combined exponential-type estimators for finite population mean in two-phase sampling, Asian Journal of Probability and Statistics. 2023; 21(2):44-58.

[15] Kumar M, Vishwakarma GK. Generalized classes of regression-cum-ratio estimators of population mean in stratified random sampling. Proceedings of the National Academy of Sciences, India Section A: Physical Sciences. 2020;90(5):933-939.

[16] Zeeshan SM, Vishwakarma GK, Kumar M. An efficient variant of dual to product and ratio estimators in sample surveys. Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics. 2021;70(2):997-1010.

[17] Rather KUI, Kocyiğit EG, Onyango R, Kadilar C. Improved regression in ratio type estimators based on robust M-estimation. PLOS ONE. 2022;17(12):1-16.

[18] Rather KUI, Kadılar C. Improved estimators for the population mean under non-response. Gazi University Journal of Science. 2023;36(2):920-931.

[19] Rather KUI, Bouza CN, Rizvi SEH, Sharma M, Bhat MIJ. Modified ratio cum product type exponential estimator of population mean in stratified random sampling. Revista Investigación Operacional. 2022;43(5):653-663.

[20] Kawathekar DM, Ajagaonkar SGP. A modified ratio estimator based on the coefficient of variation in double sampling. Journal of Indian Statistical Association. 1984;36(2):47-50.

[21] Singh HP, Tailor R, Tailor R, Estimation of finite population mean in two-phase sampling with known coefficient of variation of an auxiliary character. Statistica. 2012;72(1):111-126.

[22] Sahoo LN, Swain AKPC. Unbiased ratio-cum-product estimator. Sankhya. 1980;42:56-62.

[23] Das AK. Contributions to the theory of sampling strategies based on auxiliary information. Ph. D. Thesis, Bidhan Chandra Krishi Vishwavidyalaya, West Bengal, India; 1988.

[24] Maddala GS. Econometrics, Mc. Graw Hills Publication, New York. 1977.