AIMS Mathematics

AIMS Mathematics, 10(4): 8907–8925.
DOI: 10.3934/math.2025408
Received: 05 March 2025
Revised: 05 April 2025
Accepted: 09 April 2025
Published: 18 April 2025

https://www.aimspress.com/journal/Math

Research article

Developing and evaluating efficient estimators for finite population mean in two-phase sampling

Khazan Sher[1], Muhammad Ameeq[2,*], Sidra Naz[2], Basem A. Alkhaleel[3], Muhammad Muneeb Hassan[2] and Olayan Albalawi[4]

[1] Department of Statistics University of Peshawar, Pakistan
[2] Department of Statistics The Islamia University Bahawalpur, Punjab Pakistan
[3] Department of Industrial Engineering, King Saud University Riyadh 12372, Saudi Arabia
[4] Department of Statistics, Faculty of Science, University of Tabuk Saudi Arabia

* **Correspondence:** Email: ameeq7777@gmail.com; Tel: +923347874989.

Abstract: The estimator development process is more efficient when additional information is used. However, occasionally, it is necessary to use information regarding unknown population parameters. In these cases, we chose two-phase sampling by substituting the population mean of the supplemental variable with the sample mean from first-phase sampling. The goal of this project was to develop effective estimators of the finite population mean in a two-phase sampling scenario with a single auxiliary variable. Under certain conditions, the recommended estimators outperform the current estimators, producing biased and Mean Square Error (MSE) expressions. Empirical and theoretical comparisons of the proposed families were conducted using real and simulated data. We found that the proposed families were more effective in the two-phase sampling situation than in all-population mean estimators.

Keywords: two-phase samplings; population parameters; mean square error; finite population mean; estimator efficiency
Mathematics Subject Classification: 62DXX

1. Introduction

The utilization of additional information during sampling has proven effective in accurately estimating the population mean. Incorporating a supplementary variable during the estimation phase is a valuable strategy for enhancing the precision of estimating the population mean for the study variables. When there is a positive or negative correlation between the study and auxiliary variables, we typically employ the ratio and product estimation methods. Survey statisticians have developed exponential-type estimators for various sample scenarios, as evidenced in the works of Singh and Vishwakarma [1], Grover et al. [2], Noor-ul-Amin and Hanif [3]. The latest and efficient estimators were produced by Sher et al. [4], Muneer et al. [5] and Choudhury and Singh [6], which are hybrid type estimators. For a more comprehensive understanding, Sabat et al. [7], Di Gravio et al. [8], and Zaman and Kadilar [9], demonstrate the common use of two-phase sampling in cases where collecting data on variables of interest is cost-prohibitive and data on variables that correlate with the variables of interest is more economical. For example, conducting on-site assessments of forest surveys in remote locations can be both challenging and expensive. However, aerial photography is a more cost-effective alternative, yielding comparable results regarding the type of forest without the need for costly ground visits. However, in many important situations, before the survey, the population mean $\bar{X}$ of the supplementary variable is not known. In such a situation, the sample mean $\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i$ is used as auxiliary information obtained through an initial sample of n_1 units $(n_1 < N)$, taken through a simple random sampling w.o.r scheme. At the second stage, a sample of n $(n < n_1)$ units obtained in the same fashion, and sample means of both the study variable $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and auxiliary variable $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, are obtained from Tato and Singh [10] and Misra [11].

Suppose a population consists of N units $\Omega = \{\upsilon_1, \upsilon_2, \upsilon_3, \ldots, \upsilon_N\}$, and let a first-phase sample of n_1 units be drawn from where the estimated mean value of auxiliary variable is obtained. Then, in a second phase, a sample of n $(n<n_1)$ units is drawn where both the study and auxiliary variables are measured. To obtain the expressions for MSE and bias, let us introduce the following terms:

Notations:

$\xi_0 = \frac{\bar{y} - \bar{Y}}{\bar{Y}} \rightarrow$ is a relative error term for y; $\xi_1 = \frac{\bar{x} - \bar{X}}{\bar{X}} \rightarrow$ is a relative error term for x, $C_y \rightarrow$ is the population coefficient of variation (c.v.) for y; $C_x \rightarrow$ is the population coefficient of variation (c.v.) for x; ρ → is the population correlation coefficient between y and x; and λ and λ_1 are the finite population correction (fpc) factors for first- and second-phase sampling, respectively.

$\bar{y} = \bar{Y}(1+\xi_0)$, $\bar{x} = \bar{X}(1+\xi_1)$, and $\bar{x}_1 = \bar{X}(1+\xi_2)$, so that $E(\xi_0) = E(\xi_1) = E(\xi_2) = 0$, $E(\xi_0^2) = \lambda C_y^2$, $E(\xi_1^2) = \lambda C_x^2$, $E(\xi_2^2) = \lambda_1 C_x^2$, and $E(\xi_0\xi_1) = \lambda C_{yx}$, $E(\xi_0\xi_2) = \lambda_1 C_{yx}$, $E(\xi_1\xi_2) = \lambda_1 C_x^2$, where $\lambda = \frac{N-n}{Nn}$, $\lambda_1 = \frac{N-n_1}{n_1 N}$ and $R = \frac{\bar{Y}}{\bar{X}}$.

2. Some existing estimators

Kumar and Bahl [12] proposed the usual ratio estimator of the population mean in two-phase sampling, as follows:

$$t_R = \bar{y}\left(\frac{\bar{x}_1}{\bar{x}}\right). \tag{1}$$

The ratio estimator is typically chosen when there is a positive correlation between the study and auxiliary variables. The mean square error (MSE) of their proposed estimator up to the first order of approximation is given as

$$MSE\left(t_R\right) = \bar{Y}^2\left(\lambda C_y^2 + \lambda_1 C_x^2(1-2\psi)\right), \tag{2}$$

where $\psi = \rho_{yx}\dfrac{C_y}{C_x}$.

In situations when there is a negative correlation between the study variable and the auxiliary variable, the product estimator is typically preferred over the ratio estimator. In two-phase sampling, the classical product estimator of population mean is given as

$$t_P = \bar{y}\left(\frac{\bar{x}}{\bar{x}_1}\right). \tag{3}$$

The MSE of their proposed estimator up to the first order of approximation is given as

$$MSE\left(t_P\right) = \bar{Y}^2\left(\lambda C_y^2 + \lambda_1 C_x^2\left(1+2\psi\right)\right). \tag{4}$$

In two-phase sampling, Singh and Vishwakarma [1] suggested the exponential type of ratio and product estimators of the population mean of the research variable as

$$t_{SVR} = \bar{y}\exp\left(\frac{\bar{x}_1-\bar{x}}{\bar{x}_1+\bar{x}}\right) \tag{5}$$

and

$$t_{SVP} = \bar{y}\exp\left(\frac{\bar{x}-\bar{x}_1}{\bar{x}_1+\bar{x}}\right). \tag{6}$$

The MSEs of the above estimators up to the first order of approximation are given as

$$MSE\left(T_{SVR}\right) = \bar{Y}^2\left[\lambda C_y^2 + \lambda_1 C_x^2\left(\frac{1}{4}-\psi\right)\right] \tag{7}$$

and

$$MSE\left(t_{SVP}\right) = \bar{Y}^2\left[\lambda C_y^2 + \lambda' C_x^2\left(\frac{1}{4}+\psi\right)\right]. \tag{8}$$

Yadav et al. [13] proposed the following exponential ratio and product estimators:

$$t_{GR} = \alpha\bar{y} + \left(1-\alpha\right)\bar{y}\exp\left(\frac{\bar{x}_1-\bar{x}^*}{\bar{x}_1+\bar{x}^*}\right) \tag{9}$$

and

$$t_{GP} = \delta\bar{y} + (1-\delta)\,\bar{y}\exp\left(\frac{\bar{x}^* - \bar{x}_1}{\bar{x}^* + \bar{x}_1}\right). \tag{10}$$

The MSE up to the first order of approximation of the above estimators for optimum values of $\alpha = \frac{g-2\psi}{g}$ and $\delta = \frac{g+2\psi}{g}$ are given as

$$MSE(t_{GR}) = \bar{Y}^2\left[\lambda C_y^2 + \lambda'' g C_x^2\left(\frac{g}{4} - \psi\right) - \lambda''\frac{A^2}{4}\right] \tag{11}$$

and

$$MSE(t_{GP}) = \bar{Y}^2\left[\lambda C_y^2 + \lambda'' g C_x^2\left(\frac{g}{4} + \psi\right) - \lambda''\frac{B^2}{4}\right], \tag{12}$$

where $A = g - 2\psi$, $B = g + 2\psi$, and $g = \frac{n}{n_1 - n}$, $\lambda'' = \lambda - \lambda_1 = \frac{1}{n} - \frac{1}{n_1}$.

The classical unbiased regression estimator is used in two-phase sampling to estimate the population mean when the auxiliary variable and the research variable are correlated. It works like this:

$$t_{\mathrm{Re}g} = \bar{y} + \beta(\bar{x} - \bar{x}_1), \tag{13}$$

where β is the regression coefficient of the main study variable y regressed on auxiliary variable x. The MSE up to the first order of approximation for the above linear regression estimator is obtained as

$$MSE(t_{\mathrm{Re}g}) = \bar{Y}^2 C_y^2\left(\lambda - \lambda_1\rho^2\right). \tag{14}$$

By combining Singh and Vishwakarma [1] and regression estimator with a linear transformation, in Ozgul and Cingi [14], the following estimator was proposed:

$$t_{OC} = \left[k_1\bar{y} + k_2(\bar{x}_1 - \bar{x})\right]\exp\left(\frac{\bar{z}_1 - \bar{z}}{\bar{z}_1 + \bar{z}}\right), \tag{15}$$

where $\bar{z}_1 = u\bar{x}_1 + v$ and $\bar{z} = u\bar{x} + v$. and u, v are the generalized constants, meaning that for different values of u and v, one can obtain different estimators with different MSE. The minimum MSE up to the first order of approximation for $k_1 = 1 - \frac{2 - \lambda_1\theta^2 C_x^2}{1 + \left(\lambda - \lambda_1\rho^2\right)}$ and $k_2 = R\left[(\theta - 1) + \frac{2 - \lambda_1\theta^2 C_x^2}{1 + \left(\lambda - \lambda_1\rho^2\right)}(2\theta - \psi)\right]$ is given as

$$MSE(t_{OC}) = \bar{Y}^2\,\frac{C_y^2\left(\lambda - \lambda_1\rho^2\right)\left(1 - \lambda_1\theta^2 C_x^2\right) - \frac{\lambda_1^2\theta^4 C_x^4}{4}}{\left[1 + C_y^2\left(\lambda - \lambda_1\rho^2\right)\right]} \tag{16}$$

or

$$MSE\left(t_{OC}\right)=\bar{Y}^2\,\frac{MSE\left(t_{\mathrm{Re}g}\right)\left(1-\lambda_1\theta^2C_x^2\right)-\dfrac{\lambda_1^2\bar{Y}^2\theta^4C_x^4}{4}}{\left(\bar{Y}^2+MSE\left(t_{\mathrm{Re}g}\right)\right)}, \tag{17}$$

where $\theta=\dfrac{u\bar{X}}{2\left(u\bar{X}+v\right)}$.

3. The proposed class of estimator

The use of auxiliary information commonly leads to improved precision of the estimator. Also, the development of novel estimators for the finite population mean in two-phase sampling is not only a theoretical advancement but also a practical necessity. These novel estimators, developed under existing constraints, are expected to make a substantial contribution to the field of survey sampling by providing more reliable and precise population estimates. Motivated by these objectives, we propose the following two families of estimators under a two-phase sampling scheme using a single auxiliary variable.

3.1. First proposed estimator

Motivated by Muneer et al. [5] and Shabbir et al. [15], we propose the following class of estimators for the population mean:

$$t_{\mathrm{Pro}1}=\left(k_1\bar{y}+k_2\right)\left[\omega\frac{\bar{x}_1}{\bar{x}}\left\{2-\exp\left(\frac{u\left(\bar{x}-\bar{x}_1\right)}{u\left(\bar{x}_1+\bar{x}\right)+2v}\right)\right\}+\left(1-\omega\right)\frac{\bar{x}}{\bar{x}_1}\exp\left(\frac{u\left(\bar{x}_1-\bar{x}\right)}{u\left(\bar{x}_1+\bar{x}\right)+2v}\right)\right]. \tag{18}$$

A number of estimators can be generated from the above estimator by assigning different values of *u*, *v*, and *ω*. Here, k_1 and k_2 are the minimizing constants whose values are determined by minimizing the MSE, and *u, v, and ω,* $0\le\omega\le1$, are the generalizing constants that can assume any suitable value or any known parameter of the population. The bias and MSE of the estimator are given as:

Theorem 1. An estimator for the population mean defined in Eq (18) in the case of two-phase sampling with single auxiliary variables will have its MSE equation given as

$$MSE\left(t_{\mathrm{Pro}1}\right)_{\min}\cong\bar{Y}^2\left(1-\frac{A_{pr}D_{pr}^2+B_{pr}C_{pr}^2-2C_{pr}D_{pr}E_{pr}}{A_{pr}B_{pr}-E_{pr}^2}\right). \tag{19}$$

3.2. Second proposed estimator

Motivated by Shabbir et al. [15], Gupta and Shabbir [16], and Shabbir and Gupta [17], we propose the following improved class of estimators for a finite population mean in two-phase sampling:

$$t_{pro2}=\left(k_3\bar{y}+k_4\right)\left[\omega\frac{\bar{x}_1}{\bar{x}}+\left(1-\omega\right)\frac{\bar{x}}{\bar{x}_1}\right]\exp\left[\frac{u\left(\bar{x}_1-\bar{x}\right)}{u\left(\bar{x}_1+\bar{x}\right)+2v}\right], \tag{20}$$

where

$$A_{pr}=1+\lambda C_y^2+\left(\Delta_1\lambda+\Delta_2\lambda_1-\Delta_3\lambda_1\right)C_x^2+4\vartheta_1\left(\lambda-\lambda_1\right)C_{yx},$$

$$B_{pr}=1+\left(\Delta_1\lambda+\Delta_2\lambda_1-\Delta_3\lambda_1\right)C_x^2,\quad C_{pr}=1+\left(\vartheta_3\lambda+\vartheta_4\lambda_1-\vartheta_2\lambda_1\right)C_x^2+\vartheta_1\left(\lambda-\lambda_1\right)C_{yx},$$

$$D_{pr}=1+\left(\vartheta_3\lambda+\vartheta_4\lambda_1-\vartheta_2\lambda_1\right)C_x^2,\ \text{and}\ \ E_{pr}=1+\left(\Delta_1\lambda+\Delta_2\lambda_1-\Delta_3\lambda_1\right)C_x^2+2\vartheta_1\left(\lambda-\lambda_1\right)C_{yx}.$$

Description. In Table 1, estimators 1–3 represent product exponential estimators obtained using the value $\omega=0$. These estimators vary based on different values of *u* and *v*, derived from known parameter values. Estimators 4–6 are ratio exponential estimators obtained by setting $\omega=1$. Estimators 7–11 are obtained by setting $\omega=\frac{1}{2}$ in combination with different values of *u* and *v*. The final group of estimators are of a hybrid type, incorporating both ratio and product exponential forms, making them effectively efficient in both cases. Hence, this list of estimators covers a range of possible settings. Although many other possible estimators are possible, we have listed only a selected few.

Table 1. Estimators deduced for different values of *ω*, *u*, and *v*.

Estimator	u	v	ω	Estimators from first class	Estimators from second class
1	1	0	1	$t_{Pro1}=[k_1\bar{y}+k_2]\left[\frac{\bar{x}_1}{\bar{x}}\exp\left[\frac{(\bar{x}_1-\bar{x})}{(\bar{x}_1+\bar{x})}\right]\right]$	$t_{pro2}=[k_3\bar{y}+k_4]\left[\frac{\bar{x}_1}{\bar{x}}\right]\exp\left[\frac{\bar{x}_1-\bar{x}}{\bar{x}_1+\bar{x}}\right]$
2	1	C_x	1	$t_{Pro1}=[k_1\bar{y}+k_2]\left[\frac{\bar{x}_1}{\bar{x}}\exp\left[\frac{(\bar{x}_1-\bar{x})}{(\bar{x}_1+\bar{x})+2C_x}\right]\right]$	$t_{pro2}=[k_3\bar{y}+k_4]\left[\frac{\bar{x}_1}{\bar{x}}\right]\exp\left[\frac{(\bar{x}_1-\bar{x})}{(\bar{x}_1+\bar{x})+2C_x}\right]$
3	1	β_x	1	$t_{Pro1}=[k_1\bar{y}+k_2]\left[\frac{\bar{x}_1}{\bar{x}}\exp\left[\frac{(\bar{x}_1-\bar{x})}{(\bar{x}_1+\bar{x})+2\beta_x}\right]\right]$	$t_{pro2}=[k_3\bar{y}+k_4]\left[\frac{\bar{x}_1}{\bar{x}}\right]\exp\left[\frac{(\bar{x}_1-\bar{x})}{(\bar{x}_1+\bar{x})+2\beta_x}\right]$
4	1	0	0	$t_{Pro1}=[k_1\bar{y}+k_2]\left[\frac{\bar{x}}{\bar{x}_1}\exp\left[\frac{(\bar{x}_1-\bar{x})}{(\bar{x}_1+\bar{x})}\right]\right]$	$t_{pro2}=[k_3\bar{y}+k_4]\left[\frac{\bar{x}}{\bar{x}_1}\right]\exp\left[\frac{(\bar{x}_1-\bar{x})}{(\bar{x}_1+\bar{x})}\right]$
5	1	C_x	0	$t_{Pro1}=[k_1\bar{y}+k_2]\left[\frac{\bar{x}}{\bar{x}_1}\exp\left[\frac{(\bar{x}_1-\bar{x})}{(\bar{x}_1+\bar{x})+2C_x}\right]\right]$	$t_{pro2}=[k_3\bar{y}+k_4]\left[\frac{\bar{x}}{\bar{x}_1}\right]\exp\left[\frac{(\bar{x}_1-\bar{x})}{(\bar{x}_1+\bar{x})+2C_x}\right]$
6	1	β_x	0	$t_{Pro1}=[k_1\bar{y}+k_2]\left[\frac{\bar{x}}{\bar{x}_1}\exp\left[\frac{(\bar{x}_1-\bar{x})}{(\bar{x}_1+\bar{x})+2\beta_x}\right]\right]$	$t_{pro2}=[k_3\bar{y}+k_4]\left[\frac{\bar{x}}{\bar{x}_1}\right]\exp\left[\frac{(\bar{x}_1-\bar{x})}{(\bar{x}_1+\bar{x})+2\beta_x}\right]$
7	1	0	½	$t_{Pro1}=[k_1\bar{y}+k_2]\frac{1}{2}\left[\frac{\bar{x}_1}{\bar{x}}\left\{2-\exp\left[\frac{(\bar{x}-\bar{x}_1)}{(\bar{x}_1+\bar{x})}\right]\right\}+\frac{\bar{x}}{\bar{x}_1}\exp\left[\frac{(\bar{x}_1-\bar{x})}{(\bar{x}_1+\bar{x})}\right]\right]$	$t_{pro2}=[k_3\bar{y}+k_4]\left[\frac{1}{2}\left\{\frac{\bar{x}_1}{\bar{x}}+\frac{\bar{x}}{\bar{x}_1}\right\}\right]\exp\left[\frac{\bar{x}_1-\bar{x}}{\bar{x}_1+\bar{x}}\right]$
8	1	C_x	½	$t_{Pro1}=[k_1\bar{y}+k_2]\frac{1}{2}\left[\frac{\bar{x}_1}{\bar{x}}\left\{2-\exp\left[\frac{(\bar{x}-\bar{x}_1)}{(\bar{x}_1+\bar{x})+2C_x}\right]\right\}+\frac{\bar{x}}{\bar{x}_1}\exp\left[\frac{(\bar{x}_1-\bar{x})}{(\bar{x}_1+\bar{x})+2C_x}\right]\right]$	$t_{pro2}=[k_3\bar{y}+k_4]\left[\frac{1}{2}\left\{\frac{\bar{x}_1}{\bar{x}}+\frac{\bar{x}}{\bar{x}_1}\right\}\right]\exp\left[\frac{(\bar{x}_1-\bar{x})}{(\bar{x}_1+\bar{x})+2C_x}\right]$
9	1	β_x	½	$t_{Pro1}=[k_1\bar{y}+k_2]\frac{1}{2}\left[\frac{\bar{x}_1}{\bar{x}}\left\{2-\exp\left[\frac{(\bar{x}-\bar{x}_1)}{(\bar{x}_1+\bar{x})+2\beta_x}\right]\right\}+\frac{\bar{x}}{\bar{x}_1}\exp\left[\frac{(\bar{x}_1-\bar{x})}{(\bar{x}_1+\bar{x})+2\beta_x}\right]\right]$	$t_{pro2}=[k_3\bar{y}+k_4]\left[\frac{1}{2}\left\{\frac{\bar{x}_1}{\bar{x}}+\frac{\bar{x}}{\bar{x}_1}\right\}\right]\exp\left[\frac{(\bar{x}_1-\bar{x})}{(\bar{x}_1+\bar{x})+2\beta_x}\right]$
10	ρ	β_x	½	$t_{Pro1}=[k_1\bar{y}+k_2]\frac{1}{2}\left[\frac{\bar{x}_1}{\bar{x}}\left\{2-\exp\left[\frac{\rho(\bar{x}-\bar{x}_1)}{\rho(\bar{x}_1+\bar{x})+2\beta_x}\right]\right\}+\frac{\bar{x}}{\bar{x}_1}\exp\left[\frac{\rho(\bar{x}_1-\bar{x})}{\rho(\bar{x}_1+\bar{x})+2\beta_x}\right]\right]$	$t_{pro2}=[k_3\bar{y}+k_4]\left[\frac{1}{2}\left\{\frac{\bar{x}_1}{\bar{x}}+\frac{\bar{x}}{\bar{x}_1}\right\}\right]\exp\left[\frac{\rho(\bar{x}_1-\bar{x})}{\rho(\bar{x}_1+\bar{x})+2\beta_x}\right]$
11	ρ	C_x	½	$t_{Pro1}=[k_1\bar{y}+k_2]\frac{1}{2}\left[\frac{\bar{x}_1}{\bar{x}}\left\{2-\exp\left[\frac{\rho(\bar{x}-\bar{x}_1)}{\rho(\bar{x}_1+\bar{x})+2C_x}\right]\right\}+\frac{\bar{x}}{\bar{x}_1}\exp\left[\frac{\rho(\bar{x}_1-\bar{x})}{\rho(\bar{x}_1+\bar{x})+2C_x}\right]\right]$	$t_{pro2}=[k_3\bar{y}+k_4]\left[\frac{1}{2}\left\{\frac{\bar{x}_1}{\bar{x}}+\frac{\bar{x}}{\bar{x}_1}\right\}\right]\exp\left[\frac{\rho(\bar{x}_1-\bar{x})}{\rho(\bar{x}_1+\bar{x})+2C_x}\right]$

The MSE of the proposed estimator is given in the following theorem.

Theorem 2. An estimator for the population mean defined in Eq (20) in the case of two-phase sampling with single auxiliary variables will have its MSE equation given as

$$MSE\left(t_{pro2}\right)_{\min} \cong \bar{Y}^2\left(1-\frac{A_pD_p^2+B_pC_p^2-2C_pD_pE_p+B_p+2B_pC_p+D_p^2-2D_pE_p}{A_pB_p-E_p^2+B_p}\right), \tag{21}$$

where

$$A_p=\lambda C_y^2+\left\{\lambda\left(\delta_1^2+2\delta_3\right)-\lambda_1\left(\delta_1^2+2\delta_2-2\delta_4\right)\right\}C_x^2+4\delta_1\left(\lambda-\lambda_1\right)C_{yx},$$

$$B_p=1+\left\{\lambda\left(\delta_1^2+2\delta_3\right)-\lambda_1\left(\delta_1^2+2\delta_2-2\delta_4\right)\right\}C_x^2,\ \ C_p=\delta_1\left(\lambda-\lambda_1\right)C_{yx}+\left\{\delta_3\lambda-\lambda_1\left(\delta_2-\delta_4\right)\right\}C_x^2,$$

$$D_p=1+\left\{\delta_3\lambda-\lambda_1\left(\delta_2-\delta_4\right)\right\}C_x^2 \ \text{ and}$$

$$E_p=1+\left\{\lambda\left(\delta_1^2+2\delta_3\right)-\lambda_1\left(\delta_1^2+2\delta_2-2\delta_4\right)\right\}C_x^2+2\delta_1\left(\lambda-\lambda_1\right)C_{yx}.$$

4. Efficiency comparison

4.1. For the first proposed estimator

Our proposed estimator Tpro is better than the existing estimators subject to the following conditions:

Condition (i): By comparing Eqs (2) and (19), $MSE\left(t_{pro1}\right)<MSE\left(t_R\right)$ if

$$\lambda C_y^2+\lambda_1C_x^2(1-2\psi)+\Re_1-1>0, \tag{22}$$

where $\Re_1=\dfrac{A_{pr}D_{pr}^2+B_{pr}C_{pr}^2-2C_{pr}D_{pr}E_{pr}}{A_{pr}B_{pr}-E_{pr}^2}$.

Condition (ii): By comparing Eqs (4) and (19), $MSE\left(t_{pro1}\right)<MSE\left(t_P\right)$ if

$$\lambda C_y^2+\lambda_1C_x^2(1+2\psi)+\Re_1-1>0. \tag{23}$$

Condition (iii): By comparing Eqs (7) and (19), $MSE\left(t_{pro1}\right)<MSE\left(t_{SVR}\right)$ if

$$\lambda C_y^2+\lambda_1C_x^2\left(\frac{1}{4}-\psi\right)+\Re_1-1>0. \tag{24}$$

Condition (iv): By comparing Eqs (8) and (19), $MSE\left(t_{pro1}\right)<MSE\left(t_{SVP}\right)$ if

$$\lambda C_y^2+\lambda_1C_x^2\left(\frac{1}{4}+\psi\right)+\Re_1-1>0. \tag{25}$$

Condition (v): By comparing Eqs (11) and (19), $MSE\left(t_{pro1}\right)<MSE\left(t_{GR}\right)$ if

$$\lambda C_y^2+\lambda'' gC_x^2\left(\frac{g}{4}-\psi\right)-\lambda''\frac{A^2}{4}+\Re_1-1>0. \tag{26}$$

Condition (vi): By comparing Eqs (12) and (19), $MSE\left(t_{pro1}\right)<MSE\left(t_{GP}\right)$ if

$$\lambda C_y^2 + \lambda'' g C_x^2 \left(\frac{g}{4} + \psi\right) - \lambda'' \frac{A^2}{4} + \Re_1 - 1 > 0\,. \tag{27}$$

Condition (vii): By comparing Eqs (14) and (19), $MSE\left(t_{pro1}\right) < MSE\left(t_{\mathrm{Re}g}\right)$ if

$$C_y^2\left(\lambda - \lambda_1 \rho^2\right) + \Re_1 - 1 > 0\,. \tag{28}$$

Condition (viii): By comparing Eqs (16) and (19), $MSE\left(t_{pro1}\right) < MSE\left(t_{OC}\right)$ if

$$\nabla_1 + \Re_1 - 1 > 0\,, \tag{29}$$

where $\nabla_1 = \dfrac{C_y^2\left(\lambda - \lambda_1\rho^2\right)\left(1 - \lambda_1\theta^2 C_x^2\right) - \dfrac{\lambda_1^2\theta^4 C_x^4}{4}}{1 + C_y^2\left(\lambda - \lambda_1\rho^2\right)}$.

4.2. For the second proposed estimator

Our proposed estimator t_{pro2} is better than the existing estimators subject to the following conditions:

Condition (i): By comparing Eqs (2) and (21), $MSE\left(t_{pro2}\right) < MSE\left(t_R\right)$ if

$$\lambda C_y^2 + \lambda_1 C_x^2 (1 - 2\psi) + \frac{\Re_2}{\Re_3} - 1 > 0\,, \tag{30}$$

where $\Re_2 = A_p D_p^2 + B_p C_p^2 - 2C_p D_p E_p + B_p + 2B_p C_p + D_p^2 - 2D_p E_p$ and $\Re_3 = A_p B_p - E_p^2 + B_p$.

Condition (ii): By comparing Eqs (4) and (21), $MSE\left(t_{pro2}\right) < MSE\left(t_P\right)$ if

$$\lambda C_y^2 + \lambda_1 C_x^2 (1 + 2\psi) + \frac{\Re_2}{\Re_3} - 1 > 0\,. \tag{31}$$

Condition (iii): By comparing Eqs (7) and (21), $MSE\left(t_{pro2}\right) < MSE\left(t_{SVR}\right)$ if

$$\lambda C_y^2 + \lambda_1 C_x^2 \left(\frac{1}{4} - \psi\right) + \frac{\Re_2}{\Re_3} - 1 > 0\,. \tag{32}$$

Condition (iv): By comparing Eqs (8) and (21), $MSE\left(t_{pro2}\right) < MSE\left(t_{SVP}\right)$ if

$$\lambda C_y^2 + \lambda_1 C_x^2 \left(\frac{1}{4} + \psi\right) + \frac{\Re_2}{\Re_3} - 1 > 0\,. \tag{33}$$

Condition (v): By comparing Eqs (11) and (21), $MSE\left(t_{pro2}\right) < MSE\left(t_{GR}\right)$ if

$$\lambda C_y^2 + \lambda'' g C_x^2 \left(\frac{g}{4} - \psi\right) - \lambda'' \frac{A^2}{4} + \frac{\Re_2}{\Re_3} - 1 > 0\,. \tag{34}$$

Condition (vi): By comparing Eqs (12) and (21), $MSE\left(t_{pro2}\right) < MSE\left(t_{GP}\right)$ if

$$\lambda C_y^2 + \lambda'' g C_x^2 \left(\frac{g}{4} + \psi \right) - \lambda'' \frac{A^2}{4} + \frac{\Re_2}{\Re_3} - 1 > 0. \quad (35)$$

Condition (vii): By comparing Eqs (14) and (21), $MSE\left(t_{pro2}\right) < MSE\left(t_{\mathrm{Re}g}\right)$ if

$$C_y^2\left(\lambda - \lambda_1 \rho^2\right) + \frac{\Re_2}{\Re_3} - 1 > 0. \quad (36)$$

Condition (viii): By comparing Eqs (16) and (21), $MSE\left(t_{pro2}\right) < MSE\left(t_{OC}\right)$ if

$$\nabla_1 + \frac{\Re_2}{\Re_3} - 1 > 0. \quad (37)$$

5. Numerical study

In this section, the MSEs of the proposed estimators, along with other existing estimators for real data sets, are given.

In all these data sets, assume the following notations: N → population size; n_1 → first-phase sample size; n → second-phase sample size; Y → population mean of variable y; X → population mean of variable x; C_y → population coefficient of variation of y; C_x → population coefficient of variations of x; ρ → population correlation coefficient between y and x.

Population 1. [18] Where X is the population (in 1000s) in 1920 and Y is the population (in 1000s) in 1930 of 49 US cities.

N	n1	n	Y	X	C_y	C_x	ρ
49	24	10	127.7959	103.1429	0.9634	1.0122	0.9817

Population 2. [19] Y: The total acreage planted with wheat in 34 communities in 1974. X: The total area planted with wheat (in acres) across 34 villages in 1971.

N	n1	n	Y	X	C_y	C_x	ρ
34	13	6	856.41	208.88	0.8561	0.7205	0.4491

Population 3. [20] Y: The number of tube wells. X: The 69 communities' net irrigated area in hectares.

N	n1	n	Y	X	C_y	C_x	ρ
69	15	7	135.26	345.75	0.842	0.848	0.922

Population 4. [20] Y: The average number of hours spent sleeping. X: The person's age.

N	n1	n	Y	X	C_y	C_x	ρ
30	10	4	6.377	66.933	0.163	0.144	-0.8669

Population 5. [21]. Y: The number of agriculture workers in 1971. X: The number of agriculture workers in 1961.

N	n1	n	Y	X	C_y	C_x	P
278	25	12	39.06	25.11	1.4451	1.6198	0.7213

Table 2 shows the MSE values of various estimators of the population mean in the two-phase sampling with a single auxiliary. The software R version 4.4.2 was utilized for all numerical analysis. Three specific scenarios are introduced—the first with no transformation (1,0), the second utilizing Cz (1, Cz), and the third incorporating both Ryz and Cz (Ryz, Cz)—for each of the two groups of estimators. As shown in the table, the MSEs for all estimators in the proposed groups were considerably lower than those of the classical, ratio, and other estimators. Therefore, we can conclude that the proposed groups of estimators for the population mean in two-phase sampling with an auxiliary variable exhibit lower variances than all the existing estimators for the same parameters.

Table 3 shows the percentage relative efficiency (PRE) values of various estimators for the population mean in two-phase sampling with a single auxiliary variable. We introduce three distinct scenarios—the first with no transformation (1,0), the second utilizing Cz (1,Cz), and the third incorporating both Ryz and Cz (Ryz, Cz)—for each of the two groups of estimators. As shown in the table, the PREs of all estimators belonging to the proposed groups exhibit a significantly higher level of efficiency compared to the classical, ratio, and other estimators available in the literature. Thus, we can confidently conclude that the proposed groups of estimators for the population mean in two-phase sampling with an auxiliary variable offer exceptional efficiency when compared to all existing estimators for the same parameters.

Table 2. MSEs of the estimators under different populations in two-phase sampling with a single auxiliary variable.

Estimators	Pop-I	Pop-II	Pop-III	Pop-IV	Pop-V
t_0	1206.468	73780.58	1665.785	0.47943	254.1521
t_R	897.4455	72564.16	1094.546	0.5554184	212.3113
t_P	2226.915	111176.3	3609.537	2.211102	587.5549
t_{SVR}	963.0284	68649.96	1208.601	0.2914667	196.7865
t_{SVP}	1627.763	87956.02	2466.097	1.119308	384.4083
t_{GR}	954.4859	67885.15	1387.843	0.1654431	194.0551
t_{GP}	1367.202	64018.73	1717.514	0.2167253	357.6551
$t_{\mathrm{Re}g}$	895.9134	68629.51	1089.727	0.2898695	193.7843
t_{OC}	844.7858	62373.91	1018.88	0.28241	168.1552
$\left(t_{pro1}\right)_{1,0}^{0.5}$	48.68735	6124.312	95.30145	0.01487197	18.65086
$\left(t_{pro1}\right)_{1,C_x}^{0.5}$	47.50149	6081.405	94.78652	0.01480804	16.12708
$\left(t_{pro1}\right)_{\rho,C_x}^{0.5}$	47.47971	6029.342	94.74336	0.01494622	15.25888
$\left(t_{pro2}\right)_{1,0}^{0.5}$	57.78406	6751.424	117.1208	0.01497448	28.7724
$\left(t_{pro2}\right)_{1,C_x}^{0.5}$	51.81201	6705.387	116.5365	0.01491032	25.35458
$\left(t_{pro2}\right)_{\rho,C_x}^{0.5}$	51.70824	6649.523	116.4875	0.01504901	24.18399

Table 3. Percentage relative efficiency (PRE) values of the estimators under different populations in two-phase sampling with a single auxiliary variable.

Estimators	Pop-I	Pop-II	Pop-III	Pop-IV	Pop-V
t_0	100	100	100	100	100
t_R	134.4335	101.6763	152.1896	58.58052	119.7073
t_P	54.17664	66.36359	46.14954	130.0708	43.25589
t_{SVR}	125.2785	107.4736	137.8275	77.27718	129.1512
t_{SVP}	74.11813	83.88349	67.54742	121.2234	66.11516
t_{GR}	126.3997	108.6844	120.0269	6.364461	130.9691
t_{GP}	88.24353	115.2484	96.98812	10.89628	71.06068
t_{Reg}	134.6634	107.5056	152.8626	130.0789	131.1521
t_{OC}	142.8134	118.2876	163.4918	130.7012	151.1414
$\left(t_{pro1}\right)_{1,0}^{0.5}$	2477.99	1204.716	1747.911	1579.749	1362.683
$\left(t_{pro1}\right)_{1,C_x}^{0.5}$	2539.852	1213.216	1757.407	1586.569	1575.934
$\left(t_{pro1}\right)_{\rho,C_x}^{0.5}$	2541.017	1223.692	1758.208	1571.9	1665.602
$\left(t_{pro2}\right)_{1,0}^{0.5}$	2087.89	1092.815	1422.279	1568.934	883.3193
$\left(t_{pro2}\right)_{1,C_x}^{0.5}$	2133.748	1100.318	1429.411	1575.686	1002.391
$\left(t_{pro2}\right)_{\rho,C_x}^{0.5}$	2134.61	1109.562	1430.012	1561.164	1050.911

6. Simulation study

In two-phase sampling with a single auxiliary variable, the population was divided into two phases. In the first phase, a sample was selected based on the auxiliary variable. In the second phase, one sample was selected from the first-phase sample. This design is commonly used to reduce sampling costs and increase efficiency.

Several estimators, including regression, ratio, and difference estimators, have been proposed to estimate the population mean by using this architecture. Simulations were performed to assess estimator performance. The general procedure for performing simulations of estimators of the finite population mean in two-phase sampling using a single auxiliary is presented below.

Algorithm: Simulation for estimator evaluation

i. Initialization of parameters
   Define N,
   Define auxiliary variables and their distribution functions,
   Define population mean and variance covariance matrix.
ii. Generate population
   Based on the above parameters we generate a data set of size N.

iii. Draw a first-phase sample and sample statistics
From this dataset randomly select an initial sample S_1 called first-phase sample.
Calculate means, correlation coefficients, coefficient of variations, and other statistics for both variables y and x.

iv. Draw a second-phase sample
Draw a second-phase sample $S_2<S_2$.

v. Calculate population estimators' values
We calculate the values for the estimators of the population mean from this sample utilizing the statistics obtained from the first-phase sample as auxiliary information.
Store these values of estimators for performance evaluation.

vi. Repeat steps iii-v $\Re$ times to obtain the distribution of the estimated population mean.

vii. Evaluate estimators' performance
Calculate the distribution of estimators over $\Re$ repetition.
And obtain the MSE of the estimators

$$MSE(t_i) = \frac{1}{\Re}\sum_{r=1}^{\Re}(t_i - \bar{Y})^2.$$

viii. Analyze results
Repeat steps iii-vii for different sample sizes.
Compare MSEs across different scenarios to evaluate estimators' performance.

By conducting simulations, one can compare the performance of different estimators and choose the estimator that performs best under various scenarios.

Table 4 displays the MSE values of various estimators for the finite population mean in two-phase sampling with a single auxiliary variable on the simulated data. In this study, we introduced two novel families of estimators, tpro1 and tpro2, and presented three special cases for each of these estimators.

An analysis of the results indicates that our proposed estimators exhibit significantly lower MSEs compared to all estimators selected from the existing literature for the same parameter in similar scenarios. Moreover, for both of our proposed estimators, the estimators with u=Ryx and v=Cx consistently demonstrate smaller variances across all selected sample sizes. In addition, the values of the proposed estimators remain relatively stable across varying sample sizes.

Based on these findings, we confidently conclude that our proposed families of estimators are highly efficient and consistent for the population mean in two-phase sampling when compared with traditional estimators such as ratio, product, and other similar estimators.

Table 4. MSEs of the estimators for finite population mean in two-phase sampling.

Estimators	n=10	n=20	n=50	n=100	n=200
t_0	0.11765	0.0674538	0.0297463	0.0146399	0.006835741
t_R	0.112473	0.0595570	0.0258601	0.0129036	0.005915756
t_P	0.356272	0.2496402	0.1294851	0.0631248	0.0314097
t_{SVR}	0.085407	0.0415785	0.0157659	0.0079163	0.003416362
t_{SVP}	0.207807	0.1366718	0.0676182	0.0330339	0.01616443
t_{GR}	0.031958	0.0443500	0.0156862	0.0078855	0.003399093
t_{GP}	0.031805	0.0445138	0.0157166	0.0078912	0.003400891
$t_{\mathrm{Re}g}$	0.085769	0.0415185	0.0156956	0.0078885	0.003398686
t_{OC}	0.076553	0.0390543	0.0181793	0.0085051	0.004243778
$\left(t_{pro1}\right)_{1,0}^{0.5}$	0.023453	0.0146365	0.0065068	0.0031587	0.001479118
$\left(t_{pro1}\right)_{1,C_x}^{0.5}$	0.019830	0.0125145	0.0056139	0.0027609	0.001276948
$\left(t_{pro1}\right)_{\rho,C_x}^{0.5}$	0.0196495	0.0120682	0.0056766	0.0027526	0.001261685
$\left(t_{pro2}\right)_{1,0}^{0.5}$	0.022643	0.0143328	0.0064367	0.0031449	0.001475519
$\left(t_{pro2}\right)_{1,C_x}^{0.5}$	0.019288	0.0122986	0.0055556	0.0027491	0.001273876
$\left(t_{pro2}\right)_{\rho,C_x}^{0.5}$	0.019124	0.0118482	0.0056223	0.0027394	0.00125787

Table 5 displays the PRE values of estimators for the finite population mean in two-phase sampling with a single auxiliary variable based on simulated data. We introduce two families of estimators, t_{pro1} and t_{pro2}, from which numerous estimators can be obtained by applying different values of the generalizing constants u, v, and ω. However, we provide only three unique cases for each estimator. Upon examining the results, it is clear that the proposed estimators outperform all previously established estimators in the literature for the same parameter under comparable circumstances. Furthermore, the estimators with $u=\rho_{yx}$ and $v=C_x$ from both proposed families demonstrate smaller variances across all chosen sample sizes. Additionally, the estimators' values exhibit consistency across varying sample sizes, signifying that the proposed families are efficient and consistent estimators compared with conventional ratios, products, and other estimators for the population mean in two-phase sampling. In conclusion, the proposed estimators are highly efficient and consistent compared to the previously presented estimators.

Table 5. PREs of the estimators relative to the usual estimator of the mean.

Estimators	n=10	n=20	n=50	n=100	n=200
t_0	100	100	100	100	100
t_R	104.6039	113.2593	15.0276	113.4563	115.5514
t_P	33.02292	27.02043	22.97275	23.19201	21.76316
t_{SVR}	137.753	162.2327	188.6743	184.9327	200.0883
t_{SVP}	56.61566	49.35462	43.99152	44.31777	42.2888
t_{GR}	368.1474	152.0943	189.6341	185.6548	201.1049
t_{GP}	367.1043	151.5348	189.2668	185.5208	200.9986
$t_{\mathrm{Re}g}$	137.1721	162.4671	189.5195	185.5857	201.1289
t_{OC}	153.6846	172.7181	163.6273	172.1314	161.0768
$\left(t_{pro1}\right)_{1,0}^{0.5}$	501.6541	460.5439	457.1578	465.9208	462.1497
$\left(t_{pro1}\right)_{1,C_x}^{0.5}$	593.2857	536.4649	517.8296	532.5756	543.2696
$\left(t_{pro1}\right)_{\rho,C_x}^{0.5}$	598.7415	558.9393	531.9137	531.8565	567.5748
$\left(t_{pro2}\right)_{1,0}^{0.5}$	519.6022	470.3019	462.1383	467.9549	463.277
$\left(t_{pro2}\right)_{1,C_x}^{0.5}$	609.9509	545.885	523.2615	534.8668	544.5798
$\left(t_{pro2}\right)_{\rho,C_x}^{0.5}$	615.1905	569.3159	537.0446	534.4118	569.2958

7. Conclusions

The utilization of auxiliary information has proven to be highly effective in improving efficiency; however, such information is not always available during the design and estimation processes or can be prohibitively expensive. Two-phase sampling procedures are frequently employed to overcome this challenge. In this approach, a sample is first obtained from the population, and auxiliary information is collected from this sample. In the second phase, a smaller sample is drawn from the first-phase sample, and both the primary and auxiliary variables are measured in this subsample. Various estimators such as ratio and product estimators are commonly used to estimate the finite population mean in two-phase sampling scenarios with a single auxiliary variable. Inspired by the work of [7,8,15], we have proposed two classes of estimators for the population mean and derived their expressions for bias and mean squared error (MSE) up to the first order of approximation. We identified the theoretical conditions under which the proposed estimators are more efficient than some existing estimators in terms of having less variance. The proposed estimator families were compared with existing mean estimators on both real and simulated data based on their MSE and percentage relative efficiency (PRE).

The results of our comparison strongly support the claim that the proposed estimator families are significantly more efficient than existing methods for estimating the finite population mean in a two-phase sampling scenario.

Author contributions

Khazan Sher: Conceptualization, project administration, writing–original draft, writing–review and editing; Muhammad Ameeq, Sidra Naz, Basem A. Alkhaleel: Conceptualization, project administration, investigation, writing–original draft, writing–review and editing; Muhammad Muneeb Hassan, Olyan Albalawi: Conceptualization, project administration, investigation, writing–original draft, writing–review and editing. All authors have read and agreed to the published version of the manuscript.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Researchers Supporting Project number (RSPD2025R630), King Saud University, Riyadh, Saudi Arabia.

Conflict of interest

The authors declare that there are no conflicts of interest.

References

1. H. P. Singh, G. K. Vishwakarma, Modified exponential ratio and product estimators for finite population mean in double sampling, *Aust. J. Stat.*, **36** (2007), 217–225. https://doi.org/10.17713/ajs.v36i3.333
2. L. K. Grover, P. Kaur, G. K. Vishawkarma, Product type exponential estimators of population mean under linear transformation of auxiliary variable in simple random sampling, *Appl. Math. Comput.*, **219** (2012), 1937–1946. https://doi.org/10.1016/j.amc.2012.08.036
3. M. Noor-ul-Amin, M. Hanif, Some exponential estimators in survey sampling, *Pak. J. Stat.*, **28** (2012).
4. K. Sher, M. Ameeq, M. M. Hassan, O. Albalawi, A. Afzal, Development of improved estimators of finite population mean in simple random sampling with dual auxiliaries and its application to real world problems, *Heliyon*, **10** (2024). https://doi.org/10.1016/j.heliyon.2024.e30991
5. S. Muneer, J. Shabbir, A. Khalil, Estimation of finite population mean in simple random sampling and stratified random sampling using two auxiliary variables, *Commun. Stat.-Theor. M.*, **46** (2017), 2181–2192. https://doi.org/10.1080/03610926.2015.1035394
6. S. Choudhury, B. K. Singh, A class of product-cum-dual to product estimators of the population mean in survey sampling using auxiliary information, *Asian J. Math. Stat.*, **6** (2012). https://doi.org/10.3923/ajms.2012

7. A. K. Sabat, R. K. Sahoo, L. N. Sahoo, Composite ratio estimators in a two-phase sampling using multiple additional supplementary variables, *Far East J. Appl. Math.*, **117** (2024), 49–66. https://doi.org/10.17654/0972096024003
8. C. Di Gravio, J. S. Schildcrout, R. Tao, Efficient designs and analysis of two-phase studies with longitudinal binary data, *Biometrics*, **80** (2024). https://doi.org/10.1093/biomtc/ujad010
9. T. Zaman, C. Kadilar, New class of exponential estimators for finite population mean in two-phase sampling, *Commun. Stat.-Theor. M.*, **50** (2021), 874–889. https://doi.org/10.1080/03610926.2019.1643480
10. Y. Tato, B. K. Singh, Exponential chain dual to ratio cum dual to product estimator for finite population mean in double sampling scheme, *Appl. Appl. Math.*, **12** (2017), 4.
11. P. Misra, Regression type double sampling estimator of population mean using auxiliary information, *J. Reliab. Stat. Stud.*, 2018, 21–28.
12. M. Kumar, S. Bahl, Class of dual to ratio estimators for double sampling, *Stat. Pap.*, **47** (2006), 319. https://doi.org/10.1007/s00362-005-0291-6
13. S. K. Yadav, S. Gupta, S. S. Mishra, A. K. Shukla, Modified ratio and product estimators for estimating population mean in two-phase sampling, *Am. J. Oper. Res.*, **6** (2016), 61–68. https://doi.org/10.5923/j.ajor.20160603.02
14. N. Ozgul, H. Cingi, A new class of exponential regression cum ratio estimator in two phase sampling, *Hacet. J. Math. Stat.*, **43** (2014), 131–140.
15. J. Shabbir, S. Gupta, R. Onyango, On using the conventional and nonconventional measures of the auxiliary variable for mean estimation, *Math. Probl. Eng.*, **2021** (2021), 3741620. https://doi.org/10.1155/2021/3741620
16. S. Gupta, J. Shabbir, On improvement in estimating the population mean in simple random sampling, *J. Appl. Stat.*, **35** (2008), 559–566. https://doi.org/10.1080/02664760701835839
17. J. Shabbir, S. Gupta, On estimating finite population mean in simple and stratified random sampling, *Commun. Stat.-Theor. M.*, **40** (2010), 199–212. https://doi.org/10.1080/03610920903411259
18. W. G. Cochran, *Sampling techniques*, 3Eds., New York: John Wiley & Sons, 1977.
19. D. Singh, F. S. Chaudhary, *Theory and analysis of sample survey design*, New Delhi: New Age International Publishers, 1986.
20. R. Singh, N. S. Mangat, *Elements of survey sampling*, Springer Science & Business Media, **15** (2013).
21. A. K. Das, *Contributions to the theory of sampling strategies based on auxiliary information*, Ph.D. Thesis, Bidhan Chandra Krishi Vishwavidyalaya, 1988.

Appendix

Proof 1. To obtain the MSE of the estimator given in Eq (18), we can express Eq (18) in terms of error as follows:

$$t_{\text{Pro1}} = \left[k_1\bar{Y}\left(1+\xi_0\right)+k_2\right]\left[\begin{array}{l}\omega\dfrac{\bar{X}\left(1+\xi_2\right)}{\bar{X}\left(1+\xi_1\right)}\left\{2-\exp\left[\dfrac{u\left(\bar{X}\left(1+\xi_1\right)-\bar{X}\left(1+\xi_2\right)\right)}{u\left(\bar{X}\left(1+\xi_1\right)+\bar{X}\left(1+\xi_2\right)\right)+2v}\right]\right\}\\+\left(1-\omega\right)\dfrac{\bar{X}\left(1+\xi_1\right)}{\bar{X}\left(1+\xi_2\right)}\exp\left[\dfrac{u\left(\bar{X}\left(1+\xi_2\right)-\bar{X}\left(1+\xi_1\right)\right)}{u\left(\bar{X}\left(1+\xi_1\right)+\bar{X}\left(1+\xi_2\right)\right)+2v}\right]\end{array}\right] \tag{38}$$

or

$$t_{\text{Pro1}} = \left[k_1\bar{Y}\left(1+\xi_0\right)+k_2\right]\left[\begin{array}{l}\omega\left(1-\xi_1+\xi_2-\xi_1\xi_2+\xi_1^2\right)\left\{2-\exp\left[\frac{\eta\left(\xi_1-\xi_2\right)}{2}-\frac{\eta^2\left(\xi_1^2-\xi_2^2\right)}{4}\right]\right\}\\ +\left(1-\omega\right)\left(1+\xi_1-\xi_2-\xi_1\xi_2+\xi_2^2\right)\exp\left[\frac{\eta\left(\xi_2-\xi_1\right)}{2}-\frac{\eta^2\left(\xi_2^2-\xi_1^2\right)}{4}\right]\end{array}\right], \tag{39}$$

where $\eta = \frac{u\bar{X}}{u\bar{X}+v}$.

Now, expanding with the exponential series and Taylor series, we have

$$t_{\text{Pro1}} = \left[k_1\bar{Y}\left(1+\xi_0\right)+k_2\right]\left(1+\vartheta_1\xi_1-\vartheta_1\xi_2-\vartheta_2\xi_1\xi_2+\vartheta_3\xi_1^2+\vartheta_4\xi_2^2\right) \tag{40}$$

or

$$\begin{aligned} t_{\text{Pro1}}-\bar{Y} = {} & k_1\bar{Y}\left(1+\xi_0+\vartheta_1\xi_1+\vartheta_1\xi_0\xi_1-\vartheta_1\xi_2-\vartheta_1\xi_0\xi_2-\vartheta_2\xi_1\xi_2+\vartheta_3\xi_1^2+\vartheta_4\xi_2^2\right)\\ & +k_2\left(1+\vartheta_1\xi_1-\vartheta_1\xi_2-\vartheta_2\xi_1\xi_2+\vartheta_3\xi_1^2+\vartheta_4\xi_2^2\right)-\bar{Y}. \end{aligned} \tag{41}$$

Taking expectation on both sides of Eq (41), the bias of the estimator is given as

$$Bias\left(t_{\text{Pro1}}\right) = k_1\bar{Y}\left[1+\vartheta_1\left(\lambda-\lambda_1\right)C_{yx}-\left(\vartheta_2\lambda_1-\vartheta_3\lambda-\vartheta_4\lambda_1\right)C_x^2\right]+k_2\left[1-\left(\vartheta_2\lambda_1-\vartheta_3\lambda-\vartheta_4\lambda_1\right)C_x^2\right]-\bar{Y}. \tag{42}$$

Now by taking the square of Eq (42), we get the following expression:

$$\begin{aligned} \left(t_{\text{Pro1}}-\bar{Y}\right)^2 = {} & k_1^2\bar{Y}^2\left[1+\xi_0^2+\left(\vartheta_1^2+2\vartheta_3\right)\xi_1^2+\left(\vartheta_1^2+2\vartheta_4\right)\xi_2^2+4\vartheta_1\xi_0\xi_1-4\vartheta_1\xi_0\xi_2-2\left(\vartheta_1^2+\vartheta_2\right)\xi_1\xi_2\right]\\ & +k_2^2\left[1+\left(\vartheta_1^2+2\vartheta_3\right)\xi_1^2+\left(\vartheta_1^2+2\vartheta_4\right)\xi_2^2-2\left(\vartheta_1^2+\vartheta_2\right)\xi_1\xi_2\right]\\ & -2k_1\bar{Y}^2\left[1+\vartheta_3\xi_1^2+\vartheta_4\xi_2^2+\vartheta_1\xi_0\xi_1-\vartheta_1\xi_0\xi_2-\vartheta_2\xi_1\xi_2\right]-2k_2\bar{Y}\left[1+\vartheta_3\xi_1^2+\vartheta_4\xi_2^2-\vartheta_2\xi_1\xi_2\right]\\ & +2k_1k_2\bar{Y}\left[1+\left(\vartheta_1^2+2\vartheta_3\right)\xi_1^2+\left(\vartheta_1^2+2\vartheta_4\right)\xi_2^2+2\vartheta_1\xi_0\xi_1-2\vartheta_1\xi_0\xi_2-2\left(\vartheta_1^2+\vartheta_2\right)\xi_1\xi_2\right]+\bar{Y}^2. \end{aligned} \tag{43}$$

Taking expectations on both sides of Eq (43), we get the MSE expression as

$$\begin{aligned} MSE\left(t_{\text{Pro1}}\right) = {} & k_1^2\bar{Y}^2\left(1+\lambda C_y^2+\Delta_1\lambda C_x^2+\Delta_2\lambda_1 C_x^2+4\vartheta_1\lambda C_{yx}\varepsilon_1-4\vartheta_1\lambda_1 C_{yx}-\Delta_3\lambda_1 C_x^2\right)\\ & +k_2^2\left[1+\Delta_1\lambda C_x^2+\Delta_2\lambda_1 C_x^2-\Delta_3\lambda_1 C_x^2\right]-2k_1\bar{Y}^2\left(1+\vartheta_3\lambda C_x^2+\vartheta_4\lambda_1 C_x^2+\vartheta_1\lambda C_{yx}\varepsilon_1-\vartheta_1\lambda_1 C_{yx}-\vartheta_2\lambda_1 C_x^2\right)\\ & -2k_2\bar{Y}\left[1+\vartheta_3\varepsilon_1^2+\vartheta_4\varepsilon_2^2-\vartheta_2\lambda_1 C_x^2\right]+2k_1k_2\bar{Y}\left[1+\Delta_1\lambda C_x^2+\Delta_2\lambda_1 C_x^2+2\vartheta_1\lambda C_{yx}\varepsilon_1-2\vartheta_1\lambda_1 C_{yx}-\Delta_3\lambda_1 C_x^2\right]+\bar{Y}^2, \end{aligned}$$

where $\Delta_1 = \vartheta_1^2+2\vartheta_3$, $\Delta_2 = \vartheta_1^2+2\vartheta_4$ and $\Delta_3 = 2\left(\vartheta_1^2+\vartheta_2\right)$.

Also, the simplified form is given as

$$\begin{aligned} MSE\left(t_{\text{Pro1}}\right) = {} & k_1^2\bar{Y}^2\left[1+\lambda C_y^2+\left(\Delta_1\lambda+\Delta_2\lambda_1-\Delta_3\lambda_1\right)C_x^2+4\vartheta_1\left(\lambda-\lambda_1\right)C_{yx}\right]\\ & +k_2^2\left[1+\left(\Delta_1\lambda+\Delta_2\lambda_1-\Delta_3\lambda_1\right)C_x^2\right]-2k_1\bar{Y}^2\left[1+\left(\vartheta_3\lambda+\vartheta_4\lambda_1-\vartheta_2\lambda_1\right)C_x^2+\vartheta_1\left(\lambda-\lambda_1\right)C_{yx}\right]\\ & -2k_2\bar{Y}\left[1+\left(\vartheta_3\lambda+\vartheta_4\lambda_1-\vartheta_2\lambda_1\right)C_x^2\right]+2k_1k_2\bar{Y}\left[1+\left(\Delta_1\lambda+\Delta_2\lambda_1-\Delta_3\lambda_1\right)C_x^2+2\vartheta_1\left(\lambda-\lambda_1\right)C_{yx}\right]+\bar{Y}^2. \end{aligned} \tag{44}$$

To find the values of k_1 and k_2, we convert the above Eq (44) in the below simplified form

$$MSE\left(t_{\text{Pro1}}\right) = k_1^2\bar{Y}^2A_{pr}+k_2^2B_{pr}-2k_1\bar{Y}^2C_{pr}-2k_2\bar{Y}D_{pr}+2k_1k_2\bar{Y}E_{pr}+\bar{Y}^2. \tag{45}$$

Now differentiating Eq (45) w.r.t. k_1 and k_2 and equating to zero, we get the following two equations:

$$\frac{\partial MSE\left(t_{pro1}\right)}{\partial k_1}=0 \text{ and } \frac{\partial MSE\left(t_{pro1}\right)}{\partial k_2}=0.$$

So, we obtain

$$k_1\bar{Y}^2A_{pr}+k_2\bar{Y}E_{pr}-\bar{Y}^2C_{pr}=0 \tag{46}$$

and

$$k_1\bar{Y}E_{pr}+k_2B_{pr}-\bar{Y}D_{pr}=0. \tag{47}$$

Solving both Eqs (46) and (47) simultaneously, we obtain the following optimum values $k_1=\frac{B_{pr}C_{pr}-D_{pr}E_{pr}}{A_{pr}B_{pr}-E_{pr}^2}$ and $k_2=\frac{\bar{Y}\left(A_{pr}D_{pr}-C_{pr}E_{pr}\right)}{A_{pr}B_{pr}-E_{pr}^2}$. With these values, the minimum MSE of the estimator adopts the following form:

$$MSE\left(t_{\mathrm{Pro1}}\right)_{\min}\cong\bar{Y}^2\left(1-\frac{A_{pr}D_{pr}^2+B_{pr}C_{pr}^2-2C_{pr}D_{pr}E_{pr}}{A_{pr}B_{pr}-E_{pr}^2}\right). \tag{48}$$

Proof 2. The MSE of the proposed estimator given in Eq (20) and its bias are obtained as follows:

$$\begin{aligned} t_{pro2}=&\left(k_3\bar{Y}\left(1+\xi_0\right)+k_4\right)\left[\omega\left(1+\xi_2\right)\left(1+\xi_1\right)^{-1}+\left(1+\omega\right)\left(1+\xi_1\right)\left(1+\xi_2\right)^{-1}\right]\\ &\exp\left[\frac{1}{2}\eta\left(\xi_2-\xi_1\right)\left(1+\frac{1}{2}\eta\left(\xi_2+\xi_1\right)\right)^{-1}\right] \end{aligned} \tag{49}$$

or

$$\begin{aligned} t_{pro2}=&k_3\bar{Y}\left(1+\xi_0+\delta_1\xi_1-\delta_1\xi_2+\delta_1\xi_0\xi_1-\delta_1\xi_0\xi_2-\delta_2\xi_1\xi_2+\delta_3\xi_1^2+\delta_4\xi_2^2\right)\\ &+k_4\left(1+\delta_1\xi_1-\delta_1\xi_2-\delta_2\xi_1\xi_2+\delta_3\xi_1^2+\delta_4\xi_2^2\right), \end{aligned} \tag{50}$$

where

$\delta_1=1-2\omega-\frac{1}{2}\eta$, $\delta_2=1-\frac{1}{4}\eta+\frac{1}{4}\eta^2$, $\delta_3=\omega-\frac{1}{2}\eta+\frac{3}{8}\eta^2$, and $\delta_4=1-\omega-\frac{1}{2}\eta-\frac{1}{8}\eta^2$.

The difference equation of the proposed estimator is given as follows:

$$\begin{aligned} t_{pro2}-\bar{Y}=&\left(k_3-1\right)\bar{Y}+k_3\bar{Y}\left(\xi_0+\delta_1\xi_1-\delta_1\xi_2+\delta_1\xi_0\xi_1-\delta_1\xi_0\xi_2-\delta_2\xi_1\xi_2+\delta_3\xi_1^2+\delta_4\xi_2^2\right)\\ &+k_4\left(1+\delta_1\xi_1-\delta_1\xi_2-\delta_2\xi_1\xi_2+\delta_3\xi_1^2+\delta_4\xi_2^2\right). \end{aligned} \tag{51}$$

After taking the expectation of Eq (51), we obtain the following bias expression

$$\begin{aligned} Bias\left(t_{pro2}\right)=&\left(k_3-1\right)\bar{Y}+k_3\bar{Y}\left[\delta_1\left(\lambda-\lambda_1\right)C_{yx}+\left\{\delta_3\lambda-\left(\delta_2-\delta_4\right)\lambda_1\right\}C_x^2\right]\\ &+k_4\left[1+\left\{\delta_3\lambda-\left(\delta_2-\delta_4\right)\lambda_1\right\}C_x^2\right]. \end{aligned} \tag{52}$$

Taking the square on both sides of Eq (52) and simplifying, we have

$$
\begin{aligned}
\left(t_{pro2}-\bar{Y}\right)^2 = & \left(k_3-1\right)^2\bar{Y}^2+k_3^2\bar{Y}^2\left(\xi_0^2+\left(\delta_1^2+2\delta_3\right)\xi_1^2+\left(\delta_1^2+2\delta_4\right)\xi_2^2+4\delta_1\xi_0\xi_1-4\delta_1\xi_0\xi_2-2\left(\delta_1^2+\delta_2\right)\xi_1\xi_2\right) \\
& +k_4^2\left\{1+\left(\delta_1^2+2\delta_3\right)\xi_1^2+\left(\delta_1^2+2\delta_4\right)\xi_2^2-2\left(\delta_1^2+\delta_2\right)\xi_1\xi_2\right\} \\
& -2k_3\bar{Y}^2\left(\delta_1\xi_0\xi_1-\delta_1\xi_0\xi_2-\delta_2\xi_1\xi_2+\delta_3\xi_1^2+\delta_4\xi_2^2\right)-2k_4\bar{Y}\left(1-\delta_2\xi_1\xi_2+\delta_3\xi_1^2+\delta_4\xi_2^2\right) \\
& +2k_3k_4\bar{Y}\left\{1+\left(\delta_1^2+2\delta_3\right)\xi_1^2+\left(\delta_1^2+2\delta_4\right)\xi_2^2+2\delta_1\xi_0\xi_1-2\delta_1\xi_0\xi_2-2\left(\delta_1^2+\delta_2\right)\xi_1\xi_2\right\}.
\end{aligned} \tag{53}
$$

Taking the expectation on both sides of the above Eq (53), we obtain the following MSE equation:

$$
\begin{aligned}
MSE\left(t_{pro2}\right) = & \left(k_3-1\right)^2\bar{Y}^2+k_3^2\bar{Y}^2\left[\lambda C_y^2+\left\{\lambda\left(\delta_1^2+2\delta_3\right)-\lambda_1\left(\delta_1^2+2\delta_2-2\delta_4\right)\right\}C_x^2+4\delta_1\left(\lambda-\lambda_1\right)C_{yx}\right] \\
& +k_4^2\left[1+\left\{\lambda\left(\delta_1^2+2\delta_3\right)-\lambda_1\left(\delta_1^2+2\delta_2-2\delta_4\right)\right\}C_x^2\right]-2k_3\bar{Y}^2\left[\delta_1\left(\lambda-\lambda_1\right)C_{yx}+\left\{\delta_3\lambda-\lambda_1\left(\delta_2-\delta_4\right)\right\}C_x^2\right] \\
& -2k_4\bar{Y}\left[1+\left\{\delta_3\lambda-\lambda_1\left(\delta_2-\delta_4\right)\right\}C_x^2\right]+2k_3k_4\bar{Y}\left[1+\left\{\lambda\left(\delta_1^2+2\delta_3\right)-\lambda_1\left(\delta_1^2+2\delta_2-2\delta_4\right)\right\}C_x^2+2\delta_1\left(\lambda-\lambda_1\right)C_{yx}\right]
\end{aligned} \tag{54}
$$

or

$$
MSE\left(t_{pro2}\right)=\left(k_3-1\right)^2\bar{Y}^2+k_3^2\bar{Y}^2A_p+k_4^2B_p-2k_3\bar{Y}^2C_p-2k_4\bar{Y}D_p+2k_3k_4\bar{Y}E_p. \tag{55}
$$

Let us differentiate Eq (55) and equate to zero to obtain optimum values of the constants k_3 and k_4

$$
\frac{\partial MSE\left(t_{pro2}\right)}{\partial k_3}=0 \quad \text{and} \quad \frac{\partial MSE\left(t_{pro2}\right)}{\partial k_4}=0.
$$

So, we obtain

$$
\left(k_3-1\right)\bar{Y}^2+k_3\bar{Y}^2A_p+k_4\bar{Y}E_p-\bar{Y}^2C_p=0 \tag{56}
$$

and

$$
k_3\bar{Y}E_p+k_4B_p-\bar{Y}D_p=0. \tag{57}
$$

Solving both Eqs (56) and (57) simultaneously, we obtain the following values of the constants k_3 and k_4

$$
k_3=\frac{B_pC_p-D_pE_p+B_p}{A_pB_p-E_p^2+B_p} \quad \text{and} \quad k_4=\frac{\bar{Y}\left(A_pD_p-C_pE_p+D_p-E_p\right)}{A_pB_p-E_p^2+B_p}.
$$

Incorporating these optimum values of the constants, the minimum MSE of the proposed estimator is given as

$$
MSE\left(t_{pro2}\right)_{\min}\cong\bar{Y}^2\left(1-\frac{A_pD_p^2+B_pC_p^2-2C_pD_pE_p+B_p+2B_pC_p+D_p^2-2D_pE_p}{A_pB_p-E_p^2+B_p}\right). \tag{58}
$$

AIMS Press