

An unbiased efficient estimator of population mean using an auxiliary attribute in stratified random sampling

Sarbjit Singh Brar [§]

Manpreet Singh [†]

Department of Statistics

Punjabi University

Patiala 147002

Punjab

India

S. C. Malik ^{*}

Department of Statistics

Maharashi Dayanand University

Rohtak 124001

Haryana

India

Abstract

In this present paper, an unbiased estimator of population mean is suggested by using an auxiliary attribute within stratified simple random sampling (SSRS), the variance of the estimator has been theoretically derived, considering a first-order approximation. When compared to other comparable existing estimators, the proposed one has the smallest mean square error. The simulation study is employed to the artificial generated populations to verify the theoretical findings.

Subject Classification: 62D05, 62D99.

Keywords: Efficiency, Estimation, Mean square error, Simulation, Stratified random sampling.

[§] E-mail: sarbjitbrar@pbi.ac.in

[†] E-mail: luckybhathal91@gmail.com

^{*} E-mail: sc_malik@rediffmail.com (Corresponding Author)

1. Introduction

Information on auxiliary variable is incorporated in sample surveys to improve the efficiency of the estimators. Sometimes, in practical situations, an auxiliary variable is not available but, alternatively, a related auxiliary attribute can be utilized, as it demonstrates a strong association with the study variable. For example, in a study about the health of the human population, the gender of a person is an auxiliary attribute. The breed of cow serves as an auxiliary attribute for estimating milk production. The source of irrigation acts as an auxiliary attribute for estimating crop yield. In the estimation of the population mean with the help of an auxiliary attribute under simple random sampling, Naik and Gupta [9] introduced ratio, product, and regression type estimators and developed the conditions, using point bi-serial correlation, under which these estimators are better than the mean per unit estimator.

Further, Shabbir and Gupta [12] introduced an estimator of mean by optimizing two stable constants and obtained an expression for minimum mean square error (MSE). Singh et al. [16] proposed different ratio estimators with sample regression coefficient and two distinct real constants or the functions of a known parameter of an auxiliary attribute that are conditionally better than the ratio estimator of Naik and Gupta [9] and mean per unit estimator. Abd-Elfattah et al. [1] used the population coefficient of kurtosis of the auxiliary attribute to suggest an estimator which demonstrate greater efficiency than the estimators proposed by Naik and Gupta [9] and Singh et al. [6]. Further, Grover and Kaur [5], Koyuncu [7], Singh and Solanki [14], Singh [15], Yadav and Adewara [19], Khan et al. [6], Malik and Singh [8], Saini and Kumar [11], Audu et. al [3], Al-Isawi et al. [2] and others also contribute in same direction.

A valuable performance of auxiliary attributes in SRS leads to its use in SSRS to estimate the population mean. Singh and Audu [17] introduced four ratio-type estimators in SSRS that utilize the coefficient of kurtosis, coefficient of variation and point bi-serial correlation coefficient of the auxiliary attribute and comparing them with the traditional combined ratio estimator. The comparison demonstrates that the proposed estimators exhibit lower bias and demonstrating greater efficiency than the combined-ratio estimator under all conditions. Khan et al. [6] suggested a general class of ratio-cum-product type estimators of mean while incorporating auxiliary attribute information. The proposed class is compared with conventional estimators, including the mean per unit estimator, the ratio estimator by Naik and Gupta [9], and the estimator by Singh et al. [16]. The

results indicate that the suggested estimators consistently outperform the existing ones. Verma et al. [18] developed a regression-type estimator for the population mean and utilize a constant to minimize the mean squared error (MSE). Under optimal conditions, this estimator outperforms the conventional ratio estimator. Zaman [20] introduced modified ratio estimators by utilizing a combination of the coefficient of kurtosis and the coefficient of variation of the auxiliary attribute. Additionally, following the approach of Prasad [10], the author developed another ratio estimator for the population mean. In all cases, the suggested estimators outperform the combined ratio estimator. In addition, the author constructs an expression for the minimum MSE of suggested estimators. Zaman [21] extended Abd-Elfattah et al. [1] estimator to develop an exponential type estimator for mean which outperforms the ratio and exponential estimators.

In this paper, a new unbiased estimator is proposed to estimate the population mean under SSRS using the auxiliary attribute.

2. Notation

Now, consider a finite population U having N units partitioned into L strata. Let Y_{hi} and Φ_{hi} denote the value of the study variable y and the auxiliary attribute ϕ for the i^{th} unit ($i=1, 2, \dots, N_h$) in the stratum h , respectively. Further, N_h denotes the size of stratum h , such that $\sum_{h=1}^L N_h = N$. A simple random sample is drawn without replacement from the stratum h of size n_h , such that $\sum_{h=1}^L n_h = n$. The population is divided based on the presence or absence of the attribute to forming distinct strata.

Let

$$\Phi_{hi} = \begin{cases} 1, & \text{if } i^{\text{th}} \text{ unit of } h^{\text{th}} \text{ stratum in the population possesses attribute,} \\ 0, & \text{otherwise.} \end{cases}$$

Let Θ be a set containing those population units which have $\Phi_{hi} = 1$, where $i=1, 2, \dots, N_h$ and $h=1, 2, \dots, L$. Then we have the following notations:

$$A = \sum_{h=1}^L \sum_{i=1}^{N_h} \Phi_{hi} : \quad \text{the total number units in the population possessing attribute } \phi.$$

- $A_h = \sum_{i=1}^{N_h} \Phi_{hi}$: the total number of units in h^{th} stratum possessing attribute ϕ .
- $P = \frac{A}{N}$: the proportion of units in the population possessing attribute ϕ .
- $Q = \frac{N-A}{N}$: the proportion of units in the population that do not possess attribute ϕ .
- $P_h = \frac{A_h}{N_h}$: the proportion of units in h^{th} stratum possessing attribute ϕ .
- $Q_h = \frac{N_h - A_h}{N_h}$: the proportion of units in h^{th} stratum that do not possess attribute ϕ .
- $\bar{Y} = \sum_{h=1}^L W_h \bar{Y}_h$: the population mean of study variable.
- $\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} \bar{Y}_{hi}$: the mean of study variable in h^{th} stratum.
- $W_h = \frac{N_h}{N}$: the weight of h^{th} stratum.
- $S_{yh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2$: the population variance of h^{th} stratum of study variable.
- $S_{\phi h}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (\Phi_{hi} - P_h)^2$: the population variance of h^{th} stratum of auxiliary attribute.
 $\approx P_h Q_h$
- $S_{y\phi h} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)(\Phi_{hi} - P_h)$: the population covariance between y and ϕ in h^{th} stratum.
 $= \rho_h S_{yh} S_{\phi h}$
 $\approx P_h Q_h (\bar{Y}_{1h} - \bar{Y}_{2h})$
- $\beta_\phi^* = \frac{\sum_{h=1}^L W_h^2 \lambda_h S_{y\phi h}}{\sum_{h=1}^L W_h^2 \lambda_h S_{\phi h}^2}$: the regression coefficient of y on ϕ .

- $\beta_{\phi h} = \frac{S_{y\phi h}}{S_{\phi h}^2}$: the regression coefficient of y on ϕ in h^{th} stratum.
 ρ_h : the point bi-serial correlation coefficient of h^{th} stratum of the study variable and auxiliary attribute.
 $R = \frac{\bar{Y}}{\bar{P}}$: the population ratio.
 $R_h = \frac{\bar{Y}_h}{\bar{P}_h}$: the population ratio of h^{th} stratum.
 $C_\phi = S_\phi / P$ and $C_y = S_y / \bar{Y}$: coefficients of variation of y and ϕ , respectively.

Let

$$\phi_{hi} = \begin{cases} 1, & \text{if } i^{\text{th}} \text{ unit of } h^{\text{th}} \text{ stratum in the sample possesses attribute,} \\ 0, & \text{otherwise.} \end{cases}$$

- $a = \sum_{h=1}^L \sum_{i=1}^{n_h} \phi_{hi}$: the total number of units in the sample possessing attribute ϕ .
 $a_h = \sum_{i=1}^{n_h} \phi_{hi}$: the total number of units in the sample drawn from h^{th} stratum possessing attribute ϕ .
 $p_h = \frac{a_h}{n_h}$: the proportion of units in the sample drawn from stratum h possessing attribute ϕ .
 $\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} \bar{Y}_{hi}$: the sample mean of study variable in h^{th} stratum.
 $w_h = \frac{n_h}{n}$: the estimated weight of h^{th} stratum.
 $\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h$: the usual estimator of the population mean.
 $p_{st} = \sum_{h=1}^L w_h p_h$: the usual estimator of the population proportion.

3. Existing estimators

Let us consider a few estimators existing in the literature under SSRS. The usual weighted estimator for mean under SSRS is given by

$$T_1 = \bar{y}_{st}$$

Variance of the above estimator is as

$$V(T_1) = \sum_{h=1}^L W_h^2 \lambda_h^2 S_{yh}^2 \quad (1)$$

$$\text{Where, } \lambda_h = \left(\frac{1}{n_h} - \frac{1}{N_h} \right).$$

3.1 Separate Estimators

The Naik and Gupta [9]'s separate estimators of $\bar{Y}$ by utilizing auxiliary attribute information under SSRS are as

$$T_{NS_1} = \sum_{h=1}^L W_h \bar{y}_h \frac{P_h}{p_h}$$

$$T_{NS_2} = \sum_{h=1}^L W_h \bar{y}_h \frac{p_h}{P_h}$$

$$T_{NS_3} = \sum_{h=1}^L W_h (\bar{y}_h + \hat{\beta}_{\phi h} (P_h - p_h))$$

Where, $\hat{\beta}_{\phi h}$ is an estimator of $\beta_{\phi h}$.

MSEs of the above estimators are given as

$$MSE(T_{NS_1}) = \sum_{h=1}^L W_h^2 \lambda_h^2 [S_{yh}^2 + R_h^2 S_{\phi h}^2 - 2R_h S_{y\phi h}] \quad (2)$$

$$MSE(T_{NS_2}) = \sum_{h=1}^L W_h^2 \lambda_h^2 [S_{yh}^2 + R_h^2 S_{\phi h}^2 + 2R_h S_{y\phi h}] \quad (3)$$

$$MSE(T_{NS_3}) = \sum_{h=1}^L W_h^2 \lambda_h^2 S_{yh}^2 (1 - \rho_h^2) \quad (4)$$

The Bahl and Tuteja [4]'s separate estimators of $\bar{Y}$ by utilizing auxiliary attribute information under SSRS are as

$$T_{BS_1} = \sum_{h=1}^L W_h \bar{y}_h \exp \left(\frac{P_h - p_h}{P_h + p_h} \right)$$

$$T_{BS_2} = \sum_{h=1}^L W_h \bar{y}_h \exp \left(\frac{p_h - P_h}{p_h + P_h} \right)$$

MSEs of the above estimators are given as

$$MSE(T_{BS_1}) = \sum_{h=1}^L W_h^2 \lambda_h \left[S_{yh}^2 + \frac{R_h^2}{4} S_{\phi h}^2 - R_h S_{y\phi h} \right] \quad (5)$$

$$MSE(T_{BS_2}) = \sum_{h=1}^L W_h^2 \lambda_h \left[S_{yh}^2 + \frac{R_h^2}{4} S_{\phi h}^2 + R_h S_{y\phi h} \right] \quad (6)$$

Khan et al. [6] proposed a generalized ratio cum product type estimator with known $\beta_{\phi h}$ given as

$$T_{kS_1} = \sum_{h=1}^L W_h^2 \left[\lambda_{h'} \left[\frac{\bar{y}_h + \beta_{\phi h} (P_h - p_h)}{A'_h p_h + B_h} (A'_h P_h + B_h) \right] \right. \\ \left. + \sum_{h=1}^L W_h^2 \left[(1 - \lambda_{h'}) \left[\frac{\bar{y}_h + \beta_{\phi h} (P_h - p_h)}{A'_h p_h + B_h} (A'_h P_h + B_h) \right] \right] \right]$$

Where A'_h and B'_h are known population parameters of the auxiliary attribute in h^{th} stratum. $\lambda'_{h'}$ is a constant. MSE of the above estimator is given as

$$MSE(T_{kS_1}) = \sum_{h=1}^L W_h^2 \lambda'_h [S_{yh}^2 + (R'_h (1 - 2\lambda'_h) - \beta_{\phi h})^2 S_{\phi h}^2] + [(R'_h (1 - 2\lambda'_h) - \beta_{\phi h}) S_{y\phi h}]$$

Where, $R'_h = \frac{\bar{y}_h A'_h}{A'_h P_h + B_h}$ and for $\lambda'_h = 1/2$,

$$MSE_{min}(T_{kS_1}) = \sum_{h=1}^L W_h^2 \lambda_h S_{yh}^2 (1 - \rho_h^2) \quad (7)$$

3.2 Combined Estimators

The Naik and Gupta [9]'s combined estimators of $\bar{Y}$ by utilizing auxiliary attribute information under SSRS are as

$$T_{NC_1} = \bar{y}_{st} \frac{P}{p_{st}}$$

$$T_{NC_2} = \bar{y}_{st} \frac{p_{st}}{P}$$

$$T_{NC_3} = \bar{y}_{st} + \hat{\beta}_{\phi}^* (P - p_{st})$$

Where, $\hat{\beta}_\phi^*$ is an estimator of β_ϕ^* .

MSEs of the above estimators are given as

$$MSE(T_{NC_1}) = \sum_{h=1}^L W_h^2 \lambda_h [S_{yh}^2 + R^2 S_{\phi h}^2 - 2RS_{y\phi h}] \quad (8)$$

$$MSE(T_{NC_2}) = \sum_{h=1}^L W_h^2 \lambda_h [S_{yh}^2 + R^2 S_{\phi h}^2 + 2RS_{y\phi h}] \quad (9)$$

$$MSE(T_{NC_3}) = \sum_{h=1}^L W_h^2 \lambda_h S_{yh}^2 (1 - \rho^{*2}) \quad (10)$$

$$\text{Where, } \rho^* = \beta_\phi^* \sqrt{\frac{\sum_{h=1}^L W_h^2 \lambda_h S_{\phi h}^2}{\sum_{h=1}^L W_h^2 \lambda_h S_{yh}^2}}$$

Sharma and Singh [13] suggested ratio and product type exponential estimators under SSRS as

$$T_{SC_1} = \bar{y}_{st} \exp\left(\frac{P - p_{st}}{P + p_{st}}\right)$$

$$T_{SC_2} = \bar{y}_{st} \exp\left(\frac{p_{st} - P}{p_{st} + P}\right)$$

$$T_{SC_3} = \bar{y}_{st} \exp\left(\frac{\eta(P - p_{st})}{P + p_{st}}\right)$$

Where η is a real constant. The expressions for the mean squared errors of the above estimators are given as follows:

$$MSE(T_{SC_1}) = \sum_{h=1}^L W_h^2 \lambda_h \left[S_{yh}^2 + \frac{R^2}{4} S_{\phi h}^2 - RS_{y\phi h} \right] \quad (11)$$

$$MSE(T_{SC_2}) = \sum_{h=1}^L W_h^2 \lambda_h \left[S_{yh}^2 + \frac{R^2}{4} S_{\phi h}^2 + RS_{y\phi h} \right] \quad (12)$$

$$MSE(T_{SC_3}) = \sum_{h=1}^L W_h^2 \lambda_h \left[S_{yh}^2 + \frac{\eta^2 R^2}{4} S_{\phi h}^2 - \eta RS_{y\phi h} \right]$$

For $\eta = 2 \frac{\sum_{h=1}^L W_h^2 \lambda_h \rho_h S_{y\phi h}}{R \sum_{h=1}^L W_h^2 \lambda_h S_{\phi h}^2}$, above equation becomes:

$$MSE_{min}(T_{SC_3}) = \sum_{h=1}^L W_h^2 \lambda_h S_{yh}^2 (1 - \rho^{*2}) \quad (13)$$

Verma et al. [18] introduced an estimator of the population mean with known β_ϕ^* as

$$T_{VC_1} = [\bar{y}_{st} + \beta_\phi^*(P - p_{st})] \left(\frac{\delta'P + (1 - \delta')p_{st}}{\delta'p_{st} + (1 - \delta')P} \right)$$

Where δ' is a constant.

The expression for the mean squared error of the above estimator is given as follows:

$$MSE(T_{VC_1}) = \sum_{h=1}^L W_h^2 \lambda_h [S_{yh}^2 + (\beta_\phi^* + (1 - 2\delta')R)^2 S_{\phi h}^2] - \sum_{h=1}^L W_h^2 \lambda_h [2(\beta_\phi^* + (1 - 2\delta'))RS_{y\phi h}]$$

$$\text{For } \delta' = \frac{\bar{Y} \sum_{h=1}^L W_h^2 \lambda_h (S_{\phi h}^2 + S_{y\phi h}) - P \beta_\phi^* \sum_{h=1}^L W_h^2 \lambda_h S_{\phi h}^2}{2\bar{Y} \sum_{h=1}^L W_h^2 \lambda_h S_{\phi h}^2}$$

$$MSE_{min}(T_{VC_1}) = \sum_{h=1}^L W_h^2 \lambda_h S_{yh}^2 (1 - \rho^{*2}) \quad (14)$$

Zaman and Kadilar [22] suggested a general class of exponential estimators under SSRS as

$$T_{ZC_k} = \bar{y}_{st} \exp \left(\frac{(E'P_{st} + F') - (E'p_{st} + F')}{(E'P_{st} + F') - (E'p_{st} + F')} \right)$$

$E'(\neq 0)$ and F' , are some real constants. Particular members of the general class for $k=1,2,\dots,9$ are presented in Table 1.

Table 1
General class of exponential estimators

Estimators	Values of	
	E'	F'
$T_{ZC_1} = \bar{y}_{st} \exp \left(\frac{(P_{st} - p_{st})}{(P_{st} + p_{st} + 2\beta_\phi^*)} \right)$	1	β_ϕ^*
$T_{ZC_2} = \bar{y}_{st} \exp \left(\frac{(P_{st} - p_{st})}{(P_{st} + p_{st} + 2C_\phi)} \right)$	1	C_ϕ

Contd...

$T_{ZC_3} = \bar{y}_{st} \exp \left(\frac{(P_{st} - p_{st})}{(P_{st} + p_{st} + 2\rho^*)} \right)$	1	ρ^*
$T_{ZC_4} = \bar{y}_{st} \exp \left(\frac{\beta'_\phi (P_{st} - p_{st})}{\beta'_\phi (P_{st} + p_{st}) + 2C_\phi} \right)$	β'_ϕ	C_ϕ
$T_{ZC_5} = \bar{y}_{st} \exp \left(\frac{C_\phi (P_{st} - p_{st})}{C_\phi (P_{st} + p_{st}) + 2\beta'_\phi} \right)$	C_ϕ	β'_ϕ
$T_{ZC_6} = \bar{y}_{st} \exp \left(\frac{C_\phi (P_{st} - p_{st})}{C_\phi (P_{st} + p_{st}) + 2\rho^*} \right)$	C_ϕ	ρ^*
$T_{ZC_7} = \bar{y}_{st} \exp \left(\frac{\rho^* (P_{st} - p_{st})}{\rho^* (P_{st} + p_{st}) + 2C_\phi} \right)$	ρ^*	C_ϕ
$T_{ZC_8} = \bar{y}_{st} \exp \left(\frac{\beta'_\phi (P_{st} - p_{st})}{\beta'_\phi (P_{st} + p_{st}) + 2\rho^*} \right)$	β'_ϕ	ρ^*
$T_{ZC_9} = \bar{y}_{st} \exp \left(\frac{\rho^* (P_{st} - p_{st})}{\rho^* (P_{st} + p_{st}) + 2\beta'_\phi} \right)$	ρ^*	β'_ϕ

Where, β'_ϕ denotes coefficient of kurtosis and C_ϕ denotes coefficient of variation.

MSEs of the above estimators are given as

$$MSE(T_{ZC_k}) = \sum_{h=1}^L W_h^2 \lambda_h [S_{yh}^2 + \theta_k'^2 S_{\phi h}^2 - 2\theta_k' S_{y\phi h}] \quad (15)$$

$$\begin{aligned} \text{Where, } \theta_1' &= \frac{\bar{Y}}{2(P + \beta'_\phi)}, \theta_2' = \frac{\bar{Y}}{2(P + C_\phi)}, \theta_3' = \frac{\bar{Y}}{2(P + \rho^*)}, \theta_4' = \frac{\beta'_\phi \bar{Y}}{2(\beta'_\phi P + C_\phi)} \\ \theta_5' &= \frac{C_\phi \bar{Y}}{2(C_\phi P + \beta'_\phi)}, \theta_6' = \frac{C_\phi \bar{Y}}{2(C_\phi P + \rho^*)}, \theta_7' = \frac{\rho^* \bar{Y}}{2(\rho^* P + C_\phi)}, \theta_8' = \frac{\beta'_\phi \bar{Y}}{2(\beta'_\phi P + \rho^*)} \\ \text{and } \theta_9' &= \frac{\rho^* \bar{Y}}{2(\rho^* P + \beta'_\phi)} \end{aligned}$$

Zaman [21] proposed the following estimators under SSRS as

$$T_{ZC_{10}} = \bar{y}_{st} \exp \left(\frac{P_{st1} - p_{st1}}{P_{st1} + p_{st1}} \right)$$

$$T_{ZC_{11}} = \bar{y}_{st} \exp \left(\frac{P_{st2} - p_{st2}}{P_{st2} + p_{st2}} \right)$$

$$T_{ZC_{12}} = \bar{y}_{st} \exp\left(\frac{P_{st3} - p_{st3}}{P_{st3} + p_{st3}}\right)$$

$$T_{ZC_{13}} = \bar{y}_{st} \exp\left(\frac{P_{st4} - p_{st4}}{P_{st4} + p_{st4}}\right)$$

$$T_{ZC_{14}} = \bar{y}_{st} \exp\left(\frac{P_{st5} - p_{st5}}{P_{st5} + p_{st5}}\right)$$

Where, $P_{st1} = \sum_{h=1}^L W_h (P_h + C_{\phi h})$, $p_{st1} = \sum_{h=1}^L W_h (p_h + C_{\phi h})$,

$$P_{st2} = \sum_{h=1}^L W_h (P_h + \beta'_{3h}), p_{st2} = \sum_{h=1}^L W_h (p_h + \beta'_{\phi h}),$$

$$P_{st3} = \sum_{h=1}^L W_h (P_h \beta'_{\phi h} + C_{\phi h}), p_{st3} = \sum_{h=1}^L W_h (p_h \beta'_{\phi h} + C_{\phi h}),$$

$$P_{st4} = \sum_{h=1}^L W_h (P_h C_{\phi h} + \beta'_{\phi h}), p_{st4} = \sum_{h=1}^L W_h (p_h C_{\phi h} + \beta'_{\phi h}),$$

$$P_{st5} = \sum_{h=1}^L W_h (P_h + \rho_h), p_{st5} = \sum_{h=1}^L W_h (p_h + \rho_h),$$

$\beta'_{\phi h}$ is coefficient of kurtosis and $C_{\phi h}$ is coefficient of variation of stratum h .

The expressions for the mean squared errors of the above estimators are given as follows:

$$MSE(T_{ZC_k}) = \sum_{h=1}^L W_h^2 \lambda_h \left[S_{yh}^2 + \frac{\theta_k^2}{4} S_{\phi h}^2 - \theta'_k S_{y\phi h} \right] \quad (16)$$

where $k = 10, 11, \dots, 14$ and

$$\theta'_{10} = \frac{\sum_{h=1}^L W_h \bar{Y}}{\sum_{h=1}^L W_h (P_h + C_{\phi h})}, \theta'_{11} = \frac{\sum_{h=1}^L W_h \bar{Y}}{\sum_{h=1}^L W_h (P_h + \beta'_{\phi h})},$$

$$\theta'_{13} = \frac{\sum_{h=1}^L W_h \bar{Y}}{\sum_{h=1}^L W_h (P_h C_{\phi h} + \beta'_{\phi h})} \& \theta'_{14} = \frac{\sum_{h=1}^L W_h \bar{Y}}{\sum_{h=1}^L W_h (P_h + \rho_h)}$$

Another improved estimator by Zaman [21] is given as

$$T_{ZC_{15}} = \bar{y}_{st} \exp\left(\frac{P_{st6} - p_{st6}}{P_{st6} + p_{st6}}\right)^\delta$$

Where δ is a suitably chosen scalar, $P_{st6} = \sum_{h=1}^L W_h (E'_h P_h + F'_h)$ and $p_{st6} = \sum_{h=1}^L W_h (E'_h p_h + F'_h)$. MSE is given as

$$MSE(T_{ZC_{15}}) = \sum_{h=1}^L W_h^2 \lambda_h \left[S_{y_h}^2 + \frac{(\delta \theta'_{15})^2}{4} S_{\phi_h}^2 - \delta \theta'_{15} S_{y\phi_h} \right]$$

$$\text{Where, } \theta'_{15} = \frac{\sum_{h=1}^L W_h \bar{Y}_h}{\sum_{h=1}^L W_h (E'_h P_h + F'_h)} \text{ For } \delta_{opt} = 2 \frac{\sum_{h=1}^L W_h^2 \lambda_h S_{y\phi_h}}{R_6 \sum_{h=1}^L W_h^2 \lambda_h S_{ph}^2}$$

$$\text{and } MSE_{min}(T_{ZC_{15}}) = \sum_{h=1}^L W_h^2 \lambda_h S_{y_h}^2 (1 - \rho^{*2}) \quad (17)$$

All of the preceding estimators are biased with the exception of T_1 .

4. Proposed Estimators

Under SSRS with the known auxiliary information on P_h for each stratum, we developed an unbiased estimator of a population mean ($\bar{Y}$) as

$$\bar{y}_{prop} = \sum_{h=1}^L W_h (P_h \bar{y}_{1h} + Q_h \bar{y}_{2h}) = \sum_{h=1}^L W_h \bar{y}'_h$$

$$\text{where, } \bar{y}'_h = P_h \bar{y}_{1h} + Q_h \bar{y}_{2h}, \bar{y}_{1h} = \frac{1}{a_h} \sum_{i \in \Theta} y_{hi} \text{ and } \bar{y}_{2h} = \frac{1}{n_h - a_h} \sum_{i \notin \Theta} y_{hi}.$$

Clearly, $\bar{y}_{prop}$ is an unbiased estimator of mean, and variance of the proposed estimator is given as

$$V(\bar{y}_{prop}) = V\left(\sum_{h=1}^L W_h \bar{y}'_h\right)$$

Now,

$$V\left(\sum_{h=1}^L W_h \bar{y}'_h\right) = \left[\sum_{h=1}^L W_h^2 V(\bar{y}'_h) \right] = \sum_{h=1}^L W_h^2 \left[\lambda_h (P_h S_{1h}^2 + Q_h S_{2h}^2) + \frac{1}{n_h^2} (Q_h S_{1h}^2 + P_h S_{2h}^2) \right]$$

After using the identity,

$$S_{yh}^2 \approx P_h S_{1h}^2 + Q_h S_{2h}^2 + P_h (\bar{Y}_{1h} - \bar{Y}_h)^2 + Q_h (\bar{Y}_{2h} - \bar{Y}_h)^2$$

in above the equation, we have

$$\begin{aligned}
V\left(\sum_{h=1}^L W_h \bar{y}_h'\right) &= \sum_{h=1}^L W_h^2 [\lambda_h (S_{yh}^2 - P_h (\bar{Y}_{1h} - \bar{Y}_h)^2 - Q_h (\bar{Y}_{2h} - \bar{Y}_h)^2)] \\
&\quad + \sum_{h=1}^L W_h^2 \left[\lambda_h \left[\frac{1}{n_h^2} (Q_h S_{1h}^2 + P_h S_{2h}^2) \right] \right] \\
&= \sum_{h=1}^L W_h^2 \left[\lambda_h (S_{yh}^2 - P_h Q_h (\bar{Y}_{1h} - \bar{Y}_{2h})^2) + \frac{1}{n_h^2} (Q_h S_{1h}^2 + P_h S_{2h}^2) \right] \\
&= \sum_{h=1}^L W_h^2 [\lambda_h [S_{yh}^2 - P_h Q_h (\bar{Y}_{1h} - \bar{Y}_{2h})^2] + O(n_h)^{-2}]
\end{aligned}$$

Hence, $V(\bar{y}_{prop})$ up to first order approximation become

$$V(\bar{y}_{prop}) = \sum_{h=1}^L W_h^2 \lambda_h [S_{yh}^2 - P_h Q_h (\bar{Y}_{1h} - \bar{Y}_{2h})^2]. \quad (18)$$

5. Comparison

To compare the proposed estimator $\bar{y}_{prop}$ with existing separate and combined estimators four general expressions are formed as follows:

The general expression for MSE of separate estimators is

$$MSE(\hat{\theta}_s) = \sum_{h=1}^L W_h^2 \lambda_h [S_{yh}^2 + \kappa^2 S_{\phi h}^2 - 2\kappa S_{y\phi h}] \quad (19)$$

When, $\hat{\theta}_s = T_{NS_1}$, $\hat{\theta}_s = T_{NS_2}$, $\hat{\theta}_s = T_{BS_1}$ and $\hat{\theta}_s = T_{BS_2}$ then $\kappa = R_h$, $\kappa = R_h$, $\kappa = R_h/2$ and $\kappa = R_h/2$ respectively.

The general expression for MSE of combined estimators is

$$MSE(\hat{\theta}_c) = \sum_{h=1}^L W_h^2 \lambda_h [S_{yh}^2 + \kappa^{*2} S_{\phi h}^2 - 2\kappa^* S_{y\phi h}] \quad (20)$$

When, $\hat{\theta}_c = T_1$, $\hat{\theta}_c = T_{NC_1}$, $\hat{\theta}_c = T_{NC_2}$, $\hat{\theta}_c = T_{SC_1}$, $\hat{\theta}_c = T_{SC_2}$ and $\hat{\theta}_c = T_{ZC_k}$ then $\kappa^* = 0$, $\kappa^* = R$, $\kappa^* = R$, $\kappa^* = R/2$, $\kappa^* = R/2$ and $\kappa^* = \theta'_k$ respectively.

After subtracting (18) from (19), we have

$$MSE(\hat{\theta}_s) - V(\bar{y}_{prop}) = \sum_{h=1}^L W_h^2 \lambda_h P_h Q_h [\kappa - (\bar{Y}_{1h} - \bar{Y}_{2h})]^2 > 0 \quad (21)$$

The proposed estimator always more efficient than T_{NS_1} , T_{NS_2} , T_{BS_1} and T_{BS_2} .

Subtract (18) from (20),

$$MSE(\hat{\theta}_c) - V(\bar{y}_{prop}) = \sum_{h=1}^L W_h^2 \lambda_h P_h Q_h [\kappa^* - (\bar{Y}_{1h} - \bar{Y}_{2h})]^2 > 0 \quad (22)$$

The proposed estimator always more efficient than $T_1, T_{NC_1}, T_{NC_2}, T_{SC_1}, T_{SC_2}$ and T_{ZC_k} .

Subtract (18) from (10),

$$MSE(T_{NC_3}) - V(\bar{y}_{prop}) = \sum_{h=1}^L W_h^2 \lambda_h P_h Q_h [\beta_{\phi h} - \beta_{\phi}^*]^2 \geq 0 \quad (23)$$

The proposed estimator always more efficient than the combined regression estimator T_{NC_3} . Similarly, one can show that the proposed estimator always more efficient than the generalized estimators T_{SC_3}, T_{VC_1} and $T_{ZC_{15}}$. However, the proposed estimator as efficient as T_{NS_3} . Further, the proposed estimator is as efficient as the optimum estimator in the class T_{KS_1} .

6. Simulation study

A simulation study was conducted using R software to verify the theoretical findings for the estimators considered in the present research. Two stratified populations are generated each of size 300 (with $N_h = 100$ for each stratum), from a normal distribution are given in Table 2 and Table 3.

Table 2
Parameters of Population-I

	Size	$\bar{Y}_h$	S_{yh}^2	P	$S_{\phi h}^2$	ρ_h	$\frac{1}{2} \frac{C_{\phi}}{C_y}$
Stratum-1	100	14.9359	44.7486	0.5	0.2525	0.6147	1.1220
Stratum-2	100	17.9156	32.7073	0.6	0.2424	0.4545	1.2853
Stratum-3	100	21.1819	62.5071	0.7	0.2121	0.8167	0.8813
Overall	300	18.0111	52.8707	0.6	0.2408	0.6450	

Table 3
Parameters of Population-II

	Size	$\bar{Y}_h$	S_{yh}^2	P	$S_{\phi h}^2$	ρ_h	$\frac{1}{2} \frac{C_{\phi}}{C_y}$
Stratum-1	100	14.9359	63.0929	0.5	0.2525	-0.7475	0.9449
Stratum-2	100	16.9157	31.3484	0.6	0.2424	-0.4150	1.2396

Contd...

Stratum-3	100	15.1819	74.9803	0.7	0.2121	-0.8500	0.5768
Overall	300	15.6778	56.8749	0.6	0.24080	-0.6704	

From the above populations, 10000 stratified random samples of sizes (n) 21, 30, and 45 are drawn from each population. Further, MSEs of various estimators are shown in Table 4 and Table 5.

Table 4
MSEs for Population-I

$MSE(\hat{\theta})$	$n_1 = 7$	$n_2 = 10$	$n_3 = 15$
T_1	1.6482	1.2682	0.8555
T_{NS_1}	7.6116	6.5535	3.8380
T_{NS_3}	1.2622	0.8258	0.5134
T_{BS_1}	1.8012	1.25602	0.7874
T_{NC_1}	5.6439	4.2576	2.7616
T_{NC_3}	1.2881	0.8460	0.5391
T_{SC_1}	1.6643	1.1214	0.7189
Y_{prop}	1.2622	0.8258	0.5134

Note 1: This is conditional simulation study for $1 < a_h < 10$.

Note 2: The estimators $T_{ks_1}, T_{SC_3}, T_{VC_1}, T_{ZC_k}, k=1,2,...,15$ are not included in the simulation study because these estimators require some known constants or parameters. However, $MSE_{min}(T_{ks_1}) = MSE(T_{NS_3})$ and the minimum MSEs of other estimators are identical to $MSE(T_{NC_3})$.

Table 5
MSEs for Population-II

$MSE(\hat{\theta})$	$n_1 = 7$	$n_2 = 10$	$n_3 = 15$
T_1	1.8722	1.4415	1.0306
T_{NS_2}	3.2948	2.9110	1.9625
T_{NS_3}	1.2622	0.8258	0.5134

Contd...

T_{BS_2}	1.3725	1.0420	0.6895
T_{NC_2}	2.9032	2.4593	1.6539
T_{NC_3}	1.3356	0.8921	0.5826
T_{SC_2}	1.2817	0.9213	0.6005
Y_{prop}	1.2622	0.8258	0.5134

The percentage relative efficiency (PRE) of an estimator ($\hat{\theta}$) relative to $\bar{y}_{st}$ can be obtain as:

$$PRE(\hat{\theta}) = \frac{V(T_1)}{MSE(\hat{\theta})} \times 100$$

PREs of these estimators are given in Table 6 and Table 7.

Table 6
PREs for Population-I

$PRE(\hat{\theta})$	$n_1 = 7$	$n_2 = 10$	$n_3 = 15$
T_1	100	100	100
T_{NS_1}	21.6546	19.3524	22.2905
T_{NS_3}	130.5816	153.5793	166.6315
T_{BS_1}	91.5067	100.9698	108.6527
T_{NC_1}	29.2041	29.7880	30.9787
T_{NC_3}	127.9544	149.9104	158.6783
T_{SC_1}	99.0341	113.0917	118.9985
Y_{prop}	130.5816	153.5793	166.6315

Table 7
PREs for Population-II

$PRE(\hat{\theta})$	$n_1 = 7$	$n_2 = 10$	$n_3 = 15$
T_1	100	100	100
T_{NS_2}	56.8228	49.5201	52.5152

Contd...

T_{NS_3}	148.323	174.5642	200.733
T_{BS_2}	136.4043	138.3344	149.4687
T_{NC_2}	64.4867	58.6154	62.3108
T_{NC_3}	140.1694	161.5733	176.8984
T_{SC_2}	146.0689	156.4565	171.6087
Y_{prop}	148.323	174.5642	200.733

The following findings emerge from the simulation studies for the generated populations:

1. Usual ratio and product type estimators are less efficient than T_1 .
2. The efficiency of T_{BS_2} and T_{SC_2} is always higher than that of T_1 , and it rises with the size of the sample.
3. The proposed estimator consistently outperforms existing ones and demonstrates equal efficiency to T_{NS_3} .
4. The efficiency of suggested estimator rises with the size of the sample.

7. Conclusion

The above discussion reveals that the variance of the proposed estimator is lower than the mean squared errors of classical weighted estimators, separate and combined ratio-type estimators, separate and combined product-type estimators, as well as separate and combined exponential ratio-type and product-type estimators. Additionally, the suggested estimator is shown to be as efficient as the regression estimator and other optimal estimators from general classes. However, while the proposed estimator is unbiased, other equally efficient estimators are biased, making the proposed estimator more advantageous compared to the existing alternatives. Lastly, the simulation study confirms that the proposed estimator consistently outperforms the current estimators. Future research could extend this study to two-phase stratified random sampling.

References

- [1] A. M. Abd-Elfattah, E. A. El-Sherpieny, S. M. Mohamed, and O. F. Abdou, "Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute," *Applied mathematics and computation*, vol. 215, no. 12, pp. 4198–4202 (2010).
- [2] J. M. A. Al-Isawi, A. S. AL-Mouel, and A. H. Ali, "Bayes quadratic unbiased estimator of variance component in multi-samples repeated measurements ANOVA model (Multi-RMM)," *Journal of Statistics and Management Systems*, vol. 25, no. 3, pp. 697–707 (2022), doi: 10.1080/09720510.2021.1962028.
- [3] A. Audu, R. Singh, S. Khare, and N. S. Dauran, "Almost unbiased estimators for population mean in the presence of non-response and measurement error," *Journal of Statistics and Management Systems*, vol. 24, no. 3, pp. 573–589 (2021), doi: 10.1080/09720510.2020.1759209.
- [4] S. Bahl and R. Tuteja, "Ratio and product type exponential estimators," *Journal of information and optimization sciences*, vol. 12, no. 1, pp. 159–164 (1991).
- [5] L. K. Grover and P. Kaur, "An improved exponential estimator of finite population mean in simple random sampling using an auxiliary attribute," *Applied mathematics and Computation*, vol. 218, no. 7, pp. 3093–3099 (2011).
- [6] Z. Khan, J. Shabbir, M. Griffiths, and B. Lausen, "On the ratio method of estimation via auxiliary attributes in simple and stratified random sampling," *International Journal of Statistics and Systems*, vol. 8, no. 2, pp. 163–180 (2013).
- [7] N. Koyuncu, "Efficient estimators of population mean using auxiliary attributes," *Applied Mathematics and Computation*, vol. 218, no. 22, pp. 10900–10905 (2012).
- [8] S. Malik and R. Singh, "Estimation of population mean using information on auxiliary attribute in two-phase sampling," *Applied Mathematics and Computation*, vol. 261, pp. 114–118 (2015).
- [9] V. D. Naik and P. C. Gupta, "A note on estimation of mean with known population proportion of an auxiliary character," *Journal of the Indian Society of Agricultural Statistics*, vol. 48, no. 2, pp. 151–158 (1996).
- [10] B. Prasad, "Some improved ratio type estimators of population mean and ratio in finite population sample surveys," *Communications in Statistics – Theory and Methods*, vol. 18, no. 1, pp. 379–392 (1989).
- [11] M. Saini and A. Kumar, "Exponential type product estimator for finite population mean with information on auxiliary attribute,"

- Applications and Applied Mathematics: An International Journal*, vol. 10, no. 1, pp. 106–113 (2015).
- [12] J. Shabbir and S. Gupta, "On estimating the finite population mean with known population proportion of an auxiliary variable," *Pakistan Journal of Statistics*, vol. 23, no. 1, pp. 1–9 (2007).
- [13] P. Sharma and R. Singh, "Efficient estimator of population mean in stratified random sampling using auxiliary attribute," *World Applied Sciences Journal*, vol. 27, no. 12, pp. 1786–1791 (2013).
- [14] H. P. Singh and R. S. Solanki, "Improved estimation of population mean in simple random sampling using information on auxiliary attribute," *Applied Mathematics and Computation*, vol. 218, no. 15, pp. 7798–7812 (2012).
- [15] R. Singh, "On estimation of population mean using information on auxiliary attribute," *Pakistan Journal of Statistics & Operation Research*, vol. 9, no. 4, pp. 361–369 (2013).
- [16] R. Singh, P. Chauhan, N. Sawan, and F. Smarandache, "Ratio estimators in simple random sampling using information on auxiliary attribute," *Pakistan Journal of Statistics & Operation Research*, vol. 4, no. 1, pp. 47–53 (2007).
- [17] R. V. K. Singh and A. Audu, "Efficiency of ratio estimators in stratified random sampling using information on auxiliary attribute," *International Journal of Engineering Science and Innovative Technology*, vol. 2, no. 1, pp. 166–172 (2013).
- [18] H. K. Verma, P. Sharma, and R. Singh, "Improved estimator of finite population mean using auxiliary attribute in stratified random sampling," *Journal of Scientific Research*, vol. 58, pp. 99–105 (2014).
- [19] S. K. Yadav and A. A. Adewara, "On improved estimation of population mean using qualitative auxiliary information," *Mathematical Theory and Modeling*, vol. 3, no. 11, pp. 42–50 (2013).
- [20] T. Zaman, "Efficient estimators of population mean using auxiliary attribute in stratified random sampling," *Advances and Applications in Statistics*, vol. 56, no. 2, pp. 153–171 (2019).
- [21] T. Zaman, "An efficient exponential estimator of the mean under stratified random sampling," *Mathematical Population Studies*, vol. 28, no. 2, pp. 104–121 (2021).
- [22] T. Zaman and C. Kadilar, "On estimating the population mean using auxiliary character in stratified random sampling," *Journal of Statistics and Management Systems*, vol. 23, no. 8, pp. 1415–1426 (2020).