On Some Improved Classes of Estimators Under Stratified Sampling Using Attribute

Shashi Bhushan¹, Anoop Kumar²,∗, Dushyant Tyagi and Saurabh Singh²

¹Department of Statistics, Lucknow University, Lucknow, India
²Department of Mathematics and Statistics, Dr. Shakuntala Misra National Rehabilitation University, Lucknow, U.P., India
E-mail: anoop.asy@gmail.com
∗Corresponding Author

Received 23 February 2022; Accepted 14 March 2022; Publication 16 April 2022

Abstract

This article establishes some improved classes of difference and ratio type estimators of population mean of study variable using information on auxiliary attribute under stratified simple random sampling. The usual mean estimator, classical ratio estimator, classical product estimator and classical regression estimator are identified as particular cases of the proposed classes of estimators for different values of the characterising scalars. The expression of mean square error of the suggested classes of estimators has been studied up to first order of approximation and their effective performances are likened with respect to the conventional as well as lately existing estimators. Subsequently, an empirical study has been carried out using a real data set in support of theoretical results. The empirical results justify the proposition of the proposed classes of estimators in terms of percent relative efficiency over

doi: 10.13052/jrss0974-8024.1518
© 2022 River Publishers
all discussed work till date. Suitable suggestions are forwarded to the survey practitioners.

**Keywords:** Auxiliary attribute, efficiency, mean square error, stratified simple random sampling.

1 Introduction

In sample surveys, the use of auxiliary information is a well-known fact to enhance the efficiency of the estimators. Several improved and modified ratio, regression, product and logarithmic type estimators have been suggested using auxiliary information under different sampling schemes by various authors including Nazir et al. (2018), Lone et al. (2021), Bhushan et al. (2020a, b, c, 2021a, b), Bhushan and Kumar (2021, 2022), etc. Many times, in real life situations, the variable of interest may not be associated with a quantitative auxiliary variable and some qualitative auxiliary characteristic might be easily available which is significantly associated with the variable of interest. For example:

- The height of person (y) may depend on sex $\phi$, i.e., the person is male or female.
- The amount of yield of paddy crop (y) may rely on a certain variety of paddy ($\phi$).
- The amount of milk produce (y) may depend on a certain breed of buffalo ($\phi$).
- The use of drugs (y) may depend on the sex ($\phi$).

Thus, taking advantage of bi-serial correlation ($\rho$) into consideration, several authors proposed various class of estimators using attribute under different sampling framework. Naik and Gupta (1996) suggested classical ratio, product and regression estimators under simple random sampling (SRS). Singh et al. (2007) introduced attribute based exponential ratio and product type estimators in SRS. Abd-Elfattah et al. (2010) used information on attribute and investigated different exponential type estimators of population mean under SRS. Zaman and Kadilar (2019) addressed a novel family of exponential estimators using information of auxiliary attribute whereas Zaman and Kadilar (2021a) considered a new class of exponential estimators for finite population mean in two-phase sampling. Zaman (2019a) proffered an improved estimators using coefficient of skewness of auxiliary attribute.
under SRS. Zaman (2020) developed a generalized exponential estimator for the finite population mean based on attribute. Bhushan and Gupta (2020) envisaged an improved log-type family of estimators using attribute in SRS.

When the nature of population is heterogeneous then a well-known stratified simple random sampling (SSRS) is to be used to estimate the population parameters. It is based on dividing the whole population into homogeneous sub-populations known as “strata” and selecting a simple random sample independently from different strata. Sharma and Singh (2013) introduced exponential type estimators under SSRS using known population proportion. Zaman (2019b) evoked an efficient estimators of population mean using auxiliary attribute in SSRS. Zaman and Kadilar (2020) and Zaman (2021) proposed various exponential type estimators for population mean using auxiliary attribute under SSRS. Zaman and Kadilar (2021b) suggested exponential ratio and product type estimators of the mean in stratified two-phase sampling. It has been observed empirically that the efficiency of the above estimators introduced by different authors is at most equal to the classical regression estimator defined on the lines of Naik and Gupta (1996) under SSRS. The above discussion put a question: “Is there any procedure of obtaining better estimator than the classical regression estimator?” In this paper, we have made an effort to answer this question by suggesting some improved classes of difference and ratio type estimators using known population proportions.

The article is organized in following sections. Section 2 considers prominent estimators suggested till date in SSRS using attribute with their properties. In Section 3, we suggested some improved classes of estimators and studied their properties. The efficiency conditions are obtained in Section 4 which are further verified in Section 5 by an empirical study and discussion of empirical results. The conclusion of the study is given in Section 6.

2 Existing Estimators

Consider a finite population $U = (U_1, U_2, \ldots, U_N)$ consist of $N$ identifiable units which is divided into $L$ homogeneous strata and a simple random sample $s$ of size $n_h$ is measured from stratum $h$ using simple random sampling without replacement scheme. Let $y_{hi}$ and $\phi_{hi}$ be the study variable $y$ and auxiliary attribute $\phi$ for unit $i$ in the stratum $h$ of population $U$. It is noted that $\phi_{i} = 1$ if the unit $i$ possess attribute $\phi$ and $\phi_{i} = 0$, otherwise. Let $A = \sum_{i=1}^{N} \phi_{i}$ and $a = \sum_{i=1}^{n_h} \phi_{i}$ be the total number of units in the population.
U and sample s respectively possessing attribute φ whereas P = (A/N) be the population proportion, \( P_h = (A_h/N_h) \) be the population proportion in stratum h and \( p_h = (a_h/n_h) \) be the sample proportion of stratum h having attribute φ. The sample mean of study variable y is \( \bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \), where \( \bar{y}_h = n_h^{-1} \sum_{i=1}^{n_h} y_{hi} \), the population mean of study variable is \( \bar{Y}_{st} = \sum_{h=1}^{L} W_h \bar{Y}_h \), where \( \bar{Y}_h = N_h^{-1} \sum_{h=1}^{N_h} y_{hi} \) and the weight of stratum is \( W_h = N_h/N \). The population mean square of study variable in stratum h is \( S^2_{y_h} = (N_h - 1)^{-1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2 \) and population mean square of auxiliary attribute in stratum h is \( S^2_{p_h} = (N_h - 1)^{-1} \sum_{i=1}^{N_h} (p_{hi} - P_h)^2 \). To obtain the mean square error (MSE) of different estimators, let us assume that \( \bar{y}_{st} = \bar{Y}(1 + e_0) \) and \( p_{st} = P(1 + e_1) \) such that \( E(e_0) = E(e_1) = 0 \), \( E(e_0^2) = \frac{1}{\bar{Y}} \sum_{h=1}^{L} W_h^2 \gamma_h S^2_{y_h}, \) \( E(e_1^2) = \frac{1}{P} \sum_{h=1}^{L} W_h^2 \gamma_h S^2_{p_h} \) and \( E(e_0 e_1) = \frac{1}{\bar{Y} P} \sum_{h=1}^{L} W_h^2 \gamma_h \rho_{y_p} S_{y_h} S_{p_h} \).

Where \( \gamma_h = (N_h - n_h)/N_h n_h \).

Now, we consider a review of some prominent attribute-based estimators under SSRS along with their properties.

The usual mean estimator under SSRS is defined as

\[
T_m = \bar{y}_{st}
\]  

(1)

The variance of the above estimator is given by

\[
V(T_m) = \sum_{h=1}^{L} W_h^2 \gamma_h S^2_{y_h}
\]  

(2)

Following Naik and Gupta (1996), the classical combined ratio, product and regression estimators for population mean \( \bar{Y} \) using auxiliary attribute can be defined under SSRS as

\[
T_r = \bar{y}_{st} \left( \frac{P}{p_{st}} \right)
\]  

(3)

\[
T_p = \bar{y}_{st} \left( \frac{p_{st}}{P} \right)
\]  

(4)

\[
T_{lr} = \left[ \bar{y}_{st} + \beta_{\phi} (P - p_{st}) \right]
\]  

(5)

where \( \beta_{\phi} \) is the regression coefficient of y on φ and \( p_{st} = \sum_{h=1}^{L} W_h p_h \) is the sample proportion. The MSE of the above estimators is respectively
given by

\[
MSE(T_r) = \sum_{h=1}^{L} W_h^2 \gamma_h \left[ S_{yh}^2 + R^2 S_{ph}^2 - 2R\rho_h S_{yh} S_{ph} \right]
\]  
(6)

\[
MSE(T_p) = \sum_{h=1}^{L} W_h^2 \gamma_h \left[ S_{yh}^2 + R^2 S_{ph}^2 + 2R\rho_h S_{yh} S_{ph} \right]
\]  
(7)

\[
MSE(T_{lr}) = \sum_{h=1}^{L} W_h^2 \gamma_h \left[ S_{yh}^2 + \beta^2 R^2 S_{ph}^2 - 2\beta\rho_h S_{yh} S_{ph} \right]
\]  
(8)

where \( R = \overline{Y} / P \) is the population ratio. Now, minimizing the MSE(\( T_{lr} \)) with respect to (w.r.t.) \( \beta \phi \), we get

\[
\beta_{\phi(\text{opt})} = \frac{\sum_{h=1}^{L} W_h^2 \gamma_h \rho_h S_{yh} S_{ph}}{R \sum_{h=1}^{L} W_h^2 \gamma_h S_{ph}^2}
\]

Putting \( \beta_{\phi(\text{opt})} \) in the MSE(\( T_{lr} \)), we get

\[
\min \text{MSE}(T_{lr}) = \sum_{h=1}^{L} W_h^2 \gamma_h \left[ S_{yh}^2 (1 - \rho_h^2) \right]
\]  
(9)

The exponential functions model a relationship in which a constant change in the independent variable gives the equal proportional change in the dependent variable. Therefore, motivated by Singh et al. (2007), Sharma and Singh (2013) investigated ratio and product exponential type estimators under SSRS as

\[
T_{ss1} = \overline{y}_{st} \exp \left( \frac{P - p_{st}}{P + p_{st}} \right)
\]  
(10)

\[
T_{ss2} = \overline{y}_{st} \exp \left( \frac{p_{st} - P}{p_{st} + P} \right)
\]  
(11)

The MSE of the above estimators is given by

\[
\text{MSE}(T_{ss1}) = \sum_{h=1}^{L} W_h^2 \gamma_h \left[ S_{yh}^2 + \frac{R^2 S_{ph}^2}{4} - R\rho_h S_{yh} S_{ph} \right]
\]  
(12)

\[
\text{MSE}(T_{ss2}) = \sum_{h=1}^{L} W_h^2 \gamma_h \left[ S_{yh}^2 + \frac{R^2 S_{ph}^2}{4} + R\rho_h S_{yh} S_{ph} \right]
\]  
(13)
Sharma and Singh (2013) introduced another exponential type estimator under SSRS as

\[ T_{ss3} = \bar{y}_{st}\exp\left(\frac{\eta(P - p_{st})}{(P + p_{st})}\right) \]

(14)

where \(\eta\) is a suitably chosen scalar.

The MSE of the above estimator is given by

\[ \text{MSE}(T_{ss3}) = \sum_{h=1}^{L} W_h^2 \gamma_h \left[ S^2_{yh} + \frac{\eta^2 R^2}{4} S^2_{ph} - \eta R \rho_h S_{yh} S_{ph} \right] \]

(15)

Minimizing the MSE\((T_{ss3})\) w.r.t. \(\eta\), we get

\[ \eta(\text{opt}) = 2 \sum_{h=1}^{L} W_h^2 \gamma_h \rho_h S_{yh} S_{ph} \]

\[ R \sum_{h=1}^{L} W_h^2 \gamma_h S^2_{ph} \]

Putting \(\eta(\text{opt})\) in the MSE\((T_{ss3})\), we get

\[ \min \text{MSE}(T_{ss3}) = \sum_{h=1}^{L} W_h^2 \gamma_h S^2_{yh}(1 - \rho_h^2) \]

(16)

which is the minimum MSE of the classical regression estimator \(T_{lr}\).

Zaman and Kadilar (2020) envisaged a family of ratio exponential estimator under SSRS as

\[ T_{zk1} = \bar{y}_{st}\exp\left[\frac{(m P_{st} + n) - (m p_{st} + n)}{(m P_{st} + n) + (m p_{st} + n)}\right] \]

(17)

The MSE of the above estimator is given by

\[ \text{MSE}(T_{zk1}) = \sum_{h=1}^{L} W_h^2 \gamma_h \left[ S^2_{yh} + \frac{R^2 \nu_h^2}{4} S^2_{ph} - R \nu_h \rho_h S_{yh} S_{ph} \right] \]

(18)

where \(\nu_h = m_h P_h / (m_h P_h + n_h)\).

On the lines of Koyuncu and Kadilar (2009), Zaman and Kadilar (2020) suggested an improved form of the above estimator under SSRS as

\[ T_{zk2} = \lambda \bar{y}_{st}\exp\left[\frac{(m P_{st} + n) - (m p_{st} + n)}{(m P_{st} + n) + (m p_{st} + n)}\right] \]

(19)

where \(\lambda\) is a suitably chosen scalar.
The MSE of the estimator $T_{zk2}$ is given by

$$
\text{MSE}(T_{zk2}) = \left[ \bar{Y}^2 (\lambda - 1)^2 + \lambda^2 \sum_{h=1}^{L} W_h^2 \gamma_h S_{yh}^2 \\
+ (3\lambda^2 - 2\lambda) \frac{R^2}{d} \sum_{h=1}^{L} W_h^2 \gamma_h v_h S_{ph}^2 \\
-(2\lambda^2 - \lambda) \sum_{h=1}^{L} W_h^2 \gamma_h v_h \rho_h S_{yh} S_{ph} \right] \quad (20)
$$

Minimizing the MSE($T_{zk2}$) w.r.t. $\lambda$, we get

$$
\lambda_{(opt)} = \frac{-(R \sum_{h=1}^{L} W_h^2 \gamma_h v_h \rho_h S_{yh} S_{ph})}{(2\bar{Y}^2 + \frac{R^2}{d} \sum_{h=1}^{L} W_h^2 \gamma_h v_h S_{ph}^2)} = \lambda^*(\text{say})
$$

Putting $\lambda_{(opt)}$ in the MSE($T_{zk2}$), we get

$$
\text{minMSE}(T_{zk2}) = \left[ (\lambda^* - 1)^2 + \lambda^* \sum_{h=1}^{L} W_h^2 \gamma_h S_{h}^2 \\
+ (3\lambda^2 - 2\lambda^*) \frac{R^2}{d} \sum_{h=1}^{L} W_h^2 \gamma_h \\
\left[ v_h S_{h}^2 - R \left(2\lambda^2 - \lambda^*\right) \sum_{h=1}^{L} W_h^2 \gamma_h v_h \rho_h S_{yh} S_{ph} \right] \right] \quad (21)
$$

On the lines of Zaman (2020), one may define an exponential ratio type estimator using auxiliary attribute under SSRS as

$$
T_z = \overline{y}_{st} \left( \frac{p_{st}}{P_{st}} \right)^{\theta} \exp \left[ \frac{(mP_{st} + n) - (mp_{st} + n)}{(mP_{st} + n) + (mp_{st} + n)} \right] \quad (22)
$$

where $\theta$ is a suitably chosen scalar.

The MSE of the estimator $T_z$ is given by

$$
\text{MSE}(T_z) = \sum_{h=1}^{L} W_h^2 \gamma_h \left[ S_{yh}^2 + \theta^2 R^2 S_{ph}^2 + v_h^2 R^2 S_{ph}^2 - 2\theta R^2 v_h S_{ph}^2 \\
+ 2\theta R \rho_h S_{yh} S_{ph} - 2R v_h \rho_h S_{yh} S_{ph} \right] \quad (23)
$$
Minimizing the MSE($T_z$) w.r.t. $\theta$, we get

$$
\theta_{(opt)} = \left( \frac{\sum_{h=1}^{L} W_h^2 \gamma_h \upsilon_h^2 S_{zh}^2}{R \sum_{h=1}^{L} W_h^2 \gamma_h S_{zh}^2} - \frac{\sum_{h=1}^{L} W_h^2 \gamma_h \upsilon_h \phi_h S_{zh} S_{ph}}{R \sum_{h=1}^{L} W_h^2 \gamma_h S_{ph}^2} \right)
$$

Putting $\theta_{(opt)}$ in the MSE($T_z$), we get

$$
\text{minMSE} = \sum_{h=1}^{L} W_h^2 \gamma_h S_{zh}^2 (1 - \rho_h^2)
$$

which is the minimum MSE of classical regression estimator $T_{lr}$.

Following Abd-Elfattah et al. (2010), Zaman (2021) envisaged following class of estimator under SSRS as

$$
T_{z1} = \overline{y}_{st \exp} \left( \frac{P_{st1} - P_{st1}}{P_{st1} + P_{st1}} \right)
$$

$$
T_{z2} = \overline{y}_{st \exp} \left( \frac{P_{st2} - P_{st2}}{P_{st2} + P_{st2}} \right)
$$

$$
T_{z3} = \overline{y}_{st \exp} \left( \frac{P_{st3} - P_{st3}}{P_{st3} + P_{st3}} \right)
$$

$$
T_{z4} = \overline{y}_{st \exp} \left( \frac{P_{st4} - P_{st4}}{P_{st4} + P_{st4}} \right)
$$

$$
T_{z5} = \overline{y}_{st \exp} \left( \frac{P_{st5} - P_{st5}}{P_{st5} + P_{st5}} \right)
$$

where $P_{st1} = \sum_{h=1}^{L} W_h (P_h + C_{ph})$, $P_{st1} = \sum_{h=1}^{L} W_h (p_h + C_{ph})$, $P_{st2} = \sum_{h=1}^{L} W_h (P_h + \beta_{2h}(\phi))$, $P_{st2} = \sum_{h=1}^{L} W_h (p_h + \beta_{2h}(\phi))$, $P_{st3} = \sum_{h=1}^{L} W_h (P_h \beta_{2h}(\phi) + C_{ph})$, $P_{st3} = \sum_{h=1}^{L} W_h (p_h \beta_{2h}(\phi) + C_{ph})$, $P_{st4} = \sum_{h=1}^{L} W_h (P_h C_{ph} + \beta_{2h}(\phi))$, $P_{st4} = \sum_{h=1}^{L} W_h (p_h C_{ph} + \beta_{2h}(\phi))$, $P_{st5} = \sum_{h=1}^{L} W_h (P_h + \rho_h)$, $P_{st5} = \sum_{h=1}^{L} W_h (p_h + \rho_h)$.

The MSE of the above estimators is given by

$$
\text{MSE} = \sum_{h=1}^{L} W_h^2 \gamma_h \left[ S_{zh}^2 + \frac{R^2}{4} S_{ph}^2 - R \rho_h S_{zh} S_{ph} \right], \ i = 1, 2, \ldots, 5
$$
where $R_1 = \sum_{h=1}^{L} W_h \bar{Y}_h / \sum_{h=1}^{L} W_h (P_h + C_{ph})$, $R_2 = \sum_{h=1}^{L} W_h \bar{Y}_h / \sum_{h=1}^{L} W_h (P_h + \beta_{2h}(\phi))$, $R_3 = \sum_{h=1}^{L} W_h \bar{Y}_h / \sum_{h=1}^{L} W_h (P_h \beta_{2h}(\phi) + C_{ph})$, $R_4 = \sum_{h=1}^{L} W_h \bar{Y}_h / \sum_{h=1}^{L} W_h (P_h C_{ph} + \beta_{2h}(\phi))$, $R_5 = \sum_{h=1}^{L} W_h \bar{Y}_h / \sum_{h=1}^{L} W_h (P_h + \rho)$.  
Zaman (2021) suggested another improved estimator under SSRS given as

$$T_{z6} = \bar{y}_{st} \exp \left( \frac{P_{st6} - p_{st6}}{P_{st6} + p_{st6}} \right)^{\alpha}$$

where $\alpha$ is a suitably chosen scalar, $P_{st6} = \sum_{h=1}^{L} W_h (mP_h + n)$, $p_{st6} = \sum_{h=1}^{L} W_h (mp_h + n)$ such that $m$ and $n$ are either real values or the function of known parameters associated with the auxiliary attribute namely, standard deviation $S_{ph}$, coefficient of correlation $\rho_h$, coefficient of kurtosis $\beta_{2h}(\phi)$, coefficient of variation $C_{ph}$, etc in stratum $h$.

The MSE of the above estimator is given by

$$\text{MSE}(T_{z6}) = \sum_{h=1}^{L} W_h^2 \gamma_h \left[ S_{yh}^2 + \frac{\alpha^2 R_6^2}{4} S_{ph}^2 - \alpha R_6 \rho_h S_{yh} S_{ph} \right]$$

(32)

where $R_6 = \sum_{h=1}^{L} W_h \bar{Y}_h / \sum_{h=1}^{L} W_h (mP_h + n)$.  
Minimizing the MSE($T_{z6}$) w.r.t. $\alpha$, we get

$$\alpha_{(\text{opt})} = \frac{2 \sum_{h=1}^{L} W_h^2 \gamma_h \rho_h S_{yh} S_{ph}}{R_6 \sum_{h=1}^{L} W_h \gamma_h S_{ph}^2}$$

Putting $\alpha_{(\text{opt})}$ in the MSE($T_{z6}$), we get

$$\text{minMSE}(T_{z6}) = \sum_{h=1}^{L} W_h^2 \gamma_h S_{yh}^2 (1 - \rho_h^2)$$

(33)

which is the minimum MSE of classical regression estimator $T_{lr}$.

3 Proposed Classes of Estimators

Motivated by Bhushan and Kumar (2020) and Bhushan et al. (2021c), we have proposed some improved classes of difference and ratio type estimators
for population mean using attribute under SSRS as

\[ T_{p1} = \alpha_1 \bar{y}_{st} + \beta_1 (p_{st} - P) \]  
(34)

\[ T_{p2} = \alpha_2 \bar{y}_{st} \left( \frac{P}{p_{st}} \right) \beta_2 \]  
(35)

\[ T_{p3} = \alpha_3 \bar{y}_{st} \left( \frac{P}{P + \beta_3 (p_{st} - P)} \right) \]  
(36)

\[ T_{p4} = \alpha_4 \bar{y}_{st} + \beta_4 (p_{st}^* - P^*) \]  
(37)

\[ T_{p5} = \alpha_5 \bar{y}_{st} \left( \frac{P^*}{p_{st}^*} \right) \beta_5 \]  
(38)

\[ T_{p6} = \alpha_6 \bar{y}_{st} \left( \frac{P^*}{P^* + \beta_6 (p_{st}^* - P^*)} \right) \]  
(39)

where \( \alpha_i, \beta_i, i = 1, 2, \ldots, 6 \) are suitably chosen scalars, \( P^* = mP + n, p_{st}^* = mp_{st} + n, m \) and \( n \) are either real values or function of parameters of auxiliary attribute \( \phi \).

**Theorem 3.1.** The minimum MES of the proposed class of estimators \( T_{pi}, i = 1, 4 \) is given by

\[ \text{minMSE}(T_{pi}) = \bar{Y}^2 \left[ 1 - \alpha_i(\text{opt}) \right] = \bar{Y}^2 \left[ 1 - \frac{B_i^2}{A_i} \right] \]  
(40)

**Proof:** Consider the estimator

\[ T_{p1} = \alpha_1 \bar{y}_{st} + \beta_1 (p_{st} - P) \]

Express the above estimator in terms of e’s, we get

\[ T_{p1} - \bar{Y} = (\alpha_1 - 1)\bar{Y} + \alpha_1 \bar{Y} e_0 + \beta_1 P e_1 \]  
(41)

Squaring both sides of (41) and taking expectation, we will get the MSE of the estimator \( T_{p1} \) up to first order of approximation as

\[ \text{MSE}(T_{p1}) = \bar{Y}^2 (\alpha_1 - 1)^2 + \sum_{h=1}^{L} W_h^2 (\alpha_1 \bar{Y} \gamma_h c_{yh}^2 + \beta_1 \bar{Y} \gamma_h c_{ph}^2 \]  
\[ + 2\alpha_1 \beta_1 \bar{Y} P \gamma_h \rho_h c_{yh} c_{ph}) \]  
(42)
The above MSE is minimized for $\alpha_1$ and $\beta_1$ as

$$\alpha_{1\text{(opt)}} = \frac{1}{1 + \sum_{h=1}^{L} \gamma_h C_{yh}^2 - \left(\frac{\sum_{h=1}^{L} \gamma_h \rho_h C_{yh} C_{ph}}{\sum_{h=1}^{L} \gamma_h C_{ph}^2}\right)^2} = \frac{B_1}{A_1} \text{(say)}$$

and

$$\beta_{1\text{(opt)}} = -\frac{\bar{Y} \sum_{h=1}^{L} \gamma_h \rho_h C_{yh} C_{ph}}{P \sum_{h=1}^{L} \gamma_h C_{ph}^2} \alpha_{1\text{(opt)}}$$

Putting $\alpha_{1\text{(opt)}}$ and $\beta_{1\text{(opt)}}$ in the MSE($T_{p_1}$), we get the minimum MSE as

$$\min\text{MSE}(T_{p_1}) = \bar{Y}^2 \left(1 - \alpha_{1\text{(opt)}}\right) = \bar{Y}^2 \left(1 - \frac{B_1^2}{A_1}\right) \quad (43)$$

The minimum MSE of the other estimator $T_{p_i}$ can be found in similar lines.

**Theorem 3.2.** The minimum MSE of the proposed classes of estimators $T_{p_i}, i = 2, 3, 5, 6$ is given by

$$\min\text{MSE}(T_{p_i}) = \bar{Y}^2 \left(1 - \frac{B_i^2}{A_i}\right) \quad (44)$$

**Proof:** Using the notations discussed earlier, the MSE of estimator $T_{p_2}$ is given by

$$\text{MSE}(T_{p_2}) = \bar{Y}^2 \left[1 + \alpha_2^2 \left\{1 + \sum_{h=1}^{L} W_h^2 \gamma_h (C_{yh}^2 + \beta_2 (2 \beta_2 + 1) C_{ph}^2) - 4 \beta_2 \rho_h C_{yh} C_{ph}\right\} - 2 \alpha_2 \left\{1 + \sum_{h=1}^{L} W_h^2 \gamma_h \left(\frac{\beta_2 (\beta_2 + 1)}{2} C_{ph}^2\right) - \beta_2 \rho_h C_{yh} C_{ph}\right\}\right]$$

which can further be written as

$$\text{MSE}(T_{p_2}) = \bar{Y}^2 [1 + \alpha_2^2 A_2 - 2 \alpha_2 B_2] \quad (45)$$
Minimizing the MSE\(T_{p2}\) w.r.t. the scalar \(\alpha_2\), we get

\[
\alpha_{2(\text{opt})} = \frac{B_2}{A_2}
\]

Putting \(\alpha_{2(\text{opt})}\) in the MSE\(T_{p2}\), we get

\[
\min \text{MSE}(T_{p2}) = Y^2 \left(1 - \frac{B_2^2}{A_2^2}\right)
\] (47)

The MSE of other estimators can be obtained in similar lines. In general, we can write

\[
\text{MSE}(T_{p_i}) = Y^2 \left[1 + \alpha_i^2 A_i - 2\alpha_i B_i\right]
\]

It is to be noted that the simultaneous minimization of \(\alpha_i\) and \(\beta_i\) of the above MSE expression is not possible so we utilize the optimum values of \(\beta_i = \beta_{i(\text{opt})}\) when \(\alpha_i = 1\) and use this within \(\alpha_i = \alpha_{i(\text{opt})}\) to obtain the MSE expressions. The optimum values of the scalars are given by

\[
\alpha_{i(\text{opt})} = \frac{B_i}{A_i}, \quad i = 2, 3, 5, 6
\]

where

\[
A_2 = 1 + \sum_{h=1}^{L} W_h^2 \gamma_h \left(C_{y_h}^2 + \beta_2 (2\beta_2 + 1) C_{p_h}^2 - 4\beta_2 \rho_h C_{y_h} C_{p_h}\right)
\]

\[
B_2 = 1 + \sum_{h=1}^{L} W_h^2 \gamma_h \left(\frac{\beta_2 (\beta_2 + 1)}{2} C_{p_h}^2 - \beta_2 \rho_h C_{y_h} C_{p_h}\right)
\]

\[
A_3 = 1 + \sum_{h=1}^{L} W_h^2 \gamma_h \left(C_{y_h}^2 + 3\beta_3^2 C_{p_h}^2 - 4\beta_3 \rho_h C_{y_h} C_{p_h}\right)
\]

\[
B_3 = 1 + \sum_{h=1}^{L} W_h^2 \gamma_h \left(\beta_3 C_{p_h}^2 - \beta_3 \rho_h C_{y_h} C_{p_h}\right)
\]

\[
A_5 = 1 + \sum_{h=1}^{L} W_h^2 \gamma_h \left(C_{y_h}^2 + \beta_5 (2\beta_5 + 1) \nu_h^2 C_{p_h}^2 - 4\beta_5 \nu_h \rho_h C_{y_h} C_{p_h}\right)
\]

\[
B_5 = 1 + \sum_{h=1}^{L} W_h^2 \gamma_h \left(\frac{\beta_5 (\beta_5 + 1)}{2} \nu_h^2 C_{p_h}^2 - \beta_5 \nu_h \rho_h C_{y_h} C_{p_h}\right)
\]
On Some Improved Classes of Estimators Under Stratified Sampling

\[ A_6 = 1 + \sum_{h=1}^{L} W_h^2 \gamma_h (C_{yh}^2 + 3\beta_6^2 v_h^2 C_{ph}^2 - 4\beta_6 v_h \rho_h C_{yh} C_{ph}) \]

\[ B_6 = 1 + \sum_{h=1}^{L} W_h^2 \gamma_h (\beta_6^2 v_h^2 C_{ph}^2 - \beta_6 v_h \rho_h C_{yh} C_{ph}) \]

Here, \( \beta_{2(\text{opt})} = \left( \sum_{h=1}^{L} \gamma_h \rho_h C_{yh} C_{ph} / \sum_{h=1}^{L} \gamma_h C_{ph}^2 \right) = \beta_{3(\text{opt})} \) and \( \beta_{5(\text{opt})} = \beta_{2(\text{opt})}/v_h = \beta_{6(\text{opt})} \) are used as optimum values when \( \alpha_i = 1 \) is considered in the corresponding estimators.

4 Efficiency Conditions

We compare the minimum MSE of the proposed estimators \( T_{pi} \), \( i = 1, 2, \ldots, 6 \) with the minimum MSE of the existing estimators and get the following efficiency conditions.

(i). On comparing minimum MSE of the proposed estimators \( T_{pi} \) with mean per unit estimator \( T_m \) from (40) and (44) with (2), we get

\[ \frac{B_i^2}{A_i} \geq 1 - \sum_{h=1}^{L} W_h^2 \gamma_h C_{yh}^2 \] \hspace{1cm} \text{(48)}

(ii). On comparing minimum MSE of proposed estimators \( T_{pi} \) with classical ratio estimator \( T_r \) from (40) and (44) with (6), we get

\[ \frac{B_i^2}{A_i} \geq 1 - \sum_{h=1}^{L} W_h^2 \gamma_h [C_{yh}^2 + C_{ph}^2 - 2\rho_h C_{yh} C_{ph}] \] \hspace{1cm} \text{(49)}

(iii). On comparing minimum MSE of proposed estimators \( T_{pi} \) with classical product estimator \( T_p \) from (40) and (44) with (7), we get

\[ \frac{B_i^2}{A_i} \geq 1 - \sum_{h=1}^{L} W_h^2 \gamma_h [C_{yh}^2 + C_{ph}^2 + 2\rho_h C_{yh} C_{ph}] \] \hspace{1cm} \text{(50)}

(iv). On comparing minimum MSE of proposed estimators \( T_{pi} \) and classical regression estimator \( T_{lr} \), Sharma and Singh (2013) estimator \( T_{ss3} \), Zaman (2020) estimator \( T \) and Zaman (2021) estimator \( T_{z6} \) from (40) and (44) with (9), (16), (24), (33) respectively, we get

\[ \frac{B_i^2}{A_i} \geq 1 - \sum_{h=1}^{L} W_h^2 \gamma_h C_{yh}^2 (1 - \rho_h^2) \] \hspace{1cm} \text{(51)}
(v). On comparing minimum MSE of proposed estimator $T_{pi}$ with Sharma and Singh (2013) estimator $T_{ss1}$ from (40) and (44) with (12), we get

$$\frac{B_i^2}{A_i} \geq 1 - \sum_{h=1}^{L} W_h^2 \gamma_h \left[ C_{yh}^2 + \frac{C_{ph}^2}{4} - \rho_h C_{yh} C_{ph} \right]$$

(vi). On comparing minimum MSE of proposed estimators $T_{pi}$ with Sharma and Singh (2013) estimator $T_{ss2}$ from (40) and (44) with (13), we get

$$\frac{B_i^2}{A_i} \geq 1 - \sum_{h=1}^{L} W_h^2 \gamma_h \left[ C_{yh}^2 + \frac{C_{ph}^2}{4} + \rho_h C_{yh} C_{ph} \right]$$

(vii). On comparing minimum MSE of proposed estimators $T_{pi}$ with Zaman (2021) estimators $T_{zi}, i = 1, 2, \ldots, 5$ from (40) and (44) with (30), we get

$$\frac{B_i^2}{A_i} \geq 1 - \frac{1}{Y^2} \sum_{h=1}^{L} W_h^2 \gamma_h \left[ S_{yh}^2 + \frac{R^2}{4} S_{ph}^2 - R \rho_h S_{yh} S_{ph} \right]$$

(viii). On comparing minimum MSE of proposed estimator $T_{pi}$ with Zaman and Kadilar (2020) estimator $T_{zk1}$ from (40) and (44) with (18), we get

$$\frac{B_i^2}{A_i} \geq 1 - \sum_{h=1}^{L} W_h^2 \gamma_h \left[ C_{yh}^2 + \frac{C_{ph}^2}{4} + \rho_h C_{yh} C_{ph} \right]$$

(ix). On comparing minimum MSE of proposed estimators $T_{pi}$ with Zaman and Kadilar (2020) estimator $T_{zk2}$ from (40) and (44) with (21), we get

$$\frac{B_i^2}{A_i} \geq 1 - \frac{1}{Y^2} \left\{ \begin{array}{l}
\sum_{h=1}^{L} W_h^2 \gamma_h S_{yh}^2 \\
+ (3\lambda^2 - 2\lambda) \frac{R^2}{4} \sum_{h=1}^{L} W_h^2 \gamma_h v_h^2 S_{ph}^2 \\
- R(2\lambda^2 - \lambda) \sum_{h=1}^{L} W_h^2 \gamma_h v_h \rho_h S_{yh} S_{ph} \end{array} \right\}$$

$$\sum_{h=1}^{L} W_h^2 \gamma_h S_{yh}^2$$
It is to be noted that only under above conditions, the proposed classes of estimators $T_{p_i}$, $i = 1, 2, \ldots, 6$ become superior than the usual mean estimator, classical ratio, product and regression estimators, Sharma and Singh (2013) estimator, Zaman and Kadilar (2020) estimator, Zaman (2020) type estimator and Zaman (2021) estimators. Subsequently, these conditions are verified through an empirical study using real data set.

5 Empirical Study

In order to verify the theoretical results of the proposed classes of estimators, we have performed an empirical study over the data set of Kadilar and Cingi (2003). The data set is based on the production of apples as study variable and number of apple trees as auxiliary variable in six regions of Turkey namely, Marmara, Agean, Mediterranean, central Anatolia, Black Sea and East and Southeast Anatolia in 1999. These six regions of Turkey are considered as sub-population/strata. Taking Neyman allocation (Cochran, 1977) into consideration, we have randomly drawn the districts from each strata using given formula:

$$n_h = n \frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h}, \quad h = 1, 2, \ldots, 6$$

(57)

Neyman allocation (Neyman, 1934) is a method to allocate samples into strata based on the strata variance and corresponding sampling cost in the strata which gives an unbiased estimator of population mean provided the total sample size. The different strata may not differ much from each other w.r.t. costs. We consider equal costs for all strata. However, the sample can often be divided into subsamples using SSRS such that for L subsamples, each of them has a sample size $n_h$, $h = 1, 2, \ldots, L$, with $n_1 + n_2 + \cdots + n_L = n$. A sample of size $n = 200$ units is drawn with the help of Neyman allocation defined in (57) and summarized in Table 1. The percent relative efficiency (PRE) of the class of estimators with respect to mean per unit estimator $T_m$ are computed using the following formula:

$$\text{PRE} = \frac{\text{MSE}(T_m)}{\text{MSE}(T)} \times 100$$

where $T = T_m$, $T_r$, $T_p$, $T_{ir}$, $T_{z_{k_1}}$, $T_{z_{k_2}}$, $T_z$, $T_{z_i}$, $i = 1, 2, \ldots, 6$ and $T_{p_i}$, $i = 1, 2, \ldots, 6$. The empirical results are displayed in Table 2 in terms of MSE and PRE. These results of Table 2 are further displayed through line diagrams given in Figures 1 and 2 by MSE and PRE respectively.
Table 1 Descriptive statistics of the population

<table>
<thead>
<tr>
<th>Known Symbol for Parameters</th>
<th>Total</th>
<th>Stratum h</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population size N</td>
<td>N = 854</td>
<td>N_h</td>
<td>106</td>
<td>106</td>
<td>94</td>
<td>171</td>
<td>204</td>
<td>173</td>
</tr>
<tr>
<td>Sample size n</td>
<td>n = 200</td>
<td>n_h</td>
<td>13</td>
<td>24</td>
<td>55</td>
<td>95</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Population mean Y</td>
<td>Y = 2930.1</td>
<td>Y_h</td>
<td>1536</td>
<td>2212</td>
<td>9384</td>
<td>5588</td>
<td>5588</td>
<td>404</td>
</tr>
<tr>
<td>Population proportion P</td>
<td>P = 0.334</td>
<td>P_h</td>
<td>0.24</td>
<td>0.29</td>
<td>0.46</td>
<td>0.48</td>
<td>0.36</td>
<td>0.11</td>
</tr>
<tr>
<td>Standard deviation S_y</td>
<td>S_y = 17105</td>
<td>S_yh</td>
<td>6425</td>
<td>11551</td>
<td>29907</td>
<td>28643</td>
<td>2389</td>
<td>945</td>
</tr>
<tr>
<td>Standard deviation S_p</td>
<td>S_p = 0.466</td>
<td>S_ph</td>
<td>0.43</td>
<td>0.45</td>
<td>0.50</td>
<td>0.50</td>
<td>0.48</td>
<td>0.32</td>
</tr>
<tr>
<td>Kurtosis coefficient β_2(p)</td>
<td>β_2(p) = -1.39</td>
<td>β_2h(p)</td>
<td>0.56</td>
<td>1.16</td>
<td>-2.03</td>
<td>2.02</td>
<td>1.68</td>
<td>3.93</td>
</tr>
<tr>
<td>Variation coefficient C_p</td>
<td>C_p = 1.46</td>
<td>C_ph</td>
<td>1.76</td>
<td>1.56</td>
<td>1.07</td>
<td>1.03</td>
<td>1.33</td>
<td>2.77</td>
</tr>
<tr>
<td>Covariance coefficient S_yp</td>
<td>S_yp = 904.7</td>
<td>S_yp_h</td>
<td>996</td>
<td>1404</td>
<td>48674</td>
<td>2743</td>
<td>449</td>
<td>204</td>
</tr>
<tr>
<td>Weight W</td>
<td>W_h</td>
<td></td>
<td>0.12</td>
<td>0.12</td>
<td>0.11</td>
<td>0.20</td>
<td>0.24</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Table 2 MSE and PRE of different estimator with respect to \( T_m \)

<table>
<thead>
<tr>
<th>Estimators</th>
<th>MSE</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_m )</td>
<td>387786.7</td>
<td>100.0000</td>
</tr>
<tr>
<td>( T_r )</td>
<td>276560.1</td>
<td>140.2179</td>
</tr>
<tr>
<td>( T_p )</td>
<td>970347.4</td>
<td>39.9637</td>
</tr>
<tr>
<td>( T^* )</td>
<td>260132.8</td>
<td>149.0726</td>
</tr>
<tr>
<td>( T_{s1} )</td>
<td>273256.6</td>
<td>141.9130</td>
</tr>
<tr>
<td>( T_{s2} )</td>
<td>620150.3</td>
<td>62.5310</td>
</tr>
<tr>
<td>( T_{s3} )</td>
<td>357499.9</td>
<td>108.4718</td>
</tr>
<tr>
<td>( T_{s4} )</td>
<td>357160.8</td>
<td>108.5748</td>
</tr>
<tr>
<td>( T_{s5} )</td>
<td>358365.0</td>
<td>108.2100</td>
</tr>
<tr>
<td>( T_{s6} )</td>
<td>358812.8</td>
<td>108.0749</td>
</tr>
<tr>
<td>( T_{z1} )</td>
<td>347115.3</td>
<td>111.717</td>
</tr>
<tr>
<td>( T_{z2} )</td>
<td>360353.6</td>
<td>107.6128</td>
</tr>
<tr>
<td>( T_{z3} )</td>
<td>333791.4</td>
<td>116.5447</td>
</tr>
<tr>
<td>( T_{p1}, i = 1,3,4,6 )</td>
<td>253704.6</td>
<td>152.8497</td>
</tr>
<tr>
<td>( T_{p2}, i = 2,5 )</td>
<td>252556.1</td>
<td>153.5447</td>
</tr>
</tbody>
</table>

where \( T^* = T_r, T_{s3}, T_{s4}, T_z \).

5.1 Results and Discussion

On comparing the outcomes of Table 2, it has been observed that:

1. the proposed classes of estimators \( T_{pj}, i = 1,2,\ldots,6 \) are highly rewarding in terms of minimum MSE than the existing estimators such as usual mean estimator \( T_m \), classical ratio estimator \( T_r \), classical product estimator \( T_p \), classical regression estimator \( T_{lr} \), Sharma and Singh (2013) estimator \( T_{ss}, i = 1,2,3 \), Zaman and Kadilar (2020) estimators \( T_{zk}, i = 1,2 \), Zaman (2020) type estimator \( T_z \) and Zaman (2021)
estimators $T_{z_i}, i = 1, 2, \ldots, 6$. This can be easily observed from line
diagram displayed in Figure 1.
2. the proposed classes of estimators $T_{p_i}, i = 1, 2, \ldots, 6$ are found to
be efficient in terms of maximum PRE providing better improvement

Figure 1 MSE of existing and proposed estimators.

Figure 2 PRE of existing and proposed estimators.
over the existing estimators such as usual mean estimator $T_m$, classical ratio estimator $T_r$, classical product estimator $T_p$, classical regression estimator $T_{lr}$, Sharma and Singh (2013) estimator $T_{ss}$, $i = 1, 2, 3$, Zaman and Kadilar (2020) type estimators $T_{zk}$, $i = 1, 2$, Zaman (2020) type estimator $T_z$ and Zaman (2021) estimators $T_{z1}$, $i = 1, 2, \ldots, 6$. This can be easily observed from line diagram displayed in Figure 2.

3. Moreover, the proposed class of estimator $T_{p}$, $i = 2, 5$ performed better than the other proposed classes of estimators in terms of minimum MSE and maximum PRE which can be also observed from line diagrams displayed in Figures 1 and 2 for MSE and PRE respectively.

6 Conclusion

This paper proposes some improved classes of estimators under stratified simple random sampling to estimate population mean with their properties. The usual mean estimator, classical ratio, product and regression estimators are special cases of the suggested classes of estimators for suitably chosen values of characterizing scalars. The proposed classes of estimators are turned out to be remunerating in terms of MSE and PRE when applied in real life scenario. These estimators are also showing their supremacy in terms of lesser MSE and greater PRE over conventional estimators such as mean estimator $T_m$, ratio estimator $T_r$, product estimator $T_p$, regression estimator $T_{lr}$, Sharma and Singh (2013) estimator $T_{ss}$, $i = 1, 2, 3$, Zaman and Kadilar (2020) estimators $T_{zk}$, $i = 1, 2$, Zaman (2020) type estimator $T_z$ and Zaman (2021) estimators $T_{z1}$, $i = 1, 2, \ldots, 6$ when empirical study has been performed over real data. The performances of different estimators can also be observed from the Figures 1 and 2 displayed for MSE and PRE respectively. The empirical results support that the proposed classes of estimators are appreciatively favorable in abating the MSE to a greater extend as compare to the conventional estimators. Hence, looking on the assured behavior of the proposed classes of estimators, survey practitioners may be encouraged to utilize the proposed classes of estimators for their practical applications.

Acknowledgement

Authors are extremely thankful to the learned referees for their valuable suggestions regarding improvement of the paper and to the editor-in-chief Prof. Dr. Mangey Ram.
References


**Biographies**

**Shashi Bhushan** received his Ph.D. degree in statistics from Lucknow University, Lucknow, India. He is currently working as a Professor in the Department of Statistics, Lucknow University, Lucknow, India. He has more than 15 years of teaching experience and 20 years of research experience. He has supervised six Ph.D till date. His research interests include Sample survey, missing data, non-response, measurement errors, etc. He has various publications in National and International journals of repute.
Anoop Kumar is pursuing his Ph.D. in Applied Statistics from the Department of Mathematics and Statistics, Dr. Shakuntala Misra National Rehabilitation University, Lucknow, India. He received his M.Sc. degree from Babasaheb Bhimrao Ambedkar University, Lucknow, India. He also qualified UGC NET twice in population studies. His research area is sampling survey, missing data, measurement errors. He has publications in various National and International journals of repute.

Dushyant Tyagi has done his M.Sc., M.Phil. and Ph.D. (Statistics) from Department of Statistics, Ch. Charan Singh University, Meerut and possess Eleven years of experience of educating in various institutions of repute like G. B. Pant University of Agriculture and Technology, Institute of Technology and Science Ghaziabad, International College of Financial Planning, New Delhi and Lady Shri Ram College for Women, New Delhi. He is currently working as an Assistant Professor at the Department of Mathematics and Statistics, Faculty of Science and Technology, Dr. Shakuntala Misra National Rehabilitation University, Lucknow. His research area is Statistical Quality Control and Computational Statistics. He held the responsibility of
Convener and resource person for three AICTE sponsored Faculty Development Program on Advance Data Analysis through Data Analysis Software’s. He delivered lectures in more than 25 research methodology workshops. He has six research paper publication in reputed International journals and one book. He has presented his research work in various National and International Conferences and attended several seminars and FDP’s of statistics and related areas.

Saurabh Singh is currently pursuing Ph.D. in Applied Statistics from the Department of Mathematics and Statistics, Dr. Shakuntala Misra National Rehabilitation University, Lucknow, India. He received his M.Phil. degree from Babasaheb Bhimrao Ambedkar University, Lucknow, India. His research area is sampling survey, missing data. He has 10 publications in various National and International journals.