Mathematics

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

AIMS Mathematics, 6(12): 13592–13607.
DOI:10.3934/math.2021790
Received: 27 May 2021
Accepted: 13 September 2021
Published: 24 September 2021

Research article

A general class of estimators on estimating population mean using the auxiliary proportions under simple and two phase sampling

Xuechen Liu[1,*] **and Muhammad Arslan**[2]

[1] Faculty of Economics, Taiyuan Normal University, Taiyuan, China
[2] Shanxi University of Finance and Economics, Taiyuan, China

* **Correspondence:** Email: liu.xuechen@tynu.edu.cn.

Abstract: This article deals with estimation of finite population mean using the auxiliary proportion under simple and two phase sampling scheme utilizing two auxiliary variables. Mathematical expressions for the mean squared errors of the proposed estimators are derived under first order of approximation. We compare the proposed class of estimators "theoretically and numerically" with the usual mean estimator of Naik and Gupta [1]. The theoretical as well as numerical findings support the superiority of our proposed class of estimator as compared to estimators available in literature.

Keywords: simple and two phase sampling; proportion; auxiliary information; mean squared error; numerical comparison
Mathematics Subject Classification: 62D99, 62F10

1. Introduction

1.1. Introduction and notation in simple random sampling

It is a well-known fact, that at large scale survey sampling, the use of several auxiliary variables improve the precision of the estimators. In survey sampling, researchers have already attempted to obtain the estimates for population parameter such as mean, median etc, that posses maximum statistical properties. For that purpose a representative part of population is needed, when population of interest is homogeneous then one can use simple random sampling (SRS) for selecting units. In some situations, information available in the form of attributes, which is positively correlated with study variables. Several authors including Naik and Gupta [1], Jhajj [2], Abd-Elfattah [3], Koyuncu [4], Solanki [5], Sharma [6] and Malik [7] proposed a set of estimators, taking the advantages of bi-serial correlation between auxiliary and study variables, utilizing information on single auxiliary attribute. Verma [8], Malik [7], Solanki et al., [9] and Sharma [10] suggested some estimators utilizing information on two auxiliary attributes in SRS, Mahdizadeh and Zamanzade [11] developed a kernel-based estimation of $P(X > Y)$ in ranked-set sampling, SinghPal and Solanki [12] developed a new class of estimators of finite population mean survey sampling and Mahdizadeh and Zamanzade [13] suggest a smooth estimation of a reliability function in ranked set sampling, further more Hussain et al., [14] and Al-Marzouki et al., [15] also work in this side.

In this article, we consider the problem of estimating the finite population mean using the auxiliary proportion under simple and two phase sampling scheme. The mathematical expression of the bias and mean squared error of the proposed estimator are derived under first order of approximation. The performance of proposed class of estimator is compared with that of the existing estimators both theoretically and numerically. In terms of percentage relative efficiency (PRE), it is found that proposed class of estimator outperforms the existing ones.

Let $U = \{u_1, u_2, ..., u_N\}$ represent a finite population of size N distinct units, assumed that a sample of size n units is drawn from this population U using simple random sampling without replacement. Let y_i and ϕ_{ij} (i=1,2) denotes the observations on variable y and ϕ_i (i=1,2) for the j^{th} unit (j=1,2,...,N).
$\phi_{ij} = 1,$ if i^{th} unit posses atrributes
$\phi_{ij} = 0,$ otherwise
$P_j = \sum_i^N \phi_{ij} = A_j/N, (j = 1, 2)$ and $p_j = \sum_i^N \phi_{ij} = a_j/n, (j = 1, 2)$ are the population and sample proportions of auxiliary variable respectively. Let $\bar{Y} = \frac{\sum_{i=1}^N y_i}{N}$, $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$ be the population and sample mean of the study variable y. $S^2_{\phi_j y} = \sum_{i=1}^N \frac{(\phi_{ij}-P_j)(y_i-\bar{Y})}{N-1}, (j = 1, 2)$ are the variations between the study and the auxiliary attributes. $S^2_{\phi_1\phi_2} = \sum_{i=1}^N \frac{(\phi_{i1}-P_1)(\phi_{i2}-P_2)}{N-1}$ are the variations between the auxiliary attributes. $\rho_{y\phi_j} = \frac{S_{\phi_j y}}{S_y S_\phi}$ represents the point bi-serial correlation between the study variable y and the two auxiliary attributes p_1 and p_2 respectively. $\rho_{\phi_1\phi_2} = \frac{S_{\phi_1\phi_2}}{S_{\phi_1} S_{\phi_2}}$ represents the point bi-serial correlation between the two auxiliary attributes p_1 and p_2 respectively.

Let us define, $e_0 = \frac{\bar{y}-\bar{Y}}{\bar{Y}}$, $e_1 = \frac{p_1-P_1}{P_1}$, $e_2 = \frac{p_2-P_2}{P_2}$,
such that, $E(e_i) = 0$ $(i = 0, 1, 2)$,
$E(e_0^2) = fC_y^2 = V_{200}$, $E(e_1^2) = fC^2_{\phi_1^2} = V_{020}$, $E(e_2^2) = fC^2_{\phi_2^2} = V_{002}$,
$E(e_0e_1) = f\rho_{y\phi_1}C_yC_{\phi_1} = V_{110}$, $E(e_0e_2) = f\rho_{y\phi_2}C_yC_{\phi_2} = V_{101}$,
$E(e_1e_2) = f\rho_{\phi_1\phi_2}C_{\phi_1}C_{\phi_2} = V_{011}$.

Where $C_y = \frac{S_y}{\bar{Y}}$, $C_{\phi_j} = \frac{S_{\phi_j}}{P_j}, (j = 1, 2)$, is the co-efficient of variation of the study and auxiliary attribute. $S_y^2 = \sum_{i=1}^N \frac{(y_i-\bar{Y})^2}{N-1}$, $S^2_{\phi_j} = \sum_{i=1}^N \frac{(\phi_{ij}-P_j)^2}{N-1}, (j = 1, 2)$, is the variance of study and auxiliary attribute. $f = \left(\frac{1}{n} - \frac{1}{N}\right)$ is the correction factor.

The rest of the paper is organized as follows. In Sections 1.1 and 1.2, introduction and notations are given for simple random sampling and two phase sampling. In Sections 2.1 and 2.3, we discussed some existing estimators of the finite population mean for both sampling designs. The proposed estimators are given in Sections 2.2 and 2.4. In Sections 3.1 and 3.2, theoretical comparisons are conducted. While in Sections 4.1 and 4.2 we focus on empirical studies. Finally, application and conclusions are drawn in Sections 5 and 6.

1.2. *Introduction and notation in two phase sampling*

The precision of estimate can be increased by using two methodologies. Firstly the precision may be increased by using using adequate sampling design for the estimated variable. Secondly the precision may be increased by using an appropriate estimation procedure, i.e. some auxiliary information which is closely associated with the variable under study. In application there exist a situation when complete auxiliary information or attribute is not available or information on that attribute is expensive. In that case, a method of two phase sampling or double sampling is used to obtain the estimates of unknown population parameters. In two phase sampling, a large preliminary sample (n') is selected by SRSWOR to obtain the estimate of unknown parameter of the auxiliary variable at first phase and the information on the auxiliary variable is collected , which is use to estimate the unknown auxiliary variable. Then a sub sample ($n < n'$) is selected at second phase and both the study and auxiliary variables are collected. Here we assume that Population proportion (P_1) is unknown and introduce an improved estimator to estimate the population mean. Kiregyera [16], Mohanty [17], Malik [7] and Haq [18] used two auxiliary variables in two phase sampling for the better estimation of mean.

An example in this context is while estimating the yield of a crop, it is likely that the area under the crop may be unknown but the area of each farm may be known. Then y, P_1 and P_2 respectively are the yield area under the crop and area under cultivation.

Consider a finite population $U = (u_1, u_2 ... u_N)$ of size N and let y_i, ϕ_{i1} and ϕ_{i2} is the information on the study variable and two auxiliary attributes associated with each unit $u_i (i = 1, 2, ..., N)$ of the population such that:
$\phi_{ij} = 1$, if the i^{th} unit in the population possesses auxiliary attribute ϕ_j, $\phi_{ij} = 0$ otherwise.

We assume that the population mean of the first auxiliary proportion P_1 is unknown but the same information is known for the second proportion. Let $p_j{}' = \frac{\sum_i^{n'} \phi_{ij}}{n} = a_j/n'$ for $j = 1, 2$ be the estimate of P_j obtained from the first phase sample of size n', drawn by using SRSWOR from the population of N units. Let $\bar{y} = \sum_i^n y_i$ and $p_1 = \frac{\sum_i^n \phi_{i1}}{n} = a_1/n$ be the estimates of $\bar{Y}$ and P_1 respectively, obtained from a second sample of size n, drawn from the first phase n' using SRSWOR.

To obtain the bias and MSE for estimators in two phase sampling we define the error terms as follows:

$e_0 = \frac{\bar{y}-\bar{Y}}{\bar{Y}}$, $e_1 = \frac{p_1 - P_1}{P_1}$, $e_{2'} = \frac{p'_2 - P_2}{P_2}$, $e'_1 = \frac{p'_1 - P_1}{P_1}$.
such that:
$E(e_0) = E(e_1) = E(e'_2) = E(e'_1) = 0$,
$E(e_0^2) = fC_y^2 = V_{200}, \quad E(e_1^2) = fC_{\phi_1}^2 = V_{020} \quad E(e_2^2) = fC_{\phi_2}^2 = V_{002}$,
$E(e_0 e_1) = f\rho_{y\phi_1} C_y C_{\phi_1} = V_{110}, \quad E(e_0 e_2) = f\rho_{y\phi_2} C_y C_{\phi_2} = V_{101}$,
$E(e_1 e_2) = f\rho_{\phi_1\phi_2} C_{\phi_1} C_{\phi_2} = V_{011}$, $E(e'_1 e_0) = f'\rho_{y\phi_1} C_y C_{\phi_1} = V_{110'}$,
$E(e'_0 e_2) = f'\rho_{y\phi_2} C_y C_{\phi_2} = V'_{101}$, $E(e_1^{2\prime}) = f'C_{\phi_1}^2 = V'_{020}$,
$E(e_2^{2\prime}) = f'C_{\phi_2^2}^2 = V'_{002}$, $f = \frac{1}{n} - \frac{1}{N}$, $f' = \frac{1}{n'} - \frac{1}{N}$,
$\bar{Y} = N\sum_{i=1}^N y_i, \quad \bar{y} = n\sum_{i=1}^n y_i$,
$S_y^2 = \sum_{i=1}^N \frac{(y_i - \bar{Y})^2}{N-1}, \quad S_{y\phi_j}^2 = \sum_{i=1}^N \frac{(\phi_{ij} - P_j)(y_i - \bar{Y})}{N-1}$,
$S_{\phi_1\phi_2}^2 = \sum_{i=1}^N \frac{(\phi_{i1} - P_1)(\phi_{i2} - P_2)}{N-1}, \quad C_y = \frac{S_y}{\bar{Y}}$,
$C_{\phi_j} = \frac{S_{\phi_j}}{P_j}$, $S_{\phi_j}^2 = \sum_{i=1}^N \frac{(\phi_{ij} - P_j)^2}{N-1}$,
$s_{\phi_j}^2 = \sum_{i=1}^n \frac{(\phi_{ij} - p_j)^2}{n-1}$, represents the sample variance of size n,

$s^2_{\phi_{j'}} = \sum_{i=1}^{n'} \frac{(\phi_{ij}-p_j)^2}{n'-1}$, represents the sample variance of size n'
$\rho_{y\phi_j} = \frac{S_{\phi_j y}}{S_y S_\phi}$ represent point bi-serial correlation between the study variable (y) and the two auxiliary attributes (P_1) and (p_2).
$\rho_{\phi_1\phi_2} = \frac{S_{\phi_1\phi 2}}{S_{\phi_1} S_{\phi_2}}$ represent point bi-serial correlation between the two auxiliary attributes (P_1) and (P_2) respectively.

2. Existing and proposed estimators

2.1. Existing estimators in simple random sampling

In order to have an estimate of the study variable, using information of population proportion P, Naik [1] proposed the following estimators respectively.

$$t_U = \bar{y}. \tag{2.1}$$

The MSE of t_U is given by

$$\text{MSE}(t_U) = \bar{Y}^2 V_{200}. \tag{2.2}$$

Naik [1] the following estimator respectively

$$t_A = \bar{y}\left(\frac{P_1}{p_1}\right), \tag{2.3}$$

$$t_B = \bar{y}\left(\frac{p_2}{P_2}\right), \tag{2.4}$$

$$t_C = \bar{y}\ exp\left(\frac{P_1 - p_1}{P_1 + p_1}\right), \tag{2.5}$$

$$t_D = \bar{y}\ exp\left(\frac{p_2 - P_2}{p_2 + P_2}\right). \tag{2.6}$$

The MSE expressions of the estimators t_A, t_B, t_C and t_D are respectively given as

$$\text{MSE}(t_A) \cong \bar{Y}^2(V_{200} - 2V_{110} + V_{020}), \tag{2.7}$$

$$\text{MSE}(t_B) \cong \bar{Y}^2(V_{200} + 2V_{101} + V_{002}), \tag{2.8}$$

$$\text{MSE}(t_C) \cong \bar{Y}^2(V_{200} - V_{110} + \frac{1}{4}V_{020}), \tag{2.9}$$

$$\text{MSE}(t_D) \cong \bar{Y}^2(V_{200} + V_{101} + \frac{1}{4}V_{002}). \tag{2.10}$$

Malik [7] proposed exponential type estimator as

$$t_{MS} = \bar{y}\ exp\left(\frac{P_1 - p_1}{P_1 + p_1}\right)^{\gamma_1} exp\left(\frac{P_2 - p_2}{P_2 + p_2}\right)^{\gamma_2} + b_1(P_1 - p_1) + b_2(P_2 - p_2), \tag{2.11}$$

where $b_1 = \frac{s_{y\phi_1}}{s^2_{\phi_1}}$ and $b_2 = \frac{s_{y\phi_2}}{s^2_{\phi_2}}$ are the sample regression coefficients. γ_1 and γ_2 are two unknown constants. The optimum values of these constants are given as:

$$\gamma_{1(opt)} = \frac{2\left\{-P_1\beta_1 C_{\phi_1}(-1+\rho^2_{\phi_1\phi_2}) + \bar{Y}C_y(-\rho_{y\phi_1} + \rho_{\phi_1\phi_2}\rho_{y\phi_2})\right\}}{\bar{Y}C_{\phi_1}(-1+\rho^2_{\phi_1\phi_2})},$$

$$\gamma_{2(opt)} = \frac{2\left\{-P_1\beta_2 C_{\phi_1}(-1+\rho^2_{\phi_1\phi_2}) + \bar{Y}C_y(-\rho_{y\phi_2} + \rho_{\phi_1\phi_2}\rho_{y\phi_1})\right\}}{\bar{Y}C_{\phi_2}(-1+\rho^2_{\phi_1\phi_2})}.$$

where $\beta_1 = \frac{S_{y\phi_1}}{S^2_{\phi_1}}$ and $\beta_2 = \frac{S_{y\phi_2}}{S^2_{\phi_2}}$, are the regression coefficients. The minimum mean squared error for the optimum values of γ_1 and γ_2 are given as:

$$\text{MSE}(t_{MS\,min}) \cong f\bar{Y}^2C^2_y(1-R^2_{y\phi_1\phi_2}), \tag{2.12}$$

where $R = \frac{\rho^2_{\phi_1 y}+\rho^2_{\phi_2 y}-2\rho_{\phi_1 y}\rho_{\phi_2 y}\rho_{\phi_1\phi_2}}{1-\rho^2_{\phi_1\phi_2}}$ is the multiple correlation of y on ϕ_1 and ϕ_2.

We used some formulas for readers to easily understand and pick-out the difficulty of long equations.

2.2. *The proposed class of estimators in simple random sampling*

We proposed generalized class of estimators for estimating mean in simple random sampling using two auxiliary attributes, as

$$t_{RPR} = k_1\bar{y} - k_2(p_1 - P_1)\left[\alpha\left\{2 - exp\left(\frac{\eta(p_2-P_2)}{\eta(p_2+P_2)+2\lambda}\right)\right\} + (1-\alpha)exp\left(\frac{\eta(P_2-p_2)}{\eta(P_2+p_2)+2\lambda}\right)\right], \tag{2.13}$$

where k_1 and k_2 are suitable constants whose values are to be determined such that MSE of t_{RPR} is minimum; η and λ are either real numbers or functions of known parameters of the auxiliary attribute ϕ_2 such as coefficient of variation (C_{ϕ_2}), coefficient of kurtosis (β_{ϕ_2}) and α is the scalar ($0 \le \alpha \le 1$) for designing different estimators. Let $\bar{Y}$ and (P_1, P_2) be the population means of the study variable and auxiliary proportions respectively. $\bar{y}$ and (p_1, p_2) be the sample means of the study variable and auxiliary proportions respectively.

Putting $\alpha = 1$ and $\alpha = 0$ in (2.13), we get the following estimators.

For $\alpha = 1$, the suggested class of estimators reduces to:

$$t_{RPR(\alpha=1)} = k_1\bar{y} - k_2(p_1 - P_1)\left[2 - exp\left\{\frac{\eta(p_2-P_2)}{\eta(p_2-P_2)+2\lambda}\right\}\right].$$

For $\alpha = 0$, the suggested class of estimators reduces to

$$t_{RPR(\alpha=0)} = k_1\bar{y} - k_2(p_1 - P_1)\left[exp\left\{\frac{\eta(P_2-p_2)}{\eta(P_2-p_2)+2\lambda}\right\}\right].$$

A set of of new estimators generated from Eq (2.13) using suitable use of α, η and λ are listed in Table 1.

Table 1. Set of estimators generated from estimator t_{RPR}.

Subset of proposed estimator	α	η	λ
$t_{RPR1} = k_1\bar{y} - k_2(p_1 - P_1)\left[exp\left\{\frac{C_{\phi_2}(P_2-p_2)}{C_{\phi_2}(P_2-p_2)+2\beta_2\phi_2}\right\}\right]$	0	C_{ϕ_2}	$\beta_{2\phi_2}$
$t_{RPR2} = k_1\bar{y} - k_2(p_1 - P_1)\left[exp\left\{\frac{P_2(P_2-p_2)}{P_2(P_2-p_2)+2}\right\}\right]$	0	P_2	1
$t_{RPR3} = k_1\bar{y} - k_2(p_1 - P_1)\left[exp\left\{\frac{(P_2-p_2)}{(P_2-p_2)+2C_{\phi_2}}\right\}\right]$	0	1	C_{ϕ_2}
$t_{RPR4} = k_1\bar{y} - k_2(p_1 - P_1)\left[exp\left\{\frac{(P_2-p_2)}{(P_2-p_2)+2}\right\}\right]$	0	1	1
$t_{RPR5} = k_1\bar{y} - k_2(p_1 - P_1)\left[2 - exp\left\{\frac{C_{\phi_2}(p_2-P_2)}{C_{\phi_2}(p_2-P_2)+2\beta_2\phi_2}\right\}\right]$	1	C_{ϕ_2}	$\beta_{2\phi_2}$
$t_{RPR6} = k_1\bar{y} - k_2(p_1 - P_1)\left[2 - exp\left\{\frac{P_2(p_2-P_2)}{P_2(p_2-P_2)+2}\right\}\right]$	1	P_2	1
$t_{RPR7} = k_1\bar{y} - k_2(p_1 - P_1)\left[2 - exp\left\{\frac{(p_2-P_2)}{(p_2-P_2)+2C_{\phi_2}}\right\}\right]$	1	1	C_{ϕ_2}
$t_{RPR8} = k_1\bar{y} - k_2(p_1 - P_1)\left[2 - exp\left\{\frac{(p_2-P_2)}{(p_2-P_2)+2}\right\}\right]$	1	1	1

Expressing Eq (2.13) in terms of e's we have

$$t_{RPR} = k_1\bar{Y}(1 + e_0) - k_2P_1e_1\left[\alpha\left\{2 - \left(1 - \gamma e_2 - \frac{1}{2}\gamma^2e_2^2\right)\right\} + (1-\alpha)\left(1 - \gamma e_2 + \frac{3}{2}\gamma^2e_2^2\right)\right], \tag{2.14}$$

where $\gamma = \frac{\eta P_2}{2(\eta P_2+\lambda)}$.

To the first degree of approximation, we have:

$$t_{RPR} - \bar{Y} \cong k_1\bar{Y} + k_1\bar{Y}e_0 - k_2P_1e_1 - k_1\gamma\bar{Y}e_2 - \gamma\bar{Y}k_1e_2e_0 + \frac{\bar{Y}k_1e_2^2\gamma^2 3}{2} - \alpha\bar{Y}k_1e_2^2\gamma^2 + \gamma k_2P_1e_1e_2 - \bar{Y}. \tag{2.15}$$

Taking expectation of the above equation we get bias of t_{RPR}, given by:

$$Bias(t_{RPR}) \cong k_1\bar{Y} - \gamma\bar{Y}k_1V_{101} + \bar{Y}k_1V_{002}\gamma^2(\frac{3}{2} - \alpha) + \gamma k_2P_1V_{011}. \tag{2.16}$$

Squaring both sides of Eq (2.15) and taking expectations of both sides, we get the MSE of the estimator t_{RPR} to the first order of approximation, as

$$\begin{aligned} E(t_{RPR} - \bar{Y})^2 \cong\ & \bar{Y}^2 + \bar{Y}^2k_1^2\left(1 - 4\gamma V_{101} - 2\alpha V_{002}\gamma^2 + 4\gamma^2V_{002} + V_{200}\right) \\ & -k_1\bar{Y}^2\left(2 - 2\alpha V_{002}\gamma^2 - 2\gamma V_{101} + 3V_{002}\gamma^2\right) \\ & +2k_1k_2\bar{Y}\left(2\gamma P_1V_{011} - P_1V_{110}\right) \\ & - 2k_2\bar{Y}\left(\gamma P_1V_{011}\right) + k_2^2\left(P_1^2V_{020}\right), \end{aligned} \tag{2.17}$$

$$\text{MSE}t_{RPR} \cong \bar{Y}^2 + \bar{Y}^2k_1^2A - k_1\bar{Y}^2B + 2k_1k_2\bar{Y}C - 2k_2\bar{Y}D + k_2^2E. \tag{2.18}$$

where

$$A = 1 - 4\gamma V_{101} - 2\alpha V_{002}\gamma^2 + 4\gamma^2V_{002} + V_{200},$$

$$B = 2 - 2\alpha V_{002}\gamma^2 - 2\gamma V_{101} + 3V_{002}\gamma^2,$$

$$C = 2\gamma^2 P_1 V_{011} - P_1 V_{110}, \quad D = \gamma P_1 V_{011}, \quad E = P_1^2 V_{020}.$$

The optimum values of k_1 and k_2 are obtained by minimizing Eq (2.18) and is given by

$$k_1 = \frac{BE - 2CD}{2\,(AE - C^2)},$$

and

$$k_2 = \frac{\bar{Y}\,(2AD - BC)}{2\,(AE - C^2)},$$

Substituting the optimum values of k_1 and k_2 in Eq (2.18) we get the minimum MSE of t_{RPR} as:

$$\text{MSE}(t_{(RPR)min}) = \frac{\bar{Y}^2\left(1 - 4AD^2 + B^2E - 4BCD\right)}{4\,(AE - C^2)}. \tag{2.19}$$

The minimum MSE of the proposed estimator tRPR at Eq (2.19) depends upon many parametric constants, we use these constant for readers to easily understand and for notation convenient.

2.3. Existing estimators in two phase sampling

The usual mean per unit estimator in two phase sampling is:

$$t'_U = \bar{y}. \tag{2.20}$$

The MSE of t'_U is given by

$$\text{MSE}(t'_U) = \bar{Y}^2 V_{200}. \tag{2.21}$$

The Naik [1] estimators in two phase sampling are :

$$t_A{'} = \bar{y}\left(\frac{p'_1}{p_1}\right), \tag{2.22}$$

$$t_B{'} = \bar{y}\left(\frac{P_2}{p'_2}\right), \tag{2.23}$$

$$t_C{'} = \bar{y}\ exp\left(\frac{p_1{'} - p_1}{p_1{'} + p_1}\right), \tag{2.24}$$

$$t_D{'} = \bar{y}\ exp\left(\frac{P_2 - p_2{'}}{P_2 + p_2{'}}\right). \tag{2.25}$$

The MSE expressions of estimators $t_A{'}$, $t_B{'}$, $t_C{'}$ and $t_D{'}$ are respectively given as:

$$\text{MSE}(t_A{'}) \cong \bar{Y}^2\left(V_{200} + V_{020} - V'_{020} + 2V'_{110} - 2V'_{110}\right), \tag{2.26}$$

$$\text{MSE}(t_B') \cong \bar{Y}^2\left(V_{200} + V'_{002} + 2V'_{101}\right), \tag{2.27}$$

$$\text{MSE}(t_C') \cong \bar{Y}^2\left(V_{200} + V'_{110} - V_{110} - \frac{1}{4}V'_{020} + \frac{1}{4}V_{020}\right), \tag{2.28}$$

$$\text{MSE}(t_D') \cong \bar{Y}^2\left(V_{200} + \frac{1}{4}V'_{002} + V'_{101}\right). \tag{2.29}$$

Malik [7] used exponential type estimator with regression coefficients in two phase sampling which is given by:

$$t_{MS} = \bar{y}\ exp\left(\frac{p_1{}' - p_1}{p_1{}' + p_1}\right)^{\delta_1} exp\left(\frac{P_2 - p_2{}'}{P_2 + p_2{}'}\right)^{\delta_2} + b_1(P_1 - p_1) + b_2(P_2 - p_2), \tag{2.30}$$

where $b_1 = \frac{s_{y\phi_1}}{s^2_{\phi_1}}$ and $b_2 = \frac{s_{y\phi_2}}{s^2_{\phi_2}}$ are the sample regression coefficients. δ_1 and δ_2 are two unknown constants. The optimum values of these constants are given as:

$$\delta_{1(opt)} = \frac{-2P_1\beta_1}{\bar{Y}} + \frac{2C_y\rho_{y\phi_1}}{C_{\phi_1}},$$

$$\delta_{2(opt)} = \frac{-2P_2\beta_2}{\bar{Y}} + \frac{2C_y\rho_{y\phi_2}}{C_{\phi_2}},$$

where, $\beta_1 = \frac{S_{y\phi_1}}{S^2_{\phi_1}}$ and $\beta_2 = \frac{S_{y\phi_2}}{S^2_{\phi_2}}$ are the regression coefficients.

The minimum mean square error for the optimum values of δ_1 and δ_2 are given as:

$$\text{MSE}(t_{MS\,min}) \cong f\bar{Y}^2C_y^2\left\{-f(-1 + \rho^2_{y\phi_1}) + \lambda(\rho^2_{y\phi_1} - \rho^2_{y\phi_2})\right\}. \tag{2.31}$$

2.4. A generalized proposed class of estimators in two phase sampling

We suggest a generalized exponential estimator when P_1 is unknown and P_2 is known:

$$t'_{RPR} = k'_1\bar{y} - k'_2(p_1 - p'_1)\left[\alpha'\left\{2 - exp\left(\frac{\eta'(p'_2 - P_2)}{\eta'(p'_2 + P_2) + 2\lambda'}\right)\right\} + (1-\alpha')exp\left(\frac{\eta'(P_2 - p'_2)}{\eta'(P_2 + p'_2) + 2\lambda'}\right)\right]. \tag{2.32}$$

where k'_1 and k'_2 are suitable constants whose value are to be determined such that MSE of t_{RPR} is minimum. η' and λ' are either real numbers or functions of known parameters of the auxiliary attribute ϕ_2 such as coefficient of variation, coefficient of kurtosis (β_{ϕ_2}) and α' is a scalar ($0 \leq \alpha \leq$) for designing different estimators.

Putting $\alpha' = 1$ and $\alpha' = 0$ in above suggested class of estimators, we get the following estimators.

For $\alpha' = 1$, the suggested class of estimators reduces to:

$$t'_{RPR(\alpha'=1)} = k'_1\bar{y} - k'_2(p_1 - p'_1)\left[2 - exp\left\{\frac{\eta(p'_2 - P_2)}{\eta(p'_2 - P_2) + 2\lambda'}\right\}\right].$$

For $\alpha' = 0$, the suggested class of estimators reduces to:

$$t'_{RPR(\alpha'=0)} = k'_1\bar{y} - k'_2(p_1 - p'_1)\left[exp\left\{\frac{\eta(P_2 - p'_2)}{\eta(P_2 - p'_2) + 2\lambda'}\right\}\right].$$

Table 2. Set of estimators generated from estimator $t'_{RPR(\alpha'=1)}$.

Subset of proposed estimator	α'	η'	λ'
$t'_{RPR1} = k'_1\bar{y} - k'_2(p_1 - p'_1)\left[exp\left\{\frac{C_{\phi_2}(P_2-p'_2)}{C_{\phi_2}(P_2-p'_2)+2\beta_2\phi_2}\right\}\right]$	0	C_{ϕ_2}	$\beta_{2\phi_2}$
$t'_{RPR2} = k'_1\bar{y} - k'_2(p_1 - p'_1)\left[exp\left\{\frac{P_2(P_2-p'_2)}{P_2(P_2-p'_2)+2}\right\}\right]$	0	P_2	1
$t'_{RPR3} = k'_1\bar{y} - k'_2(p_1 - p'_1)\left[exp\left\{\frac{(P_2-p'_2)}{(P_2-p'_2)+2C_{\phi_2}}\right\}\right]$	0	1	C_{ϕ_2}
$t'_{RPR4} = k_1\bar{y} - k_2(p_1 - p'_1)\left[exp\left\{\frac{(P_2-p'_2)}{(P_2-p'_2)+2}\right\}\right]$	0	1	1
$t'_{RPR5} = k'_1\bar{y} - k'_2(p_1 - p'_1)\left[2 - exp\left\{\frac{C_{\phi_2}(p'_2-P_2)}{C_{\phi_2}(p_2-P'_2)+2\beta_2\phi_2}\right\}\right]$	1	C_{ϕ_2}	$\beta_{2\phi_2}$
$t'_{RPR6} = k'_1\bar{y} - k'_2(p_1 - p'_1)\left[2 - exp\left\{\frac{P_2(p'_2-P_2)}{P_2(p'_2-P_2)+2}\right\}\right]$	1	P_2	1
$t'_{RPR7} = k'_1\bar{y} - k'_2(p_1 - p'_1)\left[2 - exp\left\{\frac{(p'_2-P_2)}{(p'_2-P_2)+2C_{\phi_2}}\right\}\right]$	1	1	C_{ϕ_2}
$t'_{RPR8} = k'_1\bar{y} - k'_2(p_1 - p'_1)\left[2 - exp\left\{\frac{(p'_2-P_2)}{(p'_2-P_2)+2}\right\}\right]$	1	1	1

Expressing (2.32) in terms of errors we have,

$$t'_{RPR} = k'_1\bar{Y}(1+e_0) - k'_2P_1e_1 + k'_2P_1e'_1\left[\alpha\left\{2 - (1 + \gamma'e'_2 - \frac{1}{2}\gamma^{2\prime}e_2^{2\prime})\right\} + (1-\alpha')(1-\gamma'e'_2 + \frac{3}{2}\gamma^{2\prime}e_2^{2\prime})\right], \quad (2.33)$$

where $\gamma' = \frac{\eta' P_2}{2(\eta' P_2 + \lambda')}$.

To the first degree of approximation,

$$\begin{aligned} t'_{RPR} - \bar{Y} \cong\ & k'_1\bar{Y} + k'_1\bar{Y}e_0 - k'_2P_1e_1 + k'_2P_1e'_1 - k'_1\gamma'\bar{Y}e'_2 - \gamma'\bar{Y}k'_1e'_2e_0 \\ & + \frac{\bar{Y}k'_1e_2^{2\prime}\gamma^{2\prime}3}{2} - \alpha'\bar{Y}k'_1e_2^{2\prime}\gamma^{2\prime} + \gamma k'_2P_1e_1e'_2 - \gamma k'_2P_1e'_1e'_2 - \bar{Y}. \end{aligned} \quad (2.34)$$

Taking expectation both sides of Eq (2.34) we have:

$$Bias(t'_{RPR}) \cong \bar{Y}(k'_1 - 1) - \gamma\prime\bar{Y}k'_1V'_{101} + \bar{Y}k'_1V'_{002}\gamma^{2\prime}(\frac{3}{2} - \alpha'). \quad (2.35)$$

Squaring Eq (2.34) and neglecting higher powers, we get

$$\begin{aligned} E(t'_{RPR} - \bar{Y}^2) \cong\ & \bar{Y}^2 + k_1'^2\bar{Y}^2\left(1 - 4\gamma'V'_{101} - 2\alpha'V'_{002}\gamma^{2\prime} + 4\gamma^2V'_{002} + V_{200}\right) \\ & + k'_1\bar{Y}^2\left(-2 + 2\alpha'V'_{002}\gamma^{2\prime} + 2\gamma'V'_{101} + 3V'_{002}\gamma^{2\prime}\right) \\ & + 2k'_1k'_2\bar{Y}\left(P_1V_{110} - P_1V'_{110}\right) + 2k_2^2\prime\left(P_1^2V_{020} - P_1^2V'_{020}\right). \end{aligned}$$

$$\text{MSE}(t'_{RPR}) \cong \bar{Y}^2 + k_1^2\bar{Y}^2A' + k_1\bar{Y}^2B' + 2k_1k_2\bar{Y}C' + k_2^2D', \quad (2.36)$$

$$A' = 1 - 4\gamma V'_{101} - 2\alpha'V'_{002}\gamma^{2\prime} + 4\gamma^2V'_{002} + V_{200},$$

$$B' = k_1 - 2 + 2\alpha'V'_{002}\gamma^{2\prime} + 2\gamma'V'_{101} + 3V'_{002}\gamma^{2\prime},$$

$$C' = P_1V_{110} - P_1V_{110'}, D' = P_1^2V_{020} - P_1^2V'_{020}.$$

The optimum values of k'_1, k'_2 are obtained by minimizing Eq (2.36):

$$k'_1 = \frac{D'B'}{(A'D' - C^{2\prime})},$$

$$k'_2 = \frac{\bar{Y}(B'C')}{2(A'D' - C^{2\prime})}.$$

Substituting the optimum values of k'_1 and k'_2 in Eq (2.36) we get the minimum MSE of t'_{RPR} as:

$$\text{MSE}(t'_{(RPR)min}) \cong \frac{\bar{Y}^2\left(1 - B^{2\prime}D'\right)}{4(A'D' - C^{2\prime})}. \tag{2.37}$$

3. Efficiency comparisons

3.1. Efficiency comparisons in simple random sampling

In this section we compare theoretically the minimum MSE of the proposed parent family of estimators t_{RPR} with the MSE of existing estimators.

Comparison with usual mean per unit estimator:

(i) $\text{MSE}(t_U) - \text{MSE}(t_{(RPRi)min}) \geq 0$ $(i = 1, 2, ..., 8)$, if $\bar{Y}^2V_{200} - [\frac{\bar{Y}^2\left(1-4AD^2+B^2E-4BCD\right)}{4\left(AE-C^2\right)}] \geq 0$,

Comparison with Naik [1] estimators:

(ii) $\text{MSE}(t_A) - \text{MSE}(t_{(RPRi)min}) \geq 0$ $(i = 1, 2, ..., 8)$, if $\bar{Y}^2(V_{200} - 2V_{110} + V_{020}) - [\frac{\bar{Y}^2\left(1-4AD^2+B^2E-4BCD\right)}{4\left(AE-C^2\right)}] \geq 0$.

(iii) $\text{MSE}(t_B) - \text{MSE}(t_{(RPRi)min}) \geq 0$ $(i = 1, 2, ..., 8)$, if $\bar{Y}^2(V_{200} + 2V_{101} + V_{002}) - [\frac{\bar{Y}^2\left(1-4AD^2+B^2E-4BCD\right)}{4\left(AE-C^2\right)}] \geq 0$.

(iv) $\text{MSE}(t_C) - \text{MSE}(t_{(RPRi)min}) \geq 0$ $(i = 1, 2, ..., 8)$, if $\bar{Y}^2(V_{200} - V_{110} + \frac{1}{4}V_{020}) - [\frac{\bar{Y}^2\left(1-4AD^2+B^2E-4BCD\right)}{4\left(AE-C^2\right)}] \geq 0$.

(v) $\text{MSE}(t_D) - \text{MSE}(t_{(RPRi)min}) \geq 0$ $(i = 1, 2, ..., 8)$, if $\bar{Y}^2(V_{200} + V_{101} + \frac{1}{4}V_{002}) - \frac{\bar{Y}^2\left(1-4AD^2+B^2E-4BCD\right)}{4\left(AE-C^2\right)} \geq 0$.

(vi) $\text{MSE}(t_{MS}) - \text{MSE}(t_{(RPRi)min}) \geq 0$ $(i = 1, 2, ..., 8)$, if $f\bar{Y}^2C_y^2(1 - R^2_{y\phi_1\phi_2}) - \frac{\bar{Y}^2\left(1-4AD^2+B^2E-4BCD\right)}{4\left(AE-C^2\right)} \geq 0$.

We observed that the proposed estimators perform better than the existing estimators if above condition (i) – (vi) are satisfied.

3.2. Efficiency comparisons in two phase sampling

In this section we compare theoretically the minimum MSE of the proposed parent family of estimators t_{RPR} with the MSE of existing estimators.

Comparison with usual mean per unit estimator:

(i) $\text{MSE}(t'_U) - \text{MSE}(t'_{(RPRi)min}) \geq 0$ $(i = 1, 2, .., 8)$, if $\bar{Y}^2(V_{200}) - \frac{\bar{Y}^2\left(1-B^{2\prime}D'\right)}{4\left(A'D'-C^{2\prime}\right)} \geq 0$.

Comparison with Naik [1] estimator:

(ii) $\text{MSE}(t'_A) - \text{MSE}(t'_{(RPRi)min}) \geq 0$ $(i = 1, 2, ..., 8)$, if $\bar{Y}^2\left(V_{200} + V_{020} - V'_{020} + 2V'_{110} - 2V_{110'}\right) - \frac{\bar{Y}^2\left(1-B^{2\prime}D'\right)}{4\left(A'D'-C^{2\prime}\right)} \geq 0$.

(iii) $\text{MSE}(t'_B) - \text{MSE}(t'_{(RPRi)min}) \geq 0$ $(i = 1, 2, ..., 8)$, if $\bar{Y}^2\left(V_{200} + V'_{002} + 2V'_{101}\right) - \frac{\bar{Y}^2\left(1-B^{2\prime}D'\right)}{4\left(A'D'-C^{2\prime}\right)} \geq 0$.

(iv) $\mathrm{MSE}(t'_C) - \mathrm{MSE}(t'_{(RPRi)min}) \geq 0$ $(i = 1,2,...,8)$, if $\bar{Y}^2\left(V_{200} + V_{110}\prime - V_{110} - \frac{1}{4}V_{020}\prime + \frac{1}{4}V_{020}\right) - \frac{\bar{Y}^2\left(1-B^{2\prime}D'\right)}{4\left(A'D'-C^{2\prime}\right)} \geq 0$.

(v) $\mathrm{MSE}(t'_D) - \mathrm{MSE}(t'_{(RPRi)min}) \geq 0$ $(i = 1,2,...,8)$, if $\bar{Y}^2\left(V_{200} + \frac{1}{4}V'_{002} + V'_{101}\right) - \frac{\bar{Y}^2\left(1-B^{2\prime}D'\right)}{4\left(A'D'-C^{2\prime}\right)} \geq 0$.

(vi) $\mathrm{MSE}(t'_{MS}) - \mathrm{MSE}(t'_{(RPRi)min}) \geq 0$ $(i = 1,2,...,8)$, if $f\bar{Y}^2C_y^2\left\{-f(-1+\rho^2_{y\phi_1}) + \lambda(\rho^2_{y\phi_1} - \rho^2_{y\phi_2})\right\} - \frac{\bar{Y}^2\left(1-B^{2\prime}D'\right)}{4\left(A'D'-C^{2\prime}\right)} \geq 0$.

We observed that the proposed estimators perform better than the existing estimators if above condition (i) – (vi) are satisfied.

4. Numerical study

4.1. Numerical study in simple random sampling

Population 1. [19]

Let Y be the study variable of the cultivated area of wheat in 1964.

P_1 be the proportion of cultivated area of wheat greater than 100 acre in 1963.

P_2 be the proportion of cultivated area of wheat greater than 500 in 1961.

$N = 34$, $n = 15$, $\bar{Y} = 199.4412$, $P_1 = 0.73529$, $P_2 = 0.647059$, $S_y = 150.215$, $S_{\phi_1} = 0.4478111$, $S_{\phi_2} = 0.4850713$, $\beta_{2\phi_2} = -1.688$, $C_{\phi_1} = 0.6090231$, $\rho_{\phi_1\phi_2} = 0.6729$, $C_{\phi_2} = 0.7496556$, $C_y = 0.7531$, $\rho_y\phi_2 = 0.6281$, $\rho_y\phi_1 = 0.559$.

Population 2. [20]

Let Y be the study variable of the number of fishes caught in 1995.

P_1 be the proportion of fishes caught which is greater than 1000 in 1993.

P_2 be the proportion of fishes caught which is greater than 2000 in 1994.

$N = 69$, $n = 14$, $\bar{Y} = 4514.89$, $P_1 = 0.7391304$, $P_2 = 0.5507246$, $S_y = 6099.14$, $S_{\phi_1} = 0.4423259$, $S_{\phi_2} = 0.5010645$, $\beta_{2\phi_2} = -2.015$, $C_{\phi_1} = 0.5984409$, $\rho_{\phi_1\phi_2} = 0.6577519$, $C_{\phi_2} = 0.9098277$, $C_y = 1.350$, $\rho_{y\phi_2} = 0.538047$, $\rho_{y\phi_1} = 0.3966081$.

Population 3. [21]

Let study variable Y be the tobacco area production in hectares during the year 2009.

P_1 be the proportion of farms with tobacco cultivation area greater than 500 hectares during the year 2007.

P_2 be proportion of farms with tobacco cultivation area greater than 800 hectares during the year 2008 for 47 districts of Pakistan.

$N = 47$, $n = 10$, $\bar{Y} = 1004.447$, $P_1 = 0.4255319$, $P_2 = 0.3829787$, $s_y = 2351.656$, $s_{\phi_1} = 0.499$, $s_{\phi_2} = 0.4850713$, $\beta_{2\phi_2} = -1.8324$, $C_{\phi_1} = 1.174456$, $\rho_{\phi_1\phi_2} = 0.9153857$, $C_{\phi_2} = 1.283018$, $C_y = 2.341245$, $\rho_{y\phi_2} = 0.4661508$, $\rho_{y\phi_1} = 0.4395989$.

Population 4. [21]

Let study variable Y be the cotton production in hectares during the year 2009.

P_1 be the proportion of farms with cotton cultivation area greater than 37 hectares during the year 2007.

P_2 be proportion of farms with cotton cultivation area greater than 35 hectares during the year 200 for 52 districts of Pakistan.

$N = 52$, $n = 11$, $\bar{Y} = 50.03846$, $P_1 = 0.3846154$, $P_2 = 0.4423077$, $S_y = 71.13086$, $S_{\phi_1} =$

0.4912508, S_{ϕ_2} = 0.501506, $\beta_{2\phi_2}$ = −1.62014, C_{ϕ_1} = 1.277252, $\rho_{\phi_1\phi_2}$ = 0.8877181, C_{ϕ_2} = 1.13384, C_y = 1.421524, $\rho_{y\phi_2}$ = 0.6935718 , $\rho_{y\phi_1}$ = 0.7369579.

We use the following expression to obtain the Percentage Relative Efficiency PRE:

$$PRE = \frac{\text{MSE}(t_0)}{\text{MSE}(t_i min)} * 100, \tag{4.1}$$

where i=U, A, B, C, D, MS, RPR1, RPR2, RPR3, RPR4, RPR5, RPR6, RPR7 and RPR8.

Table 3. Percentage relative efficiency (PRE) with respect to usual mean estimator t_U.

Estimator	Data set1	Data set 2	Data set 3	Data set 4
t_U	100	100	100	100
t_A	133.37	118.36	123.36	207.04
t_B	30.84	45.95	55.04	36.46
t_C	140.5	114.40	118.75	185.29
t_D	55.39	67.77	75.01	58.40
t_{MS}	139.06	110.94	105.64	146.93
t_{RPR1}	125.98	134.42	165.4	225.72
t_{RPR2}	106.66	134.57	167.49	235.43
t_{RPR3}	111.08	137.09	167.89	235.65
t_{RPR4}	109.10	137.39	167.82	233.82
t_{RPR5}	125.75	120.83	166.47	223.72
t_{RPR6}	161.80	137.02	167.63	235.16
t_{RPR7}	168.29	137.47	168.16	235.96
t_{RPR8}	165.09	134.42	168.01	235.93

In Table 2, it is clearly shown that our suggested class of estimator t_{RPR_i} perform better than all the existing estimators t_A, t_B, t_C, t_D and t_{MS}. A significant increase is observed in the percentage relative efficiency of estimators of t_{RPR6}, t_{RPR7} and t_{RPR8}.

4.2. Empirical study in two phase sampling

Population 1. [19]

Let Y be the study variable cultivated area of wheat in 1964.

P_1 be the proportion of cultivated area of wheat greater than 100 acres in 1963.

P_2 be the proportion of cultivated area of wheat greater than 500 in 1961.

N = 34, n' = 15, n = 3, $\bar{Y}$ = 199.4412, P_1 = 0.73529, P_2 = 0.647059, S_y = 150.215, S_{ϕ_1} = 0.4478111, S_{ϕ_2} = 0.4850713, $\beta_{2\phi_2}$ = −1.688, C_{ϕ_1} = 0.6090231, $\rho_{\phi_1\phi_2}$ = 0.6729, C_{ϕ_2} = 0.7496556, C_y = 0.7531, $\rho_{y\phi_2}$ = 0.6281, $\rho_{y\phi_1}$ = 0.559.

Population 2. [20]

Let Y be the study variable, number of fishes caught in 1995.

P_1 be the proportion of fishes caught greater than 1000 in 1993.

P_2 be the proportion of fishes caught greater than 2000 in 1994.

N = 69, n' = 20, n = 7, $\bar{Y}$ = 4514.89, P_1 = 0.7391304, P_2 = 0.5507246, s_y = 6099.14, s_{ϕ_1} = 0.4423259, s_{ϕ_2} = 0.5010645, $\beta_{2\phi_2}$ = −2.015, C_{ϕ_1} = 0.5984409, $\rho_{\phi_1\phi_2}$ = 0.6577519, C_{ϕ_2} = 0.9098277,

$C_y = 1.350$, $\rho_{y\phi_2} = 0.538047$, $\rho_{y\phi_1} = 0.3966081$.

Population 3. [21]

Let Y be the study variable, tobacco area production in hectares during the year 2009.

P_1 be the proportion of farms with tobacco cultivation area greater than 500 hectares during the year 2007.

P_2 be proportion of farms with tobacco cultivation area greater than 800 hectares during the year 2008 for 47 districts of Pakistan.

$N = 47$, $n' = 15$, $n = 7$, $\bar{Y} = 1004.447$, $P_1 = 0.4255319$, $P_2 = 0.3829787$, $S_y = 2351.656$, $S_{\phi_1} = 0.49$, $S_{\phi_2} = 0.4850713$, $\beta_{2\phi_2} = -1.8324$, $C_{\phi_1} = 1.174456$, $\rho_{\phi_1\phi_2} = 0.9153857$, $C_{\phi_2} = 1.283018$, $C_y = 2.341245$, $\rho_{y\phi_2} = 0.4661508$, $\rho_{y\phi_1} = 0.4395989$.

Population 4. [21]

Let Y be the study variable, cotton production in hectares during the year 2009.

P_1 be the proportion of farms with cotton cultivation area greater than 37 hectares during the year 2007.

P_2 be proportion of farms with cotton cultivation area greater than 35 hectares during the year 2008 for 52 districts of Pakistan.

$N = 52$, $n' = 11$, $n = 3$, $\bar{Y} = 50.03846$, $P_1 = 0.3846154$, $P_2 = 0.4423077$, $S_y = 71.13086$, $S_{\phi_1} = 0.4912508$, $S_{\phi_2} = 0.501506$, $\beta_{2\phi_2} = -1.62014$, $C_{\phi_1} = 1.277252$, $\rho_{\phi_1\phi_2} = 0.8877181$, $C_{\phi_2} = 1.13384$, $C_y = 1.421524$, $\rho_{y\phi_2} = 0.6935718$, $\rho_{y\phi_1} = 0.7369579$.

We use the following expression to obtain the Percentage Relative Efficiency(PRE):

$$PRE = \frac{\text{MSE}(t_0)}{\text{MSE}(t'_{imin})} * 100, \tag{4.2}$$

where i=$U', A', B', C', D', MS', RPR1', RPR2', RPR3', RPR4', RPR5', RPR6', RPR7' and RPR8'$.

The results for data set 1–4 are given in Table 4.

Table 4. Percentage relative efficiency (PRE) with respect to usual mean estimator t'_U.

Estimator	Data set1	Data set 2	Data set 3	Data set 4
t'_U	100	100	100	100
t'_A	128.13	112.64	113.46	166.39
t'_B	78.44	75.41	76.75	71.54
t'_C	133.90	110.08	110.95	155.10
t'_D	90.33	88.37	89.01	86.01
t'_{MS}	134.80	111.05	133.6	209.34
t'_{RPR1}	149.36	131.69	178.24	225.68
t'_{RPR2}	158.43	138.50	181.06	239.47
t'_{RPR3}	160.04	139.77	181.55	242.17
t'_{RPR4}	159.63	139.58	181.78	242.08
t'_{RPR5}	149.59	131.98	178.95	227.06
t'_{RPR6}	158.56	138.59	181.13	239.69
t'_{RPR7}	160.36	139.77	181.13	242.80
t'_{RPR8}	159.40	139.99	182.10	242.70

In Table 4, it is clearly shown that our suggested class of estimator t'_{RPR} perform better than all the existing estimators of t'_A, t'_B, t'_C and t'_D and t'_{MS}. A significant increase is observed in the percentage relative efficiency of estimators of t'_{RPR6}, t'_{RPR7} and t'_{RPR8}.

There are many situations where we only interest in knowing everything about the study variable, which is too difficult. For this we can use two auxiliary variables in the form of proportion to find out the study variable. This manuscript provides us the basic tools to the problems related to proportion estimation and two-phase sampling. Here we can see that in abstract of the manuscript we just talk about the minimum MSE of proposed and existing estimators, reason behind is that we can easily compare the minimum MSE with other properties of good estimators like MLE ect., we can also see that the comparison is made in the form of percentage relative efficiency.

5. Application

Statisticians are constantly trying to develop efficient estimators and estimation methodologies to increase the efficiency of estimates. The progress is going on for estimators of population mean. In the present paper our task is to develop a new estimator for estimating the finite population mean under two different sampling schemes, which are simple random sampling and two-phase sampling. The new estimators will be proposed under the following situations:

1). The initial sample is collected through simple random sampling.

2). And then by two-phase sampling using simple random sampling.

In this article, we consider the problem of estimating the finite population mean using the auxiliary proportion under simple random sampling and two-phase sampling scheme. In general, during surveys, it is observed that information in most cases is not obtained on the first attempt even after some call-backs, in such types of issue we use simple random sampling. And when the required results are not obtained, we use two-phase sampling. These approaches are used to obtain the information as much as possible. In sample surveys, it is well known that while estimating the population parameters, i.e., Finite population (mean, median, quartiles, coefficient of variation and distribution function) the information of the auxiliary variable (Proportion) is usually used to improve the efficiency of the estimators. The main aim of studies is to find out more efficient estimators than classical and recent proposed estimators using the auxiliary information (in the form of proportion) for estimating finite population mean under simple random sampling and two-phase sampling scheme.

There are situations where our work is deemed necessary and can be used in daily life.

1). For a nutritionist, it is interesting to know the proportion of population that consumes 25% or more of the calorie intake from saturated fat.

2). Similarly, a soil scientist may be interested in estimating the distribution of clay percent in the soil.

3). In addition, policy-makers may be interested in knowing the proportion of people living in a developing country below the poverty line.

6. Conclusions

In this paper, we have proposed a generalized class of exponential ratio type estimators for estimating population mean using the auxiliary information in the form of proportions under simple and two phase sampling. We used SRS to estimate the population mean using the proportions of available auxiliary information, and when the auxiliary information is unknown, we used two phase sampling for estimation resolution. From the numerical results available in Tables 3 and 4 we can see that two phase sampling gave more efficient results than simple random sampling. Thus the use of auxiliary information in estimation processes increases the efficiency of the estimator, that's we have used two auxiliary variables as attributes. In the numerical study we showed that the proposed estimator is more efficient that t_U, t_A, t_B, t_C, t_D, t_{MS} and any other suggested family of estimators both in simple and two phase sampling schemes.

Some possible extensions of the current work are as follows:

Develop improved finite population mean estimators,

1). using supplementary information more than one auxiliary variable.

2). under stratified two-phase sampling.

3). in the presence of measurement errors.

4). under non-response with two-phase sampling.

Acknowledgments

The authors are thankful to the learned referee for his useful comments and suggestions.

Conflict of interest

The authors declare no conflict of interest.

References

1. V. D. Naik, P. C. Gupta, A note on estimation of mean with known population proportion of an auxiliary character, *J. Indain. Soc. Agric. Stat.*, **48** (1996), 151–158.

2. H. S. Jhajj, M. K. Sharma, L. K. Grover, A family of estimators of population mean using information on auxiliary attribute, *Pak. J. Stat.*, **48** (2006), 43.

3. A. M. Abd-Elfattah, E. A El-Sherpieny, S. M Mohamed, O. F Abdou, Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute, *Appl. Math. Comput.*, **215** (2010), 4198–4202.

4. N. Koyuncu, Efficient estimators of population mean using auxiliary attributes, *Appl. Math. Comput.*, **218** (2012), 10900–10905.

5. R. S. Solanki, H. P. Singh, Improved estimation of population mean using population proportion of an auxiliary character. *Chil. J. Stat.*, **4** (2013), 3–17.

6. P. Sharma, H. K. Verma, A. Sanaullah, R. Singh, Some exponential ratio-product type estimators using information on auxiliary attributes under second order approximation, *Int. J. Stat. Econ.*, **12** (2013), 58–66.

7. S. Malik, R. Singh, An improved estimator using two auxiliary attributes, *Appl. Math. Comput.*, **219** (2013), 10983–10986.

8. H. Verma, R. Singh, F. Smarandache, Some improved estimators of population mean using information on two auxiliary attributes, In: *On improvement in estimating population parameter(s) using auxiliary information*, Columbus: Educational Publishing, Beijing: Journal of Matter Regularity, 2013, 17–24.
9. R. S. Solanki, H. P. Singh, S. K. Pal, Improved estimation of finite population mean in sample surveys, *Columbia Int. Publ. J. Adv. Comput.*, **1** (2013), 70–78.
10. P. Sharma, R. Singh, Improved ratio type estimator using two auxiliary variables under second order approximation, *Math. J. Interdiscip. Sci.*, **2** (2014), 179–190.
11. M. Mahdizadeh, E. Zamanzade, Kernel-based estimation of p (x > y) in ranked set sampling, *SORT-Stat. Oper. Res. T.*, **40** (2016), 243–266.
12. H. P Singh, S. K. Pal, R. S. Solanki, A new class of estimators of finite population mean in sample surveys, *Commun. Stat. Theor. Methods*, **46** (2017), 2630–2637.
13. M. Mahdizadeh, E. Zamanzade, Smooth estimation of a reliability function in ranked set sampling, *Statistics*, **52** (2018), 750–768.
14. S. Hussain, S. Ahmad, S. Akhtar, A. Javed, U. Yasmeen, Estimation of finite population distribution function with dual use of auxiliary information under non-response, *PloS One*, **15** (2020), e0243584.
15. S. Al-Marzouki, C. Chesneau, S. Akhtar, J. A. Nasir, S. Ahmad, S. Hussain, et al., Estimation of finite population mean under pps in presence of maximum and minimum values, *AIMS Mathematics*, **6** (2021), 5397–5409.
16. B. Kiregyera, A chain ratio-type estimator in finite population double sampling using two auxiliary variables, *Metrika*, **27** (1980), 217–223.
17. S. Mohanty, J. Sahoo, A note on improving the ratio method of estimation through linear transformation using certain known population parameters, *Sankhyā: Indian J. Stat. Ser. B*, 1995, 93–102.
18. A. Haq, J. Shabbir. An improved estimator of finite population mean when using two auxiliary attributes, *Appl. Math. Comput.*, **241** (2014), 14–24.
19. M. N. Murthy, *Sampling theory and methods*, Florida: CRC Press LLC, 1967.
20. S. Singh, *Advanced sampling theory with applications*, Springer Science and Business Media, 2003.
21. A. Sharmin, J. R. Sarker, K. R. Das, Growth and trend in area, production and yield of major crops of Bangladesh. *Int. J. Econ. Financ. Manage. Sci.*, **4** (2016), 20–25.