ISSN 2229-5518

Double Sampling Product-Cum-Dual to Ratio Estimators for Finite Population Mean in Sample Surveys

B. K. Singh, Sanjib Choudhury

Abstract— This paper considers a class of product-cum- dual to ratio estimators for estimating finite population mean of the study variate in double sampling. The bias and mean square errors (MSEs) of the proposed estimator have been obtained in two different cases. The asymptotically optimum estimators (AOEs) of the class are identified along with their bias and MSEs. Theoretical and empirical studies have been done to demonstrate the efficiency of the proposed estimators over other estimators. An attempt has been made to find optimum sample sizes under a known fixed cost function.

Index Terms— Auxiliary variate; Finite population mean; Product estimator; Dual to ratio estimator; Double sampling; MSE; Optimum sample sizes, Efficiency.

1 INTRODUCTION

THE literature on survey sampling describes a great variety of techniques for using auxiliary information in order to obtain improved estimators for estimating some most common population parameters such as, population total, population mean, population proportion, population ratio, etc. More often, we are interested in the estimation of the mean of a certain characteristic of a finite population on the basis of a sample taken from the population following a specified sampling procedure.

The use of auxiliary information at the estimation stage appears to have started with the work of [22]. Cochran [22] used auxiliary information at estimation stage and proposed ratio estimator. Robson [5] and Murthy [13] envisaged product estimator and [6] used coefficient of variation of study variate. Motivated by [6], [3] utilized coefficient of variation of auxiliary variate. Srivenkataramana [20] was the first to propose dual to ratio estimator, [17] introduced dual variables for estimation of population parameters.

When the population mean $\bar{X}$ of the auxiliary variable x is unknown before start of the survey, it is estimated from a preliminary large sample on which only the auxiliary characteristic x is observed. The value of $\bar{X}$ in the estimator is then replaced by its estimate. After then a smaller second-phase sample of the variate of interest (study variate) y is then taken. This technique is known as double sampling or two-phase sampling. Neyman [9] was the first to give the concept of double sampling in connection with collecting information on the strata sizes in a stratified sampling, [7] considered ratio-product estimator in double sampling, [18] proposed a generalized class of double sampling estimator besed on ratio type estimator, [8] considered efficient estimator in double sample with subsample the non-respondents .

Consider a finite population $U=(u_1,u_2,...,u_N)$ of size N units, y and x are the study and auxiliary variates respectively. When the population mean $\bar{X}$ of x is not known, a first-phase sample of size n_1 is drawn from the population on which only the x characteristic is measured in order to furnish a good estimate of $\bar{X}$. After then a second-phase sample of size $n\ (n<n_1)$ is drawn on which both the variates y and x are measured.

The usual product estimator in double sampling is given as

$$\bar{y}_P^{(d)}=\bar{y}\frac{\bar{x}}{\bar{x}_1}$$

where $\bar{x}$ and $\bar{y}$ are the sample mean of x and y respectively based on the sample size n out of the population N units and $\bar{x}_1=\frac{1}{n_1}\sum_{i=1}^{n_1}x_i$ denote the sample mean of x based on the first-phase sample of the size n_1.

Consider a transformation $x_i^\sigma=\left(N\bar{X}-nx_i\right)/(N-n)$, $(i=1,2,3,...,N)$. Here $\bar{x}^\sigma=\left(N\bar{X}-n\bar{x}\right)/(N-n)$ is unbiased estimator for $\bar{X}$ and $Corr\left(\bar{y},\bar{x}^\sigma\right)=-\rho$, where ρ is the correlation coefficient between y and x.

Using the transformation of x_i^σ, [20] obtained dual to ratio estimator as

- B. K. Singh, Department of Mathematics, North Eastern Regional Institute of Science and Technology, Nirjuli-791109, Arunachal Pradesh, India. E-mail: bksinghnerist@gmail.com
- Sanjib Choudhury (Corresponding Author), Department of Mathematics, North Eastern Regional Institute of Science and Technology, Nirjuli-791109, Arunachal Pradesh, India, E-mail: sanjibchy07@gmail.com

$$\bar{y}_R^* = \bar{y}\,\frac{\bar{x}^{\sigma}}{\bar{X}}.$$

Utilizing the transformation $x_i^* = \left(n_1\bar{x}_1 - nx_i\right)/(n_1 - n)$, $\left(i = 1, 2, ..., N\right)$, [12] has obtained dual to ratio estimator in double sampling as

$$\bar{y}_k^{(d)} = \bar{y}\,\frac{\bar{x}^*}{\bar{x}_1}$$

where $\bar{x}^* = \left(n_1\bar{x}_1 - n\bar{x}\right)/\left(n_1 - n\right)$ be an unbiased estimator of $\bar{X}$ and $Corr\left(\bar{y}, \bar{x}^*\right) = -\rho$.

In this paper, we have proposed an estimator of a linear combination of usual product estimator and dual to ratio estimator in double sampling. Numerical illustrations are given in the support of the present study.

2 The Proposed Estimator

The proposed product-cum-dual to ratio estimators of population mean $\bar{Y}$ in double sampling is

$$\bar{y}_{PdR}^{(d)} = \bar{y}\left[\alpha\,\frac{\bar{x}}{\bar{x}_1} + \left(1 - \alpha\right)\frac{\bar{x}^*}{\bar{x}_1}\right] \qquad (1)$$

where α is a constant.

To obtain the bias (B) and mean square error (MSE) of $\bar{y}_{PdR}^{(d)}$, we write

$e_0 = \left(\bar{y} - \bar{Y}\right)/\bar{Y}$, $e_1 = \left(\bar{x} - \bar{X}\right)/\bar{X}$ and $e_1' = \left(\bar{x}_1 - \bar{X}\right)/\bar{X}$.

Expressing $\bar{y}_{PdR}^{(d)}$ in terms of e's, we obtain

$$\bar{y}_{PdR}^{(d)} = \bar{Y}\left(1 + e_0\right)\Big[\alpha\left(1 + e_1\right)\left(1 + e_1'\right)^{-1} + \left(1 - \alpha\right)\left\{\left(1 + g\right) - g\left(1 + e_1\right)\left(1 + e_1'\right)^{-1}\right\}\Big] \qquad (2)$$

where $g = n/\left(n_1 - n\right)$.

Assuming that the sample size is large enough so that $\left|e_1'\right| < 1$, therefore $\left(1 + e_1'\right)^{-1}$ is expandable.

Expanding the right hand side of (2), multiplying out and retaining terms of e's up to the second degree, we obtain

$$\bar{y}_{PdR}^{(d)} - \bar{Y} \cong \bar{Y}\left\{e_0 + \omega\left(e_1 - e_1' + e_1'^2 + e_0e_1 - e_0e_1' - e_1e_1'\right)\right\}, \qquad (3)$$

where $\omega = \left(1 + g\right)\alpha - g$.

To obtain the bias and MSE of $\bar{y}_{PdR}^{(d)}$, the following notations are used hereafter:

$C_y^2 = S_y^2/\bar{Y}^2$, $C_x^2 = S_x^2/\bar{X}^2$ and $\rho = S_{xy}/\left(S_xS_y\right)$,

where C_y and C_x are the coefficient of variation of study variate y and auxiliary variate x respectively.

$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^2$ and $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2$ are the population variances of study variate y, auxiliary variate x respectively and

$S_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)\left(x_i - \bar{X}\right)$ is the covariance between y and x.

The following two cases will be discussed separately.

Case I: When the second phase sample of size n is a subsample of the first phase sample of size n_1 and

Case II: When the second phase sample of size n is drawn independently of the first phase sample of size n_1, see [4].

3 Bias, MSE and Optimum of $\bar{y}_{PdR}^{(d)}$ in Case I

In Case I, we have

$$\left.\begin{aligned} &E(e_0) = E(e_1) = E(e_1') = 0, \\ &E(e_0^2) = \frac{1-f}{n}C_y^2, \quad E(e_1^2) = \frac{1-f}{n}C_x^2, \\ &E(e_1'^2) = \frac{1-f_1}{n_1}C_x^2, \quad E(e_0e_1) = \frac{1-f}{n}CC_x^2, \\ &E(e_0e_1') = \frac{1-f_1}{n_1}CC_x^2, \quad E(e_1e_1') = \frac{1-f_1}{n_1}C_x^2, \end{aligned}\right\} \qquad (4)$$

where $f = n/N$, $f_1 = n_1/N$ and $C = \rho\,C_y/C_x$.

Taking expectation on both the sides and using the results of (4) in (3), we get the bias of the estimator $\bar{y}_{PdR}^{(d)}$ to the first degree of approximation as

$$B\left\{\bar{y}_{PdR}^{(d)}\right\}_I = \frac{1 - f^*}{n}\bar{Y}CC_x^2\omega, \qquad (5)$$

where $f^* = n/n_1$.

The bias, $B\left\{\bar{y}_{PdR}^{(d)}\right\}_I$ in (5) is 'zero' if $\alpha = n/n_1\left(= f^*\right)$.

Thus the estimator $\bar{y}_{PdR}^{(d)}$ with $\alpha = n/n_1$ is almost unbiased.

To obtain MSE of proposed class of estimator, (3) can be written as

$$\bar{y}_{PdR}^{(d)} - \bar{Y} \cong \bar{Y}\left\{e_0 + \omega(e_1 - e_1')\right\} \qquad (6)$$

Squaring and taking expectation on both the sides of (6) and using the results of (4), we obtain the MSE of $\bar{y}_{PdR}^{(d)}$ up to the first degree of approximation as

$$M\left\{\bar{y}_{PdR}^{(d)}\right\}_I = \bar{Y}^2\left\{\frac{1-f}{n}C_y^2 + \frac{1-f^*}{n}C_x^2\omega\left(\omega + 2C\right)\right\} \qquad (7)$$

Minimization of (7) with respect to α yields its optimum value at

$$\alpha = f^* - \frac{C}{1+g} = \alpha_{Iopt} \text{ (say).} \tag{8}$$

Substituting the value of α from (8) in (1) gives the 'asymptotically optimum estimator' (AOE) as

$$\left\{\overline{y}_{PdR}^{(d_0)}\right\}_I = \overline{y}\left\{\left(f^* - \frac{C}{1+g}\right)\frac{\overline{x}}{\overline{x}_1} + \frac{1+C}{1+g}\frac{\overline{x}^*}{\overline{x}_1}\right\}$$

Therefore, the resulting bias and MSE of $\overline{y}_{PdR}^{(d_0)}$ are

$$B\left\{\overline{y}_{PdR}^{(d_0)}\right\}_I = -\frac{1-f^*}{n}\overline{Y}C^2C_x^2$$

and

$$M\left\{\overline{y}_{PdR}^{(d_0)}\right\}_I = \overline{Y}^2C_y^2\left(\frac{1-f}{n} - \frac{1-f^*}{n}\rho^2\right) \text{ respectively.} \tag{9}$$

Equation (9) shows that MSE of the proposed estimator is the same as the MSE of the linear regression estimator in double sampling $\overline{y}_{dlr} = \overline{y} + b_{yx}\left(\overline{x}_1 - \overline{x}\right)$, where b_{yx} is the sample regression coefficient of y on x.

Remark 1. *For $\alpha = 1$, the estimator $\overline{y}_{PdR}^{(d)}$ in (1) reduces to the usual product estimator $\overline{y}_P^{(d)}$ in double sampling. The bias and MSE of $\overline{y}_P^{(d)}$ can be obtained by putting $\alpha = 1$ in (5) and (7) respectively as*

$$B\left\{\overline{y}_P^{(d)}\right\}_I = \overline{Y}\frac{1-f^*}{n}CC_x^2$$

and

$$M\left\{\overline{y}_P^{(d)}\right\}_I = \overline{Y}^2\left\{\frac{1-f}{n}C_y^2 + \frac{1-f^*}{n}C_x^2\left(1+2C\right)\right\} \tag{10}$$

Remark 2. *For $\alpha = 0$, the estimator $\overline{y}_{PdR}^{(d)}$ in (1) reduces to dual to ratio estimator $\overline{y}_k^{(d)}$ in double sampling. The bias and MSE of $\overline{y}_k^{(d)}$ can be obtained by putting $\alpha = 0$ in (5) and (7) respectively as*

$$B\left\{\overline{y}_k^{(d)}\right\}_I = -\overline{Y}\frac{1-f^*}{n}gCC_x^2$$

and

$$M\left\{\overline{y}_k^{(d)}\right\}_I = \overline{Y}^2\left\{\frac{1-f}{n}C_y^2 + \frac{1-f^*}{n}C_x^2g\left(g-2C\right)\right\} \tag{11}$$

The variance of sample mean $\overline{y}$ under SRSWOR sampling scheme is given by

$$V\left(\overline{y}\right) = \overline{Y}^2\frac{1-f}{n}C_y^2 \tag{12}$$

4 Efficiency Comparisons of $\overline{y}_{PdR}^{(d)}$ and $\overline{y}_{PdR}^{(d_0)}$ in Case I

In the following Section, we have presented the comparisons of the proposed estimator with other estimators to investigate the ranges of the unknown parameter α for which the proposed estimator is better than others.

From (7), (10), (11) and (12) it is found that the proposed class of estimators $\overline{y}_{RdR}^{(d)}$ is better than:

i. the usual product estimator $\overline{y}_P^{(d)}$ in double sampling if

either $1 > \alpha$ and $-\left\{2C/\left(1+g\right) + N_1^*\right\} < \alpha$,

or $1 < \alpha$ and $-\left\{2C/\left(1+g\right) + N_1^*\right\} > \alpha$,

where $N_1^* = \left(n_1 - 2n\right)/n_1$.

ii. the dual to ratio estimator $\overline{y}_k^{(d)}$ in double sampling if

either $0 > \alpha$ and $2\left\{f^* - C/\left(1+g\right)\right\} < \alpha$,

or $0 < \alpha$ and $2\left\{f^* - C/\left(1+g\right)\right\} > \alpha$.

iii. the sample mean per unit estimator $\overline{y}$ if

either $f^* > \alpha$ and $\left\{f^* - 2C/\left(1+g\right)\right\} < \alpha$,

or $f^* < \alpha$ and $\left\{f^* - 2C/\left(1+g\right)\right\} > \alpha$.

Also from (9), (10), (11) and (12), it is observed that the 'AOE' $\overline{y}_{PdR}^{(d_0)}$ is better than:

i. the usual product estimator $\overline{y}_P^{(d)}$ in double sampling, since

$$M\left\{\overline{y}_P^{(d)}\right\}_I - M\left\{\overline{y}_{PdR}^{(d_0)}\right\}_I = \overline{Y}^2\frac{1-f^*}{n}C_x^2\left(1+C\right)^2 > 0.$$

ii. the dual to ratio estimator $\overline{y}_k^{(d)}$ in double sampling, since

$$M\left\{\overline{y}_k^{(d)}\right\}_I - M\left\{\overline{y}_{PdR}^{(d_0)}\right\}_I = \overline{Y}^2\frac{1-f^*}{n}C_x^2\left(g-C\right)^2 > 0.$$

iii. the sample mean per unit estimator $\overline{y}$, since

$$V\left(\overline{y}\right) - M\left\{\overline{y}_{PdR}^{(d_0)}\right\}_I = \overline{Y}^2\frac{1-f^*}{n}C^2C_x^2 > 0.$$

Now we state the following theorem:

Theorem 1. *To the first degree of approximation, the proposed strategy $\overline{y}_{PdR}^{(d)}$ under optimal condition (8) is always more efficient than $M\left\{\overline{y}_P^{(d)}\right\}$, $M\left\{\overline{y}_k^{(d)}\right\}$ and $V\left(\overline{y}\right)$ and equally efficient with the linear regression estimator $M\left\{\overline{y}_{dlr}\right\}$ in double sampling.*

5 Bias, MSE and Optimum of $\overline{y}_{PdR}^{(d)}$ in Case II

In Case II, we have

$$\left.\begin{aligned} &E(e_0)=E(e_1)=E(e_1')=0,\\ &E(e_0^2)=\frac{1-f}{n}C_y^2,\quad E(e_1^2)=\frac{1-f}{n}C_x^2,\\ &E(e_1'^2)=\frac{1-f_1}{n_1}C_x^2,\quad E(e_0e_1)=\frac{1-f}{n}CC_x^2,\\ &E(e_0e_1')=0,\quad E(e_1e_1')=0. \end{aligned}\right\} \tag{13}$$

Taking expectation in (3) and using the results of (13), we get the bias of $\bar{y}_{PdR}^{(d)}$ to the first degree of approximation as

$$B\left\{\bar{y}_{PdR}^{(d)}\right\}_{II}=\bar{Y}\omega C_x^2\left(\frac{1-f_1}{n_1}+\frac{1-f}{n}C\right) \tag{14}$$

which is vanished when $\alpha=n/n_1\left(=f^*\right)$. Thus, the estimators $\bar{y}_{PdR}^{(d)}$ with the value of $\alpha=n/n_1$ is almost unbiased.

Squaring and taking expectation in both the sides of (6) and using the results of (13), we obtain the MSE of $\bar{y}_{PdR}^{(d)}$ to the first degree of approximation as

$$M\left\{\bar{y}_{PdR}^{(d)}\right\}=\bar{Y}^2\left\{\frac{1-f}{n}C_y^2+C_x^2\omega\left(f^{**}\omega+2\frac{1-f}{n}C\right)\right\} \tag{15}$$

where $f^{**}=\left(1-f\right)/n+\left(1-f_1\right)/n_1$.

Minimization of (15) is obtained with optimum value of α as

$$\alpha=f^*-\left\{\frac{1-f}{n}C\div f^{**}\left(1+g\right)\right\}=\alpha_{IIopt}\text{ (say).} \tag{16}$$

Substituting the value of α from (16) in (1) gives the 'AOE' as

$$\left\{\bar{y}_{PdR}^{(d_0^*)}\right\}_{II}=\left\{\alpha_{IIopt}\frac{\bar{x}}{\bar{x}_1}+\left(1-\alpha_{IIopt}\right)\frac{\bar{x}^*}{\bar{x}_1}\right\}.$$

Thus, the resulting bias and MSE of 'AOE' $\bar{y}_{dR}^{(d_0^*)}$ are respectively given as

$$B\left\{\bar{y}_{PdR}^{(d_0^*)}\right\}_{II}=-\bar{Y}CC_x^2\frac{1-f}{n}\left(\frac{1-f_1}{n_1}+\frac{1-f}{n}C\right)\Big/f^{**}$$

and

$$M\left\{\bar{y}_{PdR}^{(d_0^*)}\right\}_{II}=\bar{Y}^2\left\{\frac{1-f}{n}C_y^2-\left(\frac{1-f}{n}\right)^2C^2C_x^2\Big/f^{**}\right\}.$$

For simplicity, we assume that the population size N is large enough as compared to the sample sizes n and n_1 so that the finite population correction (FPC) terms $1/N$ and $2/N$ are ignored.

Ignoring the FPC in (15), the MSE of $\bar{y}_{PdR}^{(d)}$ is reduces to

$$M\left\{\bar{y}_{PdR}^{(d)}\right\}_{II}=\bar{Y}^2\left[\frac{C_y^2}{n}+\omega C_x^2\left\{\left(\frac{1}{n}+\frac{1}{n_1}\right)\omega+2\frac{C}{n}\right\}\right] \tag{17}$$

which is minimized for

$$\alpha=f^*-c\mu=\alpha^*_{IIopt} \tag{18}$$

where $\mu=\left(n_1-n\right)/\left(n_1+n\right)$.

Substituting the value of α from (18) in (1), we obtain the 'AOE' as

$$\bar{y}_{PdR}^{(d_0^{**})}=\bar{y}\left[\left(f^*-C\mu\right)\frac{\bar{x}}{\bar{x}_1}+\left(\frac{1}{1+g}+C\mu\right)\left\{\left(1+g\right)-g\frac{\bar{x}}{\bar{x}_1}\right\}\right]$$

Therefore, the resulting MSE of $\bar{y}_{PdR}^{(d_0^{**})}$ is

$$M\left\{\bar{y}_{PdR}^{(d_0^{**})}\right\}_{II}=\frac{S_y^2}{n}\left(1-\lambda\rho^2\right), \tag{19}$$

where $\lambda=n_1/\left(n_1+n\right)$.

Remark 3. *For $\alpha=1$, the estimator $\bar{y}_{PdR}^{(d)}$ in (1) reduces to the usual product estimator $\bar{y}_P^{(d)}$ in double sampling. Thus putting $\alpha=1$ in (17), we get the MSE of $\bar{y}_P^{(d)}$ to the first degree of approximation as*

$$M\left\{\bar{y}_P^{(d)}\right\}_{II}=\bar{Y}^2\left[\frac{1}{n}\left\{C_y^2+C_x^2\left(1+2C\right)\right\}+\frac{1}{n_1}C_x^2\right]. \tag{20}$$

Remark 4. *For $\alpha=0$, the estimator $\bar{y}_{PdR}^{(d)}$ in (1) reduces to dual to ratio estimator $\bar{y}_k^{(d)}$ in double sampling. The MSE of $\bar{y}_k^{(d)}$ can be obtained by putting $\alpha=0$ in (17) as*

$$M\left\{\bar{y}_k^{(d)}\right\}_{II}=\bar{Y}^2\frac{1}{n}\left\{C_y^2-gC_x^2\left(2C-\frac{g}{\lambda}\right)\right\} \tag{21}$$

Ignoring the FPC, the variance of $\bar{y}$ under SRSWOR is given by

$$V\left(\bar{y}\right)_{II}=\frac{1}{n}\bar{Y}^2C_y^2 \tag{22}$$

and the MSE of the linear regression estimator $\bar{y}_{dlr}=\bar{y}+b_{yx}\left(\bar{x}_1-\bar{x}\right)$ in double sampling is given by

$$M\left(\bar{y}_{dlr}\right)_{II}=\frac{S_y^2}{n}\left(1-\frac{\rho^2}{1+g}\right). \tag{23}$$

6 Efficiency Comparisons of $\bar{y}_{PdR}^{(d)}$ and $\bar{y}_{PdR}^{(d_0^{**})}$ in Case II

From (17), (20), (21) and (22) it is found that the proposed class of estimators $\bar{y}_{RdR}^{(d)}$ has better efficiency than

i. the usual product estimator $\bar{y}_P^{(d)}$ in double sampling if

either $1>\alpha$ and $-\left(2C\mu+N_1^*\right)<\alpha$,

or $1<\alpha$ and $-\left(2C\mu+N_1^*\right)>\alpha$.

ii. the dual to ratio estimator $\bar{y}_k^{(d)}$ in double sampling if *either* $0 > \alpha$ and $2\left(f^* - \mu C\right) < \alpha$,

or $0 < \alpha$ and $2\left(f^* - \mu C\right) > \alpha$.

iii. the sample mean per unit estimator $\bar{y}$ if

either $f^* > \alpha$ and $\left(f^* - 2C\mu\right) < \alpha$,

or $f^* < \alpha$ and $\left(f^* - 2C\mu\right) > \alpha$.

Also from (19), (20), (21), (22) and (23), it is established that the 'AOE' $\bar{y}_{PdR}^{\left(d_0^{**}\right)}$ is better than

i. the usual product estimator $\bar{y}_P^{(d)}$ in double sampling, since

$$M\left\{\bar{y}_P^{(d)}\right\}_{II} - M\left\{\bar{y}_{PdR}^{\left(d_0^{**}\right)}\right\}_{II} = \bar{Y}^2 C_x^2 \lambda \left(C + 1/\lambda\right)^2 > 0.$$

ii. the dual to ratio estimator $\bar{y}_k^{(d)}$ in double sampling, since

$$M\left\{\bar{y}_k^{(d)}\right\}_{II} - M\left\{\bar{y}_{PdR}^{\left(d_0^{**}\right)}\right\}_{II} = \bar{Y}^2 C_x^2 n\lambda \left(\frac{C}{n} - \frac{1}{\mu n_1}\right)^2 > 0\cdot$$

iii. the sample mean per unit estimator $\bar{y}$, since

$$V\left(\bar{y}\right)_{II} - M\left\{\bar{y}_{PdR}^{\left(d_0^{**}\right)}\right\}_{II} = \bar{Y}^2 C^2 C_x^2 \frac{\lambda}{n} > 0.$$

iv. the linear regression estimator $\bar{y}_{dlr}$ in double sampling, since

$$M\left(\bar{y}_{dlr}\right)_{II} - M\left\{\bar{y}_{PdR}^{\left(d_0^{**}\right)}\right\}_{II} = S_y^2 \rho^2 \frac{n\lambda}{n_1^2} > 0.$$

Remark 5. *The MSE of the 'AOE' $\bar{y}_{PdR}^{\left(d_0^{**}\right)}$ is always less than that of $\bar{y}_{dlr}$. In addition to it, the 'AOE' $\bar{y}_{PdR}^{\left(d_0^{**}\right)}$ in Case II is more efficient than the 'AOE' $\bar{y}_{PdR}^{\left(d_0\right)}$ in Case I.*

7 Cost Aspect

The different estimators reported in this paper have so far been compared with respect to their mean square error. In practical applications, the cost aspect should also be taken into account. In literature, therefore, convention is to fix the total cost of the survey and then to find optimum sizes of preliminary and final samples so that the MSE of the estimators are minimized. In most of the practical situations, total cost is a linear function of samples selected at first and second phases.

In this section, we shall consider the cost of the survey and find the optimum sizes of the preliminary and second-phase samples in Case I and Case II separately.

Case I: In this case, Let us consider a cost function is given by

$$c = c_1 n + c_2 n_1 \tag{24}$$

where c = total cost of the survey.

c_1 = cost per unit of collecting information on y and

c_2 = cost per unit of collecting information on x.

Ignoring FPC, the MSE expression of 'AOE' $\bar{y}_{PdR}^{\left(d_0\right)}$ of (9) is given as

$$M\left\{\bar{y}_{PdR}^{\left(d_0\right)}\right\} = \frac{V_1}{n} + \frac{V_2}{n_1} \tag{25}$$

where $V_1 = S_y^2\left(1 - \rho^2\right)$ and $V_2 = S_y^2 \rho^2$.

The optimum values (sizes) of n and n_1 for fixed cost c, which minimizes the MSE of (25) are given by

$$n_{opt} = c\sqrt{V_1/c_1}\Big/\left(\sqrt{V_1 c_1} + \sqrt{V_2 c_2}\right)$$
$$= c\sqrt{\left(1-\rho^2\right)/c_1}\Big/\left(\sqrt{\left(1-\rho^2\right)c_1} + \rho\sqrt{c_2}\right) \text{ and}$$
$$n_{1opt} = c\sqrt{V_2/c_2}\Big/\left(\sqrt{V_1 c_1} + \sqrt{V_2 c_2}\right)$$
$$= c\rho\sqrt{1/c_2}\Big/\left(\sqrt{\left(1-\rho^2\right)c_1} + \rho\sqrt{c_2}\right).$$

Therefore, the MSE of $\bar{y}_{PdR}^{\left(d_0\right)}$ corresponding to optimal double sampling estimator is

$$M_{opt}\left\{\bar{y}_{PdR}^{\left(d_0\right)}\right\} = \left(S_y^2/c\right)\left\{\sqrt{\left(1-\rho^2\right)c_1} + \rho\sqrt{c_2}\right\}^2 \tag{26}$$

In case we do not use any auxiliary variate then the cost function is of the form $c_0 = nc_1$, where $c_0\left(=c\right)$ and c_1 are the total cost and cost per unit of collecting information on the study variate y respectively. The minimum variance of the sample mean $\bar{y}$ for a given fixed cost c_0 in case of large population is given by

$$V_{opt}\left(\bar{y}\right) = \frac{c_1}{c_0} S_y^2 \tag{27}$$

From (26) and (27), the proposed double sampling strategy would be profitable than sample mean $\bar{y}$ if

$$M_{opt}\left\{\bar{y}_{PdR}^{\left(d_0\right)}\right\} < V_{opt}\left(\bar{y}\right)$$

or equivalently,

$$\frac{c_2}{c_1} < \left(1 - \sqrt{1-\rho^2}\right)^2 \Big/ \rho^2. \tag{28}$$

Case II: In this case we assume that x is measured on $n^{**} = n + n_1$ units and y on n units. Following [19], we shall consider a simple cost function as

$$c = c_1 n + c_2^* n^{**}$$

where c_1 and c_2^* denotes the cost per unit observing y and x values respectively.

The MSE of $\bar{y}_{PdR}^{\left(d_0^{**}\right)}$ at (19) can be written as

$$M\left\{\bar{y}_{PdR}^{\left(d_0^{**}\right)}\right\} = \frac{V_1}{n} + \frac{V_2}{n^{**}} \tag{29}$$

The optimum values (sizes) of n and n^{**} for fixed cost c, which minimizes the MSE of (29) are given by

$$n_{opt} = c\sqrt{V_1/c_1}\Big/\left(\sqrt{V_1c_1}+\sqrt{V_2c_2^*}\right)$$
$$= c\sqrt{\left(1-\rho^2\right)/c_1}\Big/\left(\sqrt{\left(1-\rho^2\right)c_1}+\rho\sqrt{c_2^*}\right) \text{ and}$$
$$n_{opt}^{**} = c\sqrt{V_2/c_2^*}\Big/\left(\sqrt{V_1c_1}+\sqrt{V_2c_2^*}\right)$$
$$= c\rho\sqrt{1/c_2^*}\Big/\left(\sqrt{\left(1-\rho^2\right)c_1}+\rho\sqrt{c_2^*}\right)$$

The MSE of $\bar{y}_{PdR}^{\left(d_0^{**}\right)}$ corresponding to optimal double sampling estimator is

$$M_{opt}\left\{\bar{y}_{PdR}^{\left(d_0^{**}\right)}\right\} = \left(S_y^2/c\right)\left\{\sqrt{\left(1-\rho^2\right)c_1}+\rho\sqrt{c_2^*}\right\}^2 \qquad (30)$$

From (27) and (30), it is obtained that the double sampling estimator $\bar{y}_{PdR}^{\left(d_0^{**}\right)}$ yields less MSE than that of sample mean $\bar{y}$ for the same fixed cost if

$$M_{opt}\left\{\bar{y}_{PdR}^{\left(d_0^{**}\right)}\right\} < V_{opt}\left(\bar{y}\right).$$

or equivalently,

$$\frac{c_2^*}{c_1} < \left(1-\sqrt{1-\rho^2}\right)^2\Big/\rho^2 . \qquad (31)$$

8 Empirical Study

To analyze the performance of the proposed estimators in comparison to other estimators, six natural population data sets are being considered. The sources of the population, the nature of the variates y and x and the values of the various parameters are given as follows

Population I- Source: [14], p. 228

x : fixed capital,

y : output,

$N=80,\ n=10,\ n_1=30,\ \bar{Y}=5182.64,\ C_y=0.3542,$
$C_x=7507,\ \rho=0.9413$

Population II- Source: [14], p. 228

x : number of workers,

y : output,

$N=80,\ n=10,\ n_1=30,\ \bar{Y}=5182.64,\ C_y=0.3542,$
$C_x=0.9484,\ \rho=0.9150$

Population III- Source: [2]

x : number of agricultural labourers for 1961,

y : number of agricultural labourers for 1971,

$N=278,\ n=30,\ n_1=70,\ \bar{Y}=39.0680,\ C_y=1.4451,$
$C_x=1.6198,\ \rho=0.7213$

Population IV- Source: [15], p. 282

x : chlorine percentage,

y : log of leaf burn in sacs.

$N=30,\ n=4,\ n_1=12,\ \bar{Y}=0.6860,\ C_y=0.4803,$
$C_x=0.7493,\ \rho=-0.4996$

Population V- Source: [12], P. 324

x : population of village,

y : number of cultivators in the village.

$N=487,\ n=20,\ n_1=95,\ \bar{Y}=449.846,\ C_y=0.8871,$
$C_x=0.7696,\ \rho=0.881815.$

Population VI- Source: [1], P. 47

x : initial white blood cell count,

y : survival time leukemia patient.

$N=20,\ n=4,\ n_1=8,\ C_y=0.2017,\ C_x=0.1502,$
$\rho=-0.4074.$

To reflect the gain in the efficiency of the proposed class of estimators $\bar{y}_{PdR}^{(d)}$ over the conventional estimators $\left(\bar{y}\text{ and }\bar{y}_P^{(d)}\right)$ and dual to ratio estimator $\bar{y}_k^{(d)}$ in double sampling, the effective ranges and optimum values of α are presented in Table 1 with respect to the above population data sets.

TABLE 1 EFFECTIVE RANGES AND OPTIMUM VALUES OF α OF $\bar{y}_{PdR}^{(d)}$

Population data sets	Ranges of α under which the proposed class of estimators is better than			Optimum values of α
	$\bar{y}$	$\bar{y}_P^{(d)}$	$\bar{y}_k^{(d)}$	
For Case I				α_{Iopt}
I	(-0.26, 0.33)	(-0.93, 1.00)	(0.00, 0.07)	0.372
II	(0.12, 0.33)	(-0.79, 1.00)	(0.00, 0.21)	0.1055
III	(-0.31, 0.43)	(-0.88, 1.00)	(0.00, 0.12)	0.0609
IV	(0.33, 0.76)	(0.09, 1.00)	(0.00, 1.09)	0.5468
V	(-1.39, 0.21)	(-2.18, 1.00)	(-1.18, 0.00)	-0.5919
VI	(0.50, 1.05)	(0.55, 1.00)	(0.00, 1.55)	0.7735
For Case II				α^*_{IIopt}
I	(-0.11, 0.33)	(-0.78, 1.00)	(0.00, 0.22)	0.1113
II	(-0.01, 0.33)	(-0.68, 1.00)	(0.00, 0.32)	0.1625
III	(-0.09, 0.43)	(-0.66, 1.00)	(0.00, 0.34)	0.1712
IV	(0.33, 0.65)	(-0.01, 1.00)	(0.00, 0.99)	0.4935
V	(-1.12, 0.21)	(-1.90, 1.00)	(-0.90, 0.00)	-0.4524
VI	(0.50, 0.86)	(0.36, 1.00)	(0.00, 1.36)	0.6823

To observe the relative performance of different estimators of $\bar{Y}$, we have computed the percentage relative efficiencies (PREs) of the proposed estimator, conventional estimators (i.e. $\bar{y}$ and $\bar{y}_P^{(d)}$) and dual to ratio estimator $\bar{y}_k^{(d)}$ in double sampling with respect to usual unbiased estimator $\bar{y}$ by using the formula

$$PRE\left(t,\bar{y}\right) = \frac{M\left(\bar{y}\right)}{M\left(t\right)}\times 100,$$

where $t=\bar{y},\ \bar{y}_P^{(d)},\ \bar{y}_k^{(d)}$ and $\bar{y}_{PdR}^{(d)}$ or $\left(\bar{y}_{PdR}^{(d_0)}\text{ and }\bar{y}_{PdR}^{\left(d_0^{**}\right)}\right)$.

and the findings are presented in Table 2.

TABLE 2 PERCENTAGE RELATIVE EFFICIENCIES WITH RESPECT TO $\bar{y}$

Estimators → Population	$\bar{y}$	$\bar{y}_P^{(d)}$ Case I	$\bar{y}_P^{(d)}$ Case II	$\bar{y}_k^{(d)}$ Case I	$\bar{y}_k^{(d)}$ Case II	$\bar{y}_{PdR}^{(d)}$ or $\bar{y}_{PdR}^{(d_0)}$ Case I	$\bar{y}_{PdR}^{(d)}$ or $\bar{y}_{PdR}^{(d_0^{**})}$ Case II
I	100.00	*	*	297.97	199.08	307.77	298.09
II	100.00	*	*	200.42	106.40	276.16	268.76
III	100.00	*	*	147.96	125.49	149.98	157.28
IV	100.00	*	*	*	*	123.76	123.83
V	100.00	*	*	141.21	152.26	277.92	279.61
VI	100.00	103.38	*	*	*	111.56	112.44

*Data not applicable and percentage relative efficiency less than 100%.

9 CONCLUSIONS

Section 7 provides the optimum sample sizes under a known fixed cost function for Case I and Case II. We have also obtained conditions (referred in (28) and (31)) under which the proposed class of estimators $\bar{y}_{PdR}^{(d)}$ would be profitable in comparison to single-phase sampling for fixed cost of survey.

From Table 1, it is seen that the proposed class of estimators $\bar{y}_{PdR}^{(d)}$ is more efficient over the conventional estimators $\left(\bar{y} \text{ and } \bar{y}_P^{(d)}\right)$ and dual to ratio estimator $\left(\bar{y}_k^{(d)}\right)$ in double sampling for both the cases under the effective ranges of α as far as the MSE criterion is concerned. It is also observed from Table 1 that there is a scope for choosing α to obtain better estimators than $\bar{y}$, $\bar{y}_P^{(d)}$ and $\bar{y}_k^{(d)}$.

Table 2 shows that there is a considerable gain in efficiency by using proposed class of estimators $\bar{y}_{PdR}^{(d)}$ or $\left[\left\{\bar{y}_{PdR}^{(d_0)}\right\} \text{and} \left\{\bar{y}_{PdR}^{(d_0^{**})}\right\}\right]$ over the conventional estimators $\left(\bar{y} \text{ and } \bar{y}_P^{(d)}\right)$ and dual to ratio $\left(\bar{y}_k^{(d)}\right)$ estimator in double sampling with respect to the population data sets in both the cases. This shows that even if the scalar α deviates from its optimum values $\left(\alpha_{Iopt} \text{ and } \alpha^*_{IIopt}\right)$, the proposed class of estimators $\bar{y}_{PdR}^{(d)}$ will yield better estimates then $\bar{y}$, $\bar{y}_P^{(d)}$ and $\bar{y}_k^{(d)}$. It is further observed that the estimator $\bar{y}_{PdR}^{(d_0^{**})}$ is more efficient than $\bar{y}_{PdR}^{(d_0)}$, except for the data sets of population I and II, where the estimator $\bar{y}_{PdR}^{(d_0)}$ is slightly better than $\bar{y}_{PdR}^{(d_0^{**})}$. Thus, the use of the proposed class of estimators is preferable over others.

REFERENCES

[1] A. J. Dobson, *An Introduction to Generalized Linear Models*, 1st Edition, New York: Chapman & Hall, 1990.

[2] A. K. Das, *"Contribution to the theory of sampling strategies based on auxiliary information,"* Unpublished Ph. D. Thesis, Bidhan Chandra Krishi Vishwavidyalaya, India, 1988.

[3] B. V. S. Sisodia and V. K. Dwivedi, "A modified ratio estimator using coefficient of variation of auxiliary variable," *Jour. Ind. Soc. Agril. Statist.*, vol. 33, pp. 13-18, 1981.

[4] C. Bose, "Note on the sampling error in the method of double sampling," *Sankhyā*, vol. 6, pp. 329-330, 1943.

[5] D. S. Robson, "Applications of multivariate polykays to the theory of unbiased ratio-type estimation," *Journal of the American Statistical Association,* vol. 52, pp. 511–522, 1957.

[6] D. T. Searls, "The utilization of known coefficient of variation in the estimation procedure," *Journal of American Statistical Association*, vol. 59, pp. 1125-26, 1964.

[7] H. P. Singh and M. R. Espejo, "Double sampling ratio-product estimator of a finite population mean in sampling surveys," *Journal of Applied Statistics*, vol. 34, no. 1, pp. 71-85, 2007.

[8] H. P. Singh and S. Kumar, "Improved estimation of population mean under two-phase sampling with subsampling the non-respondents," *Journal of Statistical Planning and Inference*, vol. 140, no. 9, pp. 2536-2550, 2010.

[9] J. Neyman, "Contribution to the theory of sampling human populations," *Journal of the American Statistical Association*, vol. 33, pp. 101–116, 1938.

[10] L. Chand, *"Some ratio-type estimators based on two or more auxiliary variables,"* PhD Thesis, Ames, Iowa State University, 1975.

[11] M. A. Hidiroglou and C. E. Sarndal, "Use of auxiliary information for two-phase sampling," *Survey Methodology*, vol. 24, pp. 11–20, 1998.

[12] M. Kumar and S. Bahl, *"Class of dual to ratio estimators for double sampling,"* Statistical Papers, vol. 47, pp. 319-326, 2006.

[13] M. N. Murthy, *"Product method of estimation,"* Sankhyā A, vol. 26, pp. 69-74, 1964.

[14] M. N. Murthy, *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta: India, 1967.

[15] R. G. D. Steel and J. H. Torrie, *Principles and Procedures of Statistics*, New York: McGraw-Hill, 1960.

[16] R. R. Sitter, "Variance estimation for the regression estimator in two-phase sampling," *Journal of the American Statistical Association*, vol. 92, no. 438, pp. 780-787, 1997.

[17] S. Bandyopadhyay, "Improved ratio and product estimators," *Sankhyā Series C*. vol. 42, no 2, pp. 45-49, 1980.

[18] S. Bhushan, A. Pandey and S. Katara, "A class of estimators in double sampling using two auxiliary variables," *Journal of Reliability and Statistical Studies,* vol. 1, Issue 1, pp. 67-73, 2008.

[19] S. K. Srivastava, "A two-phase sampling estimator in sample surveys," *Australian Journal of Statistics*, vol. 2, no. 1 pp. 23–27, 1970.

[20] T. Srivenkataramana, "A dual to ratio estimator in sample surveys," *Biometrika,* vol. 67, no. 1, pp. 199-204, 1980.

[21] W. G. Cochran, *Sampling Technique*, 2nd Edition, Wiley: New York, 1963.

[22] W. G. Cochran, "The estimation of the yields of the cereal experiments by sampling for the ratio of grain to total produce," *The Journal of Agricultural Science,* vol. 30, no. 2, pp. 262–275, 1940.