Hacettepe Journal of Mathematics and Statistics
Volume 44 (5) (2015), 1257 – 1270

Study of Yadav and Kadilar's improved exponential type ratio estimator of population variance in two-phase sampling

Housila P. Singh , Ragen Goli and Tanveer A. Tarray*

Abstract

This paper presents a double sampling version of Yadav and Kadilar (2013) estimator alongwith its properties under large sample approximation. Cost aspect is also discussed. We have compared the proposed estimator with usual unbiased estimator and usual double sampling ratio estimator and shown that the proposed estimator is better than usual unbiased estimator and other existing estimators under some realistic conditions to two-phase sampling.

Keywords: Auxiliary variable, Bias, Efficiency, Mean squared error, Double sampling.

Received 21/01/2014 : Accepted 18/03/2014 Doi : 10.15672/HJMS.20158612895

1. Introduction

The use of auxiliary information has been dealt with at great length for improving estimators of population parameters in sample surveys. Various estimation procedures in sample surveys need advance knowledge of some auxiliary variable which is then used to increase the precision of estimates. For example, the ratio - type estimator due to Isaki (1983) need the advance knowledge of population variance S_x^2 of the auxiliary variable x. When the population variance $S_x^2$is not known, it is sometimes estimated from a preliminary large sample on which only the auxiliary characteristic x is observed. The value of S_x^2in the estimator is then replaced by its estimate. A smaller second phase sample of the variate under study y is then taken. This technique, known as double sampling or two-phase sampling, is especially appropriate if the x values are easily accessible and much cheaper to collect than the y_i values see. Hidiroglou and Sarandal (1998). The use of double sampling is necessary if the x - value is obtained by performing a non-destructive experiment where as to obtain a y - value of a unit, a destructive experiment has to be performed, see UnniKrishan and Kunte (1995). Double sampling is also an able alternative to simple random sampling when there are expected to be gains from using

*School of Studies in Statistics,Vikram University, Ujjain - 456010 - India
Email:tanveerstat@gmail.com

auxiliary information.

let $U = (U_1, U_2, ...U_N)$ denote the population of N units and let (y, x) be the variate defined on U taking values (y_i, x_i) on $U_i(i = 1, 2, ..., N)$. It is desired to estimate S_y^2 of the study variate y. A simple random sample of size n is drawn without replacement (SRSWOR) from the population U. The usual unbiased estimator of based on SRSWOR is given by :

$$(1.1) \qquad S_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2,$$

where $\bar{y} = \frac{1}{n}\sum_{i}^{n} y_i$ is the sample mean based on n observations.

To improve the usual unbiased estimator s_y^2, using the known population variance S_x^2 of the auxiliary variate x, Isaki (1983) suggested a ratio-type estimator for the population variance S_y^2 as

$$(1.2) \qquad t_l = s_y^2 \frac{S_x^2}{s_x^2},$$

where $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, is an unbiased estimator of the population variance s_x^2 and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean.

Singh et al. (2011) proposed the exponential ratio estimator for the population variance S_y^2 as

$$(1.3) \qquad t_s = s_y^2 exp(\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2}).$$

when the population varianceS_x^2 of the auxiliary character x, the usual linear regression estimator for population variance S_x^2 is defined by

$$(1.4) \qquad t_l r = s_y^2 + \hat{\beta}(S_x^2 - s_x^2)$$

where $\hat{\beta} = \frac{s_y^2(\hat{\lambda}_{22} - 1)}{s_x^2(\hat{\lambda}_{04} - 1)}$ is sample regression coefficient,
$\hat{\lambda}_{04} = \frac{\hat{\mu}_{04}}{\hat{\mu}_{02}^2}$, $\hat{\lambda}_{22} = \frac{\hat{\mu}_{22}}{\hat{\mu}_{20}\hat{\mu}_{02}}$,
$\hat{\mu}_{04} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4$, $\hat{\mu}_{02} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$,
$\hat{\mu}_{20} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^4$, $\hat{\mu}_{22} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2(x_i - \bar{x})^2$.
Motivated by Upadhyaya et al. (2011), Yadav and Kadilar (2013) suggested the following class of estimators of the population variance S_y^2 as

$$(1.5) \qquad t_y = s_y^2 exp[\frac{S_x^2 - s_x^2}{S_x^2 + (\alpha - 1)s_x^2}],$$

where $(\alpha \geq 0)$.

In this paper we have studied the properties of the above estimators $t_1, t_s, t_l r$ and $t_y$in the case of double sampling (i.e. when the population variance S_x^2 of the auxiliary variable x is not known). Cost aspects are also discussed. Numerical illustration is given in support of the present study.

2. Two-phase sampling estimators

When the population variance $S_x^2$of x is not known, a first phase sample of $n_1$is drawn from the population on which only the x-characteristic is measured in order to furnish a good estimate of S_x^2.Then a second phase sample of size n is drawn on which both the variates y and x are measured [see Singh and Ruiz Espejs (2007)].Let $(x_1, x_2, ..., x_{n1})$be the first phase sample drawn by simple random sampling without replacement (SRSWOR) from the given population U and only auxiliary variable x be measured.
Also, let $(y_1, y_2, ..., y_n)$ and $(x_1, x_2, ..., x_n)$,$(n < n_1)$denote respectively, the second phase sample for the study variable y and the auxiliary variable x respectively.
Let us write $\bar{x_1} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i$, $s_{x_1}^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_i - \bar{x_1})^2$, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$,
Then the two-phase sampling (or double sampling) estimators of population variance S_y^2 are given by

$$t_{ld} = s_y^2[\frac{s_{x1}^2}{s_x^2}], \tag{2.1}$$

$$t_{sd} = s_y^2 exp[\frac{s_{x1}^2 - s_x^2}{s_{x1}^2 + s_x^2}], \tag{2.2}$$

and

$$t_{yd} = s_y^2 exp[\frac{s_{x1}^2 - s_x^2}{s_{x1}^2 + (\alpha - 1)s_x^2}]. \tag{2.3}$$

It is to be mentioned that the estimators t_{ld}, t_{sd} and t_{yd} are double sampling versions of Isaki (1983) estimator, Singh et al. (2011) estimator and Yadav and Kadilar (2013) estimator. For $\alpha = 2$ in (8),t_{yd} reduces to the estimator t_{sd}.

3. The first Degree Approximation to the Biases and Variances of the Suggested Estimators.

In order to study the large sample properties of the proposed estimators, we define.
$s_y^2 = S_y^2(1 + \varepsilon_0)$, $s_x^2 = S_x^2(1 + \varepsilon_1)$,$s_{x1}^2 = S_x^2(1 + \varepsilon_2)$
such that $E(\varepsilon_0) = E(\varepsilon_1) = E(\varepsilon_2) = 0$
The following two cases will be considered separately.

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

Case - II : When the second phase sample of size n is drawn independently of the first phase sample of size n_1 see Bose (1943)

Case I - When the second phase sample of size n is a subsample of the first phase sample of size n_1 $(n < n_1)$, the expected values are :

$$E(\varepsilon_0^2) = \frac{1}{n}(\lambda_{40} - 1), E(\varepsilon_1^2) = \frac{1}{n}(\lambda_{04} - 1), E(\varepsilon_0\varepsilon_1) = \frac{1}{n}(\lambda_{22} - 1), E(\varepsilon_2^2) = \frac{1}{n_1}(\lambda_{04} - 1),$$

$$E(\varepsilon_0\varepsilon_2) = \frac{1}{n_1}(\lambda_{22} - 1), E(\varepsilon_1\varepsilon_2) = \frac{1}{n_1}(\lambda_{40} - 1), \tag{3.1}$$

where $\lambda_{rs} = \frac{\mu_{rs}}{(\mu_{20}^{r/2})(\mu_{02}^{s/2})}$, $\mu_r s = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^r(x_i - \bar{x})^s$
(r, s) being non-negative integers,
Case II - When the second phase sample of size n is independent of the first phase

sample of size n_1 , the expected value are :

$$E(\varepsilon_0^2) = \frac{1}{n}(\lambda_{40}-1), E(\varepsilon_1^2) = \frac{1}{n}(\lambda_{04}-1), E(\varepsilon_0\varepsilon_1) = \frac{1}{n}(\lambda_{22}-1),$$

$$E(\varepsilon_2^2) = \frac{1}{n_1}(\lambda_{04}-1), E(\varepsilon_0\varepsilon_2) = E(\varepsilon_1\varepsilon_2) = 0, \tag{3.2}$$

Expressing t_{ld}, t_{sd} and t_{yd} in terms of $\varepsilon_i' s, (i = 0, 1, 2)$, we have

$$t_{ld} = s_y^2(1+\varepsilon_0)(1+\varepsilon_1)^{-1}(1+\varepsilon_2) \tag{3.3}$$

$$t_{sd} = s_y^2(1+\varepsilon_0)exp[-\frac{(\varepsilon_1-\varepsilon_2)}{2}(1+\frac{\varepsilon_1+\varepsilon_2}{2})^{-1}] \tag{3.4}$$

$$t_{yd} = s_y^2(1+\varepsilon_0)exp[-\frac{(\varepsilon_1-\varepsilon_2)}{\alpha}(1+\frac{(\alpha-1)\varepsilon_1+\varepsilon_2}{\alpha})^{-1}] \tag{3.5}$$

Expanding the right hand side of (11), (12) and (13) multiplying out and neglecting terms of $e's$ having power greater than two we have

$$t_{ld} \cong S_y^2(1+\varepsilon_0+\varepsilon_2-\varepsilon_1+\varepsilon_0\varepsilon_2-\varepsilon_0\varepsilon_1-\varepsilon_1\varepsilon_2+\varepsilon_1^2)$$

or

$$(t_{ld}-S_y^2) \cong S_y^2(\varepsilon_0+\varepsilon_2-\varepsilon_1+\varepsilon_0\varepsilon_2-\varepsilon_0\varepsilon_1-\varepsilon_1\varepsilon_2+\varepsilon_1^2) \tag{3.6}$$

$$t_{sd} \cong S_y^2[(1+\varepsilon_0-\frac{(\varepsilon_1-\varepsilon_2)}{2}-\frac{(\varepsilon_0\varepsilon_1-\varepsilon_0\varepsilon_2)}{2}+\frac{(3\varepsilon_1^2-\varepsilon_2^2-2\varepsilon_1\varepsilon_2)}{8}]$$

or

$$(t_{sd}-S_y^2) \cong S_y^2[(\varepsilon_0-\frac{(\varepsilon_1-\varepsilon_2)}{2}-\frac{(\varepsilon_0\varepsilon_1-\varepsilon_0\varepsilon_2)}{2}+\frac{(3\varepsilon_1^2-\varepsilon_2^2-2\varepsilon_1\varepsilon_2)}{8}] \tag{3.7}$$

$$t_{yd} \cong S_y^2[1+\varepsilon_0-\frac{(\varepsilon_1-\varepsilon_2)}{\alpha}-\frac{(\varepsilon_0\varepsilon_1-\varepsilon_0\varepsilon_2)}{\alpha}+\frac{((2\alpha-1)\varepsilon_1^2-\varepsilon_2^2-\alpha\varepsilon_1\varepsilon_2)}{2\alpha^2}]$$

or

$$(t_{yd}-S_y^2) \cong S_y^2[\varepsilon_0-\frac{(\varepsilon_1-\varepsilon_2)}{\alpha}-\frac{(\varepsilon_0\varepsilon_1-\varepsilon_0\varepsilon_2)}{\alpha}+\frac{((2\alpha-1)\varepsilon_1^2-\varepsilon_2^2-2\varepsilon_1\varepsilon_2)}{2\alpha^2}] \tag{3.8}$$

Now squaring both sides of (14), (15) and (16) and neglecting terms of $\varepsilon' s$having power greater than two we have

$$(t_{ld}-s_y^2) = S_y^4(\varepsilon_0^2+(\varepsilon_2-\varepsilon_1)^2-2(\varepsilon_0\varepsilon_1-\varepsilon_0\varepsilon_2)] \tag{3.9}$$

$$(t_{sd}-S_y^2)^2 = S_y^4(\varepsilon_0^2+\frac{(\varepsilon_1-\varepsilon_2)^2}{4}-(\varepsilon_0\varepsilon_1-\varepsilon_0\varepsilon_2)] \tag{3.10}$$

and

$$(t_{yd}-S_y^2)^2 = S_y^4(\varepsilon_0^2+\frac{(\varepsilon_1-\varepsilon_2)^2}{\alpha^2}-\frac{(2\varepsilon_0\varepsilon_1-\varepsilon_0\varepsilon_2)}{\alpha})] \tag{3.11}$$

Taking expectations of both sides of (14), (15), (16) and (17), (18), (19) and using the results in (9), we get the biases and mean squared errors of $t_{1d}, t_{sd}$and t_{yd} to the first degree of approximation under case-I respectively as

$$B(t_{ld})_1 = (\frac{1}{n}-\frac{1}{n_1})(\lambda_{04}-1)S_y^2(1-C) \tag{3.12}$$

$$B(t_{sd})_1 = \frac{1}{8}(\frac{1}{n}-\frac{1}{n_1})(\lambda_{04}-1)S_y^2(3-4C) \tag{3.13}$$

$$(3.14)\quad B(t_{yd})_1 = \frac{(\lambda_{04}-1)}{2\alpha^2}[\frac{1}{n}[2\alpha(1-c)-1] - \frac{1}{n_1}(1+\alpha(1-2c))]$$

$$(3.15)\quad MSE(t_{1d})_1 = S_y^4[\frac{1}{n}(\lambda_{40}-1) + (\frac{1}{n}-\frac{1}{n_1})(\lambda_{04}-1)(1-2c)]$$

$$(3.16)\quad MSE(t_{sd})_1 = S_y^4[\frac{1}{n}(\lambda_{40}-1) + (\frac{1}{n}-\frac{1}{n_1})\frac{1}{4}(\lambda_{04}-1)(1-4c)]$$

$$(3.17)\quad MSE(t_{yd})_1 = S_y^4[\frac{1}{n}(\lambda_{40}-1) + (\frac{1}{n}-\frac{1}{n_1})\frac{1}{\alpha^2}(\lambda_{04}-1)(1-2\alpha c)]$$

where $c = \frac{\lambda_{22}-1}{\lambda_{04}-1}$, and $B(.)_1$ and $MSE(.)_1$stand the bias of (.) under case-I (i.e. when the second phase sample is a subsample of the first phase sample) respectively.
Now taking the expectations of both sides of (14), (15), (16) and (17), (18) and (19) and using results in (10) we get the biases and mean squared errors of the estimators t_{1d}, t_{sd} and t_{yd} to the first degree of approximation under case-II respectively as

$$(3.18)\quad B(t_{ld})_{11} = \frac{S_y^2(\lambda_{04}-1)}{n}(1-c)$$

$$(3.19)\quad B(t_{sd})_{11} = \frac{S_y^2(\lambda_{04}-1)}{8}[\frac{3-4c}{n} - \frac{1}{n_1}]$$

$$(3.20)\quad B(t_{yd})_{11} = \frac{S_y^2(\lambda_{04}-1)}{2\alpha^2}[\frac{(2\alpha-2\alpha c-1)}{n} - \frac{1}{n_1}]$$

$$(3.21)\quad MSE(t_{ld})_{11} = S_y^4[(\frac{1}{n})[(\lambda_{40}-1) + (\lambda_0 4-1)(1-2c)] + \frac{\lambda_{04}-1}{n_1}]$$

$$(3.22)\quad MSE(t_{sd})_{11} = S_y^4[(\frac{1}{n})[(\lambda_{40}-1) + \frac{(\lambda_{04}-1)}{4}(1-4c)] + \frac{\lambda_{04}-1}{4n_1}]$$

$$(3.23)\quad MSE(t_{yd})_{11} = S_y^4[(\frac{1}{n})[(\lambda_{40}-1) + \frac{(\lambda_{04}-1)}{\alpha^2}(1-2\alpha c)] + \frac{\lambda_{04}-1}{\alpha^2 n_1}]$$

where $B(.)_{11}$ and $MSE(.)_{11}$stand the bias of (.) and MSE of (.) under case-II.

4. Optimum choice of the scalar $'\alpha'$

Case - I The $MSE(t_{yd})_1$ at (25)is minimized for

$$(4.1)\quad \alpha = \frac{1}{c} = \alpha_{opt}(say)$$

Substitution (32) in (8) yields the asymptotically optimum estimator (AOE) of S_y^2 as

$$(4.2)\quad t_{yd(0)} = s_y^2 exp[\frac{c(s_{x1}^2 - s_x^2)}{cs_{x1}^2 + (1-c)s_x^2}]$$

The value of $'c'$ can be guessed quite accurately from the past data or experience gathered in due course of time see Yadav and Kadilar (2013, p. 148). In case c is not known, it is worth advisable to replace c by its consistent estimate $\hat{c} = \frac{(\hat{\lambda}_{22}-1)}{\hat{\lambda}_{04}-1}$ based on sample data at hand, where $\hat{\lambda}_{22}$ and $\hat{\lambda}_{04}$ are same as defined earlier. Thus replacing $'c'$ by its estimate $'\hat{c}'$in (33) , we get an estimator of S_y^2 based on estimated optimum as

$$(4.3)\quad \hat{t}_{yd(0)} = s_y^2 exp[\frac{\hat{c}(s_{x1}^2 - s_x^2)}{\hat{c}s_{x1}^2 + (1-\hat{c}s_x^2)}]$$

It can be shown to the first degree of approximation that

$$(4.4)\qquad MSE(t_{yd(0)})_1 = MSE(\hat{t}_{yd(0)})_1 = \frac{s_y^4}{n}[(\lambda_{40}-1)-(\frac{n_1-1}{n_1})\frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)}]$$

which equals to the approximate variance / MSE of the regression estimator

$$t_{lrd} = s_y^2 + \frac{s_y^2((\hat{\lambda})_{22}-1)}{s_x^2((\hat{\lambda})_{04}-1)}(s_{x1}^2 - s_x^2)$$

Thus the proposed $\hat{t}_{yd(0)}$ is an alternative to the regression estimator t_{lrd} It is well known under SRSWOR that to the first degree of approximation (ignoring fpc term) that

$$(4.5)\qquad V(s_y^2) = MSE(s_y^2) = \frac{1}{n}S_y^4(\lambda_{40}-1)$$

From (23), (24), (35) and (36) we have

$$(4.6)\qquad MSE(s_y^2) - MSE(\hat{t}_{yd(0)}) = (\frac{1}{n}-\frac{1}{n_1})S_y^4\frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)} \geq 0$$

$$(4.7)\qquad MSE(t_{ld}) - MSE(\hat{t}_{yd(0)}) = (\frac{1}{n}-\frac{1}{n_1})S_y^4(\lambda_{04}-1(1-c)^2 \geq 0$$

$$(4.8)\qquad MSE(t_{sd}) - MSE(\hat{t}_{yd(0)}) = (\frac{1}{n}-\frac{1}{n_1})S_y^4\frac{(\lambda_{04}-1)}{4}(1-2c)^2 \geq 0$$

It follows from (37), (38) and (39) that the proposed estimator $\hat{t}_{d(0)}$ is more efficient than the usual unbiased estimator S_y^2, t_{ld} and t_{sd}. Thus the proposed estimator $\hat{t}_{yd(0)}$ is an appropriate choice among the estimator S_y^2, t_{ld}, t_{sd} and $\hat{t}_{yd(0)}$ to be used in practice.
case - II: The $MSE(t_{yd})_{11}$ at (31) is minimized for

$$(4.9)\qquad \alpha = \frac{n+n_1}{n_1 c} = \alpha^*_{opt}$$

Substitution of (40) is (8) yields the asymptotically optimum estimator (AOE) under case-II as

$$(4.10)\qquad t^*_{yd(0)} = s_y^2 exp[\frac{c(s_{x1}^2 - s_x^2)}{cs_{x1}^2 + (\delta - c)s_x^2}]$$

where $\delta = (n+n_1)/n_1$
if c is not known, then we replace c by its consistent estimate $\hat{c}$. thus the estimator based on estimated optimum value $\hat{c}$ of c is given by

$$(4.11)\qquad (\hat{t})^*_{yd(0)} = s_y^2 exp[\frac{(\hat{c})(s_{x1}^2 - s_x^2)}{(\hat{c})s_{x1}^2 + (\delta - (\hat{c}))s_x^2)}]$$

To the first degree of approximation (ignoring fpc terms), it can be shown that

$$(4.12)\qquad MSE(t^*_{yd(0)}) = \frac{S_y^4}{n}[(\lambda_{40}-1) - \frac{n_1}{(n+n_1)}(\lambda_{04}-1)c^2]$$

From (29), (30), (36) and (43), we have

$$(4.13)\qquad MSE(s_y^2) - MSE(\hat{t}^*_{yd(0)}) = \frac{n_1}{n(n+n_1)}S_y^4(\lambda_{04}-1)c^2 \geq 0$$

$$(4.14)\qquad MSE(t_{ld})_{11} - MSE(\hat{t}^*_{yd(0)}) = \frac{S_y^4(\lambda_{04}-1n+n_1(1-c)^2)}{n(n+n_1)} \geq 0$$

$$(4.15)\quad MSE(t_{sd})_{11} - MSE(\hat{t}^*_{yd(0)}) = \frac{S_y^4(\lambda_{04}-1)(n+n_1-2n_1c^2)}{4nn_1(n+n_1)} \geq 0$$

Thus the proposed estimator $\hat{t}^*_{yd(0)}$ is more efficient than the usual unbiased estimator s_y^2, t_{ld} and t_{sd} under case - II.
From (35) and (43), we have

$$(4.16)\quad [MSE(t^*_{yd(0)})_1 - MSE(t^*_{yd(0)})_{11}] = \frac{ns_y^4(\lambda_{04}-1)c^2)}{n_1(n+n_1)} \geq 0$$

which shows that the proposed estimator $t_{yd(0)}$ under case -I is better than the proposed estimator $t^*_{yd(0)}$ under case - II.

5. EFFICIENCY COMPARISON OF THE PROPOSED ESTIMATOR WHEN THE SCALAR $\lfloor\alpha\rfloor$ DOES NOT COINCIDE EXACTLY WITH ITS OPTIMUM VALUE.

In this section we compare the proposed estimator $t_{yd}$with the estimators s_y^2, t_ld, t_sd under case - I and II.
Case - I: From (25) and (36) we have

$$(5.1)\quad MSE(s_y^2) - MSE(t_{yd})_1 = (\frac{1}{n} - \frac{1}{n_1})s_y^4\frac{1}{\alpha^2}(2\alpha c - 1)$$

which is positive if

$$2\alpha c - 1 > 0$$

i.e. if

$$(5.2)\quad \alpha > \frac{1}{2c}$$

From (23) and (25) we have

$$MSE(t_{ld})_1 - MSE(t_{yd})_1 = (\frac{1}{n} - \frac{1}{n_1})s_y^4(\lambda_{04}-1)[1 - 2c - \frac{1}{\alpha^2} + \frac{2c}{\alpha}]$$

which is positive if $[(1 - \frac{1}{\alpha^2}) - 2c(1 - \frac{1}{\alpha})] > 0$
i.e. if

$$(5.3)\quad either min.[1, \frac{1}{(2c-1)}] < \alpha < max.[1, \frac{1}{(2c-1)}], c > \frac{1}{2}$$

or

$$(5.4)\quad \alpha > 1, 0 \leq c \leq \frac{1}{2}$$

Further from (24) and (25) we have

$$MSE(t_{sd})_1 - MSE(t_{yd})_1 = (\frac{1}{n} - \frac{1}{n_1})s_y^4(\lambda_{04}-1)(\frac{1}{2} - \frac{1}{\alpha})(\frac{1}{2} + \frac{1}{\alpha} - 2c)$$

which is greater than ' zero' if

$$(\frac{1}{2} - \frac{1}{\alpha})(\frac{1}{2} + \frac{1}{\alpha} - 2c)$$

i.e. if

$$either min.[2, \frac{2}{4c-1}] < \alpha < max.[2, \frac{2}{4c-1}] \tag{5.5}$$

or

$$\alpha > 2, 0 \leq c \leq \frac{1}{4}$$

Thus we established the following theorem.
Theorem - 5.1 The proposed estimator t_{yd} in case-I is more efficient than :
(i) the usual unbiased estimators_y^2 if

$$\alpha > \frac{1}{2c}$$

(ii)the Isaki (1983) double sampling ratio estimator t_{ld} if

$$either min.[1, \frac{1}{(2c-1)}] < \alpha < max.[1, \frac{1}{(2c-1)}], c > \frac{1}{2}$$

or

$$\alpha > 1, 0 \leq c \leq \frac{1}{2} \tag{5.6}$$

(iii)the double sampling version of Singh et al (2011) estimator t_{sd} if

$$either min.[2, \frac{2}{4c-1}] < \alpha < max.[2, \frac{2}{4c-1}]$$

or

$$\alpha > 2, 0 \leq c \leq \frac{1}{4}$$

Case II-From (31) and (36) we have

$$MSE(s_y^2) - MSE(t_y d)_{11} = -s_y^4(\lambda_{04}-1)\frac{1}{\alpha^2}[\frac{1}{n}(1-2\alpha c) + \frac{1}{n_1}]$$

which is positive if

$$[\frac{1}{n}(1-2\alpha c) + \frac{1}{n_1}] \leq 0 \tag{5.7}$$

i.e. if $\alpha > \frac{\delta}{2c}$,
where $\delta = \frac{(n+n_1)}{n_1}$.
From (29) and (31) we have

$$MSE(t_{ld})_{11} - MSE(t_{yd})_{11} = S_y^4(\lambda_{04}-1)[\frac{1}{n}(1-2c) + \frac{1}{n_1} - \frac{(1-2\alpha c)}{n\alpha^2} - \frac{1}{\alpha^2 n_1}]$$

which is positive if

$$[(\frac{1}{n} + \frac{1}{n_1})(1 - \frac{1}{\alpha^2}) - \frac{2}{n}(1 - \frac{1}{\alpha})c] > 0$$

i.e. if

either $1 < \alpha < \dfrac{\delta}{2c-\delta}$ or $\dfrac{\delta}{(2c-\delta)} < \alpha < 1$
or equivalently,

$$(5.8) \qquad min.[1, \frac{\delta}{(2c-\delta)}] < \alpha < max.[1, \frac{\delta}{2c-\delta}].$$

Also the difference

$$(5.9) \qquad [MSE(t_{ld})_{11} - MSE(t_{yd})_{11}] \qquad ispositiveif \qquad \alpha > 1, c < \frac{\delta}{2}$$

From (30) and (31) we have

$$[MSE(t_{sd})_{11} - MSE(t_{yd})_{11}] = S_y^4(\lambda_{04}-1)[\frac{1}{n}[\frac{1-4c}{4} - \frac{1-2\alpha c}{\alpha^2}] + (\frac{1}{4} - \frac{1}{\alpha^2})\frac{1}{n_1}]$$

$$= S_y^4(\lambda_{04}-1)(1-\frac{2}{\alpha})[\frac{\delta}{4}(1+\frac{2}{\alpha} - c]$$

which is positive if
either $2 < \alpha < \dfrac{2\delta}{4c-\delta}$ or $\dfrac{2\delta}{(4c-\delta)} < \alpha < 2$
or equivalently,

$$(5.10) \qquad min.[2, \frac{2\delta}{(4c-\delta)}] < \alpha < max. - [2, \frac{2\delta}{4c-\delta}].$$

Now established the following theorem.
Theorem - 5.2 The proposed estimator t_{yd} under case II is more efficient than :
(i) the usual unbiased estimator s_y^2 if

$$\alpha > \frac{\delta}{2c}$$

(ii)the Isaki's (1983) ratio type double (two phase) sampling estimator t_{sd} if
either $[min.1, \dfrac{\delta}{(2c-\delta)}] < \alpha < max.[1, \dfrac{\delta}{2c-\delta}]$.
(iii) the Singh et al.'s (2011) estimator t_{sd} if
either $[2, \dfrac{2\delta}{4c-\delta}] < \alpha < max.[2, \dfrac{2\delta}{4c-\delta}]$

6. Comparison with single phase sampling

In this section following Singh and Ruiz Espejo (2007) the comparisons between double and Single-phase sampling have been made for fixed cost. We shall consider the cases separately.
Case - I - In this case we consider the following cost function:

$$(6.1) \qquad c^* = nc_1 + n_1c_2$$

where c^* equals the total cost of the survey and (c_1, c_2) are the costs per unit of collecting information on the study variate y and the auxiliary variate x respectively.
In this case, we express the minimum MSE of t_{yd}(or the MSE of $\hat{t}_{yd(0)}$) as

$$(6.2) \qquad M_y = \frac{M_{y1}}{n} + \frac{M_{y2}}{n_1}$$

(6.3) $$M_{y1} = [(\lambda_{40} - 1) - \frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)} = (\lambda_{40} - 1)(1 - \rho^{*2})S_y^4$$

(6.4) $$M_{y2} = (\frac{(\lambda_{40} - 1)^2}{(\lambda_{04} - 1)} = (\lambda_{40} - 1)\rho^{*2})S_y^4$$

where $\rho^* = \dfrac{(\lambda_{22} - 1)}{\sqrt{(\lambda_{22} - 1)(\lambda_{04} - 1)}}$

The optimum values of n and n_1 for fixed cost c^* , which minimizes the mean squared error M_y is given by

(6.5) $$n_{yopt} = \frac{C^* \sqrt{\frac{M_{y1}}{c_1}}}{\sqrt{M_{y1}}c_1 + \sqrt{M_{y2}c_2}} \qquad n_{y1opt} = \frac{C^* \sqrt{\frac{M_{y2}}{c_2}}}{\sqrt{M_{y1}}c_1 + \sqrt{M_{y2}c_2}}$$

The mean squared error of $\hat{y}_{yd(0)}$corresponding to optimal double sampling estimator is

$$MSE_{opt}(t_{yd})_1 = (\frac{1}{c^*})(\sqrt{c_1 M_{y1}} + \sqrt{c_2 M_{y2}})^2$$

(6.6) $$(\frac{S_y^4}{c^*})(\lambda_{40} - 1)(\sqrt{c_1(1 - \rho^{*2} + \rho^* \sqrt{c_2})^2}$$

Case - II In case II, we assume that x is measured on y on n^*=$n + n_1$ units and$y$ units. Motivated by Srivastava (1970) we shall consider a simple cost function:

(6.7) $$c^* = c_1 n + c_2^* n^*$$

where c_1 andc_2^* denote costs per unit of observing the study variate y and the auxiliary variate$x$ values respectively. The expression of mean squared error of$\hat{t}_{yd(0)}$ (under case II) can now be written as

(6.8) $$M_y^* = \frac{M_{y1}}{n} + \frac{M_{y2}}{n^*},$$

where $n^* = n + n_1$

To obtain the optimum allocation of sample between phases for a fixed cost c^*, we minimize equation (65) with the condition (64). It is easily obtained that this minimum is attained for

(6.9) $$\frac{n}{n^*} = (\frac{M_{y1}c_2^*}{M_{y2}c_1})^{1/2} = \frac{c_2^*(1 - \rho^{*2})}{c_1 \rho^{*2}}^{1/2}$$

Thus the minimum MSE corresponding to these optimum values of n and n_1 are given by

(6.10) $$MSE_{opt}(\hat{t}_{yd(0)})_{11} = [\frac{S_y^4(\lambda_{40} - 1)}{c^*}][\sqrt{(1 - \rho^{*2})c_1} + \rho^* \sqrt{c_2^*}]^2$$

Had all the resources been diverted towards the study variate y only, then we would have optimum sample size as given below

(6.11) $$n^{**} = \frac{c^*}{c_1}$$

Thus the variance of the usual unbiased estimator s_y^2 for a given fixed cost $c$in case of large population is given by

(6.12) $$MSE_{opt}(s_y^2) = (\frac{c_1}{c^*})s_y^4(\lambda_{40} - 1)$$

Case - I : From (63) and (69), the suggested double sampling strategy would be profitable if

$$MSE_{opt}(\hat{t}_{yd(0)}) < MSE_{opt}(S_y^2)$$

i.e. if

$$\frac{c_2}{c_1} < \frac{(1-\sqrt{1-\rho^{*2}})^2}{\rho^{*2}}$$

Thus we established the following theorem.
Theorem 6.1 The suggested double sampling strategy $\hat{t}_{yd(0)}$ would be more efficient than the strategy s_y^2 as long as

$$\frac{c_2}{c_1} < \frac{(1-\sqrt{1-\rho^{*2}})^2}{\rho^{*2}}$$

Case-II From (67) and (69) it is observed that the double sampling estimator $\hat{t}_{yd(0)}$ is better than the sample mean square s_y^2 for the same fixed cost, if

$$MSE(\hat{t}_{yd(0)})_{11} < MSE_{opt}(s_y^2)$$

i.e. if

$$\rho^{*2} > \frac{4c_1c_2^*}{(c_1+c_2^*)^2}$$

7. Empirical Study

The appropriateness of the proposed estimator has been examined with the help of the four data sets, given in $Table 1$ earlier considered by Subramani and Kumarapandiyan (2012).
We have computed the percent relative efficiencies of the estimators $s_y^2, t_{ld}, t_{sd}$and $\hat{t}_{yd(0)}$ with respect to the usual unbiased estimator s_y^2 by using the following formulae:

$$(i) PRE(t_{ld}, s_y^2)_1 = \frac{(\frac{1}{n})((\lambda_{40})-1)}{[(\frac{1}{n})((\lambda_{40})-1)+(\frac{1}{n}-\frac{1}{n_1})(\lambda_{04}-1)(1-2c)]} \times 100$$

$$(ii) PRE(t_{sd}, s_y^2)_1 = \frac{(\frac{1}{n})((\lambda_{40})-1)}{[(\frac{1}{n})((\lambda_{40})-1)+(\frac{1}{n}-\frac{1}{n_1})(\frac{1}{4})(\lambda_{04}-1)(1-4c)]} \times 100$$

$$(iii) PRE(\hat{t}_{yd(0)}, s_y^2)_1 = \frac{((\lambda_{40})-1)}{[((\lambda_{40})-1)-\frac{n_1-n}{n_1}c^2(\lambda_{04}-1)]} \times 100$$

$$(iv) PRE(t_{1d}, s_y^2)_{11} = \frac{(\frac{1}{n})((\lambda_{40})-1)}{[(\frac{1}{n})[(\lambda_{40}-1)+(\lambda_{04}-1)(1-2c)+(\frac{1}{n_1})(\lambda_{04}-1)]} \times 100$$

$$(v) PRE(t_{sd}, s_y^2)_{11} = \frac{(\frac{1}{n})((\lambda_{40})-1)}{[(\frac{1}{n})[(\lambda_{40}-1)+\frac{(\lambda_{04}-1)}{4}(1-4c)+(\frac{1}{n_1})\frac{(\lambda_{04}-1)}{4}]} \times 100$$

$$(vi)PRE(\hat{t}^*_{yd(0)}, s_y^2)_{11} = \frac{((\lambda_{40})-1)}{[((\lambda_{40})-1)+\dfrac{n_1}{n+n_1}c^2(\lambda_{04}-1)]} \times 100$$

Findings are shown in Table 2.
It is observed from Table 2 that the performance of the proposed estimator $\hat{t}_{yd(0)}(\hat{t}^*_{yd(0)})$ is more efficient than the estimators s_y^2 ,t_{ld} and t_{sd} . The percent relative efficiency of the proposed estimator $\hat{t}_{yd(0)}$ (under case I) is larger than the proposed estimator $(\hat{t}^*_{yd(0)})$.
Table 3, exhibits the range of α in which the proposed class of estimators $\hat{t}_{yd(0)}$ is more efficient than the usual unbiased estimator s_y^2 , Isaki (1983) ratio type estimator t_{id} in double sampling and the estimator t_{sd} which is double sampling version of Singh et al.'s (2011) exponential type estimator.

8. Conclusion

We have suggested an improved exponential ratio estimator for estimating the population variance in two phase sampling. It has been shown theoretically and numerically that the proposed estimator is better than the existing estimators in literature, the usual sample variance, traditional ratio estimator due to Isaki (1983), Yadav and Kadilar (2013) and Singh et al. (2011) exponential ratio estimator in the sense of having lesser mean square error. We have also given the range α of along with its optimum value for the proposed estimator to be more efficient than other competitors. Hence, the proposed estimator is recommended for its practical use for estimating the population variance when the auxiliary information is available. For the sake of completeness we have also discussed the cost aspect.

Acknowledgements
The authors are thankful to the Editor – in- Chief - Professor Cem- Kadilar, and to the anonymous learned referees for their valuable suggestions regarding improvement of the paper.

References

[1] Hidiroglou M.A and Sarandal C.E. (1988): Use of auxiliary information for two phase sampling . Survey Meth., 873-878.
[2] Isaki C. (1983): Variance estimation using auxiliary information. Jour. Amer. Statist. Assoc., 78,117-123.
[3] Shrivastav S.K. (1970): A two phase sampling estimator in sample surveys. Aust. Jour. Statist. 12 (1),23-27.
[4] Singh H.P and Ruiz- Espejo, M. (2007). Double sampling ratio product estimator of a finite population mean in sampling surveys. Jour. Appl. Statist. 34 (1), 71-85.
[5] Singh R., Chauhan P., Swan N. and Smarandache F. (2011): Improved exponential estimator for population variance using two auxiliary variables. Ital. Jour. Pur. Appl. Math. 28,101-108.
[6] Subramani J. and Kumarapandiyan G. (2012): Variance estimation using median of the auxiliary variable. Inter. Jour. Prob. Statist. 1(3), 36-40.
[7] Unnikrishan N.K. and Kunte S. (1995). Optimality of an analogue of Basu's estimator under a double sampling design. Sankhya, 57,103-111.
[8] Upadhyaya L., Singh H.P., Chatterjee S. and Yadav R. (2011): Improved ratio and product exponential type estimators. Jour. Statist. Theo. Pract. 5(2), 285-302.
[9] Yadav S.H. and Kadilar C. (2013): Improved exponential type ratio estimator of population variance. Rev. Colom. Estad., 36, 145-152.

Table 1. Parameters of the population

Parameters	Population 1	Population 2	Population 3	Population 4
N	103	103	80	49
$\bar{Y}$	626.2123	62.6212	51.8264	116.1633
$\bar{X}$	557.1909	556.5541	11.2646	98.6765
ρ	0.9936	0.7298	0.9413	0.6904
s_y	913.5488	91.3549	18.3569	98.8286
c_y	1.4588	1.4588	0.3542	0.8508
s_x	818.1117	610.1643	8.4563	102.9709
c_x	1.4683	1.0963	0.7507	1.0435
λ_{04}	37.3216	17.8738	2.8664	5.9878
λ_{40}	37.1279	37.1279	2.2667	4.9245
λ_{22}	37.2055	17.2220	2.2209	4.6977
c	0.9969	0.9635	0.7748	0.7846

Table 2. Percent relative efficiencies (PREs) of different estimators of population variance S_y^2 with respect to the unbiased estimator s_y^2 .

Estimator				PRE(., s_y^2)				
				Population				
	I	I	II	II	III	III	IV	IV
	Case I n_1=60 n=40	Case II n_1=60 n=40	Case I n_1=60 n=40	Case II n_1=60 n=40	Case I n_1=30 n=20	Case II n_1=30 n=20	Case I n_1=25 n=20	Case II n_1=25 n=20
s_y^2	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
t_{ld}	149.90	99.38	116.92	96.71	130.34	63.62	112.90	69.78
t_{sd}	133.41	199.80	112.54	127.65	128.62	153.73	112.00	140.10
$\hat{t}_{yd(0)}$	149.91	-	116.92	-	134.11	-	114.13	-
$\hat{t}^*_{yd(0)}$	-	249.73	-	135.22	-	184.23	-	152.45

Table 3. Range of α for t_{yd} to be more efficient than different estimators of the population variance S_y^2 .

Estimator				Population				
	I	I	II	II	III	III	IV	IV
	Case I n_1=60 n=40	Case II n_1=60 n=40	Case I n_1=60 n=40	Case II n_1=60 n=40	Case I n_1=30 n=20	Case II n_1=30 n=20	Case I n_1=25 n=20	Case II n_1=25 n=20
s_y^2	$\alpha > 0.50$	$\alpha > 0.84$	$\alpha > 0.52$	$\alpha > 0.87$	$\alpha > 0.65$	$\alpha > 1.08$	$\alpha > 0.64$	$\alpha > 1.15$
t_{ld}	$\alpha\varepsilon(1.00, 1.01)$	$\alpha\varepsilon(1.00, 1.68)$	$\alpha\varepsilon(1.00, 1.08)$	$\alpha\varepsilon(1.00, 1.79)$	$\alpha\varepsilon(1.00, 1.83)$	$\alpha\varepsilon(1.00, 3.03)$	$\alpha\varepsilon(1.00, 1.76)$	$\alpha\varepsilon(1.00, 3.17)$
t_{sd}	$\alpha\varepsilon(0.67, 2.01)$	$\alpha\varepsilon(1.44, 2.00)$	$\alpha\varepsilon(0.70, 2.00)$	$\alpha\varepsilon(1.52, 2.00)$	$\alpha\varepsilon(0.95, 2.00)$	$\alpha\varepsilon(2.00, 2.32)$	$\alpha\varepsilon(0.94, 2.00)$	$\alpha\varepsilon(2.00, 2.68)$
common range of α for t_{yd} to be more efficient s_y, t_{ld}, t_{sd} $\hat{t}_{yd(0)}$	$\alpha\varepsilon(1.00, 1.01)$	$\alpha\varepsilon(1.43, 1.68)$	$\alpha\varepsilon(1.00, 1.08)$	$\alpha\varepsilon(1.52, 1.79)$	$\alpha\varepsilon(1.00, 1.83)$	$\alpha\varepsilon(2.00, 3.03)$	$\alpha\varepsilon(1.00, 1.76)$	$\alpha\varepsilon(2.00, 2.68)$
Optimum value of α	1.003	1.672	1.038	1.73	1.291	2.152	1.275	2.295