DOI: 10.15415/mjis.2015.32017

A Generalised Family of Estimator for Estimating Unknown Variance Using Two Auxiliary Variables

VIPLAV KUMAR SINGH AND RAJESH SINGH

Department of Statistics, Banaras Hindu University, Varanasi -221005, India

Email: rsinghstat@gmail.com

Received: 25 March 2014| Revised: 19 February 2015| Accepted: 4 March 2015

Published online: March 30, 2015

Abstract: The present study deals with the estimation of unknown variance using two auxiliary variates under simple random sampling without replacement (SRSWOR) scheme. We have also extended our work to the case of double sampling scheme. The asymptotic expressions for the mean squared errors of the proposed estimators are derived and their optimum values have been obtained. A comparison between the existing and the proposed class of estimators of S_y^2 has been made empirically with the help of two population data taken from [1] and [10].

Keywords and phrases: Proposed estimator, Double sampling technique, Bias(B), Mean Square Error (MSE), Optimum estimator.

1. INTRODUCTION

In theory of sample surveys, we are estimating the population variance S_y^2 of study variable y. It is well known that use of auxiliary information improves the precision of proposed estimator. If information on an auxiliary variable is readily available then it is a well-known fact that the ratio-type and regression-type estimators can be used for estimation of parameters of interest, due to increase in efficiency of these estimators. The problem of estimating the population variance of S_y^2 of study variable y received a considerable attention of the statistician in survey sampling including [2], [3], [4], [5-7], [12], [13], [14], [15-16], [20] and [18] and have suggested improved estimators for estimation of S_y^2

Let φ_i (i=1,2,……..,N) be the population having N units such that y is highly correlated with the auxiliary variables x and z. We assume that a simple random sample without replacement (SRSWOR) of size n is drawn from the finite population of size N. Let $(s'^2_x, s'^2_z)$and(s_x^2, s_z^2, s_y^2) be the sample variances defined over n' and n and S_x^2, S_z^2 and S_y^2 be the population variances of variables x, z and y respectively.

Mathematical Journal of Interdisciplinary Sciences
Vol. 3, No. 2,
March 2015
pp. 193–207

To estimate the population variance S_y^2 of study variable y, consider two cases:

- The population variances S_x^2 and S_z^2 of auxiliary variable x and z are known.
- When both S_x^2 and S_z^2 are unknown.

Where, $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2, S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2, S_z^2 = \frac{1}{N-1}\sum_{i=1}^{N}(z_i - \bar{Z})^2$.

2. LARGE SAMPLE APPROXIMATIONS

Let, $s_y^2 = (1+\varepsilon_0)S_y^2, s_x^2 = (1+\varepsilon_1)S_x^2, s_x'^2 = (1+\varepsilon_1')S_x^2, s_z^2 = (1+\varepsilon_2)S_z^2$ and $s_Z'^2 = (1+\varepsilon_2')S_z^2$, such that $E(\varepsilon_i) = E(\varepsilon'_i) = 0, (i = 0,1,2)$.

Also to the first order of approximation, we have

$E(\varepsilon_0^2) = \gamma_1\nabla^*_{400}, E(\varepsilon_1^2) = \gamma_1\nabla^*_{040}, E(\varepsilon_2^2) = \gamma_1\nabla^*_{004}\; E(\varepsilon_1\varepsilon_2) = \gamma_1\nabla^*_{022}, E(\varepsilon_0\varepsilon_1) = \gamma_1\nabla^*_{220}$

$E(\varepsilon'^2_0) = \gamma_2\nabla^*_{400}, E(\varepsilon'^2_1) = \gamma_2\nabla^*_{040}, E(\varepsilon'^2_2) = \gamma_2\nabla^*_{004}, E(\varepsilon'_1\varepsilon'_2) = \gamma_2\nabla^*_{022}, E(\varepsilon'_0\varepsilon'_1) = \gamma_2\nabla^*_{220}, E(\varepsilon'_0\varepsilon'_2) = \gamma_2\nabla^*_{202}$ and $E(\varepsilon_0\varepsilon_2) = \gamma_1\nabla^*_{202}$.

Where, $\gamma_1 = \frac{1}{n}, \gamma_2 = \frac{1}{n'}, \gamma = \left(\frac{1}{n} - \frac{1}{n'}\right), \nabla^*_{pqr} = (\nabla_{pqr} - 1)$ and $\nabla_{pqr} = (\mu_{pqr} / \mu_{200}^{p/2}\mu_{020}^{q/2}\mu_{002}^{r/2})$

and $\nabla_{pqr} = (\mu_{pqr} / \mu_{200}^{p/2}\mu_{020}^{q/2}\mu_{002}^{r/2})$ and $\mu_{pqr} = \frac{1}{N}\sum\left(y_i - \bar{Y}\right)^p\left(X_i - \bar{X}\right)^q\left(z_i - \bar{Z}\right)^r$; p,q,r, being non negative integers.

3. ESTIMATION OF S_y^2 WHEN BOTH S_x^2 AND S_z^2 ARE KNOWN

In case when S_x^2 and S_z^2 are known, the conventional unbiased estimator is given as

The conventional unbiased estimator, $\hat{S}_y^2 = s_y^2$

Usual chain- ratio type estimator, respectively defined by, $t_R = s_y^2\left(\frac{S_x^2}{s_x^2}\right)\left(\frac{S_z^2}{s_z^2}\right)$

The expressions of the variances and mean square error (MSE) of $\hat{S}_y^2$ and t_R, up to first order of approximation, are given by

$$\mathrm{var}(S_y^2) = \gamma_1 S_y^4\nabla^*_{400} \tag{1}$$

$$MSE(t_R) = \gamma_1 S_y^4\left[\nabla^*_{400} + \nabla^*_{040} + \nabla^*_{004} - 2\nabla^*_{220} - 2\nabla^*_{202} + 2\nabla^*_{022}\right] \tag{2}$$

Motivated by [8], we proposed a generalised estimator for estimating the population variance S_y^2, as

$$T_{pr} = s_y^2\left[k_1\left\{\frac{S_x^2}{s_x^2}\right\}^{g_1}\exp\left\{\frac{a(s_x^2-S_x^2)}{S_x^2+b(s_x^2-S_x^2)}\right\}+k_2\left\{\frac{S_z^2}{s_z^2}\right\}^{g_2}\exp\left\{\frac{c(s_z^2-S_z^2)}{S_z^2+d(s_z^2-S_z^2)}\right\}\right] \quad (3)$$

where, k_1,k_2 are weights, g_1 and g_2 are constants a, b, c and d are either real numbers or the functions of known parameters.

Note: Here we will use the notations $T_{p(1)}$ and $T_{p(2)}$, rather than T_p for two different cases (ie: $k_1+k_2=1$ and $k_1+k_2\neq 1$**).**

Bias and MSE of the suggested method are derived under two different conditions:

- **Case 1:** When $k_1+k_2=1$, where k_1 and k_2 are weights.

Theorem 2.1: Estimator $T_{pr(1)}$ {equation (3)}in terms of $\varepsilon_i; i=0,1$ could be expressed as:

$$T_{pr(1)} = S_y^2\left[1+\varepsilon_0+k_1\left\{\varepsilon_1\eta_1+\varepsilon_1^2\eta_2+\varepsilon_0\varepsilon_1\eta_1\right\}+k_2\left\{\varepsilon_2\eta_3+\varepsilon_2^2\eta_4+\varepsilon_0\varepsilon_2\eta_3\right\}\right] \quad (4)$$

Where, $\eta_1=(a-g_1), \eta_2=\left\{\frac{a^2}{2}-ab-ag_1+\frac{g_1(g_1+1)}{2}\right\}, \eta_3=(c-g_2)$ and $\eta_4=\left\{\frac{c^2}{2}-cd-cg_2+\frac{g_2(g_2+1)}{2}\right\}$.

Proof:

$$T_{pr(1)} = s_y^2\left[k_1\left\{\frac{S_x^2}{s_x^2}\right\}^{g_1}\exp\left\{\frac{a(s_x^2-S_x^2)}{S_x^2+b(s_x^2-S_x^2)}\right\}+k_2\left\{\frac{S_z^2}{s_z^2}\right\}^{g_2}\exp\left\{\frac{c(s_z^2-S_z^2)}{S_z^2+d(s_z^2-S_z^2)}\right\}\right] \quad (5)$$

$$= (1+\varepsilon_0)S_y^2\left[k_1(1+\varepsilon_1)^{-g_1}\exp\left\{a\varepsilon_1(1+b\varepsilon_1)^{-1}\right\}+k_2(1+\varepsilon_2)^{-g_2}\exp\left\{c\varepsilon_2(1+d\varepsilon_2)^{-1}\right\}\right]$$

Here we assume $|\varepsilon_1|,|b\varepsilon_1|,|\varepsilon_2|$ and $|d\varepsilon_2| < 1$, so that the terms $(1+\varepsilon_1)^{-g_1},(1+b\varepsilon_1)^{-1},(1+\varepsilon_2)^{-g_2}$ and $(1+d\varepsilon_2)^{-1}$ are expandable. By expanding the right hand side of equation (5) and neglecting the terms of ε_i'shaving power greater than two, we have

$$T_{pr(1)} = S_y^2\left[k_1(1+\varepsilon_0)\left\{1+\varepsilon_1\eta_1+\varepsilon_1^2\eta_2\right\}+k_2(1+\varepsilon_0)\left\{1+\varepsilon_2\eta_3+\varepsilon_2^2\eta_4\right\}\right]$$
$$T_{pr(1)} = S_y^2\left[1+\varepsilon_0+k_1\left\{\varepsilon_1\eta_1+\varepsilon_1^2\eta_2+\varepsilon_0\varepsilon_1\eta_1\right\}+k_2\left\{\varepsilon_2\eta_3+\varepsilon_2^2\eta_4+\varepsilon_0\varepsilon_2\eta_3\right\}\right] \quad (6)$$

Theorem 2.2: Bias of $T_{p(1)}$ is given as

$$B[T_{pr(1)}] = \frac{S_y^2}{n}\left[k_1\left\{\eta_2\nabla_{040}^* + \eta_1\nabla_{220}^*\right\} + k_2\left\{\eta_4\nabla_{004}^* + \eta_3\nabla_{202}^*\right\}\right] \tag{7}$$

Proof: $B(T_{pr(1)}) = E[T_{pr(1)} - S_y^2]$

$$= S_y^2 E\left[\varepsilon_0 + k_1\left\{\varepsilon_1\eta_1 + \varepsilon_1^2\eta_2 + \varepsilon_0\varepsilon_1\eta_1\right\} + k_2\left\{\varepsilon_2\eta_3 + \varepsilon_2^2\eta_4 + \varepsilon_0\varepsilon_2\eta_3\right\}\right]$$

$$B(T_{pr(1)}) = \frac{S_y^2}{n}\left[k_1\left\{\eta_2\nabla_{040}^* + \eta_1\nabla_{220}^*\right\} + k_2\left\{\eta_3\nabla_{004}^* + \eta_3\nabla_{202}^*\right\}\right]$$

Theorem 2.3: Mean square error of $T_{p(1)}$, up to the first order of approximation : $O\left(\frac{1}{n}\right)$ is

$$MSE[T_{pr(1)}] = \frac{S_y^4}{n}\left[\nabla_{400}^* + k_1^2 A_1 + k_2^2 B_1 + 2k_1 C_1 + 2k_2 D_1 + 2k_1 k_2 E_1\right] \tag{8}$$

Where, $A_1 = \eta_1^2\nabla_{040}^*, B_1 = \eta_3^2\nabla_{004}^*, C_1 = \eta_1\nabla_{220}^*, D_1 = \eta_3\nabla_{202}^*$ and $E_1 = \eta_1\eta_3\nabla_{022}^*$

Proof:

$$MSE[T_{pr(1)}] = E[T_{pr(1)} - S_y^2]^2$$

$$= S_y^4 E\left[\varepsilon_0^2 + k_1^2\eta_1^2\varepsilon_1^2 + k_2^2\eta_3^2\varepsilon_2^2 + 2k_1\eta_1\varepsilon_0\varepsilon_1 + 2k_2\eta_3\varepsilon_0\varepsilon_2 + 2k_1k_2\eta_1\eta_3\varepsilon_1\varepsilon_2\right]$$

$$MSE[t_{pr(1)}] = \frac{S_y^4}{n}\left[\nabla_{400}^* + k_1^2 A_1 + (1 + k_1^2 - 2k_1)B_1 + 2k_1 C_1 + 2(1 - k_1)D_1 + 2k_1(1 - k_1)E_1\right] \tag{9}$$

Differentiating equation (9) w.r.t k_1 and equating it to zero, we get the optimum value of k_1 as $k_1^* = \frac{(B_1 + D_1 - C_1 - E_1)}{(A_1 + B_1 - 2E_1)}$

Putting the optimum value of $k_1$in equation (9), we get the minimum MSE of the suggested method $T_{p(1)}$.

- **Case 2:** When $k_1 + k_2 \neq 1$, where $k_1$and k_2 are weights.

Theorem 2.4: The estimator $\mathbf{T_{p(2)}}$ (equation.3) in terms of ε_i's, could be expressed as:

$$T_{pr(2)} = S_y^2\left[k_1\left\{1 + \varepsilon_0 + \varepsilon_1\eta_1 + \varepsilon_1^2\eta_2 + \varepsilon_0\varepsilon_1\eta_1\right\} + k_2\left\{1 + \varepsilon_0 + \varepsilon_2\eta_3 + \varepsilon_2^2\eta_4 + \varepsilon_0\varepsilon_2\eta_3\right\}\right]$$

Where, $\eta_1=(a-g_1),\eta_2=\left\{\frac{a^2}{2}-ab-ag_1+\frac{g_1(g_1+1)}{2}\right\},\eta_3=(c-g_2)$ and $\eta_4=\left\{\frac{c^2}{2}-cd-cg_2+\frac{g_2(g_2+1)}{2}\right\}$.

Proof: $T_{pr(2)}=s_y^2\left[k_1\left\{\frac{S_x^2}{s_x^2}\right\}^{g_1}\exp\left\{\frac{a(s_x^2-S_x^2)}{S_x^2+b(s_x^2-S_x^2)}\right\}+k_2\left\{\frac{S_z^2}{s_z^2}\right\}^{g_2}\exp\left\{\frac{c(s_z^2-S_z^2)}{S_z^2+d(s_z^2-S_z^2)}\right\}\right]$

$$=(1+\varepsilon_0)S_y^2\left[k_1(1+\varepsilon_1)^{-g_1}\exp\left\{a\varepsilon_1(1+b\varepsilon_1)^{-1}\right\}+k_2(1+\varepsilon_2)^{-g_2}\exp\left\{c\varepsilon_2(1+d\varepsilon_2)^{-1}\right\}\right] \quad (10)$$

Here we assume $|\varepsilon_1|,|b\varepsilon_1|,|\varepsilon_2|$ and $|d\varepsilon_2|<1$, so that the terms $(1+\varepsilon_1)^{-g_1},(1+b\varepsilon_1)^{-1},(1+\varepsilon_2)^{-g_2}$ and $(1+d\varepsilon_2)^{-1}$ are expandable. By expanding the right hand side of equation (10) and neglecting the terms of $\varepsilon_i's$ having power greater than two, we have

$$T_{pr(1)}=S_y^2\Big[k_1\left\{1+\varepsilon_0+\varepsilon_1\eta_1+\varepsilon_1^2\eta_2+\varepsilon_0\varepsilon_1\eta_1\right\} +k_2\left\{1+\varepsilon_0+\varepsilon_2\eta_3+\varepsilon_2^2\eta_4+\varepsilon_0\varepsilon_2\eta_3\right\}\Big] \quad (11)$$

Theorem 2.5: Bias of $T_{pr(2)}$ is given as:

$$B[T_{pr(2)}]=\frac{S_y^2}{n}\Big[k_1\left\{1+\eta_2\nabla^*_{040}+\eta_1\nabla^*_{220}\right\} +k_2\left\{1+\eta_4\nabla^*_{004}+\eta_3\nabla^*_{202}\right\}-1\Big] \quad (12)$$

Proof: $B[T_{pr(2)}]=E[T_{pr(2)}-S_y^2]$

$$=S_y^2E\left[k_1\left\{1+\varepsilon_0+\varepsilon_1\eta_1+\varepsilon_1^2\eta_2+\varepsilon_0\varepsilon_1\eta_1\right\}+k_2\left\{1+\varepsilon_0+\varepsilon_2\eta_3+\varepsilon_2^2\eta_4+\varepsilon_0\varepsilon_2\eta_3\right\}-1\right]$$

$$B[T_{pr(2)}]=\frac{S_y^2}{n}\left[k_1\left\{1+\eta_2\nabla^*_{040}+\eta_1\nabla^*_{220}\right\}+k_2\left\{1+\eta_4\nabla^*_{004}+\eta_3\nabla^*_{202}\right\}-1\right]$$

Theorem 2.6: Mean square error of $T_{pr(2)}$, up to the first order of approximation is:

$$MSE[T_{pr(2)}] = S_y^4\left[1 + k_1^2 A_2 + k_2^2 B_2 - 2k_1 C_2 - 2k_2 D_2 + 2k_1 k_2 E_2\right] \tag{13}$$

Where,

$$A_2 = 1 + \frac{1}{n}\left[\nabla^*_{400} + (\eta_1^2 + 2\eta_2)\nabla^*_{040} + 4\eta_1\nabla^*_{220}\right], B_2 = 1 + \frac{1}{n}\left[\nabla^*_{400} + (\eta_3^2 + 2\eta_4)\nabla^*_{004} + 4\eta_3\nabla^*_{202}\right]$$

$$C_2 = 1 + \frac{1}{n}\left[\eta_2\nabla^*_{040} + \eta_1\nabla^*_{220}\right], D_2 = 1 + \frac{1}{n}\left[\eta_4\nabla^*_{004} + \eta_3\nabla^*_{202}\right]$$

$$E_2 = 1 + \frac{1}{n}\left[\nabla^*_{400} + \eta_2\nabla^*_{040} + \eta_4\nabla^*_{004} + 2\eta_1\nabla^*_{220} + 2\eta_3\nabla^*_{202} + \eta_1\eta_3\nabla^*_{022}\right]$$

Proof: From equation (10), we have

$$[T_{pr(2)} - S_y^2]^2 = S_y^4\left[k_1\left\{1 + \varepsilon_0 + \varepsilon_1\eta_1 + \varepsilon_1^2\eta_2 + \varepsilon_0\varepsilon_1\eta_1\right\} + k_2\left\{1 + \varepsilon_0 + \varepsilon_2\eta_3 + \varepsilon_2^2\eta_4 + \varepsilon_0\varepsilon_2\eta_3\right\} - 1\right]^2$$

After expanding the above equation up to the first order of approximation ie: $O\left(\frac{1}{n}\right)$ and then taking expectations of both sides, we get

$$MSE[T_{pr(2)}] = S_y^4\left[1 + k_1^2 A_2 + k_2^2 B_2 - 2k_1 C_2 - 2k_2 D_2 + 2k_1 k_2 E_2\right]$$

Minimising equation (13) with respect to k_1 and k_2, we get the optimum values k_1 and k_2, respectively given as

$$k_1^* = \frac{B_2 C_2 - D_2 E_2}{A_2 B_2 - E_2{}^2} \text{ and } k_2^* = \frac{A_2 D_2 - C_2 E_2}{A_2 B_2 - E_2{}^2}$$

Putting this value of k_1 and k_2 in equation (13), we get the minimum MSE of the suggested estimator $T_{pr(2)}$.

4.ESTIMATION OF S_y^2 WHEN BOTH S_x^2 AND S_z^2 ARE UNKNOWN

Consider a finite population with N (< ∞) identifiable units. Let y be the variable under study taking values $y_i\,(i = 1, 2, ...N)$ for i[th] unit of the population. To estimate the population variance S_y^2 of y in the presence of two auxiliary variables, when the population variances of x and z are not known and SRSWOR scheme is used in selecting samples for both the phases. The following two phase sampling scheme may be recommended.

1) The first phase sample s' of size (n'< N) is drawn in order to observed x and z.
2) The second phase sample s of size n (n<n') is drawn in order to observe y, x and z.

Let $s_x'^2$ and $s_z'^2$ be the unbiased estimators of S_x^2 and S_z^2 respectively, based on first phase sample s'; $s_y'^2, s_x'^2$ and $s_z'^2$ be the unbiased estimators S_y^2, S_x^2 and S_x^2, respectively, based on second phase sample's'. When S_x^2 and S_z^2 unknown, then usual unbiased and ordinary ratio estimator are given as

Usual unbiased estimator $\hat{S}_y^2 = s_y^2$

Usual chain ratio-type estimator in two phase sampling, is defined as

$$t_R^{'} = s_y^2 \left(\frac{s'^2_x}{s_x^2} \right) \left(\frac{s'^2_z}{s_z^2} \right)$$

MSE's expressions for the estimators $\hat{S}_y^2$ and $t_R^{'}$, up to the first order of approximation are ,respectively, given by

$$MSE(S_y^2) = S_y^4 \gamma_1 \nabla^*_{400} \tag{14}$$

$$MSE(t_R^{'}) = S_y^4 \left[\nabla^*_{400}\gamma_1 + \gamma \nabla^*_{040} + \gamma \nabla^*_{004} - 2\gamma \nabla^*_{220} - 2\gamma \nabla^*_{202} + 2\gamma \nabla^*_{022} \right] \tag{15}$$

when S_x^2 and S_z^2 are not known a priori, then under double sampling technique, suggested estimator T_p takes the following form:

$$T_{pr}^{'} = s_y^2 \left[k_1^{'} \left\{ \frac{s'^2_x}{s_x^2} \right\}^{g_1} \exp\left\{ \frac{a(s_x^2 - s'^2_x)}{s'^2_x + b(s_x^2 - s'^2_x)} \right\} + k_2^{'} \left\{ \frac{s'^2_z}{s_z^2} \right\}^{g_2} \exp\left\{ \frac{c(s_z^2 - s'^2_z)}{s'^2_z + d(s_z^2 - s'^2_z)} \right\} \right] \tag{16}$$

where, k_1,$k_2$are weights, g_1 and g_2 are constants and a, b, c and d are either real numbers or the functions of known parameters.

Note: Here we have taken the notation $T'_{pr(1)}$ and $T'_{pr(2)}$, rather than T'_{pr} for two different cases (ie: $k_1 + k_2 = 1$ and $k_1 + k_2 \neq 1$ **).**

To obtain the bias and mean square error of the suggested estimator we have the following two conditions given as -

Case 1: ***When $k_1^{'} + k_2^{'} = 1$, where $k_1^{'}$ and $k_2^{'}$ are defined weights.***

Theorem 3.1: The suggested estimator $T_{pr(1)}^{'}$ (say equation 16) in terms of $\varepsilon_i^{'}s$, could be expressed as:

$$T_{pr(1)}^{'} = S_y^2 \begin{bmatrix} 1 + \varepsilon_0 + k_1^{'} \left\{ (a - g_1)(\varepsilon_1 - \varepsilon_1^{'}) + \varepsilon_1^2 Q_1 + \varepsilon'^2_1 Q_2 + \varepsilon_1 \varepsilon_1^{'} Q_3 + (a - g_1)(\varepsilon_0 \varepsilon_1 - \varepsilon_0 \varepsilon_1^{'}) \right\} \\ + k_2 \left\{ (c - g_2)(\varepsilon_2 - \varepsilon_2^{'}) + \varepsilon_2^2 Q_4 + \varepsilon'^2_2 Q_5 + \varepsilon_2 \varepsilon_2^{'} Q_6 + (c - g_2)(\varepsilon_0 \varepsilon_2 - \varepsilon_0 \varepsilon_2^{'}) \right\} \end{bmatrix}$$

where,

$$Q_1 = \frac{a^2}{2} - ab - g_1 a + \frac{g_1(g_1+1)}{2}, Q_2 = \frac{a^2}{2} + a - ab - ag_1 + \frac{g_1(g_1-1)}{2}$$

$$Q_3 = 2ab - a - a^2 + 2ag_1 - g_1^2, Q_4 = \frac{c^2}{2} - cd - g_2c + \frac{g_2(g_2+1)}{2}$$

$$Q_5 = c + \frac{c^2}{2} - cd - g_2c + \frac{g_2(g_2-1)}{2}, Q_6 = 2cd - c - c^2 + 2g_2c - g_2^2$$

Proof:

$$T'_{pr(1)} = s_y^2\left[k_1'\left\{\frac{s_x'^2}{s_x^2}\right\}^{g_1}\exp\left\{\frac{a(s_x^2 - s_x'^2)}{s_x'^2 + b(s_x^2 - s_x'^2)}\right\} + k_2'\left\{\frac{s_z'^2}{s_z^2}\right\}^{g_2}\exp\left\{\frac{c(s_z^2 - s_z'^2)}{s_z'^2 + d(s_z^2 - s_z'^2)}\right\}\right]$$

$$= s_y^2\left[\begin{array}{r} k_1'(1+\varepsilon_1')^{g_1}(1+\varepsilon_1)^{-g_1}\exp\left\{(a\varepsilon_1 - a\varepsilon_1')(1+\varepsilon_1' + b\varepsilon_1 - b\varepsilon_1')^{-1}\right\} + k_2'(1+\varepsilon_2')^{g_2}(1+\varepsilon_2)^{-g_2} \\ \exp\left\{(c\varepsilon_2 - c\varepsilon_2')(1+\varepsilon_2' + d\varepsilon_2 - d\varepsilon_2')^{-1}\right\} \end{array}\right]$$

After expanding and arraigning the above equation up to the first order of approximation, we have:

$$T'_{pr(1)} = S_y^2\left[\begin{array}{l} 1 + \varepsilon_0 + k_1'\left\{(a-g_1)(\varepsilon_1 - \varepsilon_1') + \varepsilon_1^2 Q_1 + \varepsilon_1'^2 Q_2 + \varepsilon_1\varepsilon_1' Q_3 + (a-g_1)(\varepsilon_0\varepsilon_1 - \varepsilon_0\varepsilon_1')\right\} \\ + k_2\left\{(c-g_2)(\varepsilon_2 - \varepsilon_2') + \varepsilon_2^2 Q_4 + \varepsilon_2'^2 Q_5 + \varepsilon_2\varepsilon_2' Q_6 + (c-g_2)(\varepsilon_0\varepsilon_2 - \varepsilon_0\varepsilon_2')\right\} \end{array}\right] \quad (17)$$

where, coefficient term of ε_i's are already defined.

Theorem 3.2: Bias of $T'_{pr(1)}$ is given as

$$B(T'_{pr(1)}) = S_y^2\left[k_1'\left\{Q_1\gamma_1\nabla^*_{040} + \gamma_2\nabla^*_{040}(Q_2 + Q_3) + (a-g_1)\gamma\nabla^*_{220}\right\} + k_2'\left\{Q_4\gamma_1\nabla^*_{040} + \gamma_2\nabla^*_{040}(Q_5 + Q_6) + (c-g_2)\gamma\nabla^*_{220}\right\}\right]$$

Proof: Subtracting S_y^2 from both sides and taking expectations of equation (17), we have

$$B(T'_{pr(1)}) = S_y^2\left[\begin{array}{l} k_1'\left\{Q_1\gamma_1\nabla^*_{040} + \gamma_2\nabla^*_{040}(Q_2 + Q_3) + (a-g_1)\gamma\nabla^*_{220}\right\} \\ + k_2'\left\{Q_4\gamma_1\nabla^*_{040} + \gamma_2\nabla^*_{040}(Q_5 + Q_6) + (c-g_2)\gamma\nabla^*_{220}\right\} \end{array}\right] \quad (18)$$

Theorem 3.3: Mean square error of $T'_{pr(1)}$, up to the first order of approximation is given as:

$$MSE(T'_{pr(1)}) = S_y^4\left[\frac{\nabla^*_{400}}{n} + k_1'^2 A + k_2{}^2 B + 2k_1'C + 2k_2'D + 2k_1'k_2'E\right]$$

Or

$$MSE(T'_{pr(1)}) = S_y^4\left[\frac{\nabla^*_{400}}{n} + k_1'^2 A_3 + (1-k_1')^2 B_3 + 2k_1'C_3 + 2(1-k_1')D_3 + 2k_1'(1-k_1')E_3\right]$$

where,

$A_3 = \eta_1^2 \gamma \nabla^*_{040}, B_3 = \eta_3^2 \gamma \nabla^*_{004}, C_3 = \eta_1 \gamma \nabla^*_{220}, D_3 = \eta_3 \gamma \nabla^*_{202}, E_3 = \eta_1 \eta_3 \gamma \nabla^*_{022}$

such that $\eta_1 = (a - g_1)$ and $\eta_3 = (c - g_2)$

Proof: from equation (18), we have

$$(T'_{pr(1)} - S_y^2)^2 = S_y^4 \left[\varepsilon_o + k_1' \eta_1 (\varepsilon_1 - \varepsilon_1') + k_2' \eta_3 (\varepsilon_2 - \varepsilon_2')\right]^2$$

After squaring and taking expectations of both sides of the above equation, up to the first order of approximation, we get

$$MSE(T'_{pr(1)}) = S_y^4 \left[\frac{\nabla^*_{400}}{n} + k_1'^2 A_3 + (1 - k_1')^2 B_3 + 2k_1' C_3 + 2(1 - k_1') D_3 + 2k_1'(1 - k_1') E_3\right] \quad (19)$$

Differentiating equation (19) partially w.r.t. to k_1', we get the optimum value of k_1' for minimum MSE as:

$$k_1'^* = \left(\frac{B_3 + D_3 - C_3 - E_3}{A_3 + B_3 - 2E_3}\right)$$

Putting the optimum value of $k_1'^*$ in (19), we get the minimum MSE of the suggested estimator $T'_{pr(1)}$.

Case.2: ***When*** $k_1' + k_2' \neq 1$***, where*** k_1' ***and*** k_2' ***are defined weights.***

$$T'_{pr(2)} = s_y^2 \left[k_1' \left\{\frac{s_x'^2}{s_x^2}\right\}^{g_1} \exp\left\{\frac{a(s_x^2 - s_x'^2)}{s_x'^2 + b(s_x^2 - s_x'^2)}\right\} + k_2' \left\{\frac{s_z'^2}{s_z^2}\right\}^{g_2} \exp\left\{\frac{c(s_z^2 - s_z'^2)}{s_z'^2 + d(s_z^2 - s_z'^2)}\right\}\right] \quad (20)$$

Theorem 3.4: The suggested estimator $T'_{pr(2)}$ in terms of ε_i 's , is written as

$$T'_{pr(2)} = S_y^2 \Big[k_1' \left\{1 + \varepsilon_0 + \eta_1(\varepsilon_1 - \varepsilon_1') + \varepsilon_1^2 Q_1 + \varepsilon_1'^2 Q_2 + \varepsilon_1 \varepsilon_1' Q_3 + \eta_1(\varepsilon_0 \varepsilon_1 - \varepsilon_0 \varepsilon_1')\right\}$$
$$+ k_2' \left\{1 + \varepsilon_0 + \eta_3(\varepsilon_2 - \varepsilon_2') + \varepsilon_2^2 Q_4 + \varepsilon_2'^2 Q_5 + \varepsilon_2 \varepsilon_2' Q_6 + \eta_3(\varepsilon_0 \varepsilon_2 - \varepsilon_0 \varepsilon_2')\right\}\Big]$$

where, $\eta_1, \eta_3, Q_1, Q_2, Q_3, Q_4, Q_5$ and Q_6 are defined in theorem 3.1 and 3.2.

Proof:

$$T'_{pr(2)} = s_y^2 \Big[k_1'(1 + \varepsilon_1')^{g_1}(1 + \varepsilon_1)^{-g_1} \exp\left\{(a\varepsilon_1 - a\varepsilon_1')(1 + \varepsilon_1' + b\varepsilon_1 - b\varepsilon_1')^{-1}\right\}$$
$$+ k_2'(1 + \varepsilon_2')^{g_2}(1 + \varepsilon_2)^{-g_2} \exp\left\{(c\varepsilon_2 - c\varepsilon_2')(1 + \varepsilon_2' + d\varepsilon_2 - d\varepsilon_2')^{-1}\right\} \quad (21)$$

After expanding and arraigning the above equation (21) up to the first order of approximation, we have

$$T'_{pr(2)} = S_y^2\left[k_1'\left\{1+\varepsilon_0+\eta_1(\varepsilon_1-\varepsilon_1')+\varepsilon_1^2Q_1+\varepsilon_1'^2Q_2+\varepsilon_1\varepsilon_1'Q_3+\eta_1(\varepsilon_0\varepsilon_1-\varepsilon_0\varepsilon_1')\right\}\right.$$
$$\left.+k_2'\left\{1+\varepsilon_0+\eta_3(\varepsilon_2-\varepsilon_2')+\varepsilon_2^2Q_4+\varepsilon_2'^2Q_5+\varepsilon_2\varepsilon_2'Q_6+\eta_3(\varepsilon_0\varepsilon_2-\varepsilon_0\varepsilon_2')\right\}\right] \quad (22)$$

where, coefficient term of ε_i's are already defined.

Theorem 3.5: Bias of suggested method $T'_{pr(1)}$ is given as

$$\mathbf{B}(T'_{pr(2)}) = S_y^2\left[\begin{array}{l} k_1'\left\{1+Q_1\gamma_1\nabla^*_{040}+\gamma_2\nabla^*_{040}Q_2+\gamma_2\nabla^*_{040}Q_3+\eta_1\gamma\nabla^*_{220}\right\} \\ \quad +k_2'\left\{1+Q_4\gamma_1\nabla^*_{004}+\gamma_2\nabla^*_{004}Q_5+\gamma_2\nabla^*_{004}Q_6+\eta_3\gamma\nabla^*_{202}\right\}\end{array}\right]$$

Proof: Subtracting S_y^2 both side of equation (22), we have

$$T'_{pr(2)} - S_y^2$$
$$= S_y^2\left[\begin{array}{l} k_1'\left\{1+\varepsilon_0+\eta_1(\varepsilon_1-\varepsilon_1')+\varepsilon_1^2Q_1+\varepsilon_1'^2Q_2+\varepsilon_1\varepsilon_1'Q_3+\eta_1(\varepsilon_0\varepsilon_1-\varepsilon_0\varepsilon_1')\right\} \\ +k_2'\left\{1+\varepsilon_0+\eta_3(\varepsilon_2-\varepsilon_2')+\varepsilon_2^2Q_4+\varepsilon_2'^2Q_5+\varepsilon_2\varepsilon_2'Q_6+\eta_3(\varepsilon_0\varepsilon_2-\varepsilon_0\varepsilon_2')\right\}-1\end{array}\right] \quad (23)$$

Taking expectations of both sides, we get the required bias of the suggested method as

$$\mathbf{B}(T'_{pr(2)}) = S_y^2\left[\begin{array}{l} k_1'\left\{1+Q_1\gamma_1\nabla^*_{040}+\gamma_2\nabla^*_{040}Q_2+\gamma_2\nabla^*_{040}Q_3+\eta_1\gamma\nabla^*_{220}\right\} \\ \quad +k_2'\left\{1+Q_4\gamma_1\nabla^*_{004}+\gamma_2\nabla^*_{004}Q_5+\gamma_2\nabla^*_{004}Q_6+\eta_3\gamma\nabla^*_{202}\right\}\end{array}\right]$$

Theorem 3.6: Mean square error of the suggested estimator $T'_{pr(2)}$, is given by

$$MSE(T'_{pr(2)}) = S_y^4\left[1+k_1'^2A+k_2'^2B-2k_1'C-2k_2'D+2k_1'k_2'E\right]$$

where

$$A = \left[1+\frac{\nabla^*_{400}}{n}+\nabla^*_{040}\left\{\eta_1^2\gamma+\frac{2Q_1}{n}+\frac{2Q_2}{n'}+\frac{2Q_3}{n'}\right\}+4\eta_1\gamma\nabla^*_{220}\right],$$

$$B = \left[1+\frac{\nabla^*_{400}}{n}+\nabla^*_{004}\left\{\eta_3^2\gamma+\frac{2Q_4}{n}+\frac{2Q_5}{n'}+\frac{2Q_6}{n'}\right\}+4\eta_3\gamma\nabla^*_{202}\right],$$

$$C = \left[1+\nabla^*_{040}\left\{\frac{Q_1}{n}+\frac{Q_2}{n'}+\frac{Q_3}{n'}\right\}+\eta_1\gamma\nabla^*_{220}\right], D = \left[1+\nabla^*_{004}\left\{\frac{Q_4}{n}+\frac{Q_5}{n'}+\frac{Q_6}{n'}\right\}+\eta_3\gamma\nabla^*_{202}\right]$$

$$E = \left[\begin{array}{l} 1+\frac{\nabla^*_{400}}{n}+\nabla^*_{040}\left\{\eta_1^2\gamma+\frac{Q_1}{n}+\frac{Q_2}{n'}+\frac{Q_3}{n'}\right\}+\nabla^*_{004}\left\{\eta_3^2\gamma+\frac{Q_4}{n}+\frac{Q_5}{n'}+\frac{Q_6}{n'}\right\} \\ \qquad +2\eta_3\gamma\nabla^*_{202}+2\eta_1\gamma\nabla^*_{220}+\eta_1\eta_3\gamma\nabla^*_{022}\end{array}\right]$$

Proof: we have

$$T'_{pr(2)} - S_y^2 = S_y^2 \begin{bmatrix} k_1' \left\{1 + \varepsilon_0 + \eta_1(\varepsilon_1 - \varepsilon_1') + \varepsilon_1^2 Q_1 + \varepsilon_1'^2 Q_2 + \varepsilon_1 \varepsilon_1' Q_3 + \eta_1(\varepsilon_0 \varepsilon_1 - \varepsilon_0 \varepsilon_1')\right\} \\ + k_2' \left\{1 + \varepsilon_0 + \eta_3(\varepsilon_2 - \varepsilon_2') + \varepsilon_2^2 Q_4 + \varepsilon_2'^2 Q_5 + \varepsilon_2 \varepsilon_2' Q_6 + \eta_3(\varepsilon_0 \varepsilon_2 - \varepsilon_0 \varepsilon_2')\right\} - 1 \end{bmatrix}$$

After squaring and taking expectations 6f both sides of the above equation up to the first order of approximation, MSE of the estimator

$$MSE(T'_{pr(2)}) = S_y^4 \left[1 + k_1'^2 A + k_2'^2 B - 2k_1' C - 2k_2' D + 2k_1' k_2' E\right] \qquad (24)$$

The optimum value of k_1' and k_2' is obtained by partially differentiating equation (24) w.r.t. k_1' and k_2' for minimum MSE, as

$$k^{*'}_1 = \left(\frac{BC - DE}{AB - E^2}\right) \text{and } k^{*'}_2 = \left(\frac{AD - CE}{AB - E^2}\right)$$

Putting the optimum values of k_1^{*} and k_2^{*} in (19), we get the minimum $MSE(T'_{pr(2)})$.

6. EMPIRICAL STUDY

Population 1: For empirical study we have taken the data of [10], which contain the data of 34 villages. The variables are:

y- Area under wheat in 1964.

x- Area under wheat in 1963.

z- Cultivated area in 1961.

Also,

$$\nabla_{400} = 3.726, \nabla_{040} = 2.912, \nabla_{004} = 2.808, \nabla_{220} = 3.105, \nabla_{202} = 2.979, \nabla_{022} = 2.738$$

$$S_y^2 = 22564.56, S_x^2 = 197095.3, S_z^2 = 2652.05, S_{yx} = 60304.01, S_{yz} = 22158.05$$

n = 7, $n' = 15$ N = 34

Population 2: The data for the empirical study are taken from [1]. The population consists of 340 villages. The variables are:

y- Number of literate persons.

x- Number of household.

z- Total population in the village.

To estimate the variance of y, we have used x and z as the prior information. From the data set, we have

Table 1: Percentage Relative Efficiency (PRE) for population-1.

Estimator	a	b	c	d	g_1	g_2	PRE
S_y^2							100.00
t_R							135.2941
$T_{pr(1)}$	1	1	-1	1	1	1	658.7851
	0	1	-1	1	1	1	927.2479
	0	1	1	1	1	1	897.380
$T_{pr(2)}$	1	1	-1	1	1	1	4417.39
	1	1	-1	1	1	0	5256.36
	0	1	-1	1	1	1	1365.2191
$T'_{pr(1)}$	1	1	1	1	1	-1	182.6071
	-1	1	-1	1	1	-1	190.078
	-1	1	-1	1	1	0	186.5093
$T'_{pr(2)}$	1	-1	-1	1	-1	1	404.3805
	1	0	-1	1	-1	1	347.8099
	1	-1	-1	1	0	1	342.1697

Note: Here the notations $T_{pr(1)}, T_{pr(2)}$ stands for suggested estimators in single phase and respectively $T'_{pr(1)}, T'_{pr(2)}$ is for two phase sampling.

$$N = 340, n' = 120, n = 50, S_y^2 = 71379.47, S_x^2 = 11838.85, S_z^2 = 691820.23$$

$$\nabla^*_{400} = 9.90334289, \nabla^*_{040} = 7.05448224, \nabla^*_{004} = 8.2552346, \nabla^*_{220} = 6.31398563$$

$$\nabla^*_{202} = 8.12904924, \nabla^*_{022} = 6.13646859$$

The percentage relative efficiency (PRE) of estimator is defined as

$$PRE(*) = \frac{VAR(S_y^2)}{MSE(*)} \times 100$$

Table 2: Percentage Relative Efficiency (PRE) for Population-2.

Estimator	a	b	c	d	g_1	g_2	PRE
S_y^2							100.00
t_R							115.1561
$T_{pr(1)}$	1	1	-1	1	1	1	521.62661
	0	1	-1	1	1	1	342.12
	0	1	- 1	1	0	1	409.4370
$T_{pr(2)}$	1	1	-1	1	1	1	868.2550
	1	0	-1	1	1	1	2494.1349
	1	1	1	0	1	-1	709.3792
$T'_{pr(1)}$	1	1	-1	1	1	1	189.21
	01	-1	1	1	1		170.3568
	1	1	-1	1	-1	1	169.3268
$T'_{pr(2)}$	1	-1	-1	1	1	1	315.53
	1	-1	-1	1	0	1	342.16
	1	-1	-1	1	-1	1	404.38

Note: Here the notations $T_{pr(1)}, T_{pr(2)}$ stands for suggested estimators in single phase and $T'_{pr(1)}, T'_{pr(2)}$ stands for suggested estimators in two phase sampling.

CONCLUSION

From the above Table 1 and Table 2, we observed that, the suggested estimators $T_{pr(1)}$, $T_{pr(2)}$ (in single phase sampling) and $T'_{pr(1)}, T'_{pr(2)}$ (in two phase sampling) perform much better than the usual unbiased and ratio estimator.

REFERENCES

[1] Ahmed M.S.: Some Estimation Procedure Using Multivariate Auxiliary Information in Sample Surveys. Ph.D thesis, Department of Statistics & Operations Research, Aligarh Muslim University, Aligarh-202002, India (1995).

[2] Grover, L. K .: A correction note on improvement in variance estimation using auxiliary information. Commun. Stat. Theo. Meth. **39**:753–764 (2010). http://dx.doi.org/10.1080/03610920902785786

[3] Isaki, C.T.: Variance estimation using auxiliary information. J. Amer. Statist. Assoc.**78**:117–123 (1983). http://dx.doi.org/10.1080/01621459.1983.10477939

[4] Jhajj, H .S., Sharma, M. K. and Grover, L. K.: An efficient class of chain estimators of population variance under sub-sampling scheme. J. Japan Stat. Soc., **35**(2), 273-286 (2005). http://dx.doi.org/10.14490/jjss.35.273

[5] Kadilar, C., Cingi, H.: Improvement in variance estimation using auxiliary information. Hacettepe J. Math. Statist. **35**(1):111–115 (2006).

[6] Kadilar, C., Cingi, H. : Improvement in variance estimation in simple random sampling. Commun. Statist. Theory. Meth. **36**:2075–2081(2007). http://dx.doi.org/10.1080/03610920601144046

[7] Kadilar, C., Cingi, H.: A new estimator using two auxiliary variables. Applied mathematics and computation **162** 901-908 (2005). http://dx.doi.org/10.1016/j.amc.2003.12.130

[8] Koyuncu, N.: Improved estimation of population mean in stratified random sampling using information on auxiliary attribute. Proceeding of 59-th ISI world Statistics Congress, Hong Kong, China, August, pp: 25-30 (2013).

[9] Lu, J., Yan, Z. ,Ding, C., Hong, Z.: The chain ratio estimator using two auxiliary information. International Conference on Computer and Communication Technologies in Agriculture Engineering 586-589 (2010).

[10] Murthy, M.N.: Sampling Theory and Methods (Statistical Publishing Society, India) (1967).

[11] Shabbir, J. and Gupta, S. : On improvement in variance estimation using auxiliary information. Commun. Statist. Theor. Meth. **36**(12):2177–2185 (2007). http://dx.doi.org/10.1080/03610920701215092

[12] Singh, H.P. and Singh, R.: Improved ratio-type estimator for variance using auxiliary information. J Indian Soc Agric Stat **54**(3):276–287 (2001).

[13] Singh, H. P. and Singh, R. : Estimation of variance through regression approach in two phase sampling. Aligarh Journal of Statistics, **23**, 13-30 (2003).

[14] Singh, H. P. and Solanki, R.S.: A new procedure for variance estimation in simple random sampling using auxiliary information. Stat. Paps. DOI 10.1007/s00362-012-0445-2 (2012). http://dx.doi.org/10.1007/s00362-012-0445-2

[15] Singh, R., Chauhan, P., Sawan, N.: On linear combination of Ratio-product type exponential estimator for estimating finite population mean. Statistics in Transition **9**(1) -115 (2008a).

[16] Singh, R., Chauhan, P., Sawan, N. and Smarandache,F. : Almost unbiased ratio and product type estimator of finite population variance using the knowledge of kurtosis of an auxiliary variable in sample surveys. Octogon Mathematical Journal, Vol. 16, No. **1**, 123-130 (2008b).

[17] Singh, R., Kumar, M. , Smarandache, F. : Almost Unbiased Estimator for Estimating Population Mean Using Known Value of Some Population Parameter(s); Pak.j.stat.oper.res. **76** (2), 63-76 (2008).

[18] Singh, R., Chauhan, P., Sawan, N. and Smarandache, F.: Improved Exponential Estimator for Population Variance Using Two Auxiliary Variables. Italian Jour. Of Pure and Appld. Math., **28**, 103-110(2011).

[19] Singh, V. K., Singh, R. and Florentin, S.: Some improved estimators for population variance using two auxiliary variables in double sampling. on improvement in estimating population parameter(s) using auxiliary information. Educational Publishing & Journal of Matter Regularity (Beijing) (2013).

[20] Singh, V. K., Singh, R.: Predictive estimation of finite population mean using generalised family of estimators , accepted in JTSA(2014).