



Using Rank of the Auxiliary Variable in Estimating Variance of the Stratified Sample Mean

Javid Shabbir¹ and Sat Gupta²

¹Department of Statistics, Quaid-i-Azam University, Islamabad, 45320, Pakistan.

²Department of Mathematics and Statistics, The University of North Carolina at Greensboro, NC 27412, USA.

Received October 26, 2018, Revised June 15, 2019, Accepted September 16, 2019, Published November 1, 2019

Abstract: We propose a generalized class of estimators for finite population variance using the auxiliary variable as well as rank of the auxiliary variable in stratified sampling. We identify many estimators as special cases of the proposed generalized class of estimators. We discuss the properties of all considered estimators up to first order of approximation. A real data set is used to observe the performances of estimators. It is observed that the proposed generalized class of estimators is more efficient than usual sample variance estimator, traditional ratio estimator, Bahl and Tuteja (1991) exponential ratio type estimator, usual difference estimator and Rao (1991) difference-type estimator.

Keywords: Stratified sampling, Auxiliary variable, Ranking, Bias, Mean squared error (*MSE*), Efficiency.

1. INTRODUCTION

The problem of estimation of finite population variance is an important issue where in application it is difficult to control the variability. In agriculture and biological experiments, researches face this problem deeply and consequently target results seem to be uncontrollable. The use of auxiliary information in an appropriate way may increase the precision of estimators. In this paper we use rank of the auxiliary variable in addition to the auxiliary variable for further improvement as discussed by Yaqub (2017) and Haq et al. (2017). To some extent we can control the variability by adopting the stratified sampling at planning, designing as well at estimation stages. In stratified sampling, heterogeneous population is divided into homogenous groups called strata and samples may be drawn from each stratum separately. Many authors who have contributed in estimating the population variance are Das and Tripathi (1978), Isaki (1983), Kadilar and Cingi (2007), Unyazici (2008), Unyazici and Cingi (2008), Shabbir and Gupta (2010), Shukla et al. (2015), Mishra et al. (2017) and Bhat et al. (2018).

In this paper, we propose a general class of estimators for finite population variance using the auxiliary information twice i.e. auxiliary variable as well as rank of the auxiliary variable. Expressions for biases and mean squared errors (*MSEs*) are derived up to the first order of approximation. A real data set is used to assess the performances of the estimators.

a) We divide the heterogeneous population of N units into L homogenous strata. Let N_h be the size of the h^{th} ($h=1,2,\dots,L$) stratum such that $\sum_{h=1}^L N_h = N$. Let y_{hi} , x_{hi} and $r_{(x)hi}$ be the values of the study variable (y), the auxiliary variable (x) and rank of the auxiliary variable ($r_{(x)i}$) respectively, on the i^{th} unit in the h^{th} stratum. Let

$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$, $\bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}$ and $\bar{r}_{(x)h} = \frac{1}{n_h} \sum_{i=1}^{n_h} r_{(x)hi}$ be the sample means corresponding to population means



$\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y_{hi}$, $\bar{X}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} x_{hi}$ and $\bar{R}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} r_{(x)hi}$ respectively in the h th stratum. Let
 $s_{yh}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$, $s_{xh}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2$ and $s_{r(x)h}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (r_{(x)hi} - \bar{r}_{(x)h})^2$
 be the sample variances corresponding to the population variances

$S_{r(x)h}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (r_{(x)hi} - \bar{R}_{(x)h})^2$
 and $S_{yh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2$, $S_{xh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)^2$
 respectively.

To obtain the biases and MSEs of estimators, we use the following error terms.

Let $\Lambda_{0h} = \frac{s_{yh}^2 - S_{yh}^2}{S_{yh}^2}$, $\Lambda_{1h} = \frac{s_{xh}^2 - S_{xh}^2}{S_{xh}^2}$, $\Lambda_{2h} = \frac{s_{r(x)h}^2 - S_{r(x)h}^2}{S_{r(x)h}^2}$ such that $E(\Lambda_{ih}) = 0$ ($i = 0, 1, 2$). Also

$E(\Lambda_{0h}^2) = n_h^{-1} Q_{400h}^*$, $E(\Lambda_{1h}^2) = n_h^{-1} Q_{040h}^*$, $E(\Lambda_{2h}^2) = n_h^{-1} Q_{004h}^*$, where $Q_{cdeh}^* = (Q_{cdeh} - 1)$,

$$Q_{cdeh} = \frac{\mu_{cdeh}}{\mu_{200h}^{c/2} \mu_{020h}^{d/2} \mu_{002h}^{e/2}} \text{ and } \mu_{cdeh} = \frac{\sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^c (x_{hi} - \bar{X}_h)^d (r_{(x)hi} - \bar{R}_{(x)h})^e}{N_h - 1}.$$

We ignore the finite population correction term for ease of computation.

2. Estimators in Literature

The variance of usual estimator $\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h$ in stratified sampling, is given by

$$\text{Var}(\bar{y}_{st}) = \sum_{h=1}^L n_h^{-1} (W_h S_{yh})^2 = S_{st(y)}^2 \text{ (say)}, \quad (1)$$

where $W_h = N_h / N$ are known stratum weights.

The unbiased estimator $\hat{S}_{st(y)}^2$ of $S_{st(y)}^2$, is given by

$$\hat{S}_{st(y)}^2 = \sum_{h=1}^L n_h^{-1} W_h^2 s_{yh}^2. \quad (2)$$

The variance of $\hat{S}_{st(y)}^2$, is given by

$$\text{Var}(\hat{S}_{st(y)}^2) = \sum_{h=1}^L n_h^{-3} (W_h S_{yh})^4 Q_{400h}^* = \text{MSE}(\hat{S}_{st(y)}^2). \quad (3)$$

The traditional ratio estimator for population variance in stratified sampling is:

$$\hat{S}_{st(R)}^2 = \sum_{h=1}^L n_h^{-1} W_h^2 s_{yh}^2 \left(\frac{S_{xh}^2}{S_{xh}^2} \right), \quad (4)$$

where S_{xh}^2 is the known population variance.



The bias and MSE of $\hat{S}_{st(R)}^2$ to first order of approximation, are given by

$$Bias(\hat{S}_{st(R)}^2) \approx \sum_{h=1}^L n_h^{-2} (W_h S_{yh})^2 \{Q_{040h}^* - Q_{220h}^*\}$$

and

$$MSE(\hat{S}_{st(R)}^2) \approx \sum_{h=1}^L n_h^{-3} (W_h S_{yh})^4 \{Q_{400h}^* + Q_{040h}^* - 2Q_{220h}^*\}. \tag{6}$$

From (3) and (6), the $MSE(\hat{S}_{st(R)}^2)$ is more efficient than $MSE(\hat{S}_{st(y)}^2)$ if

$$\sum_{h=1}^L n_h^{-3} (W_h S_{yh})^4 \{2Q_{220h}^* - Q_{040h}^*\} > 0.$$

Bahl and Tuteja (1991) exponential ratio estimator for population variance is:

$$\hat{S}_{st(E)}^2 = \sum_{h=1}^L n_h^{-1} W_h^2 s_{yh}^2 \exp\left(\frac{S_{xh}^2 - s_{xh}^2}{S_{xh}^2 + s_{xh}^2}\right). \tag{7}$$

The bias and MSE of $\hat{S}_{st(E)}^2$ to first order of approximation, are given by

$$Bias(\hat{S}_{st(E)}^2) \approx \sum_{h=1}^L n_h^{-2} (W_h S_{yh})^2 \left\{ \frac{3}{8} Q_{040h}^* - Q_{220h}^* \right\} \tag{8}$$

and $MSE(\hat{S}_{st(E)}^2) \approx \sum_{h=1}^L n_h^{-3} (W_h S_{yh})^4 \left\{ Q_{400h}^* + \frac{1}{4} Q_{040h}^* - Q_{220h}^* \right\} \tag{9}$

From (3) and (9), the $MSE(\hat{S}_{st(E)}^2)$ is more efficient than $MSE(\hat{S}_{st(y)}^2)$ if

$$\sum_{h=1}^L n_h^{-3} (W_h S_{yh})^4 \left\{ Q_{220h}^* - \frac{1}{4} Q_{040h}^* \right\} > 0.$$

The usual difference estimator is:

$$\hat{S}_{st(D)}^2 = \sum_{h=1}^L n_h^{-1} W_h^2 \{s_{yh}^2 + m_{0h} (S_{xh}^2 - s_{xh}^2)\}, \tag{10}$$

where m_{0h} is the constant.

The minimum variance of $\hat{S}_{st(D)}^2$, is given by

$$Var(\hat{S}_{st(D)}^2)_{\min} = \sum_{h=1}^L n_h^{-3} (W_h S_{yh})^4 Q_{400h}^* \left(1 - \rho_{s_{yh}^2, s_{xh}^2}^2\right) = MSE(\hat{S}_{st(D)}^2)_{\min}, \tag{11}$$

where $\rho_{s_{yh}^2, s_{xh}^2} = \frac{Q_{220h}^*}{\sqrt{Q_{400h}^*} \sqrt{Q_{040h}^*}}$ and the optimum value is $m_{0h(opt)} = \frac{S_{yh}^2 Q_{220h}^*}{S_{xh}^2 Q_{040h}^*}$.

From (3) and (11), the $MSE(\hat{S}_{st(D)}^2)_{\min}$ performs better than $MSE(\hat{S}_{st(y)}^2)$ if



$$\sum_{h=1}^L n_h^{-3} (W_h S_{yh})^4 Q_{400h}^* \rho_{s_{yh}^2, s_{xh}^2}^2 > 0,$$

which is always true.

Following Rao (1991), the difference type estimator is:

$$\hat{S}_{st(Rao)}^2 = \sum_{h=1}^L n_h^{-1} W_h^2 \left\{ m_{1h} s_{yh}^2 + m_{2h} (S_{xh}^2 - s_{xh}^2) \right\}, \quad (12)$$

where m_{1h} and m_{2h} are constants.

The bias and minimum MSE of $\hat{S}_{st(Rao)}^2$, are given by

$$Bias(\hat{S}_{st(Rao)}^2) \approx \sum_{h=1}^L n_h^{-2} (W_h S_{yh})^2 (m_{1h} - 1) \quad (13)$$

and

$$MSE(\hat{S}_{st(Rao)}^2)_{\min} \approx \sum_{h=1}^L n_h^{-3} (W_h S_{yh})^4 \frac{1}{Y_h} Q_{400h}^* \left(1 - \rho_{s_{yh}^2, s_{xh}^2}^2 \right), \quad (14)$$

$$\text{where } Y_h = 1 + n_h^{-1} Q_{400h}^* \left(1 - \rho_{s_{yh}^2, s_{xh}^2}^2 \right).$$

The optimum values of m_{1h} and m_{2h} are $m_{1h(opt)} = 1/Y_h$ and $m_{2h(opt)} = \frac{S_{yh}^2 Q_{220h}^*}{S_{xh}^2 Q_{040h}^*}$.

From (11) and (14), the $MSE(\hat{S}_{st(Rao)}^2)_{\min}$ performs better than $MSE(\hat{S}_{st(D)}^2)_{\min}$ if

$$\sum_{h=1}^L n_h^{-4} (W_h S_{yh})^4 \frac{1}{Y_h} Q_{400h}^{*2} \left(1 - \rho_{s_{yh}^2, s_{xh}^2}^2 \right)^2 > 0,$$

which is always true. This indicates $MSE(\hat{S}_{st(Rao)}^2)_{\min}$ is always better than $MSE(\hat{S}_{st(y)}^2)$, $MSE(\hat{S}_{st(R)}^2)$ and $MSE(\hat{S}_{st(E)}^2)$.

3. Proposed Estimator

Following Yaqoob (2017) and Haq et al. (2017), when there exists a relationship between the study variable and the auxiliary variable, the ranks of the auxiliary variable are also correlated with the study variable as well as with the auxiliary variable. Using the above idea, we propose the following generalized class of estimators for population variance as:

$$\begin{aligned} \hat{S}_{st(Prop)}^2 &= \sum_{h=1}^L n_h^{-1} W_h^2 \left\{ m_{1h} s_{yh}^2 + m_{2h} (S_{xh}^2 - s_{xh}^2) + m_{3h} (S_{r(x)h}^2 - s_{r(x)h}^2) \right\} \\ &\quad \times \left[\left(\frac{S_{xh}^2}{s_{xh}^2} \right)^{\alpha_{1h}} \left\{ \exp \alpha_{2h} \left(\frac{S_{xh}^2 - s_{xh}^2}{S_{xh}^2 + s_{xh}^2} \right) \right\} \right], \end{aligned} \quad (15)$$

where m_{ih} ($i = 1, 2, 3$) are constants; α_{ih} ($i = 1, 2$) are scalar quantities.



We can get existing estimators from a generalized class of estimators by substituting different values of α_{ih} ($i = 1, 2$) and m_{ih} ($i = 1, 2, 3$) in (15).

- (i) For $\alpha_{1h} = \alpha_{2h} = m_{2h} = m_{3h} = 0$ and $m_{1h} = 1$ in (15), we get $\hat{S}_{st(y)}^2$ (see (2)).
- (ii) For $\alpha_{1h} = m_{1h} = 1$ and $\alpha_{2h} = m_{2h} = m_{3h} = 0$ in (15), we get $\hat{S}_{st(R)}^2$ (see (4)).
- (iii) For $\alpha_{2h} = m_{1h} = 1$ and $\alpha_{1h} = m_{2h} = m_{3h} = 0$ in (15), we get $\hat{S}_{st(E)}^2$ (see (7)).
- (iv) For $\alpha_{1h} = \alpha_{2h} = m_{3h} = 0$, $m_{2h} = m_{0h}$ and $m_{1h} = 1$ in (15), we get $\hat{S}_{st(D)}^2$ (see (10)).
- (v) For $\alpha_{1h} = \alpha_{2h} = m_{3h} = 0$ and $m_{2h} = m_{0h}$ in (15), we get $\hat{S}_{st(Rao)}^2$ (see (12)).

Solving (15) in terms of errors, we have

$$\hat{S}_{st(Prop)}^2 - S_y^2 \cong \sum_{h=1}^L n_h^{-1} W_h^2 \left[(m_{1h} - 1) S_{yh}^2 + m_{1h} S_{yh}^2 \left\{ \Lambda_{0h} - \frac{1}{2} k_{1h} \Lambda_{1h} + \frac{1}{8} k_{2h} \Lambda_{1h}^2 - \frac{1}{2} k_{1h} \Lambda_{0h} \Lambda_{1h} \right\} - m_{2h} S_{xh}^2 \left\{ \Lambda_{1h} - \frac{1}{2} k_{1h} \Lambda_{1h}^2 \right\} - m_{3h} S_{r(x)h}^2 \left\{ \Lambda_{2h} - \frac{1}{2} k_{1h} \Lambda_{1h} \Lambda_{2h} \right\} \right],$$

where $k_{1h} = 2\alpha_{1h} + \alpha_{2h}$ and $k_{2h} = 4\alpha_{1h}(1 + \alpha_{1h} + \alpha_{2h}) + \alpha_{2h}(\alpha_{2h} + 2)$.

The bias and MSE of $\hat{S}_{st(Prop)}^2$ to first order of approximation, are given by

$$Bias(\hat{S}_{st(Prop)}^2) \cong \sum_{h=1}^L n_h^{-1} W_h^2 \left[(m_{1h} - 1) S_{yh}^2 + 2m_{1h} S_{yh}^2 D_h + m_{2h} S_{xh}^2 E_h + m_{3h} S_{r(x)h}^2 F_h \right] \tag{17}$$

and

$$MSE(\hat{S}_{st(Prop)}^2) \cong \sum_{h=1}^L n_h^{-2} W_h^4 \left[(m_{1h} - 1)^2 S_{yh}^4 + m_{1h}^2 S_{yh}^4 A_h + m_{2h}^2 S_{xh}^4 B_h + m_{3h}^2 S_{r(x)h}^4 C_h - m_{1h} S_{yh}^4 D_h - m_{2h} S_{yh}^2 S_{xh}^2 E_h - m_{3h} S_{yh}^2 S_{r(x)h}^2 F_h - 2m_{1h} m_{2h} S_{yh}^2 S_{xh}^2 G_h - 2m_{1h} m_{3h} S_{yh}^2 S_{r(x)h}^2 H_h + 2m_{2h} m_{3h} S_{xh}^2 S_{r(x)h}^2 I_h \right], \tag{18}$$

where

$$A_h = n_h^{-1} \left\{ Q_{400h}^* + \frac{1}{4} Q_{040h}^* (k_{1h}^2 + k_{2h}) - 2k_{1h} Q_{220h}^* \right\}, B_h = n_h^{-1} Q_{040h}^*, C_h = n_h^{-1} Q_{004h}^*,$$

$$D_h = n_h^{-1} \left\{ \frac{1}{4} k_{2h} Q_{040h}^* - k_{1h} Q_{220h}^* \right\}, E_h = n_h^{-1} k_{1h} Q_{040h}^*, F_h = n_h^{-1} k_{1h} Q_{022h}^*, G_h = n_h^{-1} \left\{ Q_{220h}^* - k_{1h} Q_{040h}^* \right\},$$

$$H_h = n_h^{-1} \left\{ Q_{202h}^* - k_{1h} Q_{022h}^* \right\}, I_h = n_h^{-1} Q_{022h}^*.$$

From (18), the optimum values of m_{ih} ($i = 1, 2, 3$) are:

$$m_{1h(opt)} = \frac{\Delta_{1h}}{2\Delta_{2h}}, m_{2h(opt)} = \frac{S_{yh}^2 \Delta_{3h}}{2S_{xh}^2 \Delta_{2h}} \text{ and } m_{3h(opt)} = \frac{S_{yh}^2 \Delta_{4h}}{2S_{r(x)h}^2 \Delta_{2h}},$$

where



$$\begin{aligned}\Delta_{1h} &= B_h C_h D_h + B_h F_h H_h + C_h E_h G_h - D_h I_h^2 \\ &\quad - E_h H_h I_h + 2B_h C_h - 2I_h^2, \\ \Delta_{2h} &= A_h B_h C_h - A_h I_h^2 - B_h H_h^2 - C_h G_h^2 \\ &\quad + 2G_h H_h I_h + B_h C_h - I_h^2, \\ \Delta_{3h} &= A_h C_h E_h - A_h F_h I_h + C_h D_h G_h - D_h H_h I_h \\ &\quad - E_h H_h^2 + F_h G_h H_h + C_h E_h + 2C_h G_h \\ &\quad - F_h I_h - 2H_h I_h, \\ \Delta_{4h} &= A_h B_h F_h - A_h E_h I_h + B_h D_h H_h - D_h G_h I_h \\ &\quad + E_h G_h H_h - F_h G_h^2 + B_h F_h + 2B_h H_h \\ &\quad - E_h I_h - 2G_h I_h.\end{aligned}$$

Substituting the optimum values of m_{ih} ($i=1,2,3$) in (18), we get the minimum MSE of $\hat{S}_{st(Prop)}^2$ to first degree of approximation, given by

$$MSE(\hat{S}_{st(Prop)}^2)_{\min} \approx \sum_{h=1}^L n_h^{-2} (W_h S_{yh})^4 \left(1 - \frac{\Delta_{5h}}{4\Delta_{2h}} \right), \quad (19)$$

where

$$\begin{aligned}\Delta_{5h} &= A_h B_h F_h^2 + A_h C_h E_h^2 - 2A_h E_h F_h I_h + B_h C_h D_h^2 \\ &\quad + 2B_h D_h F_h H_h + 2C_h D_h E_h G_h - D_h^2 I_h^2 \\ &\quad - 2D_h E_h H_h I_h - 2D_h F_h G_h I_h - E_h^2 H_h^2 \\ &\quad + 2E_h F_h G_h H_h - F_h^2 G_h^2 + 4B_h C_h D_h \\ &\quad + B_h F_h^2 + 4B_h F_h H_h + C_h E_h^2 \\ &\quad + 4C_h E_h B_h - 4D_h I_h^2 - 2E_h F_h I_h \\ &\quad - 4E_h H_h I_h - 4D_h G_h J_h \\ &\quad + 4B_h C_h - 4I_h^2.\end{aligned}$$

Note: From (15), we have generated 4 new more difference type estimators which are given in (20), (22), (25) and (28).

(i) Put $\alpha_{1h} = \alpha_{2h} = 0$ and $m_{1h} = 1$ in (15), we get the following difference type estimator:

$$\hat{S}_{st(Prop1)}^2 = \sum_{h=1}^L n_h^{-1} W_h^2 \left\{ S_{yh}^2 + m_{2h} (S_{xh}^2 - s_{xh}^2) + m_{3h} (S_{r(x)h}^2 - s_{r(x)h}^2) \right\}. \quad (20)$$

The minimum MSE of $\hat{S}_{st(Prop1)}^2$ to first degree of approximation, is given by

$$MSE(\hat{S}_{st(Prop1)}^2)_{\min} \approx \sum_{h=1}^L n_h^{-3} (W_h S_{yh})^4 Q_{400h}^* \left(1 - R_{s_{yh}^2, s_{xh}^2, s_{r(x)h}^2}^2 \right), \quad (21)$$



where $R^2_{s_{yh}^2, s_{xh}^2, s_{r(x)h}^2} = \frac{\Pi_1}{\Pi_2}$ is the multiple correlation coefficient.;

$$\Pi_1 = \rho_{s_{yh}^2, s_{xh}^2}^2 + \rho_{s_{yh}^2, s_{r(x)h}^2}^2 - 2\rho_{s_{yh}^2, s_{xh}^2} \rho_{s_{yh}^2, s_{r(x)h}^2} \rho_{s_{xh}^2, s_{r(x)h}^2}, \Pi_2 = 1 - \rho_{s_{xh}^2, s_{r(x)h}^2}^2, \rho_{s_{yh}^2, s_{xh}^2}^* = \frac{Q_{220h}^*}{\sqrt{Q_{400h}^*} \sqrt{Q_{040h}^*}},$$

$$\rho_{s_{yh}^2, s_{r(x)h}^2}^* = \frac{Q_{202h}^*}{\sqrt{Q_{400h}^*} \sqrt{Q_{004h}^*}} \text{ and } \rho_{s_{xh}^2, s_{r(x)h}^2}^* = \frac{Q_{022h}^*}{\sqrt{Q_{040h}^*} \sqrt{Q_{004h}^*}}.$$

The optimum values $m_{ih} (i = 2, 3)$, are given by

$$m_{2h(opt)} = \frac{S_{yh}^2}{S_{xh}^2} \left\{ \frac{Q_{004h}^* Q_{220h}^* - Q_{022h}^* Q_{202h}^*}{Q_{004h}^* Q_{040h}^* - Q_{022h}^{*2}} \right\} \text{ and } m_{3h(opt)} = \frac{S_{yh}^2}{S_{r(x)h}^2} \left\{ \frac{Q_{040h}^* Q_{202h}^* - Q_{022h}^* Q_{220h}^*}{Q_{004h}^* Q_{040h}^* - Q_{022h}^{*2}} \right\}.$$

(ii) Put $\alpha_{1h} = \alpha_{2h} = 0$ in (15), we get the following difference type estimator:

$$\hat{S}_{st(Prop2)}^2 = \sum_{h=1}^L n_h^{-1} W_h^2 \left\{ m_{1h} s_{yh}^2 + m_{2h} (S_{xh}^2 - s_{xh}^2) + m_{3h} (S_{r(x)h}^2 - s_{r(x)h}^2) \right\}. \tag{22}$$

The bias and minimum MSE of $\hat{S}_{st(Prop2)}^2$ to first degree of approximation, are given by

$$Bias(\hat{S}_{st(Prop2)}^2) \approx \sum_{h=1}^L n_h^{-2} W_h^2 S_{yh}^2 Q_{400h}^* (m_{1h} - 1) \tag{23}$$

and

$$MSE(\hat{S}_{st(Prop2)}^2)_{min} \approx \sum_{h=1}^L n_h^{-3} (W_h S_{yh})^4 \left\{ \frac{Q_{400h}^* (1 - R^2_{s_{yh}^2, s_{xh}^2, s_{r(x)h}^2})}{1 + n_h^{-1} Q_{400h}^* (1 - R^2_{s_{yh}^2, s_{xh}^2, s_{r(x)h}^2})} \right\}. \tag{24}$$

The optimum values $m_{ih} (i = 1, 2, 3)$ are

$$m_{1h(opt)} = \frac{\delta_{1h}}{\delta_{2h}}, m_{2h(opt)} = \frac{S_{yh}^2 \delta_{3h}}{S_{xh}^2 \delta_{2h}}, m_{3h(opt)} = \frac{S_{yh}^2 \delta_{4h}}{S_{r(x)h}^2 \delta_{2h}},$$

where $\delta_{1h} = Q_{040h}^* Q_{004h}^* - Q_{022h}^{*2}$, $\delta_{2h} = n_h^{-1} \left\{ Q_{040h}^* Q_{004h}^* Q_{004h}^* - Q_{040h}^* Q_{202h}^{*2} - Q_{400h}^* Q_{022h}^{*2} + 2Q_{022h}^* Q_{220h}^* Q_{202h}^* - Q_{004h}^* Q_{220h}^{*2} \right\} + \delta_{1h}$,

$\delta_{3h} = Q_{022h}^* Q_{202h}^* - Q_{004h}^* Q_{220h}^*$, and $\delta_{4h} = Q_{040h}^* Q_{202h}^* - Q_{022h}^* Q_{220h}^*$.

(iii) Put $\alpha_{1h} = 1$ and $\alpha_{2h} = 0$ in (15), we get the following difference type estimator:

$$\hat{S}_{st(Prop3)}^2 = \sum_{h=1}^L n_h^{-1} W_h^2 \left\{ m_{1h} s_{yh}^2 + m_{2h} (S_{xh}^2 - s_{xh}^2) + m_{3h} (S_{r(x)h}^2 - s_{r(x)h}^2) \right\} \left(\frac{S_{xh}^2}{s_{xh}^2} \right). \tag{25}$$

The bias and minimum MSE of $\hat{S}_{st(Prop3)}^2$ to first degree of approximation, are given by

$$Bias(\hat{S}_{st(Prop3)}^2) \approx \sum_{h=1}^L n_h^{-1} W_h^2 \left\{ (m_{1h} - 1) S_{yh}^2 + m_{1h} S_{yh}^2 n_h^{-1} (Q_{040h}^* - Q_{220h}^*) + m_{2h} S_{xh}^2 n_h^{-1} Q_{040h}^* + m_{3h} S_{r(x)h}^2 n_h^{-1} Q_{022h}^* \right\} \tag{26}$$

and



$$MSE(\hat{S}_{st(Prop3)}^2)_{\min} \approx \sum_{h=1}^L n_h^{-3} (W_h S_{yh})^4 \left\{ \frac{(1 - n_h^{-1} Q_{040h}^*) (1 - R_{s_{yh}^2, s_{xh}^2, s_{r(x)h}^2}^2)}{(1 - n_h^{-1} Q_{040h}^*) + n_h^{-1} Q_{400h}^* (1 - R_{s_{yh}^2, s_{xh}^2, s_{r(x)h}^2}^2)} \right\}. \quad (27)$$

The optimum values m_{ih} ($i = 1, 2, 3$) are: $m_{1h(opt)} = \frac{\eta_{1h}}{\eta_{2h}}$, $m_{2h(opt)} = -\frac{S_{yh}^2 \eta_{3h}}{S_{xh}^2 \eta_{2h}}$, and $m_{3h(opt)} = \frac{S_{yh}^2 \eta_{4h}}{S_{r(x)h}^2 \eta_{2h}}$,

where $\eta_{1h} = (Q_{040h}^* Q_{004h}^* - Q_{022h}^{*2}) (n_h^{-1} Q_{040h}^* - 1)$, $\eta_{2h} = n_h^{-1} \{ Q_{040h}^* (Q_{040h}^* Q_{004h}^* - Q_{022h}^{*2} - Q_{400h}^* Q_{004h}^* + Q_{202h}^{*2}) + Q_{022h}^* (Q_{022h}^* Q_{400h}^* - 2Q_{220h}^* Q_{202h}^*) + Q_{004h}^* Q_{220h}^{*2} \} - Q_{040h}^* Q_{004h}^* + Q_{022h}^{*2}$,

$\eta_{3h} = n_h^{-1} \{ Q_{040h}^* (Q_{040h}^* Q_{004h}^* - Q_{022h}^{*2} + Q_{022h}^* Q_{202h}^* + Q_{400h}^* Q_{004h}^* - Q_{220h}^* Q_{004h}^* - Q_{202h}^{*2}) - Q_{022h}^* Q_{004h}^* + 2Q_{022h}^* Q_{220h}^* Q_{202h}^* - Q_{220h}^* Q_{004h}^* \} - Q_{040h}^* Q_{004h}^* + Q_{022h}^{*2} - Q_{022h}^* Q_{202h}^* + Q_{220h}^* Q_{004h}^*$,

$\eta_{4h} = (Q_{040h}^* Q_{202h}^* - Q_{022h}^* Q_{220h}^*) (n_h^{-1} Q_{040h}^* - 1)$.

(iv) Put $\alpha_{1h} = 0$ and $\alpha_{2h} = 1$ in (15), we get the following difference type estimator:

$$\hat{S}_{st(Prop4)}^2 = \sum_{h=1}^L n_h^{-1} W_h^2 \{ m_{1h} S_{yh}^2 + m_{2h} (S_{xh}^2 - s_{xh}^2) + m_{3h} (S_{r(x)h}^2 - s_{r(x)h}^2) \} \exp \left(\frac{S_{xh}^2 - s_{xh}^2}{S_{xh}^2 + s_{xh}^2} \right) \quad (28)$$

The bias and minimum MSE of $\hat{S}_{st(Prop4)}^2$ to first degree of approximation, are given by

$$Bias(\hat{S}_{st(Prop4)}^2) \approx \sum_{h=1}^L n_h^{-2} W_h^2 \{ (m_{1h} - 1) S_{yh}^2 + m_{1h} S_{yh}^2 n_h^{-1} \left(\frac{3}{8} Q_{040h}^* - \frac{1}{2} Q_{220h}^* \right) + m_{2h} S_{xh}^2 n_h^{-1} Q_{040h}^* + m_{3h} S_{r(x)h}^2 n_h^{-1} Q_{022h}^* \} \quad (29)$$

and

$$MSE(\hat{S}_{st(Prop4)}^2)_{\min} \approx \sum_{h=1}^L n_h^{-3} (W_h S_{yh})^4 \frac{\Psi_{1h}}{\Psi_{2h}}, \quad (30)$$

where

$$\Psi_{1h} = Q_{400h}^* (1 - R_{s_{yh}^2, s_{xh}^2, s_{r(x)h}^2}^2) - (64n_h)^{-1} Q_{040h}^{*2} - (4n_h)^{-1} Q_{400h}^* Q_{040h}^* (1 - R_{s_{yh}^2, s_{xh}^2, s_{r(x)h}^2}^2)$$

and

$$\Psi_{2h} = 1 + n_h^{-1} Q_{400h}^* (1 - R_{s_{yh}^2, s_{xh}^2, s_{r(x)h}^2}^2).$$

The optimum values m_{ih} ($i = 1, 2, 3$) are

$$m_{1h(opt)} = -\frac{\gamma_{1h}}{8\gamma_{2h}}, \quad m_{2h(opt)} = \frac{S_{yh}^2 \gamma_{3h}}{8S_{xh}^2 \gamma_{2h}}, \quad m_{3h(opt)} = -\frac{S_{yh}^2 \gamma_{4h}}{8S_{r(x)h}^2 \gamma_{2h}},$$

where $\gamma_{1h} = (Q_{040h}^* Q_{004h}^* - Q_{022h}^{*2}) (n_h^{-1} Q_{040h}^* - 8)$, $\gamma_{2h} = n_h^{-1} \{ Q_{004h}^* Q_{040h}^* Q_{400h}^* - Q_{004h}^* Q_{220h}^{*2} - Q_{040h}^* Q_{202h}^{*2} - Q_{400h}^* Q_{022h}^{*2} + 2Q_{022h}^* Q_{220h}^* Q_{202h}^* \} + Q_{004h}^* Q_{040h}^* - Q_{022h}^{*2}$,

$\gamma_{3h} = n_h^{-1} \{ Q_{004h}^* (Q_{040h}^{*2} + 4Q_{040h}^* Q_{400h}^* - Q_{040h}^* Q_{220h}^* - 4Q_{220h}^{*2}) + Q_{040h}^* (Q_{022h}^* Q_{202h}^* - Q_{022h}^{*2} - 4) \}$



$$\gamma_{4h} = \left(Q_{040h}^* Q_{202h}^* - Q_{022h}^* Q_{220h}^* \right) \left(n_h^{-1} Q_{040h}^* - 8 \right) \cdot$$

$$\left. -Q_{202h}^{*2} - 4Q_{022h}^* Q_{400h}^* + 8Q_{022h}^* Q_{220h}^* Q_{202h}^* \right\} - 4Q_{040h}^* Q_{004h}^* + 8Q_{004h}^* Q_{220h}^* + Q_{022h}^{*2} - 8Q_{022h}^* Q_{202h}^*,$$

Theoretical comparison of estimators is very complex, so we use the following data set for numerical comparison

4. Data set

We use the Neyman allocation to obtain the sample size in each stratum.

Population: [Source: Freud et al. (2006)]

Let y be the field goals made (FGM) and x be the attempted field goals (FGAT) in 4 regions (1-4).

$$N = 66, N_1 = 15, N_2 = 18, N_3 = 17, N_4 = 16, n = 30, n_1 = 7, n_2 = 8, n_3 = 8, n_4 = 7, \bar{Y}_1 = 3537.4,$$

$$\bar{Y}_2 = 3564.889, \bar{Y}_3 = 3583.882, \bar{Y}_4 = 3622.625, \bar{X}_1 = 7547.6, \bar{X}_2 = 7569.111, \bar{X}_3 = 7560.353,$$

$$\bar{X}_4 = 7565.75, \bar{R}_{x1} = 8.0, \bar{R}_{x2} = 9.4444, \bar{R}_{x3} = 8.9412, \bar{R}_{x4} = 8.5, S_{y1}^2 = 32448.4, S_{y2}^2 = 25293.6340,$$

$$S_{y3}^2 = 26828.1103, S_{y4}^2 = 18506.65, S_{x1}^2 = 53746.6857, S_{x2}^2 = 33420.3398, S_{x3}^2 = 74601.2426,$$

$$S_{x4}^2 = 30767.6667, S_{r(x)1}^2 = 20.0, S_{r(x)2}^2 = 28.4967, S_{r(x)3}^2 = 24.6838, S_{r(x)4}^2 = 22.6667,$$

$$Q_{4001} = 3.5287, Q_{4002} = 2.8881, Q_{4003} = 2.7864, Q_{4004} = 1.8248, Q_{0401} = 2.0520, Q_{0402} = 2.3936,$$

$$Q_{0403} = 1.9201, Q_{0404} = 1.8923, Q_{0041} = 1.67, Q_{0042} = 1.6940, Q_{0043} = 1.6840, Q_{0044} = 1.6786,$$

$$Q_{2201} = 1.4871, Q_{2202} = 1.7137, Q_{2203} = 1.51347, Q_{2204} = 1.1087, Q_{0221} = 1.7396, Q_{0222} = 1.8085,$$

$$Q_{0223} = 1.6648, Q_{0224} = 1.7393, Q_{2021} = 1.6556, Q_{2022} = 1.3703, Q_{2023} = 1.3290, Q_{2024} = 1.0368.$$

The results based on this data set are given in Table 1.

We use the following expression to obtain the percent relative efficiency (PRE) of different estimators:

$$PRE = \frac{MSE(\hat{S}_{st(y)}^2)}{MSE(\hat{S}_{st(i)}^2) \text{ or } MSE(\hat{S}_{st(i)\min}^2)} \times 100, \quad (i = y, R, E, D, Rao, Prop (i = 1, 2, 3, 4)).$$

Table 1. MSE values and PRE of different estimators with respect to $\hat{S}_{st(y)}^2$.

Estimator	MSE	Efficiency
$\hat{S}_{st(y)}^2$	47660.68	100.000
$\hat{S}_{st(R)}^2$	49715.46	95.867
$\hat{S}_{st(E)}^2$	41932.95	113.659
$\hat{S}_{st(D)}^2$	41459.55	114.957
$\hat{S}_{st(Rao)}^2$	33353.19	142.897
$\hat{S}_{st(Prop1)}^2$	35843.68	132.968
$\hat{S}_{st(Prop2)}^2$	29892.70	159.437
$\hat{S}_{st(Prop3)}^2$	29060.73	164.004
$\hat{S}_{st(Prop4)}^2$	28746.36	165.797



In Table 1, we see that all proposed estimators perform very well in a given data except $\hat{S}_{st(Prop1)}^2$, where MSE of $\hat{S}_{st(Prop1)}^2$ is larger than MSE of $\hat{S}_{st(Rao)}^2$.

5. Conclusion

We proposed a generalized class of estimators for finite population variance in stratified sampling. The ratio estimator \hat{S}_R^2 performs poorly in this data set. Among all proposed estimators, the performance of regression-in-exponential ratio type estimator \hat{S}_{Prop4}^2 is the best. By using the idea of Yaqub (2017) and Haq et al. (2017), overall proposed generalized class of estimators is found to be better except the estimator $\hat{S}_{st(Prop1)}^2$ which is inferior than the Rao (1991) type estimator $\hat{S}_{st(Rao)}^2$.

Acknowledgements:

The authors are thankful to the learned referees for their valuable constructive comments.

References

- [1] S. Bahl and R.K. Tuteja, "Ratio and product type exponential estimators", *Journal of Information and Optimization Sciences*, vol. 12(1), pp. 159-163, 1991.
- [2] M. A. Bhat, S. Maqbool, S. A. Saraf, S. H. Malik, Y. Ismail and M. Subzar, "Improved and robust estimators for finite population variance using linear combination of probability weighted moment and quartiles as auxiliary information", *Journal of Scientific and Research Reports*, vol. 18(4), pp. 1-6, 2018.
- [3] A. K. Das and T. P. Tripathy, "Use of auxiliary information in estimating the finite population variance", *Sankhya*, vol.40, pp.139-148, 1978.
- [4] R. J. Freund, W. J. Wilson and S. A. Ping, *Regression Analysis: Statistical Modeling of a Response Variable*, Second edn., Academic Press, London, 2006.
- [5] A. Haq, M. Khan, and Z. Hussain, "A new estimator of finite population mean based on the dual use of the auxiliary information", *Communications in Statistics-Theory and Methods*, vol. 46,(9), pp. 4425-4436, 2017.
- [6] C. T. Isaki, "Variance estimation using auxiliary information", *Journal of the American Statistical Association*, vo. 78, pp. 117-123, 1983.
- [7] C. Kadilar and H. Cingi, "Improvement in variance estimation in simple random sampling", *Communications in Statistics-Theory and Methods*", vol. 36(11), pp. 2075-2081, 2007.
- [8] S. Mishra, K. Dipika and D. K. Yadav, "Estimation of population variance utilizing auxiliary information, *International Journal of Statistics and Systems*, 12(2), 303-309, 2017.
- [9] T. Rao, "On certain methods of improving ratio and regression estimators, *Communications in Statistics-Theory and Methods*", vol. 20(10), pp. 3325-3340, 1991.
- [10] G. Ozel, H. Cingi and M. Oguz, "Separate ratio estimators for the population variance in stratified sampling, *Communications in Statistics-Theory and Methods*", vol. 43, pp. 4766-4779, 2014
- [11] J. Shabbir and S. Gupta, "Some estimators of finite population variance of stratified sample mean, *Communications in Statistics-Theory and Methods*", vol. 39, pp. 3001-3008, 2010
- [12] A. K. Shukla, S. K. Yadav and V. Tiwari, "Improved variance function in cluster sampling, *International Journal of Computational and Theoretical Statistics*", vol. 2(1), pp. 25-30, 2015.
- [13] Y. Unyazici, "*Variance Estimation Methods in Some Sampling Designs*": Unpublished PhD Thesis, Faculty of Science, Hacettepe University, Ankara, Turkey, 2008.



- [14] Y. Unyazici and H. Cingi, "New generalized estimators for the population variance using auxiliary information, *Hacettepe Journal of Mathematics and Statistics*", vol. 37(2), pp. 177-184, 2008.
- [15] M. Yaqub, J. Shabbir and S. Gupta, "Estimation of population mean based on dual use of auxiliary variable in non-response", *Communications in Statistics-Theory and Methods*, vol.46 (24), pp. 12130-12151, 2017.