IMPROVED ESTIMATION OF VARIANCE UNDER COMPLETE AND INCOMPLETE INFORMATION

R. R. Sinha[1] and Vinod Kumar[2]
Department of Mathematics, Dr. B. R. Ambedkar
National Institute of Technology, Jalandhar, India

ABSTRACT
The present paper deals with the efficiency improvement of the estimators for finite population variance under complete and incomplete information emerging due to non-response. Using the information of auxiliary character, improved regression cum exponential estimators are proposed for two different cases. The bias and mean square error (MSE) of the proposed estimators are derived and theoretical conditions are obtained to make the suggested estimators more efficient than the unbiased, ratio, regression, generalized and more over classes of estimators used in practice. Theoretical as well as empirical studies are carried out to illustrate the efficiency of the proposed estimators.

KEYWORDS: Variance; Bias; Mean square error; Auxiliary character; Non-response.

MSC: 62D05

RESUMEN
El presente paper trata del incremento en la eficiencia de estimadores de la varianza para poblaciones finitas bajo la existencia de información completa e incompleta que proviene del las no-respuestas.. Usando la información de una auxiliar característica, se proponen estimadores mejorados de los tipos "regression cum exponential" para dos diferentes casos. El sesgo y el error cuadrático medio (MSE) de estos estimadores son derivados y se obtienen condiciones teóricas para hacerles más eficientes que el insesgado, el de razón , el de regresión, el generalizado y otros más en las clases de estimadores usados en la práctica. Estudios teóricos y empíricos se desarrollaron para ilustrar la eficiencia de los estimadores propuestos.

PALABRAS CLAVE: Varianza; Sesgo; Error Cuadrático Medio; Característica Auxiliar; No-respuesta.

1. INTRODUCTION

To measures the amount of variability among the study character, estimation of finite population variance is a subject of great interest and assignment for the researchers of various fields. Over the years, researchers are interested to utilize the information available on the auxiliary character on both at the selection and estimation stages to improve the precisions of estimators of population parameters. In the contest of complete information available on study and auxiliary character(s), Singh et al. (1973) discussed the utilization of a coefficient of kurtosis while Das and Tripathi (1978) suggested the utilization of coefficient of variation and known population variance of the auxiliary character in estimation of variance of study character. Isaki (1983) first suggested the ratio estimator for population variance using known population variance of the auxiliary character. Prasad and Singh (1990) revised the Isaki (1983) estimator for estimating population variance. Many more authors Garcia and Cebran (1996), Upadhyaya and Singh (2001), Jhajj et al. (2005), Kadilar and Cingi (2007), Turgut and Cingi (2008), Gupta and Shabbir (2008), Dubey and Sharma (2008), Subramani and Kumarapandiyan (2012), Singh and Solanki (2013a, b), Solanki and Singh (2013), Yadav et al. (2013) and Bauza et al. (2019) have given their tremendous efforts in suggesting different type of estimators for estimating the population variance. Further Yadav and Kadilar (2013) proposed an improved exponential type ratio estimator for the estimation of population variance while Adichwal et al. (2013), Asghar et al. (2014), Yasmeen et al. (2015, 18) and Sanuallah et al. (2016) suggested some class/generalized exponential-type estimators using the information of auxiliary character for estimating the population variance.
Consider a finite population $\xi_N = \{\xi_i;\ i = 1,2,\dots,N\}$ of N identifiable units. Let y and x are highly correlated study and auxiliary characters observed on ξ_i $(i = 1,2,\dots,N)$ having non-negative i^{th} value with population mean $\bar{Y}$ and $\bar{X}$ respectively. Let a sample ξ_n of size 'n' is drawn from ξ_N and assume ξ_{n_1} of n_1

[1] E-mail: raghawraman@gmail.com (Corresponding author)
[2] E-mail: vinod7865@gmail.com

units respond and ξ_{n_2} of $n_2 = (n - n_1)$ units do not respond for the study character 'y'. Suppose the population ξ_N be constituted by two unknown non-overlapping strata ξ_{N_1} and ξ_{N_2} of responding and non-responding soft core groups [see Khare and Srivastava (1996)] with stratum weights $\mathcal{W}_1 (= N_1/N)$ for responding and $\mathcal{W}_2 (= N_2/N)$ for non-responding groups. Let $\omega_1 = n_1/n$ and $\omega_2 = n_2/n$ are the estimates of stratum weights of $\mathcal{W}_1$ and $\mathcal{W}_2$ respectively. Since the estimation of parameters based on only the responding n_1 units, creates bias because it cannot be treated as the representative of whole population. So, to reduce the amount of bias, a sub-sample $\xi_{n(2m)}$ of size $m = n_2\alpha^{-1}(\alpha > 1)$ units from ξ_{n_2}at second stage has been drawn using simple random sampling without replacement (SRSWOR) method and information on these 'm' units has been collected through personal interview. Let $\bar{y}_1$ and $\bar{y}_{n(2)}$ are the sample means of study character 'y_i' based on 'n_1' and 'm' units respectively. Hansen and Hurwitz (1946) proposed an unbiased estimator for population mean $\bar{Y}\left(=\frac{1}{N}\sum_{U_N} y_i\right)$ on total available information $(n_1 + m)$ units of the study character 'y' as

$$\bar{y}^* = \omega_1\bar{y}_1 + \omega_2\bar{y}_{n(2)}. \tag{1}$$

The variance of $\bar{y}^*$ up to the order (n^{-1}) is given by

$$V(\bar{y}^*) = \pi\mathbb{S}_y^2 + \gamma_2\mathbb{S}_{y(2)}^2\ , \tag{2}$$

where $\mathbb{S}_y^2 = (N-1)^{-1}\sum_{\xi_N}(y_i - \bar{Y})^2$, $\mathbb{S}_{y(2)}^2 = (N_2-1)^{-1}\sum_{\xi_{N_2}}\left(y_i - \bar{Y}_{(2)}\right)^2$,

$\bar{Y}_{(2)} = (N_2)^{-1}\sum_{\xi_{N_2}} y_i$, $\pi = \frac{N-n}{Nn}$, $\gamma_2 = \frac{\mathcal{W}_2(\alpha-1)}{n}$.

Okafor and Lee (2000) discussed an estimator to estimate the population variance $\mathbb{S}_y^2$ of study character y in presence of non-response as

$$T_1^* = s_y^{*2} = (n-1)^{-1}\left\{\sum_{\xi_{n_1}} y_i^2 + \alpha\sum_{\xi_{n(2m)}} y_i^2 - n\bar{y}^{*2}\right\}, \tag{3}$$

such that $E(T_1^*) = \mathbb{S}_y^2$ and its variance up to the order (n^{-1}) is given by

$$V_y^* = V\left(s_y^{*2}\right) = \pi\{\beta_2(y) - 1\}\mathbb{S}_y^4 + \gamma_2\left\{\beta_{2(2)}(y) - 1\right\}\mathbb{S}_{y(2)}^4, \tag{4}$$

where $\beta_2(y) = \frac{\mu_{40}}{\mu_{20}^2}$, $\beta_{2(2)}(y) = \frac{\mu_{40(2)}}{\mu_{20(2)}^2}$ and $\mu_{rs} = (N)^{-1}\sum_{i=1}^{N}(y_i - \bar{Y})^r(x_i - \bar{X})^s$.

Since we have considered the incident that the questionnaire remains empty for some elements in the sample i.e. no answers are obtained from the non-responding individuals [See Bethlehem et al. (2011), page 209], therefore, an unbiased estimator for estimating the population variance $\mathbb{S}_x^2$ of auxiliary character may be given by

$$T_2^* = s_x^{*2} = (n-1)^{-1}\left(\sum_{\xi_{n_1}} x_i^2 + \alpha\sum_{\xi_{n(2m)}} x_i^2 - n\bar{x}^{*2}\right), \tag{5}$$

where
$$\bar{x}^* = p_1\bar{x}_1 + p_2\bar{x}_{n(2)} \tag{6}$$

with $\bar{x}_1 = (n_1)^{-1}\sum_{\xi_{n_1}} x_i$ and $\bar{x}_{n(2)} = (m_1)^{-1}\sum_{\xi_{n(2m)}} x_i$. The estimator T_2^* and $\bar{x}^*$ are unbiased estimators of $\mathbb{S}_x^2 = (N-1)^{-1}\sum_{\xi_N}(x_i - \bar{X})^2$ and $\bar{X}\left\{= (N^{-1})\sum_{\xi_N} x_i\right\}$ respectively and their variance up to the order of (n^{-1}) are given by

$$V_x^* = V\left(s_x^{*2}\right) = \pi\{\beta_2(x) - 1\}\mathbb{S}_x^4 + \gamma_2\left\{\beta_{2(2)}(x) - 1\right\}\mathbb{S}_{x(2)}^4\ , \tag{7}$$

$$V(\bar{x}^*) = \pi\mathbb{S}_x^2 + \gamma_2(\alpha - 1)\mathbb{S}_{x(2)}^2 \tag{8}$$

where $\beta_2(x) = \frac{\mu_{04}}{\mu_{02}^2}$, $\beta_{2(2)}(x) = \frac{\mu_{04(2)}}{\mu_{02(2)}^2}$,

$\mathbb{S}_x^2 = (N-1)^{-1}\sum_{\xi_N}(x_i - \bar{X})^2$,

$\mathbb{S}_{x(2)}^2 = (N_2-1)^{-1}\sum_{\xi_{N_2}}\left(x_i - \bar{X}_{(2)}\right)^2$

and $\bar{X}_{(2)} = (N_2)^{-1}\sum_{\xi_{N_2}} x_i$.

For the case of highly positive correlation between study and auxiliary characters, the estimator proposed by Isaki (1983) for estimating the finite population variance $\mathbb{S}_y^2$ under complete information on study and auxiliary characters with known population variance $\mathbb{S}_x^2\left(= (N-1)^{-1}\sum_{\xi_N}(x_i - \bar{X})^2\right)$ of auxiliary character 'x' is

$$t_1 = s_y^2\left(\frac{\mathbb{S}_x^2}{s_x^2}\right) \tag{9}$$

where $s_x^2 = (n-1)^{-1}\sum_{\xi_n}(x_i - \bar{x})^2$ is an unbiased estimator of $\mathbb{S}_x^2$ with sample mean $\bar{x} = \frac{1}{n}\sum_{U_n} x_i$.

The mean square error (MSE) of t_1 up to the first order of approximation is given by

$$MSE(t_1) = \pi\left[\beta_2(y) + \beta_2(x) - 2\frac{\mu_{22}}{\mu_{20}\mu_{02}}\right]\mathbb{S}_y^4 \tag{10}$$

For the same situation, Garcia and Cebran (1996) proposed a generalized estimator for $\mathbb{S}_y^2$ as

$$t_2 = s_y^2\left(\frac{\mathbb{S}_x^2}{s_x^2}\right)^{\alpha} \tag{11}$$

where α is the constant and through this way the estimator proposed by Isaki (1983) becomes a particular member of t_2 for $\alpha = 1$.

Further, Upadhaya and Singh (2001) suggested a difference type estimator for estimating $\mathbb{S}_y^2$ as

$$t_3 = s_y^2 + b(\mathbb{S}_x^2 - s_x^2) \tag{12}$$

where b is a constant.

An important property of the estimators t_2 and t_3 is that both the estimators attain their minimum MSE for optimum value of the constants α and b and is given by

$$MSE(t_2)_{min.} = MSE(t_3)_{min.} = \pi\left[\{\beta_2(y)-1\} - \frac{\left\{\frac{\mu_{22}}{\mu_{20}\mu_{02}}-1\right\}^2}{\{\beta_2(x)-1\}}\right]\mathbb{S}_y^4 \tag{13}$$

If we compare the MSE of t_2 and t_3 with the variance of s_y^2 i.e. $V(s_y^2) = \pi\,[\beta_2(y)-1]S_y^4$, we find

$$V(s_y^2) - MSE(t_2)_{min.} = V(s_y^2) - MSE(t_3)_{min.} = \frac{\pi\left\{\frac{\mu_{22}}{\mu_{20}\mu_{02}}-1\right\}^2\mathbb{S}_y^4}{\{\beta_2(x)-1\}} \geq 0 \tag{14}$$

Following Garcia and Cebran (1996) and Upadhaya and Singh (2001), generalized and difference type estimator for estimating $\mathbb{S}_y^2$ may be suggested for two different cases **(I)** when non-response occurs on both study and auxiliary characters, and **(II)** when non-response occurs only on study and character [cases discussed in Rao (1986)], which are given by

For case (I)

$$t_{1(1)} = s_y^{*2}\left(\frac{S_x^{*2}}{\mathbb{S}_x^2}\right)^{a_{1(1)}}$$

$$t_{2(1)} = s_y^{*2} + a_{2(1)}(\mathbb{S}_x^2 - s_x^{*2})$$

For case (II)

$$t_{1(2)} = s_y^{*2}\left(\frac{S_x^2}{\mathbb{S}_x^2}\right)^{a_{1(2)}}$$

$$t_{2(2)} = s_y^{*2} + a_{2(2)}(\mathbb{S}_x^2 - s_x^2)$$

Some other parametric functions like $t_{i(j)};\ i = j = 1,2$ may also be formed to estimate $\mathbb{S}_y^2$ for case (I) and case (II) as follows:

For case (I)

$$t_{3(1)} = s_y^{*2}\left(\frac{\mathbb{S}_x^2 + a_{3(1)}}{s_x^{*2} + a_{3(1)}}\right)$$

$$t_{4(1)} = s_y^{*2}\left(\frac{a_{4(1)} + 1}{a_{4(1)} + \frac{s_x^{*2}}{\mathbb{S}_x^2}}\right)$$

$$t_{5(1)} = s_y^{*2}\left[a_{5(1)} + (1 - a_{5(1)})\frac{s_x^{*2}}{\mathbb{S}_x^2}\right]$$

$$t_{6(1)} = \frac{s_y^{*2} + a_{6(1)}(\mathbb{S}_x^2 - s_x^{*2})}{\left(\ell_{1(1)}s_x^2 + \ell_{2(1)}\right)}\ \left(\ell_{1(1)}\mathbb{S}_x^2 + \ell_{2(1)}\right)$$

For case (II)

$$t_{3(2)} = s_y^{*2}\left(\frac{\mathbb{S}_x^2 + a_{3(2)}}{s_x^2 + a_{3(2)}}\right)$$

$$t_{4(2)} = s_y^{*2}\left(\frac{a_{4(2)} + 1}{a_{4(2)} + \frac{s_x^2}{\mathbb{S}_x^2}}\right)$$

$$t_{5(2)} = s_y^{*2}\left[a_{5(2)} + (1 - a_{5(2)})\frac{s_x^2}{\mathbb{S}_x^2}\right]$$

$$t_{6(2)} = \frac{s_y^{*2} + a_{6(2)}(\mathbb{S}_x^2 - s_x^2)}{\left(\ell_{1(2)}s_x^2 + \ell_{2(2)}\right)}\ \left(\ell_{1(2)}\mathbb{S}_x^2 + \ell_{2(2)}\right)$$

etc., where $a_{i(j)}; i = 1,2,\ldots 6;\ j = 1,2$ are suitably chosen constants and $\ell_{i(j)}; i = j = 1,2$ are real numbers or the functions of the known parameters.

As all the estimators $t_{i(j)}; i = 1,2,\ldots 6;\ j = 1,2$ are the function of s_y^{*2} and either $u_1 = \frac{s_x^{*2}}{\mathbb{S}_x^2}$ or $u_2 = \frac{s_x^2}{\mathbb{S}_x^2}$, therefore wider classes of estimators for estimating the population mean $\mathbb{S}_y^2$ may be defined as

$$T_{c(i)} = f^{(i)}(z, u_i);\ i = 1,2 \tag{15}$$

such that $f^{(i)}(\mathbb{S}_y^2, 1) = \mathbb{S}_y^2$ and $\left(\frac{\partial T_{c(i)}}{\partial z}\right)_{(\bar{Y},1)} = 1$,

where $z = s_y^{*2}$, $u_1 = \frac{s_x^{*2}}{\mathbb{S}_x^2}$ and $u_2 = \frac{s_x^2}{\mathbb{S}_x^2}$.

To expend and obtain the mean square error of $T_{c(i)}$, it is assumed that the first and second partial derivatives of the function $f^{(i)}(z,u_i)$ exist and are continuous and bounded in closed convex sub set D_i, it is also supposed that the function $f^{(i)}(z,u_i)$ satisfies necessary regularity conditions for its expansion about the point $(z,u_i)=(\mathbb{S}_y^2,1)$. Following Khare and Sinha (2012) and using Taylor's series expansion, the minimum mean square error of $T_{c(i)}; i=1,2$ can be obtained and is given by

$$MSE(T_{c(i)})_{min.} = V(z) - \frac{[Cov(z,u_i)]^2}{V(u_i)}, \; i=1,2 \tag{16}$$

for respective value of $\left(\frac{\partial f^{(i)}}{\partial u_i}\right)_{(1,1)} = -\frac{Cov(z,u_i)}{V(u_i)}; i=1,2.$

It is pertinent to mention here that all the estimators considered earlier utilize the same information of known population variance of auxiliary character and all are particular members of the general classes of estimators $T_{c(i)}$ either for $i=1$ or 2. Hence, they will be either less efficient or at the most equally efficient than corresponding $T_{c(i)}; i=1,2$ up to the first order of approximation under specified conditions.

In this paper, the aim is to propose such types of estimators for estimating the population variance $\mathbb{S}_y^2$ which are more efficient than the usual unbiased estimator and the classes of estimators $T_{c(i)}; i=1,2$ under the said conditions, therefore, improved regression-cum-exponential estimators for two different cases are proposed, which are not the members of $T_{c(i)}; i=1,2$ and their properties are studied. Theoretical and empirical comparisons are carried out to show that the proposed estimators are always more efficient than all the existing estimators considered in this paper.

2. THE ESTIMATOR AND ITS PROPERTIES

Following Sinha and Kumar (2017), the two different proposed estimators for two different cases using the known value of $\mathbb{S}_x^2$ are as follows:

For **case (I):** when non-response occurs on both study and auxiliary characters

$$T_1 = \left[c_1 s_y^{*2} + d_1(\mathbb{S}_x^2 - s_x^{*2})\right] exp[(\mathbb{S}_x^2 - s_x^{*2})(\mathbb{S}_x^2 + s_x^{*2})^{-1}] \tag{17}$$

For **case (II):** when non-response occurs only on study and character

$$T_2 = \left[c_2 s_y^{*2} + d_2(\mathbb{S}_x^2 - s_x^{2})\right] exp[(\mathbb{S}_x^2 - s_x^{2})(\mathbb{S}_x^2 + s_x^{2})^{-1}] \tag{18}$$

where c_1, c_2, d_1 and d_2 are the scalars.

To obtain the bias and mean square error of $T_i, i=1,2$, it has been assumed that the population ξ_N is large enough compared to the sample ξ_n so that the following approximations may define

$$s_y^{*2} = \mathbb{S}_y^2(1+\bar{\epsilon}_0),\; s_x^{*2} = \mathbb{S}_x^2(1+\bar{\epsilon}_1),\; s_x^2 = \mathbb{S}_x^2(1+\bar{\epsilon}_3) \text{ such that } E(\bar{\epsilon}_i)=0,\; i=0,1,2.$$

Now, the proposed estimator T_1 is defined in terms of $\bar{\epsilon}_i$ as

$$T_1 = \left[c_1\mathbb{S}_y^2(1+\bar{\epsilon}_0) - d_1\mathbb{S}_x^2\bar{\epsilon}_1\right] exp[-\bar{\epsilon}_1(2+\bar{\epsilon}_1)^{-1}]$$

or

$$T_1 = \left[c_1\mathbb{S}_y^2(1+\bar{\epsilon}_0) - d_1\mathbb{S}_x^2\bar{\epsilon}_1\right] exp\left[-\frac{\bar{\epsilon}_1}{2}\left(1 - \frac{\bar{\epsilon}_1}{2} + \frac{\bar{\epsilon}_1^2}{4} - \cdots \ldots \ldots \ldots\right)\right]$$

Neglecting the terms of order higher than (n^{-1}), we have

$$T_1 - \mathbb{S}_y^2 = \mathbb{S}_y^2\left[(c_1-1) - \frac{c_1\bar{\epsilon}_1}{2} + \frac{3c_1\bar{\epsilon}_1^2}{8} + c_1\bar{\epsilon}_0 - \frac{c_1\bar{\epsilon}_0\bar{\epsilon}_1}{2}\right] - d_1\mathbb{S}_x^2\left(\bar{\epsilon}_1 - \frac{\bar{\epsilon}_1^2}{2}\right). \tag{19}$$

Similarly, we may express the estimator T_2 in terms of $\bar{\epsilon}_i$'s is

$$T_2 - \mathbb{S}_y^2 = \mathbb{S}_y^2\left[(c_2-1) - \frac{c_2\bar{\epsilon}_3}{2} + \frac{3c_2\bar{\epsilon}_3^2}{8} + c_2\bar{\epsilon}_0 - \frac{c_2\bar{\epsilon}_0\bar{\epsilon}_3}{2}\right] - d_2\mathbb{S}_x^2\left(\bar{\epsilon}_3 - \frac{\bar{\epsilon}_3^2}{2}\right) \tag{20}$$

On taking the expectations on both sides of equations (19) and (20), the expressions of bias of T_1 and T_2 up to the first order of approximation are given by

$$Bias(T_1) = \mathbb{S}_y^2\left[(c_1-1) + \frac{3c_1}{8}\frac{V(s_x^{*2})}{\mathbb{S}_x^4} - \frac{c_1}{2}\frac{Cov(s_x^{*2},s_y^{*2})}{\mathbb{S}_x^2\mathbb{S}_y^2}\right] + d_1\mathbb{S}_x^2\frac{V(s_x^{*2})}{\mathbb{S}_x^4} \tag{21}$$

and

$$Bias(T_2) = \mathbb{S}_y^2\left[(c_2-1) + \frac{3c_2}{8}\frac{V(s_x^{2})}{\mathbb{S}_x^4} - \frac{c_1}{2}\frac{Cov(s_x^{2},s_y^{*2})}{\mathbb{S}_x^2\mathbb{S}_y^2}\right] + d_2\mathbb{S}_x^2\frac{V(s_x^{2})}{\mathbb{S}_x^4} \tag{22}$$

Now squaring both sides of equations (19) and (20) and taking expectation, the expressions of mean square error of T_1 and T_2 up to the first order approximation are given by

$$MSE(T_1) = \mathbb{S}_y^4\left\{(c_1-1)^2 + c_1^2Q_1 - \frac{c_1}{2}\left[Q_2 + \frac{V(s_x^{*2})}{2\mathbb{S}_x^4}\right] - \frac{Cov(s_x^{*2},s_y^{*2})}{\mathbb{S}_x^2\mathbb{S}_y^2}\right\} + d_1^2V(s_x^{*2})$$
$$+2d_1\mathbb{S}_x^2\mathbb{S}_y^2\left[c_1Q_2 - \frac{V(s_x^{*2})}{2\mathbb{S}_x^4}\right] \tag{23}$$

and $$MSE(T_2) = \mathbb{S}_y^4 \left\{(c_2 - 1)^2 + c_2^2 Q_1^* - \frac{c_2}{2}\left[Q_2^* + \frac{V(s_x^2)}{2\mathbb{S}_x^4}\right] - \frac{Cov(s_x^2, s_y^{*2})}{\mathbb{S}_x^2 \mathbb{S}_y^2}\right\} + d_2^2 V(s_x^2)$$
$$+2d_2 \mathbb{S}_x^2 \mathbb{S}_y^2 \left[c_2 Q_2^* - \frac{V(s_x^2)}{2\mathbb{S}_x^4}\right], \tag{24}$$

where, $Cov(s_x^{*2}, s_y^{*2}) = \pi \mathbb{S}_x^2 \mathbb{S}_y^2 \left[\frac{\mu_{22}}{\mu_{20}\mu_{02}} - 1\right] + \frac{W_2(k-1)}{n} \mathbb{S}_{x(2)}^2 \mathbb{S}_{y(2)}^2 \left[\frac{\mu_{22(2)}}{\mu_{20(2)}\mu_{02(2)}} - 1\right]$,

$Cov(s_x^2, s_y^{*2}) = \pi \mathbb{S}_x^2 \mathbb{S}_y^2 \left[\frac{\mu_{22}}{\mu_{20}\mu_{02}} - 1\right]$,

$Q_1 = \frac{V(s_y^{*2})}{\mathbb{S}_y^4} + \frac{V(s_x^{*2})}{\mathbb{S}_x^4} - \frac{2Cov(s_x^{*2}, s_y^{*2})}{\mathbb{S}_x^2 \mathbb{S}_y^2}$, $\quad Q_2 = \frac{V(s_x^{*2})}{\mathbb{S}_x^4} - \frac{2Cov(s_x^{*2}, s_y^{*2})}{\mathbb{S}_x^2 \mathbb{S}_y^2}$

$Q_1^* = \frac{V(s_y^{*2})}{\mathbb{S}_y^4} + \frac{V(s_x^2)}{\mathbb{S}_x^4} - \frac{2Cov(s_x^2, s_y^{*2})}{\mathbb{S}_x^2 \mathbb{S}_y^2}$, $\quad Q_2^* = \frac{V(s_x^2)}{\mathbb{S}_x^4} - \frac{2Cov(s_x^2, s_y^{*2})}{\mathbb{S}_x^2 \mathbb{S}_y^2}$.

Now, differentiate partially equations (23) and (24) with respect to c_1, c_2, d_1 and d_2 and solving them, we have the optimum value of c_1, c_2, d_1 and d_2.as follows

$$c_1 = \frac{1}{(Q_1 + 1)}\left[1 - d_1 \frac{\mathbb{S}_x^2}{\mathbb{S}_y^2} Q_2 + \frac{1}{4}\left(Q_2 + \frac{V(s_x^{*2})}{2\mathbb{S}_x^4} - \frac{Cov(s_x^{*2}, s_y^{*2})}{\mathbb{S}_x^2 \mathbb{S}_y^2}\right)\right], \tag{25}$$

$$d_1 = \frac{\mathbb{S}_y^2}{2\mathbb{S}_x^2} - c_1 \frac{Q_2 \mathbb{S}_x^2 \mathbb{S}_y^2}{V(s_x^{*2})}, \tag{26}$$

$$c_2 = \frac{1}{(Q_1^* + 1)}\left[1 - d_2 \frac{\mathbb{S}_x^2}{\mathbb{S}_y^2} Q_2^* + \frac{1}{4}\left(Q_2^* + \frac{V(s_x^2)}{2\mathbb{S}_x^4} - \frac{Cov(s_x^2, s_y^{*2})}{\mathbb{S}_x^2 \mathbb{S}_y^2}\right)\right], \tag{27}$$

$$d_2 = \frac{\mathbb{S}_y^2}{2\mathbb{S}_x^2} - c_2 \frac{Q_2^* \mathbb{S}_x^2 \mathbb{S}_y^2}{V(s_x^2)}. \tag{28}$$

Generally in practice, it is difficult to obtain the optimum values of c_1, c_2, d_1 and d_2 due to the involvement of various unknown quantities like population variance and covariance of the variables, therefore from the practical point of view, Reddy (1978) suggested that up to the first order of approximation, there would not be any effect on the efficiency of the estimator if one uses prior information available from past data or good presumed values of these parameters from a survey conducted in recent past or by conducting a preliminary survey. However, if no information is available from the past data, Srivastava and Jhajj (1983) have shown that there would not be any effect on the efficiency of the estimators up to the terms of order (n^{-1}) if the unknown parameters are replaced with the corresponding consistent estimate.

Putting the value of (c_1, d_1) from equations (25) and (26) in equation (23) and (c_2, d_2) from equations (27) and (28) in equation (24), we get the minimum mean square error of T_1 and T_2 respectively as

$$MSE(T_1)_{min.} = \frac{\left[V(s_y^{*2}) - \frac{[Cov(s_x^{*2}, s_y^{*2})]^2}{V(s_x^{*2})}\right]}{\left[1 + \frac{1}{\mathbb{S}_y^4}\left[V(s_y^{*2}) - \frac{[Cov(s_x^{*2}, s_y^{*2})]^2}{V(s_x^{*2})}\right]\right]} - \frac{\frac{V(s_x^{*2})}{\mathbb{S}_x^4}\left[4\left\{V(s_y^{*2}) - \frac{[Cov(s_x^{*2}, s_y^{*2})]^2}{V(s_x^{*2})}\right\} + \frac{\mathbb{S}_y^4}{4\mathbb{S}_x^4} V(s_x^{*2})\right]}{16\left[1 + \frac{1}{\mathbb{S}_y^4}\left\{V(s_y^{*2}) - \frac{[Cov(s_x^{*2}, s_y^{*2})]^2}{V(s_x^{*2})}\right\}\right]} \tag{29}$$

and $$MSE(T_2)_{min.} = \frac{\left[V(s_y^{*2}) - \frac{[Cov(s_x^2, s_y^{*2})]^2}{V(s_x^2)}\right]}{\left[1 + \frac{1}{\mathbb{S}_y^4}\left[V(s_y^{*2}) - \frac{[Cov(s_x^2, s_y^{*2})]^2}{V(s_x^2)}\right]\right]} - \frac{\frac{V(s_x^{*2})}{\mathbb{S}_x^4}\left[4\left\{V(s_y^{*2}) - \frac{[Cov(s_x^2, s_y^{*2})]^2}{V(s_x^{*2})}\right\} + \frac{\mathbb{S}_y^4}{4\mathbb{S}_x^4} V(s_x^2)\right]}{16\left[1 + \frac{1}{\mathbb{S}_y^4}\left\{V(s_y^{*2}) - \frac{[Cov(s_x^2, s_y^{*2})]^2}{V(s_x^{*2})}\right\}\right]}. \tag{30}$$

3. CONCLUDING REMARKS

To get the conclusion over efficiency of our proposed estimators T_1 and T_2 with the different relevant estimators, we compared the mean square error of the proposed estimators T_1 and T_2 with the other relevant estimators and we get

a) From equation (16)

$$V(s_y^{*2}) - MSE(T_{c(1)})_{min.} = \frac{[Cov(s_x^{*2}, s_y^{*2})]^2}{V(s_x^{*2})} > 0 \tag{31}$$

and $$V(s_y^{*2}) - MSE(T_{c(2)})_{min.} = \frac{[Cov(s_x^2, s_y^{*2})]^2}{V(s_x^2)} > 0. \tag{32}$$

b) From equations (29) and (30)

$$MSE(T_{c(1)})_{min.} - MSE(T_1)_{min.} = \frac{\frac{\left[MSE(T_{c(1)})_{min.}\right]^2}{\mathbb{S}_y^4}}{1+\frac{MSE(T_{c(1)})_{min.}}{\mathbb{S}_y^4}} + \frac{\left[MSE(T_{c(1)})_{min.}+\frac{\mathbb{S}_y^4}{16\mathbb{S}_x^4}V(s_x^{*2})\right]}{4\left[1+\frac{MSE(T_{c(1)})_{min.}}{\mathbb{S}_y^4}\right]} > 0 \qquad (33)$$

and
$$MSE(T_{c(2)})_{min.} - MSE(T_2)_{min.} = \frac{\frac{\left[MSE(T_{c(2)})_{min.}\right]^2}{\mathbb{S}_y^4}}{1+\frac{MSE(T_{c(2)})_{min.}}{\mathbb{S}_y^4}} + \frac{\left[MSE(T_{c(2)})_{min.}+\frac{\mathbb{S}_y^4}{16\mathbb{S}_x^4}V(s_x^{2})\right]}{4\left[1+\frac{MSE(T_{c(2)})_{min.}}{\mathbb{S}_y^4}\right]} > 0 \qquad (34)$$

c) From equations (31) , (32), (33) and (34)

$$MSE(T_1)_{min.} < MSE(T_{c(1)})_{min.} < V\left(s_y^{*2}\right) \qquad (35)$$

and
$$MSE(T_2)_{min.} < MSE(T_{c(2)})_{min.} < V\left(s_y^{*2}\right) \qquad (36)$$

Hence, we conclude from equations (35) and (36) that the proposed estimators T_1 and T_2 are always more efficient than all the existing estimators in the circumstances discussed in the paper.

4. EMPIRICAL STUDY

[Source: Singh and Chaudhary, (1986), page 142] In this data set, the 34% villages (i.e. 10 villages) from bottom of the list have been considered as non-responding group of the population. Here we have considered area under guava crops as study character (y) and area under fresh fruits as auxiliary character (x). The parameters under study are as follows:

$N = 47$	$n = 14$	$\bar{Y} = 23.83$	$\bar{X} = 29.75$
$\mathbb{S}_y^2 = 742.39$	$\mathbb{S}_x^2 = 1474.72$	$\mu_{04} = 3569542.53$	$\mu_{40} = 15216459.36$
$\mathbb{S}_{y(2)}^2 = 36.93$	$\mathbb{S}_{x(2)}^2 = 83.70$	$\mu_{04(2)} = 3567.71$	$\mu_{40(2)} = 18100.73$
	$\mu_{22} = 7047068.18$	$\mu_{22(2)} = 7373.25$	

To examine the gain in efficiency of our proposed estimators T_1 and T_2, the relative efficiency $(in\ \%)$ of $T_i; i = 1,2$ and $T_{C(i)}; i = 1,2$ with respect to s_y^{*2} at different level of sub-sampling fraction$1/\alpha$ are computed by

$$R.E.(.) = \frac{V\left(s_y^{*2}\right)}{MSE(.)} \times 100$$

and values are given in Table 1 along with their mean square error.

Table 1: *Mean Square error (.) and* $R.E.$ *(in %) of the estimators under fixed sample approach* $(N = 47$ and $n = 14)$

$1/\alpha$	Estimators				
	$V\left(s_y^{*2}\right)$	$MSE(T_{c(1)})_{min.}$	$MSE(T_1)_{min.}$	$MSE(T_{c(2)})_{min.}$	$MSE(T_2)_{min.}$
$\alpha = 10$	151832 (100.00)*	15259.3 (995.01)	12964.7 (1171.12)	15588 (974.03)	13262.0 (1144.87)
$\alpha = 9$	151782 (100.00)	15245.5 (995.58)	12953.6 (1171.74)	15537.6 (976.86)	13217.9 (1148.31)
$\alpha = 8$	151731 (100.00)	15231.6 (996.16)	12942.5 (1172.35)	15487.3 (982.59)	13173.7 (1151.77)
$\alpha = 7$	151681 (100.00)	15217.8 (996.74)	12931.3 (1172.97)	15436.9 (985.48)	13129.6 (1155.26)
$\alpha = 6$	151631 (100.00)	15203.9 (997.314)	12920.2 (1173.59)	15386.5 (985.48)	13085.4 (1158.77)
$\alpha = 5$	151580 (100.00)	15190.1 (997.89)	12909.1 (1174.21)	15336.1 (988.39)	13041.3 (1162.31)
$\alpha = 4$	151530 (100.00)	15176.2 (998.47)	12897.9 (1174.84)	15285.8 (991.31)	12997.1 (1165.88)
$\alpha = 3$	151479 (100.00)	15162.3 (999.05)	12886.8 (1175.46)	15235.4 (994.26)	12952.9 (1169.46)

$\alpha = 2$	151429 (100.00)	15148.5 (999.63)	12875.7 (1176.09)	15185.0 (997.23)	12908.7 (1173.08)

*Figures in parenthesis show the R. E. (in %) of the estimator

5. CONCLUSION

From Table 1, we observe that the proposed estimators $T_i; i = 1,2$ are more efficient than usual unbiased estimator s_y^{*2} and corresponding classes of estimators $T_{C(i)}; i = 1,2$ at all the level of sub-sampling fractions $1/\alpha$. The mean square error of $T_i; i = 1,2$ is decreasing by increasing the value of sub-sampling fraction $1/\alpha$ while $R.E.$ of $T_i; i = 1,2$ with respect to usual unbiased estimator s_y^{*2} is increasing correspondingly because the variance $\left[V\left(s_y^{*2}\right)\right]$ is decreasing at a faster rate compared to $T_i; i = 1,2$. Therefore, on the basis of theoretical and empirical studies, it has been clear that the proposed estimators $T_i; i = 1,2$ will give the better estimate and precision in large sample surveys under the situation discussed in the text.

RECEIVED: JANUARY, 2020.
REVISED: MAY, 2020.

REFERENCES

[1] ADICHWAL, N.K., SHARMA, P., VERMA, H. K. and SINGH, R. (2015): Generalized class of estimators for estimating population variances using auxiliary attribute. **International Journal of Applied and Computational Mathematics,** 1, 1–10.
[2] ASGHAR, A., SANAULLAH, A. and HANIF, M. (2014): Generalized exponential-type estimator for population variance in survey sampling. **Revista Colombiana de Estadistica,** 37, 213–224.
[3] BOUZA, C. N., GARCIA, J. F., VISHWAKARMA, G. K. AND ZEESHAN, S. M. (2019): Variance estimation of persons infected with AIDS under ranked set sampling. **Clinical Epidemiology and Global Health,** 7, 17-21.
[4] BETHLEHEM, J. G., COBBEN, F. and SCHOUTEN, B. (2011): **Handbook of Non-response in Household Surveys**. Wiley, Hoboken, New Jersey.
[5] DAS, A. K. and TRIPATHI, T. P. (1978): Use of auxiliary information in estimating the finite population variance. **Sankhya,** 4(C), 139-148.
[6] DUBEY, V. and SHARMA, H. K. (2008): On estimation population variance using auxiliary information. **Statistics in Transition-new series,** 9, 7-18.
[7] GARCIA, M. R. and CEBRIAN, A. A. (1996): Repeated substitution method-the ratio estimator for the population variance. **Metrika,** 43, 101-105.
[8] GUPTA, S. and SHABBIR, J. (2008): Variance estimation in simple random sample using auxiliary information. **Hacettepe Journal of Mathematics and Statistics,** 37, 57-67.
[9] HANSEN, M. H. and HURWITZ, W. N. (1946): The problem of non-response in sample surveys. **Journal of the American Statistical Association,** 41, 517-529.
[10] ISAKI, C. T. (1983): Variance estimation using auxiliary information. **Journal of the American Statistical Association,** 78, 117-123.
[11] JHAJJ, H. S., SHARMA, M. K. and GROVER, L. K. (2005): An efficient class of chain estimators of population variance under sub-sampling scheme. **Journal of the Japan Statistical Society,** 35, 273-286.
[12] KADILAR, C. and CINGI, H. (2007): Improvement in variance estimation in simple random sampling. **Communications in Statistics - Theory and Methods,** 36, 2075-2081.
[13] KHARE, B. B. and SINHA, R. R. (2012): Combined Class of Estimators for Ratio and Product of Two Population Means in Presence of Non-Response. **International Journal of Statistics and Economics,** 8, 86-95.
[14] KHARE, B. B. and SRIVASTAVA, S. (1996): Transformed product type estimators for population mean in presence of softcore observations. **Proc. Math. Soc.,** 12, 29-34.
[15] OKAFOR, F. C. and LEE, H. (2000): Double sampling for ratio and regression estimation with sub sampling the non-respondent**. Survey Methodology,** 26, 183-188.
[16] PRASAD, B. and SINGH, H. P. (1990): Some improved ratio-type estimators of finite population variance using auxiliary information in sample surveys. **Communications in Statistics - Theory and Methods,** 19, 1127-1139.

[17] RAO, P. S. R. S. (1986): Ratio estimation with subsampling the nonrespondents. **Survey Methodology,** 12, 217-230.
[18] REDDY, V. N. (1978): A study of use of prior knowledge on certain population parameters in estimation. **Sankhya,** 40(C), 29-37.

[19] SANAULLAH, A., HANIF, M., ASHAR, A. (2016): Generalized exponential estimators for population variance under two-phase sampling. **International Journal of Computational Mathematics,** 2, 75-84.
[20] SINGH, D. and CHAUDHARY, F. S. (1986): **Theory and analysis of sample survey design**. Wiley, New York.
[21] SINGH, H. P. and SOLANKI R. S. (2013a): Improved estimation of finite population variance using auxiliary information. **Communications in Statistics - Theory and Methods,** 42, 2718-2730.
[22] SINGH, H. P. and SOLANKI, R. S. (2013b): A new procedure for variance estimation in simple random sampling using auxiliary information. **Statistical Paper,** 54, 479-497.
[23] SINGH, J., PANDEY, B. N. and HIRANO, K. (1973): On the utilization of a known coefficient of kurtosis in estimation procedure of variance. **Annals of the Institute of Statistical Mathematics,** B, 51-55.
[24] SINHA, R. R. and KUMAR, V. (2017): Regression cum exponential estimators for finite population mean under incomplete information. **Journal of Statistics and Management Systems,** 20, 355-268.
[25] SOLANKI, R. S. and SINGH, H. P. (2013): An improved class of estimators for the population variance. **Model Assisted Statistics and Applications,** 8, 229–238.
[26] SRIVASTAVA, S. K. and JHAJJ, H. S. (1983): A class of estimators of the population mean using multi-auxiliary information. **Calcutta Statistical Association Bulletin,** 32, 47-56.
[27] SUBRAMANI, J. and KUMARAPANDIYAN, G. (2012): Variance estimation using quartiles and their functions of an auxiliary variable. **International Journal of Statistics and Applications,** 2, 67–72.
[28] TURGUT, Y. and CINGI, H. (2008): New generalized estimators for the population variance using auxiliary information. **Hacettepe Journal of Mathematics and Statistics,** 37, 177-184.
[29] UPADHYAYA, L. N. and SINGH, H. P. (2001): Estimation of the population standard deviation using auxiliary information. **American Journal of Mathematical and Management Sciences,** 21, 345-358.
[30] YADAV, R., UPADHYAYA, L. N., SINGH, H. P. and CHATTERJEE, S. (2013): A generalized family of transformed ratio-product estimators for variance in sample surveys. **Communications in Statistics - Theory and Methods,** 42, 1839–1850.
[31] YADAV, S. K. and KADILAR, C. (2013): Improved exponential type ratio estimator of population variance. **Revista Colombiana de Estadistica,** 36, 145–152.
[32] YASMEEN, U., UL AMIN, M. N. and HANIF, M. (2015): Generalized exponential estimators of finite population mean using transformed auxiliary variables. **International Journal of Applied and Computational Mathematics,** 1, 1–10.
[33] YASMEEN, U., UL AMIN, M. N. and HANIF, M. (2018): Exponential Estimators of Finite Population Variance Using Transformed Auxiliary Variables**. Proceedings of the National Academy of Sciences, India, Section A,** DOI 10.1007/s40010-017-0410-5.