Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online):2229-5666
Vol. 8, Issue 1 (2015): 77-93

ESTIMATION OF RATIO OF POPULATION VARIANCES IN ABSENCE AND PRESENCE OF NON - RESPONSE

A. Bandyopadhyay[1] and G. N. Singh[2]
[1]Department of Mathematics, Asansol Engineering College,
Asansol, India
[2]Department of Applied Mathematics, Indian School of Mines,
Dhanbad, India.
E Mail: [1]arnabbandyopadhyay4@gmail.com;
[2] gnsingh_ism@yahoo.com

Received December 11, 2014
Modified May 13, 2015
Accepted May 25, 2015

Abstract

The present investigation deals with the problem of estimation of ratio of population variances under two different realistic situations of complete response and random non-response in the sampled units. Using information on an auxiliary variable, two classes of Jack-Knife estimators of ratio of population variances are proposed separately for these two situations and it is shown that the proposed classes of Jack-Knife estimators are unbiased up to the first order of approximations. Properties of the proposed classes of estimators have been studied and their respective optimality conditions are discussed. Proposed classes of estimators are empirically compared with the usual sample estimator of ratio of population variances under the similar realistic situations and their performances have been demonstrated through numerical illustration and graphical interpretation which are followed by suitable recommendations.

Mathematics Subject Classification: 62D05

Key Words: Ratio of Population Variances, Study Variable, Auxiliary Variable, Random Non - Response, Jack-Knife Technique, Unbiased, Variance.

1. Introduction

The problem of estimation of population variance arises in many practical situations. For example, an agriculturist needs an adequate understanding of the variations in climatic factors especially from place to place (or time to time) to be able to plan on when, how and where to plant his crop. The variance estimation technique using auxiliary variable was first considered by Das and Tripathi (1978). Further this was extended by Srivastava and Jhaji (1980), Isaki (1983), Upadhyay and Singh (1983), Tripathi et al. (1988) and Ahamed et al. (2003) among others. In many situations, information on an auxiliary variable may be readily available on all unit of the population; for example, tonnage (or seat capacity) of each vehicle or ship is known in survey sampling of transportation and number of beds in different hospitals may be known in hospital surveys.

However in some practical situations, it is common experience in sample surveys that data cannot always be collected from all the units selected in the sample. For example, the selected families may not be at home at the first attempt and some of them may refuse to cooperate with the interviewer even if contacted. As many respondents do not reply, available sample of returns is incomplete. The resulting incompleteness is called non-response and is sometimes so large that can completely vitiate the results. Statisticians have long known that failure to account for the stochastic nature of incompleteness can damage the actual conclusion. An obvious problem, that one needs to justify, arises when ignoring the incomplete mechanism. Rubin (1976) advocated three concepts: missing at random (MAR), observed at random (OAR), and parameter distribution (PD). Rubin defined: “The data are MAR if the probability of the observed missingness pattern, given the observed and unobserved data, does not depend on the value of the unobserved data”. Singh and Joarder (1998) studied the estimation procedures of population variance by defining one discrete probability distribution model in the presence of random non-response situations in the sampled units and afterward this model was adopted in the population variance estimation procedure by Singh et al. (2012).

It is to be noted that the estimation procedures of population variance have been developed by several authors; however, no efforts have been made to estimate the ratio of population variances. An estimate of the ratio of variances of two characters of a population may be of considerable interest and its huge important real life impacts may be presented. For example, if the ratios of variations of income and expenditure are known from previous few years, then one may estimate their yearly ratio and plan for suitable investment for the current year and if the ratio of variations of body temperature and pulse rate of a patient during diseases estimated properly, then the doctors may prescribe adequate medicine for him. Similarly, the consistent performance of a company may be determined from the study of variation of NAV (in SENSEX or other equivalent indexing system) and its monthly production. Any one of them alone is not sufficient to determine it. By computing their monthly ratio for some period of time, we can easily estimate how consistently the company has performed. In order to provide an in-depth presentation of the proposed work, an illustrative scenario is provided. Consider the case of player’s selection process in international cricket. It is often seen that the all-rounder cricketers with consistent performances get the priority. In order to judge the same the ratio of batting variances and balling variances of the players plays an important role. To estimate this ratio, the data of batting and balling performances of the players are collected from previous few years/ months where it may be seen that some players were out of international cricket for certain period of time for several reasons such as lack of good performances and injury cases and these lack of information cases may be treated as non–response situations. Motivated with the above arguments, in the present work an attempt has been made to estimate the ratio of variances of two characters of a population. In some practical situations, it is seen that the bias becomes a serious draw back. Therefore, inspired with the Jack-Knife technique of unbiased estimation adopted by Quenouille (1956) and Gray and Schucancy (1972), we have proposed two classes of estimators separately for estimating the ratio of population variances for two different realistic situations of complete response and random non-response in the sampled units. The dominance of the proposed classes of estimators over the sample estimator of ratio of population variances under the similar realistic situations have been established through numerical illustration and graphical interpretation.

2. Formulation of estimators

Consider a finite population $U = (U_1, U_2, U_3, \ldots, U_N)$ of N units, y and x are the variables under study and z be the auxiliary variable. Let y_k, x_k and z_k be the values of y, x and z for the k-th (k = 1, 2, . . ., N) unit in the population. Our purpose is to estimate the ratio population variances of the study variables y and x in presence of the auxiliary variable z. Let a sample S of size n (n =2m; m being integer) is drawn by simple random sampling without replacement scheme (SRSWOR) from the entire population U. The sample S of size n is split up at random into two sub samples S_1 and S_2 of size m each.

Hence onwards, we use the following notations for the population parameters:

$\bar{Y}, \bar{X}, \bar{Z}$: Population mean of the variables y, x and z respectively.

$S_y^2 = (N-1)^{-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$: Population variance of the study variable y.

S_x^2, S_z^2 : Population variances of the study variable x and the auxiliary variable z respectively.

$R = \dfrac{S_y^2}{S_x^2}$: Ratio of population variances of the study variables y and x.

$\mu_{abc} = \dfrac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^a (x_i - \bar{X})^b (z_i - \bar{Z})^c$; (a, b, c) being non negative integers,

$$\lambda_{abc} = \mu_{abc} \Big/ \left\{\mu_{200}^{a/2}\ \mu_{020}^{b/2}\ \mu_{002}^{c/2}\right\},$$

$$C_0 = \sqrt{(\lambda_{400} - 1)},\ C_1 = \sqrt{(\lambda_{040} - 1)},\ C_2 = \sqrt{(\lambda_{004} - 1)},$$

$$\rho_{01} = (\lambda_{220} - 1)\Big/\sqrt{(\lambda_{400} - 1)(\lambda_{040} - 1)},\ \rho_{02} = (\lambda_{202} - 1)\Big/\sqrt{(\lambda_{400} - 1)(\lambda_{004} - 1)},$$

$$\rho_{12} = (\lambda_{022} - 1)\Big/\sqrt{(\lambda_{040} - 1)(\lambda_{004} - 1)},\ f_N = -\frac{1}{N}.$$

It is to be noted that ρ_{01} is the correlation between $\left(y - \bar{Y}\right)^2$ and $\left(x - \bar{X}\right)^2$. Similarly ρ_{12} is the correlation between $\left(x - \bar{X}\right)^2$ and $\left(z - \bar{Z}\right)^2$ and ρ_{02} is the correlation between $\left(y - \bar{Y}\right)^2$ and $\left(z - \bar{Z}\right)^2$; see for instance Upadhyaya and Singh (2006).

Now, to estimate the ratio of population variances R we consider the following two different realistic situations separately as:

Situations I: Units of the samples S, S_1 and S_2 give complete response for the study variables y, x and the auxiliary variable z.

Situations II: Random non-response situations occur for the variables y, x and z in the units of the samples S, S_1 and S_2.

We have suggested two classes of estimators separately to estimate R under the above different realistic situations and present them below.

2.1. Proposed class of estimators applicable for situation I

When the units of the sample S, S_1 and S_2 give complete response for the variables y, x and z, the usual sample estimator of ratio of population variances R may be considered as

$$R_n = \frac{s_{y_n}^2}{s_{x_n}^2} \tag{1}$$

where

$s_{y_n}^2 = (n-1)^{-1}\sum_{i=1}^{n}(y_i - \bar{y}_n)^2$ and $s_{x_n}^2 = (n-1)^{-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$ are the sample variances of the study variables y and x respectively based on sample S of size n,

$\bar{y}_n$ and $\bar{x}_n$ are the sample means of the variables y and x respectively based on the sample S.

However, the aim of a statistician is to develop superior of the method of estimation over the usual one. Accordingly, inspired with the estimation procedures suggested by Bahl and Tuteja (1991) which discussed exponential ratio and product type structures for estimating population mean, one may suggest a class of estimators of R based on the sample S of size n as

$$t_n = R_n \exp\left(c\,\frac{S_z^2 - s_{z_n}^2}{S_z^2 + s_{z_n}^2}\right) \tag{2}$$

where c is suitably chosen real constant,

$\bar{z}_n$ and $s_{z_n}^2 = (n-1)^{-1}\sum_{i=1}^{n}(z_i - \bar{z}_n)^2$ are the sample mean and variance of the auxiliary variable z respectively based on the sample S of size n.

It may be seen that the proposed class of estimators t_n converges to (i) R_n when c = 0 (ii) exponential ratio-type estimator when c = 1 and (iii) exponential product-type estimator when c = -1. It is also noted that the estimators R_n and t_n are biased which becomes a serious drawback for their practical applications. Therefore, an unbiased estimator of ratio of population variances R is more desirable.

Motivated with the above arguments and following the Jack-Knife unbiased estimation technique adopted by Quenouille (1956) and Gray and Schucancy (1972), we propose the class of estimators of ratio of population variances R as

$$T_1 = \frac{t_n - r_c t_m}{1 - r_c} \tag{3}$$

where $t_m = \frac{1}{2}\left(t_1 + t_2\right)$, (4)

$t_1 = R_{1m}\exp\left(c\frac{S_z^2 - s_{z_{1m}}^2}{S_z^2 + s_{z_{1m}}^2}\right)$ and $t_2 = R_{2m}\exp\left(c\frac{S_z^2 - s_{z_{2m}}^2}{S_z^2 + s_{z_{2m}}^2}\right)$ are the classes of exponential type estimators of R based on the sample S_1 and S_2 respectively,

$s_{y_{jm}}^2 = \frac{1}{(m-1)}\sum_{i=1}^{m}(y_i - \bar{y}_{jm})^2$, $s_{x_{jm}}^2 = \frac{1}{(m-1)}\sum_{i=1}^{m}(x_i - \bar{x}_{jm})^2$ and $s_{z_{jm}}^2 = \frac{1}{(m-1)}\sum_{i=1}^{m}(z_i - \bar{z}_{jm})^2$ are the sample variances and $\bar{y}_{jm}$, $\bar{x}_{jm}$ and $\bar{z}_{jm}$ are the sample means of the respective variables based on the samples $S_j\,(j = 1, 2)$ of size m,

$R_{1m} = \frac{s_{y_{1m}}^2}{s_{x_{1m}}^2}$ and $R_{2m} = \frac{s_{y_{2m}}^2}{s_{x_{2m}}^2}$.

It is to be noted that the classes of estimators $t_i\,(i = 1, 2)$ and t_n are biased and we have considered $r_c = \frac{B(t_n)}{B(t_m)}$ where $B(t)$ denote the bias of a class of estimators t.

2.2. Proposed class of estimators applicable for situation II

In this section, we have considered that the occurrences of the random non-response situations of the study variables y, x and the auxiliary variable z on the samples S, S_1 and S_2 follow discrete probability distribution models as presented below.

If $r\{r = 0, 1, 2, \ldots, (n-2)\}$ denotes the number of units of the sample S on which information could not be obtained due to random non-responses, then the observations of the variables y, x and z can be taken from the remaining $(n - r)$ responding units of the sample S. Since we are interested in the problem of unbiased estimation of ratio of the population variances, it is assumed that r is less than $(n - 1)$, that is, $0 \le r \le (n-2)$. We also assume that if p denotes the probability of a non-response among the $(n - 2)$ possible values of non-responses, then r has the following discrete probability distribution given by

$$P(r) = \frac{(n-r)}{nq + 2p}\,{}^{n-2}C_r\, p^r q^{n-2-r} \tag{5}$$

where $q = 1 - p$ and ${}^{n-2}C_r$ denote the total number of ways of obtaining r non-responses out of the total possible $(n - 2)$ responses; see for instance Singh and Joarder (1998) and Singh et al. (2012).

Similarly, if the information on $r_1\{r_1 = 0, 1, 2, \ldots, (m-1)\}$ and $r_2\{r_2 = 0, 1, 2, \ldots, (m-1)\}$ numbers units of the samples S_1 and S_2 respectively could not be obtained due random non-responses, then the required observations of the variables y, x and z can be taken from remaining $(m - r_1)$ and $(m - r_2)$ responding units of the respective samples S_1 and S_2. As the sample S is divided at random into two sub sample S_1 and S_2 of equal size, therefore, it is assumed that $0 \le r_1 \le (m-1)$ and $0 \le r_2 \le (m-1)$. We have denoted p_1 and p_2 as the probability of non-response among the (m - 1) possible values of non-responses from the sample

S_1 and S_2 respectively. Thus, r_1 and r_2 have the following discrete probability distributions as

$$P(r_1)=\frac{(m-r_1)}{mq_1+p_1}\,{}^{m-1}C_{r_1}\,p_1^{r_1}q_1^{m-1-r_1}, \tag{6}$$

and $$P(r_2)=\frac{(m-r_2)}{mq_2+p_2}\,{}^{m-1}C_{r_2}\,p_2^{r_2}q_2^{m-1-r_2} \tag{7}$$

where $q_1 = 1-p_1$, $q_2 = 1-p_2$, ${}^{m-1}C_{r_1}$ and ${}^{m-1}C_{r_2}$ are the total number of ways of obtaining r_1 and r_2 non-responses respectively from the total possible $(m-1)$ responses. Since the number of non-responding units in the sample S remains same even after splitting of S into two sub samples S_1 and S_2, therefore, the relation among r, r_1 and r_2 can be defined as

$$r = r_1 + r_2 . \tag{8}$$

It is to be noted that the non-response probability models defined in equations (5)-(7) are free from actual data values; hence, can be considered as a model suitable for MAR situation.

We have defined following sample parameters based on the responding units of the samples S, S_1 and S_2 as:

$\bar{x}_n^*=\frac{1}{n-r}\sum_{i=1}^{n-r}x_i,\ \bar{y}_n^*=\frac{1}{n-r}\sum_{i=1}^{n-r}y_i,\ \bar{z}_n^*=\frac{1}{n-r}\sum_{i=1}^{n-r}z_i$: Sample means of the respective variables based on the responding units of the sample S.

$\bar{x}_{jm}^*=\frac{1}{m-r_j}\sum_{i=1}^{m-r_j}x_i,\ \bar{y}_{jm}^*=\frac{1}{m-r_j}\sum_{i=1}^{m-r_j}y_i,\ \bar{z}_{jm}^*=\frac{1}{m-r_j}\sum_{i=1}^{m-r_j}z_i$: Sample means of the respective variables based on the responding units of the samples $S_j\,(j=1,2)$.

$s_{y_n}^{*2}=\frac{1}{(n-r-1)}\sum_{i=1}^{n-r}(y_i-\bar{y}_n^*)^2,\ s_{x_n}^{*2}=\frac{1}{(n-r-1)}\sum_{i=1}^{n-r}(x_i-\bar{x}_n^*)^2,\ s_{z_n}^{*2}=\frac{1}{(n-r-1)}\sum_{i=1}^{n-r}(z_i-\bar{z}_n^*)^2$: Sample variances of the respective variables based on the responding units of the sample S.

$s_{y_{jm}}^{*2}=\frac{1}{(m-r_j-1)}\sum_{i=1}^{m-r_j}(y_i-\bar{y}_{jm}^*)^2,\ s_{x_{jm}}^{*2}=\frac{1}{(m-r_j-1)}\sum_{i=1}^{m-r_j}(x_i-\bar{x}_{jm}^*)^2,\ s_{z_{jm}}^{*2}=\frac{1}{(m-r_j-1)}\sum_{i=1}^{m-r_j}(z_i-\bar{z}_{jm}^*)^2$:

Sample variances of the respective variables based on the responding units of the samples $S_j\,(j=1,2)$.

Following the work of section 2.1 and considering the random non-response probability models discussed above, we propose the class of estimators for estimating the ratio of population variances R based on the responding units of the samples S, S_1 and S_2 as

$$T_2=\frac{t_n^*-r_{c^*}t_m^*}{1-r_{c^*}} \tag{9}$$

where

$t_n^* = R_n^* \exp\left(c^* \frac{S_z^2 - s_{z_n}^{*2}}{S_z^2 + s_{z_n}^{*2}} \right)$, $t_m^* = \frac{1}{2}\left(t_1^* + t_2^*\right)$,

$R_n^* = \frac{s_{y_n}^{*2}}{s_{x_n}^{*2}}$ is the sample estimator of ratio of population variances R based on the responding units of the sample S,

$t_1^* = R_{1m}^* \exp\left(c^* \frac{S_z^2 - s_{z_{1m}}^{*2}}{S_z^2 + s_{z_{1m}}^{*2}} \right)$, $t_2^* = R_{2m}^* \exp\left(c^* \frac{S_z^2 - s_{z_{2m}}^{*2}}{S_z^2 + s_{z_{2m}}^{*2}} \right)$ are the classes of exponential type estimators of R based on the responding units of the samples S_1 and S_2 respectively,

$R_{1m}^* = \frac{s_{y_{1m}}^{*2}}{s_{x_{1m}}^{*2}}$, $R_{2m}^* = \frac{s_{y_{2m}}^{*2}}{s_{x_{2m}}^{*2}}$,

c^* is the suitably chosen real constant and we have considered $r_{c^*} = \frac{B\left(t_n^*\right)}{B\left(t_m^*\right)}$.

3. Variances of the proposed classes of estimators T_1 and T_2

3. 1. Variance of the class of estimators T_1

If the units of the samples S, S_1 and S_2 give complete response for the variables y, x and z, then to derive the expression of variance of the proposed class of estimators T_1 to the first order approximations we use the following transformations under large sample approximations as

$s_{y_{1m}}^2 = S_y^2\left(1+e_0\right)$, $s_{x_{1m}}^2 = S_x^2\left(1+e_1\right)$, $s_{z_{1m}}^2 = S_z^2\left(1+e_2\right)$, $s_{y_{2m}}^2 = S_y^2\left(1+e_3\right)$, $s_{x_{2m}}^2 = S_x^2\left(1+e_4\right)$

$s_{z_{2m}}^2 = S_z^2\left(1+e_5\right)$ $s_{y_n}^2 = S_y^2\left(1+e_6\right)$, $s_{x_n}^2 = S_x^2\left(1+e_7\right)$, $s_{z_n}^2 = S_z^2\left(1+e_8\right)$.

Such that $\left|e_i\right| < 1 \;\; \forall$ (i = 0, 1, . . ., 8).

Thus, we have the following expectations.

$$\left.\begin{array}{l}
E(e_0^2)=E(e_3^2)=f_m C_0^2,\ E(e_1^2)=E(e_4^2)=f_m C_1^2,\ E(e_2^2)=E(e_5^2)=f_m C_2^2,\ E(e_2e_5)=f_N C_2^2,\\
E(e_6^2)=E(e_0e_6)=E(e_3e_6)=f_n C_0^2,\ E(e_7^2)=E(e_1e_7)=E(e_4e_7)=f_n C_1^2,\ E(e_1e_4)=f_N C_1^2,\\
E(e_8^2)=E(e_2e_8)=E(e_5e_8)=f_n C_2^2,\ E(e_0e_3)=f_N C_0^2,\ E(e_0e_4)=E(e_1e_3)=f_N\rho_{01}C_0C_1,\\
E(e_0e_5)=E(e_2e_3)=f_N\rho_{02}C_0C_2,\ E(e_1e_5)=E(e_2e_4)=f_N\rho_{12}C_1C_2,\ E(e_2e_7)=E(e_5e_7)=f_n\rho_{12}C_1C_2\\
E(e_0e_2)=E(e_3e_5)=f_m\rho_{02}C_0C_2, E(e_1e_2)=E(e_4e_5)=f_m\rho_{12}C_1C_2,\ E(e_1e_8)=E(e_4e_8)=f_n\rho_{12}C_1C_2\\
E(e_1e_6)=E(e_4e_6)=E(e_6e_7)=E(e_0e_7)=E(e_3e_7)=f_n\rho_{01}C_0C_1, E(e_0e_1)=E(e_3e_4)=f_m\rho_{01}C_0C_1,\\
E(e_7e_8)=f_n\rho_{12}C_1C_2,\ E(e_2e_6)=E(e_5e_6)=E(e_6e_8)=E(e_0e_8)=E(e_3e_8)=f_n\rho_{02}C_0C_2,\\
E(e_0e_7)=E(e_3e_7)=f_n\rho_{01}C_0C_1,
\end{array}\right\} \quad (10)$$

where

$$f_m=\left(\frac{1}{m}-\frac{1}{N}\right) \text{ and } f_n=\left(\frac{1}{n}-\frac{1}{N}\right).$$

Under the above transformations t_1, t_2 and t_n take the following forms

$$t_1=R(1+e_0)(1+e_1)^{-1}\exp\left\{\left(\frac{-ce_2}{2}\right)\left(1+\frac{e_2}{2}\right)^{-1}\right\}, \quad (11)$$

$$t_2=R(1+e_3)(1+e_4)^{-1}\exp\left\{\left(\frac{-ce_5}{2}\right)\left(1+\frac{e_5}{2}\right)^{-1}\right\} \quad (12)$$

and $t_n=R(1+e_6)(1+e_7)^{-1}\exp\left\{\left(\frac{-ce_8}{2}\right)\left(1+\frac{e_8}{2}\right)^{-1}\right\}$ (13)

Expanding the above terms binomially and using the results from equation (10), we have obtained bias of the classes of estimators t_1, t_2 and t_n to the first order of approximations as

$$B(t_1)=E(t_1-R)=f_m A \quad (14)$$

$$B(t_2)=E(t_2-R)=f_m A \quad (15)$$

and

$$B(t_n)=E(t_n-R)=f_n A \quad (16)$$

where $A=\left[C_1^2+\frac{(c^2+2c)}{8}C_2^2-\rho_{01}C_0C_1+\frac{c}{2}(\rho_{12}C_1-\rho_{02}C_0)C_2\right]R.$

Substituting the above results in equation (3) and taking expectations up to first order of approximations, it is found that

$$E(T_1)=R \quad (17)$$

and $r_c=\frac{f_n}{f_m}$ (18)

which indicates that the class of estimators T_1 is unbiased for estimating R up to the first order of approximations.

Thus, we have the following theorem.

Theorem 1: Variance of the class of estimators T_1 to the first order of approximations are obtained as

$$V(T_1)=f_n\left[\left(C_0^2+C_1^2-2\rho_{01}C_0C_1\right)+\frac{c^2}{4}C_2^2-c\left(\rho_{02}C_0-\rho_{12}C_1\right)C_2\right]R^2 \qquad (19)$$

Proof. The variance of the class of estimators T_1 is given by

$$V(T_1)=E(T_1-R)^2=E\left(\frac{t_n-r_ct_m}{1-r_c}-R\right)^2=\frac{1}{(1-r_c)^2}E\left\{t_n-r_ct_m-(1-r_c)R\right\}^2$$

$$=\frac{1}{(1-r_c)^2}E\left[(t_n-R)-r_c(t_m-R)\right]^2$$

$$=\frac{1}{(1-r_c)^2}E(t_n-R)^2+\frac{r_c^2}{4(1-r_c)^2}E\left[(t_1-R)+(t_2-R)\right]^2$$

$$-\frac{r_c}{(1-r_c)^2}E\left[(t_n-R)(t_1-R)+(t_n-R)(t_2-R)\right]$$

Using the expansions of t_1, t_2 and t_n given in equations (11)-(13) and results from equation (10), we have the expression for the variance of the class of estimators T_1 to the first order of approximations as presented in equation (19).

3. 2. Variance of the class of estimators T_2

When random non-response situations occur for the study variables y, x and the auxiliary variable z on the samples S, S_1 and S_2, we derive the variance of the class of estimators T_2 up to first order of approximations using following transformations under large sample approximations as.

$s_{y_{1m}}^{*2}=S_y^2\left(1+e_0^*\right)$, $s_{x_{1m}}^{*2}=S_x^2\left(1+e_1^*\right)$, $s_{z_{1m}}^{*2}=S_z^2\left(1+e_2^*\right)$, $s_{y_{2m}}^{*2}=S_y^2\left(1+e_3^*\right)$, $s_{x_{2m}}^{*2}=S_x^2\left(1+e_4^*\right)$,

$s_{z_{2m}}^{*2}=S_z^2\left(1+e_5^*\right)$ $s_{y_n}^{*2}=S_y^2\left(1+e_6^*\right)$, $s_{x_n}^{*2}=S_x^2\left(1+e_7^*\right)$, $s_{z_n}^{*2}=S_z^2\left(1+e_8^*\right)$.

Such that $\left|e_i^*\right|<1$ $\forall$ (i = 0, 1, . . ., 8).

Thus, we have following expectations.

$$\left.\begin{array}{l}
E\left(e_0^{*2}\right)=f_1^*C_0^2,\ E\left(e_1^{*2}\right)=f_1^*C_1^2,\ E\left(e_2^{*2}\right)=f_1^*C_2^2,\ E\left(e_3^{*2}\right)=f_2^*C_0^2,\ E\left(e_4^{*2}\right)=f_2^*C_1^2,\ E\left(e_5^{*2}\right)=f_2^*C_2^2,\\
E\left(e_6^{*2}\right)=E\left(e_0^*e_6^*\right)=E\left(e_3^*e_6^*\right)=f_n^*C_0^2,\ E\left(e_7^{*2}\right)=E\left(e_1^*e_7^*\right)=E\left(e_4^*e_7^*\right)=f_n^*C_1^2,\ E\left(e_2^*e_5^*\right)=f_NC_2^2,\\
E\left(e_8^{*2}\right)=E\left(e_2^*e_8^*\right)=E\left(e_5^*e_8^*\right)=f_n^*C_2^2,\ E\left(e_0^*e_3^*\right)=f_NC_0^2,\ E\left(e_1^*e_4^*\right)=f_NC_1^2,\ E\left(e_0^*e_1^*\right)=f_1^*\rho_{01}C_0C_1,\\
E\left(e_0^*e_2^*\right)=f_1^*\rho_{02}C_0C_2,\ E\left(e_1^*e_2^*\right)=f_1^*\rho_{12}C_1C_2, E\left(e_6^*e_7^*\right)=f_n^*\rho_{01}C_0C_1,\ E\left(e_3^*e_4^*\right)=f_2^*\rho_{01}C_0C_1,\\
E\left(e_3^*e_5^*\right)=f_2^*\rho_{02}C_0C_2,\ E\left(e_4^*e_5^*\right)=f_2^*\rho_{12}C_1C_2,\ E\left(e_6^*e_8^*\right)=f_n^*\rho_{02}C_0C_2, E\left(e_7^*e_8^*\right)=f_n^*\rho_{12}C_1C_2,\\
E\left(e_0^*e_4^*\right)=E\left(e_1^*e_3^*\right)=f_N\rho_{01}C_0C_1,\ E\left(e_0^*e_5^*\right)=E\left(e_2^*e_3^*\right)=f_N\rho_{02}C_0C_2,\\
E\left(e_1^*e_6^*\right)=E\left(e_4^*e_6^*\right)=E\left(e_0^*e_7^*\right)=E\left(e_3^*e_7^*\right)=f_n^*\rho_{01}C_0C_1, E\left(e_1^*e_5^*\right)=E\left(e_2^*e_4^*\right)=f_N\rho_{12}C_1C_2\\
E\left(e_2^*e_6^*\right)=E\left(e_5^*e_6^*\right)=E\left(e_0^*e_8^*\right)=E\left(e_3^*e_8^*\right)=f_n^*\rho_{02}C_0C_2,\\
E\left(e_2^*e_7^*\right)=E\left(e_5^*e_7^*\right)=E\left(e_1^*e_8^*\right)=E\left(e_4^*e_8^*\right)=f_n^*\rho_{12}C_1C_2,
\end{array}\right\} \quad (20)$$

where $f_1^*=\left(\dfrac{1}{mq_1+p_1}-\dfrac{1}{N}\right)$, $f_2^*=\left(\dfrac{1}{mq_2+p_2}-\dfrac{1}{N}\right)$ and $f_n^*=\left(\dfrac{1}{nq+2p}-\dfrac{1}{N}\right)$.

Proceeding as section 3.1 and using the results from equation (20), it can be observed that the class of estimators T_2 is unbiased for the ratio of population variances R up to first order of approximations $\left[\text{i. e. } E\left(T_2\right)=R\right]$ and we have obtained the expression for the variance of the class of estimators T_2 to the first order of approximations as

$$V\left(T_2\right)=\frac{f_n^*+r_{c^*}^2f^o-2r_{c^*}f_n^*}{\left(1-r_{c^*}\right)^2}\left[\left(C_0^2+C_1^2-2\rho_{01}C_0C_1\right)+\frac{c^{*2}}{4}C_2^2-c^*\left(\rho_{02}C_0-\rho_{12}C_1\right)C_2\right] \quad (21)$$

where

$$r_{c^*}=\frac{2f_n^*}{\left(f_1^*+f_2^*\right)} \text{ and } f^o=\frac{1}{2}\left(\frac{1}{nq_1+2p_1}+\frac{1}{nq_2+2p_2}-\frac{2}{N}\right).$$

It is to be noted that if $p=p_1=p_2=0$ (there is no non-response), the variance of the class of estimators T_2 coincide with the variance of the class of estimators T_1 as given in equation (19).

4. Minimum variances of the proposed classes of estimators T_1 and T_2

It is obvious from the equations (19) and (21) that the variances of the proposed classes of estimators $T_i\left(i=1,2\right)$ depend on the different values of the constants c and c^*. Therefore, we desire to minimize their variances.

The optimality conditions under which proposed classes of estimators T_1 and T_2 have minimum variances are obtained as

$$c=c^*=\frac{2\left(\rho_{02}C_0-\rho_{12}C_1\right)}{C_2} \quad (22)$$

Substituting these optimum values of the constants c and c^* in equations (19) and (21), we have obtained the expressions of minimum variances of the classes of estimators $T_i\left(i=1,2\right)$ as

$$\text{Min. } V\left(T_1\right)=f_n\left\{C_0^2+C_1^2-2\rho_{01}C_0C_1-\left(\rho_{02}C_0-\rho_{12}C_1\right)^2\right\}R^2 \quad (23)$$

and $$\text{Min. } V\left(T_2\right)=\frac{f_n^*+r_{c^*}^2f^o-2r_{c^*}f_n^*}{\left(1-r_{c^*}\right)^2}\left\{C_0^2+C_1^2-2\rho_{01}C_0C_1-\left(\rho_{02}C_0-\rho_{12}C_1\right)^2\right\}R^2 \quad (24)$$

Remark 4.1: It may be observed from the optimality conditions in equation (22) that the optimum values of constants c and c^* of the proposed classes of estimators $T_i\left(i=1,2\right)$ depend on unknown population parameters such as C_0, C_1, C_2, ρ_{12} and ρ_{02}. Thus, to use such estimators one has to use guessed or estimated values of them. Guessed values of population parameters can be obtained either from past data or experience gathered over time. A scatter diagram for at least a part of current data will help in this regard; for instance see Murthy (1967) and Tracy et al. (1996). If the guessed values are not known then it is advisable to use sample data to estimate these parameters as suggested by Singh et al. (2007) and Gupta and Shabbir (2008). Replacement of the population parameters by their respective sample estimates may turn the proposed classes of estimators to be biased. However, they can be converted to the classes of unbiased estimators up to first order of approximations by using Jack-Knife technique as suggested in this paper. It could be seen that the variances of the proposed classes of estimators remain same up to the first order of approximations, even if population parameters are replaced by their respective sample estimates.

5. Efficiency comparisons of the proposed classes of estimators T_1 and T_2

To examine the performances of our proposed classes of estimators T_1 and T_2, we have compared their efficiencies with usual sample estimator of ratio of population variances under the similar realistic situations. Since the sample estimator R_n and R_n^* defined in sections 2.1 and 2.2 respectively are biased, therefore, we have obtained their mean square errors up to the first order of approximations as

$$M\left(R_n\right)=f_n\left(C_0^2+C_1^2-2\rho_{01}C_0C_1\right)R^2 \quad (25)$$

and

$$M\left(R_n^*\right)=f_n^*\left(C_0^2+C_1^2-2\rho_{01}C_0C_1\right)R^2 \text{ respectively.} \quad (26)$$

The proofs of the mean square errors of the estimators R_n and R_n^* defined above can be derived in similar way as suggested in sections 3.1 and 3.2 respectively. The performances of the proposed classes of estimators T_1 and T_2 under their respective optimality conditions are empirically compared with R_n and R_n^* and we have demonstrated their merits through numerical illustration and graphical interpretation.

5.1. Numerical Illustration

We have chosen four natural population data sets to illustrate the efficacious performances of our proposed classes of estimators T_1 and T_2. The source of the populations, the nature of the variables y, x, z and the values of the various parameters are given as follows.

Population I-Source: Cochran (1977, Page- 182)
y: Number of 'placebo' children.
x: Number of paralytic polio cases in the placebo group.
z: Number of paralytic polio cases in the 'not inoculated' group.
N= 34, $C_0 = 2.3219$, $C_1 = 1.8268$, $C_2 = 2.0887$, $\rho_{01} = 0.6661$, $\rho_{02} = 0.5657$, $\rho_{12} = 0.6005$.
It is to be noted that this population was also considered by several authors including Choudhury and Singh (2012).

Population II-Source: Murthy (1967, Page- 399)
y: Area under wheat in 1964.
x: Area under wheat in 1963.
z: Cultivated area in 1961.
N= 34, $C_0 = 1.6510$, $C_1 = 1.3828$, $C_2 = 1.3447$, $\rho_{01} = 0.9218$, $\rho_{02} = 0.8914$, $\rho_{12} = 0.9346$.
This population was considered as numerical evidence in the works of several authors including Jhajj et al. (2005) and Choudhury and Singh (2012).

Population III- Source: Sukhatme (1970, Page- 185)
y: Area under wheat in 1937.
x: Area under wheat in 1936.
z:Total cultivated area in 1931.
N= 34, $C_0 = 1.5959$, $C_1 = 1.5105$, $C_2 = 1.3200$, $\rho_{01} = 0.6251$, $\rho_{02} = 0.8007$, $\rho_{12} = 0.5342$.
It is to be noted that this population was presented to justify the works of by several authors including Agrawal and Roy (1999).

Population IV-Source: Murthy (1967, Page- 288)
y: Output.
x: Fixed Capital
z: Number of workers.
N= 80, $C_0 = 1.1255$, $C_1 = 1.6065$, $C_2 = 1.3662$, $\rho_{01} = 0.7319$, $\rho_{02} = 0.7940$, $\rho_{12} = 0.9716$.
This population was also considered as numerical evidence in the works of several authors including Jhajj et al. (2005).

Using the above data sets, we have compared the efficiencies of our proposed classes of estimators T_i $(i = 1, 2)$ under their respective optimality conditions with the sample estimator of ratio of population variances and the findings are displayed in tables 1-2. It is to be noticed from equation (24) that the minimum variance expression of the class of estimators T_2 depend on values of the non-response rates p, p_1 and p_2, which are difficult to obtain. Therefore, to examine the performance of the class of estimators T_2 we make use of maximum likelihood estimators of p, p_1 and p_2 say $\hat{p}$, $\hat{p}_1$ and $\hat{p}_2$ respectively, which are presented in Lemma 5.1. Replacing the non response rates by their respective likelihood estimates, we have the following expression for minimum variance of the class of estimator T_2 up to the first order of approximations as

$$\text{Min. } V\left(\hat{T}_2\right)=\frac{\hat{f}_n^* + \hat{r}_{c^*}^2 \hat{f}^o - 2\hat{r}_{c^*}\hat{f}_n^*}{\left(1-\hat{r}_{c^*}\right)^2}\left\{C_0^2 + C_1^2 - 2\rho_{01}C_0C_1 - \left(\rho_{02}C_0 - \rho_{12}C_1\right)^2\right\}R^2 \qquad (27)$$

where

$$\hat{f}_1^* = \left(\frac{1}{m\hat{q}_1 + \hat{p}_1} - \frac{1}{N}\right),\ \hat{f}_2^* = \left(\frac{1}{m\hat{q}_2 + \hat{p}_2} - \frac{1}{N}\right),\ \hat{f}_n^* = \left(\frac{1}{n\hat{q} + 2\hat{p}} - \frac{1}{N}\right),$$

$$\hat{r}_{c^*} = \frac{2\,\hat{f}_n^*}{\left(\hat{f}_1^* + \hat{f}_2^*\right)} \text{ and } \hat{f}^o = \frac{1}{2}\left(\frac{1}{n\hat{q}_1 + 2\hat{p}_1} + \frac{1}{n\hat{q}_2 + 2\hat{p}_2} - \frac{2}{N}\right).$$

We have designated the percent relative efficiencies (PREs) of our proposed classes of estimators with respect to estimators R_n and R_n^* as

$$E = \frac{M(R_n)}{\text{Min. } V(T_1)}\times 100 \text{ and } E^* = \frac{M(R_n^*)}{\text{Min. } V(\hat{T}_2)}\times 100 \text{ respectively.}$$

Lemma 5.1. We have obtained the expressions of likelihood estimates of non-response probability of p, p_1 and p_2 as

$$\hat{p} = \frac{(n-1+r) - \sqrt{\left(n-1+r\right)^2 - 4rn\left(n-3\right)/\left(n-2\right)}}{2\left(n-3\right)}, \qquad (28)$$

$$\hat{p}_1 = \frac{(m+r_1) - \sqrt{\left(m+r_1\right)^2 - 4r_1\left(m+1\right)\left(m-2\right)/\left(m-1\right)}}{2\left(m-2\right)}, \qquad (29)$$

and

$$\hat{p}_2 = \frac{(m+r_2) - \sqrt{\left(m+r_2\right)^2 - 4r_2\left(m+1\right)\left(m-2\right)/\left(m-1\right)}}{2\left(m-2\right)} \text{ respectively} \qquad (30)$$

(see for instance Singh and Joarder (1998) and singh et al (2012))

where the numbers of non-responding units of the samples S, S_1 and S_2 satisfy the linear relationship between them as indicated in equation (8).

Population	E
Population I	101.5480
Population II	108.1028
Population III	113.9210
Population IV	158.9146

Table 1: PRE of the class of estimators T_1 with respect to R_n.

Population I															
n	R	r_1	E^*	n	R	r_1	E^*	n	R	r_1	E^*	n	R	r_1	E^*
20	3	1	100.9	22	4	2	101.1	24	5	2	100.9	26	6	3	101.1
		2	100.9			3	100.5			3	100.9			4	100.6
	4	2	100.9		5	2	100.7		6	3	100.9		7	3	100.8
		3	100.1			3	100.7			4	100.3			4	100.8
Population II															
n	R	r_1	E^*	n	R	r_1	E^*	n	R	r_1	E^*	n	R	r_1	E^*
20	3	1	107.4	22	4	2	107.6	24	5	2	107.5	26	6	3	107.6
		2	107.4			3	106.9			3	107.5			4	107.1
	4	2	107.4		5	2	107.2		6	3	107.5		7	3	107.3
		3	106.5			3	107.2			4	106.8			4	107.3
Population III															
n	R	r_1	E^*	n	R	r_1	E^*	n	R	r_1	E^*	n	R	r_1	E^*
20	3	1	113.2	22	4	2	113.4	24	5	2	113.2	26	6	3	113.4
		2	113.2			3	112.7			3	113.2			4	112.9
	4	2	113.2		5	2	113.0		6	3	113.2		7	3	113.1
		3	112.3			3	113.0			4	112.6			4	113.1
Population IV															
n	R	r_1	E^*	n	R	r_1	E^*	n	R	r_1	E^*	n	R	r_1	E^*
30	4	2	158.4	40	6	3	158.5	50	8	5	158.4	60	10	4	158.6
		3	157.7			4	158.2			6	157.9			6	158.6
	6	3	158.0		8	5	157.9		10	5	158.5		12	5	158.5
		4	157.2			6	156.8			6	158.3			7	158.5

Table 2: PRE of the class of estimators T_2 with respect to R_n^* for different choices of n, r and r_1.

5.2. Graphical interpretation

We have established the dominance of our proposed classes of estimators T_1 and T_2 over the usual sample estimator of ratio of population variances by pictorial representation. In this section, we have considered N = 500, n = 100, r = 20, r_1 = 12, $C_0 \cong C_1 \cong C_2$ and compare the efficiencies of our proposed classes of estimators for different choices of correlations ρ_{01}, ρ_{02} and ρ_{12}. This could not only improve the readability of the results but also allow the comparison of a much denser grid of different correlation values.

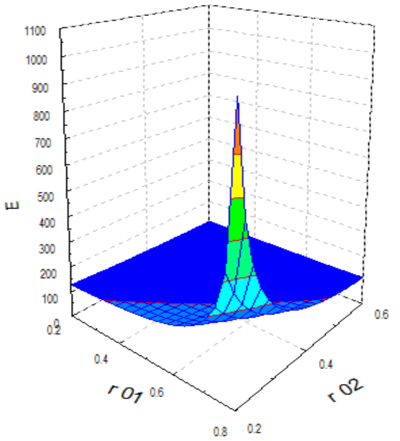

Figure 1: PRE of T_1 **(**$\rho_{12} = 0.8$**)**

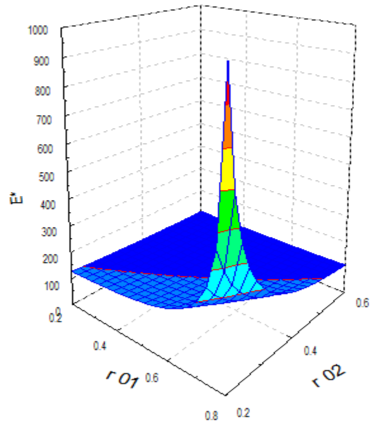

Figure 2: PRE of T_2 **(**$\rho_{12} = 0.8$**)**

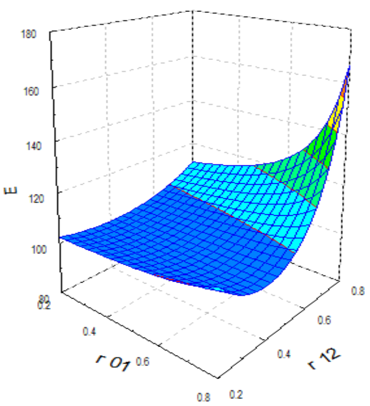

Figure 3: PRE of T_1 **(**$\rho_{02} = 0.4$**)**

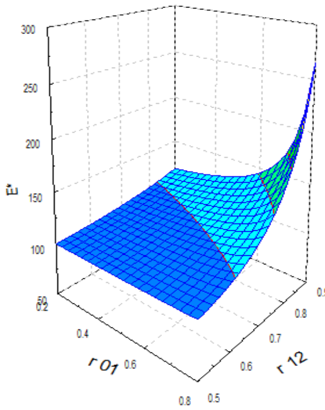

Figure 4: PRE of T_2 **(**$\rho_{02} = 0.4$**)**

Note: r 01, r 02 and r 12 denote ρ_{01}, ρ_{02} and ρ_{12} respectively in the figures 1- 4.

6. Conclusions

The following conclusions may be drawn from the present study.

1. From tables 1- 2, it is observed that:

(a) Our proposed class of estimators T_1 under its optimality condition is preferable over the usual sample estimator R_n.

(b) For the different choices of sample size n and the number of non-responding units (i. e. r and r_1), proposed class of estimators T_2 under its optimality condition is more efficient than the sample estimator R_n^* which indicates that the class of estimators T_2 can handle the problem of random non-responses effectively.

2. From figures 1-4, it is noticed that:

(a) For fixed value of ρ_{12}, the percent relative efficiencies of the proposed classes of estimators T_i $(i = 1, 2)$ under their respective optimality condition are increasing with the increasing values of ρ_{01}.

(b) For fixed value of ρ_{02}, the percent relative efficiencies of the proposed classes of estimators T_i $(i = 1, 2)$ under their respective optimality condition are increasing with the increasing values of ρ_{12}.

These phenomena indicate that our proposed classes of estimators could perform significantly, if a high positively correlated auxiliary variable is available and the study variables have strong correlation between them.

Thus, the superiority of the suggested classes of estimators over the usual sample estimator of ratio of population variances has been established and it is found that the use of an auxiliary variable is highly rewarding in terms of the proposed classes of estimators. Moreover, it may be noted that the proposed classes of estimators are unbiased up to first order of approximations, which indicate profoundness of their practical applications. Hence, the proposals of the classes of estimators in the present study are justified as they unify several results and therefore, they may be recommended to the survey statisticians and practitioners for their usage in real life problems.

References

1. Agrawal, M. C. and Roy, D. C. (1999). Efficient estimators of population variance with regression-type and ratio-type predictor-inputs, Metron, LVII (3), p. 169-178.
2. Ahmed, M. S., Abu-Dayyeh, W. and Hurairah, A. A. O. (2003). Some estimators for population variance under two phase sampling, Statistics in Transition 6 (1), p. 143–150.
2. Bahl, S. and Tuteja, R. K. (1991). Ratio and product type exponential estimator, Information and Optimization Sciences. 12, p. 159-163.
3. Cochran, W. G. (1977). Sampling techniques. New-York. John Wiley and Sons.
4. Choudhury, S. and Singh, B. K. (2012). A class of chain ratio–product type estimators with two auxiliary variables under double sampling scheme, Journal of the Korean Statistical Society, 41, p. 247–256.
5. Das, A. K. and Tripathi, T. P. (1978). Use of auxiliary information in estimating the finite population variance, Sankhya, C, 40, p. 139–148.
6. Gray, H. L. and Schucany W.R. (1972). The Generalized Jack-knife Statistic, Marcel Dekker, New York.
7. Gupta, S. and Shabbir, J. (2008). On improvement in estimating the population mean in simple random sampling, Journal of Applied Statistics. 35 (5), p. 559–566.
8. Isaki, C. T. (1983). Variance estimation using auxiliary information, Journal of American Statistical Association, 78, p. 117–123.
9. 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 Japan Statistical Society, 35(2), p. 273–286.

10. Murthy, M. N. (1967). Sampling Theory and Methods, Statistical Publishing Society. Calcutta, India.
11. Quenouille, M.H. (1956). Notes on bias in estimation, Biometrika, 43, p. 353-360.
12. Rubin, D. B. (1976). Inference and missing data, Biometrika, 63(3), 581–592.
13. Sukhatme, P. V. and Sukhatme, B. V. (1970). Sampling Theory of Surveys with Applications, Second Edition, Asia Publishing House, London.
14. Srivastava, S. K. and Jhajj, H. S. (1980). A class of estimators using auxiliary information for estimating finite population variance. Sankhya C, 42, p. 87–96.
15. Singh, S. and Joarder, A. H. (1998). Estimation of finite population variance using random non-response in survey sampling, Metrika, 98, p. 241–249.
16. Singh, H. P., Chandra, P., Joarder, A. H. and Singh, S. (2007). Family of estimators of mean, ratio and product of a finite population using random non-response, Test, 16. p. 565–597.
17. Singh, H. P., Tailor, R., Kim, J. M. and Singh, S. (2012). Families of estimators of finite population variance using a random non-response in survey sampling, The Korean Journal of Applied Statistics, 25(4), p. 681-695.
18. Tripathi, T. P., Singh, H. P. and Upadhyaya, L. N. (1988). A generalized method of estimation in double sampling. Journal of Indian Statistical Association, 26, p. 91– 101.
19. Tracy, D. S., Singh, H. P. and Singh, R. (1996). An alternative to the ratio-cum-product estimator in sample surveys, Journal of Statistical Planning and Inference. 53, p. 375- 387.
20. Upadhyaya, L. N. and Singh, H. P. (1983). Use of auxiliary information in the estimation of population variance, Mathematical Forum, 6, (2), p. 33–36.
21. Upadhyaya, L. N. and Singh, H. P. (2006). Almost unbiased ratio and product-type estimators of finite population variance in sample surveys, Statistics in Transition. 7(5), p. 1087-1096.