STATISTICA, anno LXXIV, n. 4, 2014

ONE PARAMETER FAMILY OF ESTIMATORS OF POPULATION MEAN IN TWO-OCCASION SUCCESSIVE SAMPLING

Garib Nath Singh
Department of Applied Mathematics, Indian School of Mines, Dhanbad
Alok Kumar Singh [1]
Department of Applied Mathematics, Indian School of Mines, Dhanbad

1. Introduction

Successive sampling often applied to gather information in applied sciences and socio-economic researches where characters are liable to change over the time. Government agencies such as National Bureau of Statistics and other research based institutions collect information on regular basis to estimate important population parameters and to find the patterns of variations in these parameters over the period of time.

The theory of successive sampling appears to have started with the work of Jessen (1942). He utilized entire information collected on the previous occasion in order to produce the reliable estimates on current occasion. This theory was further extended by Patterson (1950), Rao and Graham (1964), Gupta (1979), Das (1982) and Chaturvedi and Tripathi (1983), among others. Sen (1971) developed estimators of the population mean on the current occasion using information on two auxiliary variables which were readily available on previous occasion. Sen (1972, 1973) extended his work for several auxiliary variables. Singh *et al.* (1991), and Singh and Singh (2001) used the auxiliary information on current occasion for estimating the current population mean in two occasion successive sampling. Singh (2001) extended his work for h-occasion successive sampling.

In many practical situations, information on an auxiliary variable may be readily available on the first as well as on the second occasion, for example, tonnage (or seat capacity) of each vehicle or ship may be known in transportation survey, many other examples may be cited where the information on auxiliary variables are readily available on both the occasion in two-occasion successive sampling. Utilizing the auxiliary information on both the occasions , Feng and Zou (1997), Biradar and Singh (2001), Singh (2005), Singh and Priyanka (2006, 2007), Singh and Priyanka (2008, 2010), Singh and Karna (2009), Singh and Vishwakarma

[1] Corresponding Author. E-mail: aksingh.ism@gmail.com

(2009), Singh and Priyanka (2010), Singh *et al.* (2011), Singh and Prasad (2013), Singh and Homa (2013) have proposed varieties of estimators of population mean on current (second) occasions in two occasion successive sampling.

Singh and Shukla (1987) proposed an one parameter family of factor type ratio estimators of population mean in single occasion sample survey and presented its nice behaviours. The main attraction of the proposed class of estimators was, it produces the precise estimate and at the same time controls the bias as well. This type of the behaviour was not found in any other class of estimators. Motivated with these points, we wish to extend the possible aspects of one parameter family of factor type ratio estimators in two occasion successive sampling. Hence, in this paper we propose two classes of estimators of current population mean in two-occasion successive sampling under the assumption that the information on a stable auxiliary variable is readily available on both occasions. Properties of the proposed classes of estimators have been studied and suitable recommendations are made.

2. Formulation of the Classes of Estimators

Let $U = (U_1, U_2, ..., U_N)$ be the finite population of N units,which has been sampled over two occasions. The character under study be denoted by $x(y)$ on the first (second) occasion respectively. It is assumed that the information on an auxiliary variable z (stable over occasion) whose population mean known and closely related to x and y are available on first (second) occasion respectively. Let a simple random sample (without replacement) of size n is drawn on the first occasion. A random sub sample m of $m = n\lambda$ units is retained (matched) for its use on the second occasion. While a fresh simple random sample (without replacement) of size u of $u = (n - m) = n\mu$ is drawn on the second occasion from the entire population so that the sample size on the current (second) occasion is also n. Here λ and μ $(\lambda + \mu = 1)$ are the fractions of matched and fresh samples respectively at the current (second) occasion. The values of λ or μ would be choosen optimally. The following notations have been considered for the further use:

$\bar{X}$, $\bar{Y}$: The population means of the study variables $x(y)$ on the first (second) occasion respectively.

$\bar{Z}$: The population mean of the auxiliary variable z.

$\bar{x}_n, \bar{x}_m, \bar{y}_u, \bar{y}_m\bar{y}_u, \bar{z}_n, \bar{z}_u$: The sample means of the respective variables based on the sample sizes shown in suffices.

$\rho_{yx}, \rho_{yz}, \rho_{xz}$: The population correlation coefficients between the variables shown in suffices.

$S_x^2 = (N-1)^{-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$: The Population variance of the variable x.

S_y^2, S_z^2: The population variances of the variables y and z respectively.

β_{yx}: The population regression coefficient of y on x.

C_x, C_y, C_z: The coefficient of variations of the variables shown in suffices.

To estimate the population mean $\bar{Y}$ on the current (second) occasion, two non overlapping samples of size u and m are available. Therefore, we suggest some relevant estimators based on these two samples.Knowing the attractive behaviours of one parameter family of factor type ratio estimators of Singh and Shukla (1987), we considered this estimator for estimation of population mean $\bar{Y}$ on current occasion based on fresh sample of size u.

$$T_u(d) = \bar{y}_u \left[\frac{(A+C)\bar{Z} + fB\bar{z}_u}{(A+fB)\bar{Z} + C\bar{z}_u} \right] \tag{1}$$

where $A = (d-1)(d-2),\ B = (d-1)(d-4),\ C = (d-2)(d-3)(d-4),\ f = \frac{n}{N}$ and d is non negative constant identified to minimize the mean square error of the class of estimator $T_u(d)$.

REMARK 1. *The classes of estimators $T_u(d)$ reduce to following form of estimators for different value of d.*

Value of d	*Estimator*	*Remark*
$d = 1$	$T_u(1) = \bar{y}_u \left(\frac{\bar{Z}}{\bar{z}_u}\right)$	*Ratio Estimator*
$d = 2$	$T_u(2) = \bar{y}_u \left(\frac{\bar{z}_u}{\bar{Z}}\right)$	*Product Estimator*
$d = 3$	$T_u(3) = \bar{y}_u \left(\frac{\bar{z}_u^*}{\bar{Z}}\right)$	*Dual to Ratio Estimator*
$d = 4$	$T_u(4) = \bar{y}_u$	*Sample Mean*
$d \to \infty$	$T_u(d) = T_u(1)$	*Ratio Estimator*

where

$$\bar{z}_u^* = \bar{y}_u \left[\frac{N\bar{Z} - n\bar{z}_u}{(N-n)\bar{Z}} \right]$$

Again we considered two more estimators $T_{jm}(j = 1, 2)$ proposed by Chand (1975) and Kiregyera (1980) for the estimation of current population mean $\bar{Y}$ based on the matched sample of size m and presented as

$$T_{1m} = \frac{\bar{y}_m}{\bar{x}_m} \left(\frac{\bar{x}_n}{\bar{z}_n} \bar{Z} \right) \tag{2}$$

and

$$T_{2m} = \frac{\bar{y}_m}{\bar{x}_m} [\bar{x}_n + b_{xz}^{(n)} (\bar{Z} - \bar{z}_n)] \tag{3}$$

where $b_{xz}^{(n)}$ is the sample regression coefficient between the variable shown in suffix. As per the philosophy of successive sampling, the class of estimators T_u is suitable to estimate the population mean on each occasion, while the estimators $T_{jm}(j = 1, 2)$ are more appropriate for estimating the change over two occasions. To address both the problems simultaneously, a suitable combination of T_u and T_{jm} are required.

Motivated with the above arguments the final estimators of current population mean $\bar{Y}$ is consider convex linear combinations of T_u and T_{jm} and presented as

$$T_j(d, \varphi_j) = \varphi_j T_u(d) + (1 - \varphi_j) T_{jm}\ ;\ (j = 1, 2) \tag{4}$$

where $\varphi_j (0 \le \varphi_j \le 1)$ is an unknown constant (scalar) to be determined under certain criterion.

3. Properties of the Proposed Classes of Estimators

3.1. Bias and Mean Square Error

Since the estimator $T_u(d)$ and T_{jm} $(j=1,2;d>0)$ are one-parameter family of factor-type ratio estimators and chain-type ratio, and chain-type ratio to regression estimators respectively and they are biased estimators of the population mean $\bar{Y}$. Therefore, the resulting classes of estimators are also biased estimators of $\bar{Y}$, Singh and Shukla (1987) and Cochran (1977). The bias B(.) and mean square errors M(.) of the estimators $T_j(d,\varphi_j)$ are derived under large sample assumption and up to the first order of approximations using the following transformations:
$\bar{y}_u=(1+e_0)\bar{Y}, \bar{y}_m=(1+e_1)\bar{Y}, \bar{x}_n=(1+e_2)\bar{X}, \bar{x}_m=(1+e_3)\bar{X}$,
$\bar{z}_u=(1+e_4)\bar{Z}, \bar{z}_n=(1+e_5)\bar{Z}, s_{xz}(n)=(1+e_6)S_{xz}, s_z^2(n)=(1+e_7)S_z^2$
such that $E(e_k)=0$ and $|e_k|\leq 1 \ \forall \ k=1,2,...,5$.

Under the above transformations, the estimators $T_u(d)$ and T_{jm} $(j=1,2;d>0)$ take the following forms:

$$T_u(d)=\bar{Y}(1+e_0)\left[\frac{(A+C)\bar{Z}+fB\bar{Z}(1+e_4)}{(A+fB)\bar{Z}+C\bar{Z}(1+e_4)}\right]$$

$$T_u(d)=\bar{Y}(1+e_0)(1+\theta_1e_4)(1+\theta_2e_4)^{-1} \tag{5}$$

where $\theta_1(d)=\frac{fB}{A+fB+C}$ and $\theta_2(d)=\frac{C}{A+fB+C}$

$$T_{1m}=\frac{\bar{Y}(1+e_1)}{\bar{X}(1+e_3)}\left[\frac{\bar{X}(1+e_2)}{\bar{Z}(1+e_5)}\bar{Z}\right]$$

$$T_{1m}=\bar{Y}(1+e_1)(1+e_3)^{-1}(1+e_2)(1+e_5)^{-1} \tag{6}$$

$$T_{2m}=\frac{\bar{Y}(1+e_1)}{\bar{X}(1+e_3)}[\bar{X}(1+e_2)+(1+e_6)(1+e_7)^{-1}\bar{Z}\beta_{xz}(-e_5)]$$

$$T_{2m}=\bar{Y}(-e_5)(1+e_1)(1+e_3)^{-1}(1+e_6)(1+e_7)^{-1}(1+e_1)(1+e_3)^{-1}\frac{\bar{Y}\bar{Z}\beta_{xz}}{\bar{X}} \tag{7}$$

Thus we have the following theorems:

Theorem 2. *Bias of the estimators $T_j(d)$ $(j=1,2;d>0)$ to the first order of approximations is obtained as*

$$B(T_j(d))=\varphi_jB(T_u(d))+(1-\varphi_j)B(T_{jm})\ ;\ (j=1,2) \tag{8}$$

where

$$B(T_u(d))=\bar{Y}\left(\frac{1}{u}-\frac{1}{N}\right)(\theta_2-\theta_1)(\theta_2C_z^2-\rho_{yz}C_yC_z) \tag{9}$$

$$B(T_{1m})=\bar{Y}\left[\left(\frac{1}{m}-\frac{1}{n}\right)(C_x^2-\rho_{yx}C_yC_x)+\left(\frac{1}{n}-\frac{1}{N}\right)(C_z^2-\rho_{yz}C_yC_z)\right] \tag{10}$$

$$B(T_{2m}) = \bar{Y}\left[\left(\frac{1}{m}-\frac{1}{n}\right)(C_x^2-\rho_{yx}C_yC_x)+\left(\frac{1}{n}-\frac{1}{N}\right)(\rho_{xz}^2C_z^2-\rho_{xz}\rho_{yz}C_yC_x)\right] \quad (11)$$

PROOF. The bias of the estimators $T_j(d)$ are given by

$$B(T_j(d)) = E[T_j(d)-\bar{Y}] = \varphi_j E(T_u(d)-\bar{Y}) + (1-\varphi_j)E(T_{jm}-\bar{y})$$

$$B(T_j(d)) = \varphi_j B(T_u(d)) + (1-\varphi_j)B(T_{jm}) \quad (12)$$

where $B(T_u(d)) = E[T_u(d)-\bar{Y}]$ and $B(T_{jm}) = E[T_{jm}-\bar{Y}]$

To derive the $B(T_u(d))$, we proceed as follows:

$$E[T_u(d)-\bar{Y}] = E[\bar{Y}(1+e_0)(1+\theta_1e_4)(1+\theta_2e_4)^{-1})] \quad (13)$$

Now we expand the right hand side of equation (13) binomially, taking expectations and retaining the terms up-to the first order of approximations, we have the expression of the bias of the estimator $T_u(d)$ as given in equation (9).

Similarly, the bias of the estimators T_{jm} ;(j=1,2) are written as

$$E[T_{1m}-\bar{Y}] = E[\bar{Y}(1+e_1)(1+e_3)^{-1}(1+e_2)(1+e_5)^{-1}-\bar{Y}] \quad (14)$$

$$\begin{aligned} E[T_{2m}-\bar{Y}] &= E[\bar{Y}(1+e_1)(1+e_3)^{-1}(1+e_2) \\ &- e_5(1+e_6)(1+e_7)^{-1}(1+e_1)(1+e_3)^{-1}\frac{\bar{Y}\bar{Z}\beta_{xz}}{\bar{X}}-\bar{Y})] \end{aligned} \quad (15)$$

Expanding the right hand side of equation (14) and (15) binomially, taking expectations retaining the terms up-to the first order of approximations, we have the expressions of the bias of the estimators T_{1m} and T_{2m} as shown in equations (10) and (11) respectively.

THEOREM 3. *Mean square error of the estimators $T_j(d)$ (j=1, 2) to the first degree of approximation are obtained as*

$$M(T_j(d)) = \varphi_j^2 M(T_u(d))_{min} + (1-\varphi_j)^2 M(T_{jm}); \quad (j=1,2) \quad (16)$$

where

$$M(T_u(d))_{min} = \left(\frac{1}{u}-\frac{1}{N}\right)(1-\rho_{yz}^2) \quad (17)$$

$$M(T_{1m}) = S_y^2\left(\frac{1}{m}(2-2\rho_{yx})+\frac{1}{n}(2\rho_{yx}-2\rho_{yz})-\frac{1}{N}(2-2\rho_{yz})\right) \quad (18)$$

$$M(T_{2m}) = S_y^2\left(\frac{1}{m}(2-2\rho_{yx})+\frac{1}{n}(2\rho_{yx}-1-\rho_{yz}^2)-\frac{1}{N}(1-\rho_{xz}^2)\right) \quad (19)$$

PROOF. Since the estimators $T_u(d)$ and T_{jm} (j=1, 2) are based on two different non overlapping samples, the covariance terms are of order $O(N^{-1})$, therefore, they are ignored for large population size. It is obvious that mean square errors of the estimators $T_j(d)$ (j=1, 2) are given by

$$M(T_j(d)) = E[T_j(d) - \bar{Y}]^2 = E[\varphi_j(T_u(d) - \bar{Y}) + (1 - \varphi_j)(T_{jm} - \bar{y})]^2$$

$$M(T_j(d)) = \varphi_j^2 M(T_u(d)) + (1 - \varphi_j)^2 M(T_{jm}); \quad (j = 1, 2) \tag{20}$$

where $M(T_u(d)) = E(T_u(d) - \bar{Y})^2$ and $M(T_{jm}) = E(T_{jm} - \bar{y})^2$
To derive the $M(T_u(d))$, we proceed as follows:

$$E(T_u(d) - \bar{Y})^2 = E(\bar{Y}(e_0 + \theta_1 e_4 - \theta_2 e_4) - \bar{Y})^2 \tag{21}$$

Now we expand the right hand side of equation (21) binomially, taking expectations and retaining the terms up-to the first order of approximations, we have the expression of the mean square error of the estimator $T_u(d)$ given as

$$M(T_u(d)) = \bar{Y}^2 \left(\frac{1}{u} - \frac{1}{N}\right)(C_y^2 + \theta^2 C_z^2 + 2\theta\rho_{yz}C_yC_z) \tag{22}$$

where $\theta(d) = \theta_1(d) - \theta_2(d)$.

The expression for mean square error of estimator $T_u(d)$ derive in equation (22) is a function of unknown constant d. To derive the optimum value of d, differentiating $M(T_u(d))$ with respect to d and writing θ' for the differential coefficient of and equate it to zero, we have

$$\frac{\partial M(T_u(d))}{\partial d} = 2\left(\frac{1}{u} - \frac{1}{N}\right)\bar{Y}^2\theta'[\theta C_z^2 + \rho_{yz}C_yC_z] = 0 \tag{23}$$

where $\theta' = \frac{\partial \theta}{\partial d}$.
Further from equation (24), we get

$$\frac{fB - C}{A + fB + C} = -V \tag{24}$$

where $V = \rho\frac{C_y}{C_x}$.
On simplifying the above equation, we have the cubic equation in the form of d. This equation can be solved for d in order to get optimum value say $d = d_0$. Since there will be three possible values of d. Although MSE will be minimum for all these three values of d a criterion for selecting suitable d can be set as follows: Out of all the possible values of optimum d, select the $d = d_0$ as the most suitable choice, which makes $|B(T_u(d))|$ smallest. Using equation (24) the minimum mean square error of the estimator $M(T_u(d))_{min}$ is given in (17).
Similarly the mean square errors of the estimators T_{jm}; $(j = 1, 2)$ is written as

$$E(T_{1m} - \bar{Y})^2 = E[\bar{Y}(e_1 + e_2 - e_3 - e_5) - \bar{Y}]^2 \tag{25}$$

$$E(T_{2m}-\bar{Y})^2 = E\left[\bar{Y}\left((e_1+e_2-e_3)-\frac{\bar{Z}}{\bar{X}}\beta_{xz}e_5\right)-\bar{Y}\right]^2 \tag{26}$$

Expanding the right hand side of equations (25) and (26) binomially, taking expectations both sides, retaining the terms up-to the first order of approximations, we have the expression of the mean square error of the estimator T_{1m} and T_{2m} as shown in equation (18) and (19).

Remark 4. *The above results are derived under the assumptions that the coefficients of variation of variables x,y and z are approximately equal. Reddy (1978) as described that the coefficients of variation is a stable quantity over the period of time. Since x and y are same study variable over two occasions and z is an auxiliary variables correlated to x and y, therefore their coefficient of variations are assumed equal and the results are derived under these assumption.*

3.2. Minimum mean square errors of the classes of estimator

Since the mean square errors of the estimator $T_j(d)$; $(j=1,2)$ in equation (16) are the functions of the unknown constants (scalar) φ_j , therefore, we have minimized them with respect to φ_j and subsequently the optimum value of φ_j is obtained as

$$\varphi_{j_{opt}} = \frac{M(T_{jm})}{M(T_u(d))+M(T_{jm})}; \quad (j=1,2\ ;\ d>0) \tag{27}$$

From equation (27), substituting the value of $\varphi_{j_{opt}}$ in equation (16) we get the optimum mean square errors of the estimator T_j as

$$M(T_j(d))_{opt} = \frac{M(T_u(d))_{min}M(T_{jm})}{M(T_u(d))_{min}+M(T_{jm})}; \quad (j=1,2\ ;\ d>0) \tag{28}$$

Further substituting the values from equations (17) - (19) in equations (27) and (28), the simplified values of $\varphi_{j_{opt}}$ and $M(T_j(d))_{opt}(j=1,2)$ are obtained as

$$\varphi_{1_{opt}} = \frac{\mu_1(A_6-\mu_1A_5)}{A_1-\mu_1A_9-\mu_1^2A_8} \tag{29}$$

$$M(T_1)_{opt} = \left[\frac{A_{14}+\mu_1^2A_{13}-\mu_1A_{15}}{A_1-\mu_1^2A_8-\mu_1A_9}\right]\frac{S_y^2}{n} \tag{30}$$

$$\varphi_{2_{opt}} = \frac{\mu_2(A_{18}+\mu_2A_{17})}{A_1-\mu_2A_{20}+\mu_2^2A_{20}} \tag{31}$$

$$M(T_2)_{opt} = \left[\frac{A_{23}-\mu_2^2A_{22}+\mu_2A_{24}}{A_1+\mu_2^2A_{19}-\mu_2A_{20}}\right]\frac{S_y^2}{n} \tag{32}$$

where $\mu_j = \frac{u_j}{n}(j=1,2)$ are the fractions of fresh sample.

$A_1 = 1-\rho yz^2, A_2 = (2-2\rho_{yx}), A_3 = (2\rho_{yx}-2\rho_{yz}), A_4 = (2-2\rho_{yx}), A_5 = A_3-fA_4,$

$$A_6 = A_2 + A_5, A_7 = A_1 + A_4, A_8 = A_3 - fA_7, A_9 = A_1 - A_2 - A_8, A_{10} = A_1A_2,$$
$$A_{11} = A_1A_3, A_{12} = A_1A_4, A_{13} = fA_{11} - f^2A_{12}, A_{14} = A_{10} + A_{11} - fA_{12},$$
$$A_{15} = A_{11} - fA_{12} + fA_{10} + A_{13}, A_{16} = (1 + \rho_{yz}^2 - 2\rho_{yx}), A_{17} = A_{16} + fA_1,$$
$$A_{18} = A_2 - A_{17}, A_{19} = A_{16} + 2fA_1, A_{20} = A_1 - A_2 + A_{19}, A_{21} = A_1A_{16},$$
$$A_{22} = fA_{21} + f^2A_{21}, A_{23} = A_{10} - A_{21} - fA_1^2, A_{24} = A_{21} - fA_{10} + fA_1^2 + A_{22},$$
$$f = \frac{n}{N}$$

4. Optimum replacement strategies of the classes of estimators $T_j(d)$

The optimum mean square errors $M(T_j)_{opt}$ (j=1, 2) in equations (30) and (32) are functions of μ_j (j=1, 2) (fractions of sample to be drawn afresh at the second occasion) which play important role in reducing the cost of the survey, therefore, to determine the optimum values of μ_j so that $\bar{Y}$ may be estimated with maximum precision and minimum cost, we minimize $M(T_1)_{opt}$ and $M(T_2)_{opt}$ with respect to μ_1 and μ_2 respectively which results in a quadratic equations in μ_1 and μ_2 , which are shown as

$$\mu_1^2 D_1 - 2\mu_1 D_2 - D_3 = 0 \tag{33}$$

and

$$\mu_2^2 D_4 - 2\mu_2 D_5 + D_6 = 0 \tag{34}$$

Solving the equation (33) and (34) for μ_1 and μ_2 respectively, the solutions of μ_j (say $\hat{\mu}_j$) $(j = 1, 2)$ are given as

$$\hat{\mu}_1 = \frac{D_2 \pm \sqrt{D_2^2 + D_1D_3}}{D_1} \tag{35}$$

and

$$\hat{\mu}_2 = \frac{D_5 \pm \sqrt{D_5^2 + D_4D_6}}{D_4} \tag{36}$$

where, $D_1 = (A_9A_{13} + A_8A_{15}), D_2 = (A_1A_{13} + A_8A_{14}), D_3 = (A_9A_{14} - A_1A_{15}),$ $D_4 = (A_{20}A_{22} - A_{19}A_{24}), D_5 = (A_1A_{22} + A_{19}A_{23}), D_6 = (A_1A_{24} + A_{20}A_{23}).$

From equations (35) and (36) it is clear that the real values of $\hat{\mu}_j$ $(j = 1, 2)$ exist, iff, the quantities under square roots are greater than or equal to zero. For any combinations of correlations, which satisfy the condition of real solutions, two real values of $\hat{\mu_j}$ $(j = 1, 2)$ are possible. Hence, while choosing the values of $\hat{\mu}_j$, it should be remembered that $0 \leq \hat{\mu}_j \leq 1$, all other values of $\hat{\mu}_j$ are said to be inadmissible. If both the values of $\hat{\mu}_j$ are admissible, the lowest one is the best choice as it reduces the cost of the survey. From equation (35) and (36), substituting the admissible value of $\hat{\mu}_j$ $(j = 1, 2)$ (say $\hat{\mu}_j^0$) in equation (30) and (32), we have the optimum values of mean square errors of the classes of estimator $T_j(d)$, which is shown below:

$$M(T_1)^*_{opt} = \left[\frac{A_{14} + \mu_1^{(0)2} A_{13} - \mu_1^0 A_{15}}{A_1 - \mu_1^{(0)2} A_8 - \mu_1^0 A_9} \right] \frac{S_y^2}{n} \tag{37}$$

$$M(T_2)^*_{opt} = \left[\frac{A_{23} - \mu_2^{(0)2} A_{22} + \mu_2^0 A_{24}}{A_1 + \mu_2^{(0)2} A_{19} - \mu_2^0 A_{20}} \right] \frac{S_y^2}{n} \tag{38}$$

5. Efficiency Comparison

The percent relative efficiencies of the classes of estimators $T_j(d, \varphi_j)(j = 1, 2)$ with respect to (i) sample mean estimator $\bar{y}_n$ when there is no matching and (ii) natural successive sampling estimator $\hat{\hat{Y}} = \varphi^* \bar{y}_u + (1 - \varphi^*)\bar{y}_m^{'}$ when no auxiliary information is used at any occasion where, $\bar{y}_m^{'} = \bar{y}_m + \beta_{yx}(\bar{x}_n - \bar{x}_m)$ have been computed for different choices of correlations and presented in Tables 1-2. Since $\bar{y}_n$ and $\hat{\hat{Y}}$ are unbiased estimators of $\bar{Y}$, therefore, following Sukhatme *et al.* (1984), the variance of $\bar{y}_n$ and optimum variance of $\hat{\hat{Y}}$ are given by

$$V(\bar{y}_n) = \left(\frac{1}{n} - \frac{1}{N} \right) S_y^2 \tag{39}$$

$$V(\hat{\hat{Y}}) = \left[1 + \sqrt{1 - \rho_{yx}^2} \right] \frac{S_y^2}{2n} - \frac{S_y^2}{N} \tag{40}$$

For different choices of correlations ρ_{yz}, ρ_{yx} and f Tables 1-2 present the optimum values of $\mu_j \quad (j = 1, 2)$ and the percent relative efficiencies E_1 and E_2 of T with respect to $\bar{y}_n$ and $\hat{\hat{Y}}$ respectively, where $E_1 = \frac{V(\bar{y}_n)}{M(T_1)^*_{opt}} \times 100$ and $E_2 = \frac{V(\hat{\hat{Y}})}{M(T_2)^*_{opt}} \times 100$.

6. Interpretation of results

From Table 1 it is clear that

(a) For fixed value of ρ_{yz} and different values of f, the values of μ_1, E_1 and E_2 are increasing with the increasing values of ρ_{yx} . This behaviour is in agreement with Sukhatme *et al.* (1984), results which explain that more the value of ρ_{yx} , more the fraction of fresh sample required at the current occasion.

(b) For fixed value of ρ_{yx} , the values of E_1 and E_2 are increasing while the values of μ_1 do not follow any definite pattern with the increasing value of ρ_{yz}.

(c) Minimum value of is observed as 0.2222, which indicates that the fraction to be replaced at the current occasion is as low as about 22 percent of the total sample size, which leads to appreciable amount of the survey cost.

From Table 2 it is observed that

TABLE 1

Optimum value of μ_1 and PREs of $T_1(d)$ with respect to $(\bar{y}_n)$ and $(\hat{\hat{Y}})$

f	ρ_{yz}	ρ_{yx}	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.6	μ_1	0.2755	0.2222	*	*	0.7945	0.7481
		E_1	116.34	121.44	**	**	161.07	182.20
		E_2	110.95	112.40	**	**	125.28	125.10
	0.7	μ_1	0.3282	0.3277	0.3049	0.1142	0.9315	0.7162
		E_1	139.79	147.38	156.51	166.16	196.31	222.28
		E_2	133.31	136.41	139.12	139.78	152.69	152.62
	0.8	μ_1	0.3282	0.3429	0.3578	0.3657	0.2498	0.6884
		E_1	179.94	190.82	204.47	222.38	246.53	297.06
		E_2	171.60	176.62	181.75	187.07	191.74	203.96
	0.9	μ_1	0.2786	0.2963	0.3185	0.3474	0.3865	0.3782
		E_1	275.62	294.37	318.32	350.69	398.77	484.55
		E_2	262.84	272.46	282.95	295.00	310.15	332.69
0.2	0.6	μ_1	0.3215	0.3077	0.2492	*	0.9656	0.7660
		E_1	109.14	114.75	121.13	**	156.32	174.86
		E_2	103.44	105.14	105.98	**	117.24	113.21
	0.7	μ_1	0.3449	0.3532	0.3545	0.3173	*	0.7432
		E_1	130.06	137.58	146.86	158.56	**	212.49
		E_2	123.28	126.06	128.51	130.23	**	137.57
	0.8	μ_1	0.3334	0.3500	0.3689	0.3881	0.3815	0.7790
		E_1	166.32	176.82	190.07	207.69	232.94	283.10
		E_2	157.64	162.01	166.31	170.59	174.70	183.294
	0.9	μ_1	0.2796	0.2976	0.3202	0.3500	0.3920	0.4423
		E_1	252.54	270.29	293.07	324.06	370.52	455.39
		E_2	239.37	247.66	256.43	266.16	277.89	294.83
0.3	0.6	μ_1	0.3504	0.3529	0.3421	0.2642	*	0.8004
		E_1	101.50	106.49	113.41	121.65	**	166.60
		E_2	**	**	**	**	**	101.47
	0.7	μ_1	0.3571	0.3703	0.3827	0.3849	0.2741	0.8103
		E_1	119.31	126.61	135.76	147.72	163.57	201.86
		E_2	112.20	114.50	116.36	117.56	116.83	120.52
	0.8	μ_1	0.3377	0.3556	0.3769	0.4021	0.4254	*
		E_1	151.55	161.55	174.26	191.32	216.31	**
		E_2	142.51	146.09	149.36	152.25	154.51	**
	0.9	μ_1	0.2805	0.2986	0.3216	0.3521	0.3959	0.4637
		E_1	228.00	244.57	265.94	295.23	339.58	422.13
		E_2	214.40	221.16	227.95	234.95	242.56	252.04

Note: * indicate μ_1 do not exist and ** indicate no gain.

TABLE 2
Optimum value of μ_2 and PREs of $T_2(d)$ with respect to $(\bar{y}_n)$ and $(\hat{\bar{Y}})$

f	ρ_{yz}	ρ_{yx}	0.4	0.5	0.6	0.7	0.8	0.9
	0.6	μ_2	0.4221	0.4444	0.4721	0.5081	0.5585	0.6414
		E_1	123.96	130.93	139.62	151.00	167.16	194.30
		E_2	118.21	121.19	124.11	127.02	130.01	133.41
0.1	0.7	μ_2	0.3946	0.4166	0.4440	0.4797	0.5303	0.6149
		E_1	144.92	153.44	164.12	178.20	198.41	232.85
		E_2	138.20	142.02	145.88	149.90	154.32	159.88
	0.8	μ_2	0.3539	0.3750	0.4015	0.4365	0.4868	0.5729
		E_1	183.10	194.57	209.06	228.36	256.47	305.51
		E_2	174.61	180.09	185.83	192.09	199.47	209.76
	0.9	μ_2	0.2846	0.3036	0.3277	0.3601	0.4080	0.4936
		E_1	276.54	295.64	320.11	353.31	402.89	493.14
		E_2	263.71	273.64	284.54	297.20	313.36	338.59
	0.6	μ_2	0.4221	0.4444	0.4721	0.5081	0.5585	0.6414
		E_1	115.36	122.46	131.41	143.33	160.66	190.93
		E_2	109.34	112.20	114.98	117.72	120.50	123.61
0.2	0.7	μ_2	0.3946	0.4166	0.4440	0.4797	0.5303	0.6149
		E_1	134.07	142.63	153.48	168.02	189.36	227.15
		E_2	127.07	130.68	134.29	138.00	142.02	147.06
	0.8	μ_2	0.3539	0.3750	0.4015	0.4365	0.4868	0.5729
		E_1	167.96	179.26	193.69	213.20	242.21	294.73
		E_2	159.20	164.25	169.48	175.11	181.66	190.81
	0.9	μ_2	0.2846	0.3036	0.3277	0.3601	0.4080	0.4936
		E_1	250.23	268.49	292.09	324.50	373.80	466.48
		E_2	237.18	246.01	255.58	266.53	280.35	302.01
	0.6	μ_2	0.4221	0.4444	0.4721	0.5081	0.5585	0.6414
		E_1	105.99	113.35	122.83	135.79	155.41	192.15
		E_2	101.56	102.50	105.28	108.06	111.00	114.73
0.3	0.7	μ_2	0.3946	0.4166	0.4440	0.4797	0.5303	0.6149
		E_1	122.10	130.80	142.04	157.48	181.00	225.54
		E_2	114.82	118.28	121.75	125.32	129.28	134.66
	0.8	μ_2	0.3539	0.3750	0.4015	0.4365	0.4868	0.5729
		E_1	151.12	162.29	176.78	196.81	227.62	286.91
		E_2	142.11	146.76	151.52	156.62	162.58	171.30
	0.9	μ_2	0.2846	0.3036	0.3277	0.3601	0.4080	0.4936
		E_1	221.00	238.28	260.89	292.50	341.86	439.49
		E_2	207.82	215.48	223.62	232.77	244.18	262.40

(a) For fixed value of ρ_{yz} and different values of f, the values of μ_2, E_1 and E_2 are increasing with increasing value of . This behaviour is in agreement with Sukhatme *et al.* (1984), results which explain that more the value of ρ_{yx}, more the fraction of fresh sample required at the current occasion.

(b) For fixed value of ρ_{yx} , the values of E_1 and E_2 are increasing while the values of μ_2 are decreasing with the increasing values of ρ_{yz} . This behaviour is highly desirable in terms of percent relative efficiencies as well as the cost of the survey. It concludes that if the information on highly correlated auxiliary variable is available, its use at estimation stage not only pay in terms of enhance precision of estimates but also reduces the cost of the survey as well.

(c) Minimum value of μ_2 is observed as 0.2846, which indicates that the fraction to be replaced at the current occasion is as low as about 28 percent of the total sample size, which leads to reduction in cost to an appreciable amount.

(d) For fixed values of ρ_{yx} and ρ_{yz} , the values of μ_2 are equal for different values of f.

7. Conclusions

It is clear from the above interpretation the use of an auxiliary character is highly rewarding in terms of the proposed classes of estimators. It is also clear that if a highly correlated auxiliary variable is used, relatively, only a smaller fraction of the sample on the current (second) occasion is required to be replaced by a fresh sample, which is reducing the cost of the survey. If one has to make choice between $T_1(d)$ and $T_2(d)$ it is visible from empirical results that for same combination of correlation, the PRE of $T_2(d)$ is more than the PRE of $T_1(d)$, which produces more reliable estimates than $T_1(d)$, hence it may be preferred for practical application.

Acknowledgements

Authors are thankful to the Reviewer for his valuable suggestions. Authors are also thankful to the Indian School of Mines, Dhanbad for providing financial assistance and necessary infrastructure to carry out the present research work.

References

R. S. Biradar, H. P. Singh (2001). *Successive sampling using auxiliary information on both occasions.* Calcutta Statistical Association Bulletin, 51, pp. 243–251.

D. K. Chaturvedi, T. P. Tripathi (1983). *Estimation of population ratio on two occasions using multivariate auxiliary information.* Journal of Indian Statistical Association, 21, pp. 113–120.

W. G. Cochran (1977). *Sampling Techniques, 3rd edition.* John Wiley & Sons, New York.

A. K. Das (1982). *Estimation of population ratio on two occasions.* Journal of Indian Society of Agricultural Statistics, 34, pp. 1–9.

S. Feng, G. Zou (1997). *Sample rotation method with auxiliary variable.* Comunication in Statistics -Theory and Methods, 26, no. 6, pp. 1497–1509.

P. C. Gupta (1979). *Sampling on two successive occasions.* Journal of Statistical Research, 13, pp. 7–16.

R. J. Jessen (1942). *Statistical investigation of a sample survey for obtaining farm facts.* Iowa Agricultural Experiment Station Research Bulletin, , no. 304, pp. 1–104.

H. D. Patterson (1950). *Sampling on successive occasions with partial replacement of units.* Journal of the Royal Statistical Society, 12, pp. 217–223.

J. N. K. Rao, J. E. Graham (1964). *Rotation design for sampling on repeated occasions.* Journal of American Statistical Association, 59, pp. 492–509.

V. N. Reddy (1978). *A study on the use of prior knowledge on certain population parameters in estimation.* Sankhya C, 407.

A. R. Sen (1971). *Successive sampling with two auxiliary variables.* Sankhya, 33, no. Series B, pp. 371–378.

A. R. Sen (1972). *Successive sampling with p (p 1) auxiliary variables.* Annals of Mathematical Statistics, 43, pp. 2031–2034.

A. R. Sen (1973). *Theory and application of sampling on repeated occasions with several auxiliary variables.* Biometrics, 29, pp. 381–385.

G. N. Singh (2001). *Estimation of population mean using auxiliary information on recent occasion in h-occasion successive sampling.* Statistics in Transition, 6, pp. 523–532.

G. N. Singh (2005). *On the use of chain-type ratio estimator in successive sampling.* Statistics in Transition, 7, no. 1, pp. 21–26.

G. N. Singh, F. Homa (2013). *Effective rotation patterns in successive sampling over two occasions.* Journal of Statistical Theory and Practice, 7, no. 1, pp. 146–155.

G. N. Singh, J. P. Karna (2009). *Estimation of population mean on current occasions in two-occasion successive sampling.* Metron, 67, no. 1, pp. 69–85.

G. N. Singh, S. Prasad (2013). *Best linear unbiased estimators of population mean on current occasion in two-occasion successive sampling.* Statistics in Transition-New Series, 14, no. 1, pp. 57–74.

G. N. Singh, K. Priyanka (2006). *On the use of chain-type ratio to difference estimator in successive sampling.* International Journal of Applied Mathematics and Statistics, 5, no. S06, pp. 41–49.

G. N. Singh, K. Priyanka (2007). *On the use of auxiliary information in search of good rotation patterns on successive occasions.* Bulletin of Statistics and Economics, 1, no. A07, pp. 42–60.

G. N. Singh, K. Priyanka (2008). *On the use of auxiliary variables to improve the precision of estimates at current occasion.* Journal of Indian Society of Agricultural Statistics, 62, no. 3, pp. 253–265.

G. N. Singh, K. Priyanka (2010). *Estimation of population mean at current occasion in presence of several varying auxiliary variates in two occasion successive sampling.* Statistics in Transition- new series, 11, no. 1, pp. 105–126.

G. N. Singh, V. K. Singh (2001). *On the use of auxiliary information in successive sampling.* Journal of Indian Society of Agricultural Statistics, 54, no. 1, pp. 1–12.

H. P. Singh, R. Tailor, S. Singh, J. M. Kim (2011). *Estimation of population variance in successive sampling.* Quality and Quantity, 45, pp. 477–194.

H. P. Singh, G. K. Vishwakarma (2009). *A general procedure for estimating population mean in successive sampling.* Communications in Statistics-Theory and Methods, 38, no. 2, pp. 293–308.

V. K. Singh, D. Shukla (1987). *One parameter family of factor-type ratio estimators.* Metron, 45, no. 1-2,30, pp. 273–283.

V. K. Singh, G. N. Singh, D. Shukla (1991). *An efficient family of ratio-cum-difference type estimators in successive sampling over two occasions.* Journal of Scientific Research, 41, pp. 149–159.

P. V. Sukhatme, B. V. Shukhatme, S. Shukhatme, C. Ashok (1984). *Sampling theory of surveys with applications.* Iowa State University Press, Ames, Iowa (USA) and Indian Society of Agricultural Statistics, New Delhi., New York.

Summary

The goal of this paper is to consider the problem of estimation of the population mean on current (second) occasion in two-occasion successive sampling. Some classes of estimators have been proposed and their detailed behaviours are examined. The dominance of the suggested classes of estimators has been established over sample mean estimator when there is no matching from the previous occasion and the natural successive sampling estimator. Optimum replacement strategies have been discussed. Empirical studies are carried out to study the performances of the proposed classes of estimator and suitable recommendations have been made.

Keywords: Successive sampling; Auxiliary information; Bias; Mean square error, Optimum replacement policy.