On the Use of Multiple Auxiliary Variables in Estimation of Current Population Mean in Two-Occasion Successive (Rotation) Sampling

G. N. Singh, J. P. Karna and S. Prasad
Indian School of Mines, Dhanbad, India

ABSTRACT

The present work emphasizes the role of several stable auxiliary variables at both the occasions to improve the precision of estimates at current occasion in two-occasion successive sampling. A chain-type multiple linear regressions in ratio estimator has been proposed and its theoretical properties are examined. Relative comparison of efficiencies of the proposed estimator with the sample mean estimator, when there is no matching from the previous occasion and the natural successive sampling estimator, when no auxiliary information is used have been made. Theoretical results have been well supported with empirical illustrations.

Key words: Successive sampling, stable auxiliary variables, bias, mean square error, optimum replacement policy.

1 Introduction

Surveys often gets repeated on many occasions for estimating same characteristics at different points of time. The information collected on previous occasion may be used to study the change or the total value over occasion for the character and also in addition to study the average value for the most recent occasion. For example, in agricultural survey one may be interested in (i) estimating the average amount of yield per acre of an important crop (say wheat) in current season (ii) estimating the change in average amount of yield for a province (county) over two different seasons and (iii) estimating both parameters from (i) and (ii) simultaneously. The successive method of sampling consists of selecting sample units on different occasions such that some units are common with samples selected on previous occasions. If sampling on successive occasions is done according to a specific rule, with partial replacement of sampling units, it is known as successive sampling. A key issue is the extent to which elements sampled at a previous occasion should be retained in the sample selected at the current occasion; this is termed as optimum replacement policy.

The method of successive sampling was developed by Jessen [1] and extended by Patterson [2], Eckler [3], Rao and Graham [4], Sen [5, 6, 7], Cochran [8], Gupta [9], Singh et al. [10], Feng and Zou [11], Biradar and Singh [12], Singh and Singh [13], Singh [14], Singh [15], Singh and Priyanka [16, 17, 18, 19] and Singh and Karna [20, 21] among others.

Sometimes, information on several auxiliary variables may be readily available or may be made easily available by diverting a small amount of fund available for the survey. For instance, to study the case of public health and welfare of a state or country, several factors are available that can be treated as auxiliary variables, such as the number of beds, number of doctors and supporting staffs in different hospitals, the amount of funds available for medicine etc. may be known. Likewise, there may be several information available, which if efficiently utilized can improve the precision of estimates. Following the work of Singh and Priyanka [19], the objective of the present work is to propose estimator for estimating the population mean at current occasion using information on several stable auxiliary variables available on both the occasions. Utilizing information on p stable auxiliary variables, chain-type multiple linear regressions in ratio estimator has been proposed. Relative comparison of efficiencies of the proposed estimator with the sample mean estimator, when there is no matching from the previous occasion and the natural successive sampling estimator, when no auxiliary information is used have been made. Empirical studies show the highly significant gains for the proposed estimator.

2 Formulation of Estimator

Let $U = (U_1, U_2, \ldots, U_N)$ be the finite population of N units, which has been sampled over two occasions. The character under study is denoted by $x(y)$ on the first (second) occasion, respectively. It is assumed that the information on p-stable (non negative integer constant) auxiliary variables $z_j (j = 1, 2, ..., p)$ with known population means and correlated to x and y

on the first and second occasions respectively, are readily available on the first as well as on the second occasion. A simple random sample (without replacement) of size n is drawn on the first occasion. A random sub-sample of size $m = n\lambda$ is retained (matched) for its use on the second occasion, while a fresh simple random sample (without replacement) of size $u = (n - m) = n\mu$ is drawn on the second occasion from the entire population so that the sample size on the second occasion is also n. λ and μ ($\lambda+\mu$ =1) are the fractions of the matched and fresh samples, respectively, at the current(second) occasion. The values of λ or μ should be chosen optimally. Following notations have been considered for further use:

$\bar{X}$, $\bar{Y}$: Population means of x and y respectively.

$\bar{Z}_j$: Population means of the $j^{th}(j = 1, 2, ..., p)$ auxiliary variable.

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

$\bar{z}_{ju}$, $\bar{z}_{jn}$, $\bar{z}_{jm}$: Sample means of the $j^{th}(j = 1, 2, ..., p)$ auxiliary variable based on the sample sizes shown in suffices.

ρ_{yx}, ρ_{xzj}, ρ_{yzj}, ρ_{zjzk} : Correlation coefficients between the variables shown in suffices, where $j \neq k = 1, 2, ..., p$.

$S_x^2 = (N-1)^{-1} \sum_{i=1}^{N} (x_i - \bar{X})^2$: Population mean square of x.

S_y^2, S_{zj}^2 : Population mean squares of y and $z_j(j = 1, 2, ..., p)$ respectively.

To estimate the population mean $\bar{Y}$ on the current (second) occasion, utilizing information on p-stable auxiliary variables, two different estimators are suggested. One is a simple multiple linear regression estimator based on sample of size $u = (n\mu)$ drawn afresh on the current occasion and is given by

$$T_1 = \bar{y}_u + \sum_{j=1}^{p} b_{yzj}(u)(\bar{Z}_j - \bar{z}_{ju}) \qquad (1)$$

The second estimator is a chain-type multiple linear regressions in ratio estimator based on the sample of size $m = (n\lambda)$ common with both the occasions and is defined as

$$T_2 = \frac{\bar{y}_m^*}{\bar{x}_m^*} \bar{x}_n^* \qquad (2)$$

where $\bar{y}_m^* = \bar{y}_m + \sum_{j=1}^{p} b_{yzj}(m)(\bar{Z}_j - \bar{z}_{jm})$, $\bar{x}_m^* = \bar{x}_m + \sum_{j=1}^{p} b_{xzj}(m)(\bar{Z}_j -$

$\bar{z}_{jm})$, $\bar{x}_n^* = \bar{x}_n + \sum_{j=1}^{p} b_{xzj}(n)(\bar{Z}_j - \bar{z}_{jn})$ and $b_{yzj}(u), b_{yzj}(m), b_{xzj}(n)$ and $b_{xzj}(m)$ $(j = 1, 2, ..., p)$ are the sample regression coefficients between the variables shown in suffices and based on the sample sizes shown in braces. Considering the convex linear combination of the estimators T_1 and T_2 defined in equations (1) and (2), we have the final estimator T for estimating the population mean $\bar{Y}$ on the current (second) occasion, which is defined as:

$$T = \varphi T_1 + (1 - \varphi) T_2 \qquad (3)$$

where φ is the unknown constant to be determined under certain criterion.

3 Properties of the Estimator T

Since, T_1 and T_2 are simple multiple linear regression and chain-type multiple linear regressions in ratio estimator, they are biased for population mean $\bar{Y}$, therefore, the resulting estimator T defined in equation (3) is also biased estimator of $\bar{Y}$. The bias $B(.)$ and mean square error $M(.)$ of the estimator T up-to the first order of approximations are derived under large sample approximations using the following transformations:
$\bar{y}_u = \bar{Y}(1 + e_1)$, $\bar{y}_m = \bar{Y}(1 + e_2)$, $\bar{x}_m = \bar{X}(1 + e_3)$, $\bar{x}_n = \bar{X}(1 + e_4)$, $\bar{z}_{ju} = \bar{Z}_j(1 + e_{5j})$, $\bar{z}_{jm} = \bar{Z}_j(1 + e_{6j})$, $\bar{z}_{jn} = \bar{Z}_j(1 + e_{7j})$, $s_{yzj}(u) = S_{yzj}(1+e_{8j})$, $s^2_{zj}(u) = S^2_{zj}(1+e_{9j})$, $s_{yzj}(m) = S_{yzj}(1+e_{10j})$, $s^2_{zj}(m) = S^2_{zj}(1 + e_{11j})$, $s_{xzj}(m) = S_{xzj}(1 + e_{12j})$, $s_{xzj}(n) = S_{xzj}(1 + e_{13j})$ and $s^2_{zj}(n) = S^2_{zj}(1 + e_{14j})$; Such that $E(e_i) = 0$ and $E(e_{hj}) = 0$, $|e_i| \leq 1 \quad \forall i = 1, 2, 3, 4$ and $|e_{hi}| \leq 1 \quad \forall h = 5, 6, 7, ..., 14$. where $(j = 1,$ 2, 3, $..., p)$. Under the above transformations T_1 and T_2 take the following

forms:

$$T_1 = \bar{Y}(1+e_1) - \sum_{j=1}^{p} \beta_{yzj} \bar{Z}_j e_{5j}(1+e_{8j})(1+e_{9j})^{-1} \quad (4)$$

$$\begin{aligned} T_2 &= \left\{ \bar{Y}(1+e_2) - \sum_{j=1}^{p} \beta_{yzj} \bar{Z}_j e_{6j}(1+e_{10j})(1+e_{11j})^{-1} \right\} \\ &\times \left\{ (1+e_4) - \sum_{j=1}^{p} \frac{\beta_{xzj}\bar{Z}_j}{\bar{X}} e_{7j}(1+e_{13j})(1+e_{14j})^{-1} \right\} \\ &\times \left\{ 1 + \left(e_3 - \sum_{j=1}^{p} \frac{\beta_{xzj}\bar{Z}_j}{\bar{X}} e_{6j}(1+e_{12j})(1+e_{11j})^{-1} \right) \right\}^{-1} \end{aligned} \quad (5)$$

Thus, we have the following theorems:

Theorem 1: Bias of the estimator T to the first order of approximations is obtained as

$$B(T) = \varphi B(T_1) + (1-\varphi) B(T_2) \quad (6)$$

where

$$B(T_1) = \left(\frac{1}{u} - \frac{1}{N}\right) \sum_{j=1}^{p} \beta_{yzj} \left(\frac{\alpha_{003}}{S_{zj}^2} - \frac{\alpha_{012}}{S_{yzj}} \right) \quad (7)$$

$$\begin{aligned} B(T_2) &= \left(\frac{1}{m} - \frac{1}{N}\right) \sum_{j=1}^{p} \beta_{yzj} \left(\frac{\alpha_{003}}{S_{zj}^2} - \frac{\alpha_{012}}{S_{yzj}} \right) + \left(\frac{1}{m} - \frac{1}{n}\right) \frac{\bar{Y}}{\bar{X}} \left[\left(\frac{S_x^2}{\bar{X}} - \frac{S_{yx}}{\bar{Y}} \right) \right. \\ &+ \sum_{j=1}^{p} \beta_{xzj} \left(\frac{\alpha_{102}}{S_{xzj}} - \frac{\alpha_{003}}{S_{zj}^2} \right) + \sum_{j=1}^{p} \beta_{xzj} \left(\frac{S_{yzj}}{\bar{Y}} - \frac{S_{xzj}}{\bar{X}} \right) \\ &+ \left. \sum_{j \neq k=1}^{p} \beta_{xzk} \left(\frac{\beta_{xzj}}{\bar{X}} - \frac{\beta_{yzj}}{\bar{Y}} \right) S_{zjzk} \right] \end{aligned} \quad (8)$$

where $\alpha_{rst} = E[(x-\bar{X})^r (y-\bar{Y})^s (z_j - \bar{Z}_j)^t]$; $(r, s, t \geq 0)$ are integers and $(j = 1, 2, 3, ..., p)$. Proof: The bias of the estimator T is given by

$$\begin{aligned} B(T) &= E[T - \bar{Y}] = \varphi E(T_1 - \bar{Y}) + (1-\varphi) E(T_2 - \bar{Y}) \\ &= \varphi B(T_1) + (1-\varphi) B(T_2) \end{aligned} \quad (9)$$

where $B(T_1) = E\left[T_1 - \bar{Y}\right]$ and $B(T_2) = E\left[T_2 - \bar{Y}\right]$.
The bias of the estimators T_1 and T_2 are derived as follows:

$$B(T_1) \quad = \quad E\left[T_1 - \bar{Y}\right]$$

Substituting the expression of the estimator T_1 from equation (4) and expanding it binomially, taking expectations and retaining the terms up-to the first order of approximations, we have the bias of the estimator T_1 as shown in equation (7). Similarly, with the help of expression (5) we get the bias of the estimator T_2 as shown in equation (8). Now substituting the values of $B(T_1)$ and $B(T_2)$ in the equation (9), we get the bias of the estimator T as shown in equation (6).
Theorem 2: Mean square error of the estimator T to the first order of approximations is obtained as

$$M(T) = \varphi^2 M(T_1) + (1-\varphi)^2 M(T_2) + 2\varphi(1-\varphi)C(T_1,T_2) \tag{10}$$

where

$$M(T_1) = \left(\frac{1}{u} - \frac{1}{N}\right)\left(1 - \sum_{j=1}^{p} \rho_{yzj}^2 + \sum_{j\neq k=1}^{p} \rho_{yzj}\rho_{yzk}\rho_{zjzk}\right) S_y^2 \tag{11}$$

$$\begin{aligned} M(T_2) = & \left[\left(\frac{1}{m} - \frac{1}{N}\right)\left(1 - \sum_{j=1}^{p} \rho_{yzj}^2 + \sum_{j\neq k=1}^{p} \rho_{yzj}\rho_{yzk}\rho_{zjzk}\right) + \left(\frac{1}{m} - \frac{1}{n}\right)\right. \\ & \left. \times \left\{1 - 2\rho_{yx} + \sum_{j=1}^{p} \rho_{yzj}^2 - \sum_{j\neq k=1}^{p} \rho_{yzj}\rho_{yzk}\rho_{zjzk}\right\}\right] S_y^2 \end{aligned} \tag{12}$$

$$C(T_1,T_2) = -\frac{1}{N}\left(1 - \sum_{j=1}^{p} \rho_{yzj}^2 + \sum_{j\neq k=1}^{p} \rho_{yzj}\rho_{yzk}\rho_{zjzk}\right) S_y^2 \tag{13}$$

Proof: It is obvious that mean square error of the estimator T is given by

$$\begin{aligned} M(T) &= E[T-\bar{Y}]^2 = E[\varphi(T_1-\bar{Y}) + (1-\varphi)(T_2-\bar{Y})]^2 \\ &= \varphi^2 M(T_1) + (1-\varphi)^2 M(T_2) + 2\varphi(1-\varphi)C(T_1,T_2) \end{aligned} \tag{14}$$

where $M(T_1) = E[T_1-\bar{Y}]^2$, $M(T_2) = E[T_2-\bar{Y}]^2$ and $C(T_1,T_2) = E[(T_1-\bar{Y})(T_2-\bar{Y})]$
The mean square errors of the estimators T_1 and T_2 are derived as follows:

$$M(T_1) \quad = \quad E\left[T_1 - \bar{Y}\right]^2$$

Substituting the expression of the estimator T_1 from equation (4) and expanding it binomially, taking expectations and retaining the terms up-to the first order of approximations, we have the expression of $M(T_1)$ as shown in equation (11).

Similarly, with the help of expression (5) we get

$$M(T_2) = \Bigg[\left(\frac{1}{m}-\frac{1}{N}\right)\left(1-\sum_{j=1}^{p}\rho_{yzj}^2+\sum_{j\neq k=1}^{p}\rho_{yzj}\rho_{yzk}\rho_{zjzk}\right)+\left(\frac{1}{m}-\frac{1}{n}\right) \times \Bigg\{1-2\rho_{yx}-\sum_{j=1}^{p}\rho_{xzj}^2+2\sum_{j=1}^{p}\rho_{xzj}\rho_{yzj}+\sum_{j\neq k=1}^{p}\rho_{xzj}\rho_{xzk}\rho_{zjzk} -2\sum_{j\neq k=1}^{p}\rho_{xzj}\rho_{yzk}\rho_{zjzk}\Bigg\}\Bigg] S_y^2$$

Further, we consider $\rho_{xzj} = \rho_{yzj}, (j = 1, 2, ..., p)$, which is an intuitive assumption, considered by Cochran [8] and Feng and Zou [11]. In the light of this assumption the above equation of $M(T_2)$ takes the form as shown in equation (12).
The covariance type term between the estimators T_1 and T_2 is derived as

$$C(T_1, T_2) = E[(T_1 - \bar{Y})(T_2 - \bar{Y})]$$

Substituting the expressions of T_1 and T_2 from the equations (4) and (5), expanding binomially, taking expectations and retaining the terms up-to order n^{-1}, we have the expression of $C(T_1, T_2)$ as shown in equation (13).
Now substituting the values of $M(T_1)$, $M(T_2)$ and $C(T_1, T_2)$ in the equation (14), we get the mean square error of the estimator T as it is given in equation (10).
Remark 1: Results shown in equations (11) - (13) are derived under the assumption that the coefficients of variation of x, y and z are approximately equal.

4 Minimum Mean Square Error of the Estimator T

Since, the mean square error of the estimator T in equation (10) is a function of unknown constant φ , therefore, it is minimized with respect to φ and subsequently the optimum value of φ is obtained as

$$\varphi_{opt} = \frac{M(T_2) - C(T_1, T_2)}{M(T_1) + M(T_2) - 2C(T_1, T_2)} \tag{15}$$

Now substituting the value of φ_{opt} from equation (15) in equation (10), we get the optimum mean square error of the estimator T as

$$M(T)_{opt} = \frac{M(T_1).M(T_2) - \{C(T_1, T_2)\}^2}{M(T_1) + M(T_2) - 2C(T_1, T_2)} \tag{16}$$

Further, substituting the values from equations (11) - (13) in equations (15) and (16), we get the simplified values of φ_{opt} and $M(T)_{opt}$ as:

$$\varphi_{opt} = \frac{\mu[A_1 + \mu A_2]}{A_1 + \mu^2 A_2} \tag{17}$$

$$M(T)_{opt} = \frac{A_1}{n}\left[\frac{A_1 + \mu A_2}{A_1 + \mu^2 A_2} - f\right] S_y^2 \tag{18}$$

where $A_1 = 1 - \sum_{j=1}^{p} \rho_{yzj}^2 + \sum_{j \neq k=1}^{p} \rho_{yzj}\rho_{yzk}\rho_{zjzk}$,
$A_2 = 1 - 2\rho_{yx} + \sum_{j=1}^{p} \rho_{yzj}^2 - \sum_{j \neq k=1}^{p} \rho_{yzj}\rho_{yzk}\rho_{zjzk}$, $f = \frac{n}{N}$ and μ is the fraction of the fresh sample drawn on the current (second) occasion corresponding to the estimator T.

5 Optimum Replacement Policy

To determine the optimum value of μ so that population mean $\bar{Y}$ may be estimated with the maximum precision, we minimize $M(T)_{opt}$ in equation (18) with respect to μ, which result in quadratic equation in μ, which is shown as

$$A_2\mu^2 + 2A_1\mu - A_1 = 0 \tag{19}$$

Solving equation (19) for μ, the solutions are given as

$$\mu_{opt} = \frac{-A_1 \pm \sqrt{A_1^2 + A_1 A_2}}{A_2} \tag{20}$$

The real values of μ_{opt} exist, if $(A_1^2 + A_1 A_2) \geq 0$. For any combinations of correlations, which satisfies this condition, two real values of μ_{opt} are possible, hence, to choose a value of μ_{opt}, it should be remembered that $0 \leq \hat{\mu} \leq 1$, all other values of μ_{opt} are inadmissible. If both the values are admissible, the lower one is the best choice. Substituting the admissible values of μ_{opt} say μ_0 from equation (20) in equation (18), we have the optimum value of the mean square error of the estimator T, which is shown as

$$M(T^0)_{opt} = \frac{A_1}{n}\left[\frac{A_1+\mu_0 A_2}{A_1+\mu_0^2 A_2} - f\right] S_y^2 \tag{21}$$

6 Special Case

There may be several instances where the p-auxiliary variables are mutually uncorrelated but they are correlated to study variable. For example, in a survey of commercial products say the aim is to estimate the number of persons reading newspaper, then in that case the number of copies produced by different newspapers are different and the number of copies produced by a particular newspaper company is uncorrelated to the number of copies produced by another newspaper but both are correlated to the study variable, i.e., number of persons reading newspaper. Hence, for modeling such type of situations in the proposed estimator, we consider the case where the p-auxiliary variables are mutually uncorrelated, i.e., $\rho_{zjzk} = 0 \quad \forall j \neq k = 1, 2, ..., p$ then the expression for the optimum value of μ and $M(T^0)_{opt}$ reduces to

$$\mu_{opt} = \frac{-A_1^* \pm \sqrt{A_1^{*2} + A_1^* A_2^*}}{A_2^*}$$

and

$$M(T^0)_{opt} = \frac{A_1^*}{n}\left[\frac{A_1^*+\mu_0 A_2^*}{A_1^*+\mu_0^2 A_2^*} - f\right] S_y^2$$

where

$$A_1^* = 1 - \sum_{j=1}^{p} \rho_{yzj}^2 \quad and \quad A_2^* = 1 - 2\rho_{yx} + \sum_{j=1}^{p} \rho_{yzj}^2$$

7 Efficiency Comparison

The percent relative efficiencies of T with respect to (i) sample mean $\bar{y}_n$, when there is no matching and (ii) $\hat{\hat{Y}} = \varphi^* \bar{y}_u + (1-\varphi^*)\bar{y}_m^{'}$, 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 obtained for different choices of the correlations involved. Variance of the estimator $\bar{y}_n$ and the optimum variance of the estimator $\hat{\hat{Y}}$ are given by

$$V(\bar{y}_n) = \frac{1}{n}(1-f)\,S_y^2 \tag{22}$$

$$V(\hat{\hat{Y}})_{opt^*} = \frac{1}{2n}\left[1+\sqrt{(1-\rho_{yx}^2)}-2f\right]S_y^2 \tag{23}$$

The percent relative efficiencies E_1 and E_2 of T (under optimal condition) with respect to $\bar{y}_n$ and $\hat{\hat{Y}}$ respectively are given by

$$E_1 = \frac{V(\bar{y}_n)}{M(T^0)_{opt}} \times 100 \text{ and } E_2 = \frac{V(\hat{\hat{Y}})_{opt^*}}{M(T^0)_{opt}} \times 100$$

8 Empirical Study

The expressions of the optimum μ ($i.e. \mu_0$) and the percent relative efficiencies E_1 and E_2 are in terms of population correlation coefficients. Therefore, the values of μ_0, E_1 and E_2 have been computed for different choices of positive correlations while the value of f (sampling fraction) is chosen as 0.1. For empirical studies, cases of $p = 1, 2$ and 3 have been considered.

Case 1: For $p = 1$, the values of A_1 and A_2 take the form $A_1 = 1 - \rho_{yz}^2$ and $A_2 = 1 - 2\rho_{yx} + \rho_{yz}^2$, which is the work of Singh and Karna [21].

Case 2: For $p = 2$ and assuming that the two auxiliary variables are correlated i.e., $\rho_{z1z2} \neq 0$. The values of A_1 and A_2 are given by $A_1 = 1 - \left(\rho_{yz1}^2 + \rho_{yz2}^2\right) + 2\rho_{yz1}\rho_{yz2}\rho_{z1z2}$ and $A_2 = 1 - 2\rho_{yx} + \left(\rho_{yz1}^2 + \rho_{yz2}^2\right) - 2\rho_{yz1}\rho_{yz2}\rho_{z1z2}$. Substituting these values of A_1 and A_2 in equations (20) and (21), we have the values of optimum μ, $M(T^0)_{opt}$, E_1 and E_2. For different choices of correlations, Tables 1 - 2 show the optimum values of μ i.e., μ_0 and percent relative efficiencies E_1 and E_2 of the estimator T (under optimal condition) with respect to $\bar{y}_n$ and $\hat{\hat{Y}}$ respectively.

Case 3: For $p = 2$ and assuming that the two auxiliary variables are uncorrelated i.e., $\rho_{z1z2} = 0$. The values of A_1^* and A_2^* are given by $A_1^* = 1 - \left(\rho_{yz1}^2 + \rho_{yz2}^2\right)$ and $A_2^* = 1 - 2\rho_{yx} + \left(\rho_{yz1}^2 + \rho_{yz2}^2\right)$. Using these values in equations (20) and (21), the optimum values of μ, E_1 and E_2 are shown in Table 3.

Case 4: For $p = 3$ and assuming that the two auxiliary variables are correlated i.e., $\rho_{z1z2} \neq 0 \quad \forall j \neq k = 1, 2, 3$. In this case the values of A_1 and A_2 take the following form:

Table 1: Optimum values of μ and percent relative efficiencies of T with respect to $\bar{y}_n$ and $\hat{\hat{Y}}$ for $\rho_{yx} = 0.7$.

ρ_{yz1}		0.5			0.7			0.9		
ρ_{yz2}	ρ_{z1z2}	μ_0	E_1	E_2	μ_0	E_1	E_2	μ_0	E_1	E_2
0.5	0.5	0.527	141.6	119.1	0.502	164.6	138.5	0.446	226.2	190.2
	0.7	0.543	129.1	108.6	0.527	141.6	119.1	0.493	172.9	145.4
	0.9	0.557	118.8	**	0.549	124.7	104.9	0.527	141.6	119.1
0.7	0.5	0.502	164.6	138.5	0.479	187.2	157.5	0.425	253.8	213.5
	0.7	0.527	141.6	119.1	0.520	148.0	124.5	0.496	170.3	143.3
	0.9	0.549	124.7	104.9	0.550	123.5	103.9	0.541	130.9	110.1
0.9	0.5	0.446	226.2	190.2	0.425	253.8	213.5	0.360	367.6	309.2
	0.7	0.493	172.9	145.4	0.496	170.3	143.3	0.480	186.2	156.6
	0.9	0.527	141.6	119.1	0.541	130.9	110.1	0.541	130.4	109.7

Note: "**" indicates no gain.

Table 2: Optimum values of μ and percent relative efficiencies of T with respect to $\bar{y}_n$ and $\hat{\hat{Y}}$ for $\rho_{yx} = 0.9$.

ρ_{yz1}		0.5			0.7			0.9		
ρ_{yz2}	ρ_{z1z2}	μ_0	E_1	E_2	μ_0	E_1	E_2	μ_0	E_1	E_2
0.5	0.5	0.695	182.3	125.1	0.635	214.9	147.6	0.582	304.4	209.0
	0.7	0.673	164.7	113.14	0.659	182.3	125.1	0.628	226.8	155.7
	0.9	0.685	150.5	103.3	0.678	158.7	108.9	0.659	182.3	125.1
0.7	0.5	0.635	214.9	147.6	0.614	247.4	169.9	0.562	345.5	237.2
	0.7	0.659	182.3	125.1	0.652	191.3	131.3	0.630	223.1	153.1
	0.9	0.678	158.7	108.9	0.679	157.0	107.8	0.671	167.3	114.9
0.9	0.5	0.582	304.4	209.0	0.562	345.5	237.2	0.493	518.8	356.2
	0.7	0.628	226.8	155.7	0.630	223.1	153.1	0.615	245.9	168.8
	0.9	0.659	182.3	125.1	0.671	167.3	114.9	0.671	166.6	114.4

Table 3: Optimum values of μ and percent relative efficiencies of T with respect to $\bar{y}_n$ and $\hat{\hat{Y}}$ for $\rho_{z1z2} = 0$.

ρ_{yx}		0.5			0.7			0.9		
ρ_{yz1}	ρ_{yz2}	μ_0	E_1	E_2	μ_0	E_1	E_2	μ_0	E_1	E_2
0.3	0.3	0.475	115.2	106.6	0.539	132.6	111.5	0.669	169.6	116.4
	0.5	0.448	134.2	124.2	0.511	155.5	130.8	0.645	201.9	138.6
	0.7	0.393	182.9	169.2	0.455	214.7	180.6	0.591	287.6	197.4
0.5	0.3	0.448	134.2	124.2	0.511	155.5	130.8	0.645	201.9	138.6
	0.5	0.414	162.5	150.4	0.477	189.9	159.7	0.612	251.3	172.5
	0.7	0.337	250.7	232.0	0.397	298.5	251.1	0.532	412.8	283.4
0.7	0.3	0.393	182.9	169.2	0.455	214.7	180.6	0.591	287.6	197.4
	0.5	0.337	250.7	232.0	0.397	298.5	251.1	0.532	412.8	283.4
	0.7	0.123	1143	1058	0.154	1433	1206	0.240	2271	1559

$A_1 = 1 - \left(\rho_{yz1}^2 + \rho_{yz2}^2 + \rho_{yz3}^2\right) + 2\left(\rho_{yz1}\rho_{yz2}\rho_{z1z2} + \rho_{yz1}\rho_{yz3}\rho_{z1z3} + \rho_{yz2}\rho_{yz3}\rho_{z2z3}\right)$ and $A_2 = 1 - 2\rho_{yx} + \left(\rho_{yz1}^2 + \rho_{yz2}^2 + \rho_{yz3}^2\right) - 2\left(\rho_{yz1}\rho_{yz2}\rho_{z1z2} + \rho_{yz1}\rho_{yz3}\rho_{z1z3} + \rho_{yz2}\rho_{yz3}\rho_{z2z3}\right)$.

In this case there are seven different correlations. For few sets of these seven correlations optimum values of μ i.e., μ_0 and percent relative efficiencies E_1 and E_2 of the estimator T (under optimal condition) with respect to $\bar{y}_n$ and $\hat{\hat{Y}}$ respectively have been computed and shown below:
Set1: $\rho_{yx} = 0.3, \rho_{yz1} = 0.9, \rho_{yz2} = 0.9, \rho_{yz3} = 0.9, \rho_{z1z2} = 0.3, \rho_{z2z3} = 0.3,$
$\rho_{z1z3} = 0.3, \mu_0 = 0.1239, E_1 = 816.7343, E_2 = 795.8347$
Set2: $\rho_{yx} = 0.5, \rho_{yz1} = 0.9, \rho_{yz2} = 0.9, \rho_{yz3} = 0.9, \rho_{z1z2} = 0.3, \rho_{z2z3} = 0.3,$
$\rho_{z1z3} = 0.3, \mu_0 = 0.1433, E_1 = 948.7066, E_2 = 878.0940$
Set3: $\rho_{yx} = 0.7, \rho_{yz1} = 0.9, \rho_{yz2} = 0.9, \rho_{yz3} = 0.9, \rho_{z1z2} = 0.3, \rho_{z2z3} = 0.3,$
$\rho_{z1z3} = 0.3, \mu_0 = 0.1776, E_1 = 1184.1, E_2 = 996.0495$
Set4: $\rho_{yx} = 0.9, \rho_{yz1} = 0.9, \rho_{yz2} = 0.9, \rho_{yz3} = 0.9, \rho_{z1z2} = 0.3, \rho_{z2z3} = 0.3,$
$\rho_{z1z3} = 0.3, \mu_0 = 0.2723, E_1 = 1851.2, E_2 = 1271.1$
Case 5: For $p = 3$ and assuming that the two auxiliary variables are independent (uncorrelated) i.e., $\rho_{zjzk} = 0 \forall j \neq k = 1, 2, 3$. In this case the values of A_1 and A_2 take the following form: $A_1 = 1 - \left(\rho_{yz1}^2 + \rho_{yz2}^2 + \rho_{yz3}^2\right)$ and $A_2 = 1 - 2\rho_{yx} + \left(\rho_{yz1}^2 + \rho_{yz2}^2 + \rho_{yz3}^2\right)$ For few sets of above four correlations, the values of μ_0, E_1 and E_2 are shown below:
Set1: $\rho_{yx} = 0.3, \rho_{yz1} = 0.3, \rho_{yz2} = 0.5, \rho_{yz3} = 0.7$
$\mu_0 = 0.2584, E_1 = 288.5298, E_2 = 281.1465$
Set2: $\rho_{yx} = 0.3, \rho_{yz1} = 0.5, \rho_{yz2} = 0.5, \rho_{yz3} = 0.7$
$\mu_0 = 0.0779, E_1 = 1424.9, E_2 = 1388.5$
Set3: $\rho_{yx} = 0.5, \rho_{yz1} = 0.5, \rho_{yz2} = 0.5, \rho_{yz3} = 0.7$
$\mu_0 = 0.0909, E_1 = 1666.7, E_2 = 1542.6$
Set4: $\rho_{yx} = 0.7, \rho_{yz1} = 0.7, \rho_{yz2} = 0.7, \rho_{yz3} = 0.5$
$\mu_0 = 0.1143, E_1 = 2106.3, E_2 = 1771.8$
Set5: $\rho_{yx} = 0.9, \rho_{yz1} = 0.5, \rho_{yz2} = 0.5, \rho_{yz3} = 0.7$
$\mu_0 = 0.1827, E_1 = 3414.2, E_2 = 2344.2$

9 Conclusion

The following conclusions can be read out from the empirical studies:
(1) From Tables 1 - 2 it is vindicated that
(a) For the fixed values of ρ_{z1z2}, ρ_{yz1}, and ρ_{yz2}, the values of μ_0,E_1 and E_2 increase with the increasing values of ρ_{yx}. This behavior is in agreement

with the Sukhatme et al. [22] results, which explains that the more the value of ρ_{yx}, more the fraction of fresh sample is required on the current occasion.

(b) For the fixed values of ρ_{yx}, ρ_{z1z2} and ρ_{yz1}, the values of μ_0 decrease and E_1 and E_2 increase with the increasing values of ρ_{yz2}. Similarly, for fixed values of ρ_{yx}, ρ_{z1z2} and ρ_{yz2}, the optimum values of μ_0 decrease and E_1 and E_2 increase with the increasing values of ρ_{yz1}. These patterns indicate that smaller fresh sample on the current occasion is required, if highly correlated auxiliary variables are available.

(c) For the fixed values of ρ_{yx}, ρ_{yz1} and ρ_{yz2}, the values of μ_0 increase with the increasing values of ρ_{z1z2} while E_1 and E_2 are decreasing with increasing trends in ρ_{z1z2}, it means that auxiliary variables are quite sensitive with respect to the relation between them.

(2) From Table 3, when the auxiliary variables are uncorrelated, it can be observed that

(a) For the fixed values of ρ_{yz1} and ρ_{yz2} the values of μ_0, E_1 and E_2 increase with the increasing values of ρ_{yx}, which is a desirable result.

(b) For fixed values of ρ_{yz1} and ρ_{yx} the values of μ_0 decrease, while E_1 and E_2 increase with the increasing values of ρ_{yz2}. Similar patterns are visible for the case when the values of ρ_{yz2} and ρ_{yx} are fixed and increasing values of ρ_{yz1} are observed.

(3) For $p = 3$ and when the three auxiliary variables are mutually correlated, we observed that no specific pattern is seen as for so many combinations of correlations the optimum values of μ do not exist. This behavior suggests that the correlation between the auxiliary variable do not play significant role in terms of proposed estimator.

(4) For $p = 3$ and when the three auxiliary variables are uncorrelated then for fixed values of ρ_{yx}, ρ_{z1z3}, ρ_{z2z3}, ρ_{z1z2}, ρ_{yz2} and ρ_{yz3} the values of μ_0 decrease, while E_1 and E_2 increase with the increasing values of ρ_{yz1}. Similar patterns are observed if the case for the increasing values of ρ_{yz2} or ρ_{yz3} is taken into account.

(5) The results obtained for $p = 1$ and $p = 2$ are quite appreciable, while when the number of auxiliary variables increase, the expressions become complex due to increase in the number of correlations, hence, practically, it is more realistic to use two auxiliary variables out of several available auxiliary variables.

Thus it is clear that the use of the auxiliary variables are highly rewarding in terms of the proposed estimator. It is also clear that if highly correlated auxiliary variables are used, relatively only a smaller fraction of sample on the current (second) occasion is desired to be replaced by a fresh sample which reduces the cost of the survey. Hence it can be recommended for future use.

Acknowledgement

Authors are thankful to the referees for their valuable and inspiring suggestions. Authors are also thankful to UGC, New Delhi and the Indian School of Mines, Dhanbad for providing financial assistances to carry out the present research work.

References

1. R. J. Jessen. Statistical investigation of a sample survey for obtaining farm facts. Iowa Agricultural Experiment Station Research Bulletin, 304:1–104, 1942.

2. H. D. Patterson. Sampling on successive occasions with partial replacement of units. Journal of the Royal Statistical Society, 12:241–255, 1950.

3. A. R. Eckler. Rotation sampling. Ann. Math. Stat., 26:664–685, 1955.

4. J. N. K. Rao and J. E. Graham. Rotation design for sampling on repeated occasions. Jour. Amer. Statist. Assoc., 59:492–509, 1964.

5. A. R. Sen. Successive sampling with two auxiliary variables. Sankhya, 33(B):371–378, 1971.

6. A. R. Sen. Successive sampling with p ($p \geq 1$) auxiliary variables. Ann. Math. Statist., 43:2031–2034, 1972.

7. A. R. Sen. Theory and application of sampling on repeated occasions with several auxiliary variables. Biometrics, 29:381–385, 1973.

8. W. G. Cochran. Sampling Techniques. New York: John Wiley and Sons, 1977.

9. P. C. Gupta. Sampling on two successive occasions. Jour. Statist. Res., 13:7–16, 1979.

10. V. K. Singh, G. N. Singh, and D. Shukla. An efficient family of ratio-cum-difference type estimators in successive sampling over two occasions. J. Sci. Res., 41(C):149–159, 1991.

11. S. Feng and G. Zou. Sample rotation method with auxiliary variable. Communications in Statistics: Theory and Methods, 26(6):1497–1509, 1997.

12. R. S. Biradar and H. P. Singh. Successive sampling using auxiliary information on both occasions. Cal. Statist. Assoc. Bull, (51):243–251, 2001.

13. G. N. Singh and V. K. Singh. On the use of auxiliary information in successive sampling. J. Indian Soc. Agric. Statist., 54(1):1–12, 2001.

14. G. N. Singh. Estimation of population mean using auxiliary information on recent occasion in h-occasion successive sampling. Statistics in Transition, 6:523–532, 2003.

15. G. N. Singh. On the use of chain-type ratio estimator in successive sampling. Statistics in Transition, 7:21–26, 2005.

16. G. N. Singh and K. Priyanka. On the use of chain-type ratio to difference estimator in successive sampling. IJAMAS, 5(S06):41–49, 2006.

17. G. N. Singh and K. Priyanka. On the use of auxiliary information in search of good rotation patterns on successive occasions. Bulletin of Statistics and Economics, 1(A07)):42–60, 2007.

18. G. N. Singh and K. Priyanka. Search of good rotation patterns to improve the precision of estimates at current occasion. Communications in Statistics- Theory and Methods, 37(3):337–348, 2008.

19. G. N. Singh and K. Priyanka. On the use of several auxiliary variates to improve the precision of estimates at current occasion estimation of population mean at current occasion in presence of several varying auxiliary variates in two-occasion successive sampling. J. Indian Soc. Agric. Statist., 62(3):253–265, 2008.

20. G. N. Singh and J. P. Karna. Estimation of population mean on current occasion in two-occasion successive sampling. METRON, 67(1):69–85, 2009.

21. G. N. Singh and J. P. Karna. Search of effective rotation patterns in presence of auxiliary information in successive sample over two-occasions. Statistics in Transition-new series, 10(1):59–73, 2009.

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