

RATIO-TO-REGRESSION ESTIMATOR IN SUCCESSIVE SAMPLING USING ONE AUXILIARY VARIABLE

Zoramthanga Ralte¹, Gitasree Das²

ABSTRACT

The problem of estimation of finite population mean on the current occasion based on the samples selected over two occasions has been considered. In this paper, first a chain ratio-to-regression estimator was proposed to estimate the population mean on the current occasion in two-occasion successive (rotation) sampling using only the matched part and one auxiliary variable, which is available in both the occasions. The bias and mean square error of the proposed estimator is obtained. We proposed another estimator, which is a linear combination of the means of the matched and unmatched portion of the sample on the second occasion. The bias and mean square error of this combined estimator is also obtained. The optimum mean square error of this combined estimator was compared with (i) the optimum mean square error of the estimator proposed by Singh (2005) (ii) mean per unit estimator and (iii) combined estimator suggested by Cochran (1977) when no auxiliary information is used on any occasion. Comparisons are made both analytically as well as empirically by using real life data.

Key words: ratio-to-regression estimator, auxiliary variable, successive sampling, bias, mean square error, optimum replacement policy.

1. Introduction

The successive method of sampling consists in selecting samples of the same size on different occasions such that some units are common to samples selected on previous occasions. In successive sampling, the ratio estimator is among the most commonly adopted estimators of the population mean or total of some variables of interest of a finite population with the help of an auxiliary variable when the correlation coefficient between the two variables is positive. Patterson (1950) and Cochran (1977) suggested a number of estimation procedures on

¹ Department of Statistics, North-Eastern Hill University, Shillong, India.
E-mail: aramaralte7@gmail.com.

² Department of Statistics, North-Eastern Hill University, Shillong, India.
E-mail: gitasree22@gmail.com.

sampling over two occasions. Rao and Graham (1964), Gupta (1979), Das (1982), Sen (1971) developed estimators for the population mean on the current occasion using information on two auxiliary variables available on the previous occasion. Sen (1972, 1973) extended his work for several auxiliary variates. 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 (2003) extended their work for h-occasions successive sampling.

Utilizing the auxiliary information on both the occasions, Singh (2005), Singh and Priyanka (2006, 2007a, 2008) proposed varieties of chain-type ratio, difference and regression estimators for estimating the population mean on the current (second) occasion in two occasion successive sampling. Singh (2005) suggested two estimators for population mean using the information on an auxiliary variable in successive sampling over two occasions. Consider a character under study on the first (second) occasion is denoted by $x(y)$ respectively and the auxiliary variable z is available on both the occasions. A simple random sample (without replacement) of n units is taken on the first occasion. The suggested chain-type ratio estimator based on the sample of size $m (= n\lambda)$ common to both the occasions is given by

$$T_2 = \frac{\bar{y}_m \bar{x}_n}{\bar{x}_m \bar{z}_n} \bar{Z} \quad (1)$$

The bias $B(\cdot)$ and mean square error $M(\cdot)$ of T_2 up to the first order of approximation and for large population of size N of equation (1) are derived as

$$B(T_2) = \bar{Y} \left[\left(\frac{1}{m} - \frac{1}{n} \right) (C_x^2 - 2\rho_{yx} C_y C_x) + \frac{1}{n} (C_z^2 - 2\rho_{yz} C_y C_z) \right]$$

and

$$M(T_2) = \bar{Y}^2 \left[\frac{C_y^2}{m} + \left(\frac{1}{m} - \frac{1}{n} \right) (C_x^2 - 2\rho_{yx} C_y C_x) + \frac{1}{n} (C_z^2 - 2\rho_{yz} C_y C_z) \right]$$

A classical ratio estimator based on a sample of size $u = n - m = n\mu$ (fraction of a sample) drawn afresh on the second occasion is given by

$$T_1 = \frac{\bar{y}_u}{\bar{z}_u} \bar{Z} \quad (2)$$

Combining T_1 and T_2 , the resulting estimator of \bar{Y} is

$$T = \phi T_1 + (1 - \phi) T_2$$

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

After optimizing ϕ and μ (sampling fraction), the optimum mean square error of the estimator T is given by

$$M(T)_{opt} = \frac{(A + C)^2 + (A + C)(B - C)\mu_0}{n[(A + C) + (B - C)\mu_0^2]}$$

where $A = \bar{Y}^2 C_y^2$, $B = \bar{Y}^2 (C_x^2 - 2\rho_{yx} C_x C_y)$, $C = \bar{Y}^2 (C_z^2 - 2\rho_{yz} C_y C_z)$ and

$$\mu_0 = \frac{-(1 - \rho_{yz}) \pm \sqrt{(1 - \rho_{yz})(1 - \rho_{yx})}}{(\rho_{yz} - \rho_{yx})}$$

Following the chain-type ratio estimator proposed by Singh (2005), in the present work we propose a chain ratio-to-regression estimator for estimating population mean on the current occasion using auxiliary information that is available on both the occasions. The behaviour of the proposed estimator have been examined analytically and also through empirical means of comparison.

2. Proposed estimator

Consider a population consisting of N units. Let a character under study on the first (second) occasion be denoted by x(y), respectively. It is assumed that the information on an auxiliary variable z is available on the first as well as on the second occasion. We consider the population to be large enough, and the sample size is constant on each occasion. Using simple random sampling without replacement (SRSWOR) we select a sample of size n on the first occasion. Of these n units, a sub-sample of size $m = n\lambda$ is retained on the second occasion. This sub-sample is supplemented by selecting SRSWOR of $u = (n - m) = n\mu$ units afresh from the units that were not selected on the first occasion.

By modifying the estimator in (1), we propose a chain ratio-to-regression estimator for \bar{Y} on the second occasion which is based on a sample of size m common to both the occasions and is given by

$$T_p = \frac{\bar{y}_m}{\bar{x}_m} \left[\bar{x}_m + b_{xz} (\bar{z}_n - \bar{z}_m) \right] \frac{\bar{Z}}{\bar{z}_n} \tag{3}$$

where

\bar{Z} : Population mean of z.

$\bar{x}_m, \bar{y}_m, \bar{z}_m, \bar{z}_n$: Sample means of the respective variates with sample sizes as shown in the subscript.

b_{xz} : Regression coefficient of x on z.

The bias of the proposed estimator T_p up to the second order of approximation as obtained in Appendix A is:

$$\text{Bias}(T_p) = \frac{1}{X} \left\{ \begin{aligned} &\beta_{xz} \bar{Y} \left(-\frac{\text{Cov}(\bar{z}_n, S_z^2)}{S_z^2} + \frac{\text{Cov}(\bar{z}_n, S_{xz})}{S_{xz}} + \frac{\text{Cov}(\bar{z}_m, S_z^2)}{S_z^2} + \frac{\text{Cov}(\bar{z}_m, S_{xz})}{S_{xz}} \right) \\ &+ \beta_{xz} \left[\text{Cov}(\bar{y}_m, \bar{z}_n) - \text{Cov}(\bar{y}_m, \bar{z}_m) \right] - \frac{\bar{Y}}{X} \text{Var}(\bar{x}_m) \\ &- \beta_{xz} \frac{\bar{Y}}{X} \left[\text{Cov}(\bar{x}_m, \bar{z}_n) - \text{Cov}(\bar{x}_m, \bar{z}_m) \right] - \bar{X} \text{Cov}(\bar{y}_m, \bar{z}_n) + \frac{\bar{X}\bar{Y}}{\bar{Z}^2} \text{Var}(\bar{z}_n) \end{aligned} \right\} \quad (4)$$

And the mean square error of T_p considering $E(T_p)$ up to the first order of approximation obtained in Appendix B is:

$$M(T_p) = \bar{Y}^2 \left[\frac{C_y^2}{m} + \left(\frac{1}{m} - \frac{1}{n} \right) (\rho_{xz}^2 C_x^2 - 2\rho_{xz}\rho_{yz} C_x C_y) + \frac{1}{n} (C_z^2 - 2\rho_{yz} C_y C_z) \right] \quad (5)$$

Following Singh (2005), we combine the estimators T_1 and T_p and the final estimator of \bar{Y} on the second occasion is given as

$$T_{p_1} = \psi T_1 + (1 - \psi) T_p \quad (6)$$

where ψ is an unknown constant to be determined under certain criterion.

3. Bias and mean square error of a combined estimator T_{p_1}

Since both T_1 and T_p are biased estimators of \bar{Y} , therefore, the resulting estimator T_{p_1} is also a biased estimator of \bar{Y} . Using the notations in Section 2, the bias $B(T_{p_1})$ and mean square error $M(T_{p_1})$ up to the first order of approximation are given below:

$$B(T_{p_1}) = \psi B(T_1) + (1 - \psi) B(T_p) \quad (7)$$

$$\text{and} \quad M(T_{p_1}) = \psi^2 M(T_1) + (1 - \psi)^2 M(T_p) \quad (8)$$

where, from Cochran (1977)

$$B(T_1) = \frac{\bar{Y}}{u} (C_z^2 - \rho_{yz} C_y C_z) \tag{9}$$

and
$$M(T_1) = \frac{\bar{Y}^2}{u} (C_y^2 + C_z^2 - 2\rho_{yz} C_y C_z) \tag{10}$$

4. Minimum mean square error of T_{p_1}

Since $M(T_{p_1})$ in (8) is a function of unknown constant ψ , therefore, it is minimized with respect to ψ and subsequently the optimum value of ψ obtained in Appendix C is

$$\psi_{opt} = \frac{M(T_p)}{M(T_1) + M(T_p)} \tag{11}$$

Now, substituting the value of ψ_{opt} in equation (8) we obtain the optimum mean square error of T_{p_1} (given in Appendix C) as

$$M(T_{p_1})_{opt} = \frac{M(T_1)M(T_p)}{M(T_1) + M(T_p)} \tag{12}$$

Using equation (5) and (10), let

$$A' = \bar{Y}^2 C_y^2, \quad B' = \bar{Y}^2 (\rho_{xz}^2 C_x^2 - 2\rho_{xz}\rho_{yz} C_x C_y), \quad C' = \bar{Y}^2 (C_z^2 - 2\rho_{yz} C_y C_z)$$

Then, from Appendix C

$$M(T_{p_1})_{opt} = \frac{(A' + C')^2 + (A' + C')(B' - C')\mu}{n[(A' + C') + (B' - C')\mu^2]}$$

or
$$M(T_{p_1})_{opt} = \frac{(\alpha'_1)^2 + \alpha'_1 \alpha'_2 \mu}{n[\alpha'_1 + \alpha'_2 \mu^2]} \tag{13}$$

where $\alpha'_1 = A' + C', \quad \alpha'_2 = B' - C'$

5. Optimum replacement policy

To determine the optimum value of μ (fraction of a sample to be taken afresh at the second occasion) so that population mean \bar{Y} may be estimated with maximum precision, we minimize mean square error of T_{p_1} with respect to μ which results in quadratic equation in μ and solution of μ (from Appendix D) is given below:

$$\begin{aligned} \mu'_0 &= \frac{-\alpha'_1 \pm \sqrt{(\alpha'_1)^2 + (\alpha'_1 \alpha'_2)}}{\alpha'_2} \\ &= \frac{-2(1-\rho_{yz}) \pm \sqrt{2(1-\rho_{yz})\{2(1-\rho_{yz}) + (1-\rho_{xz})(2\rho_{yz} - \rho_{xz} - 1)\}}}{(1-\rho_{xz})(2\rho_{yz} - \rho_{xz} - 1)} \end{aligned} \quad (14)$$

From equation (14), it is obvious that the real value of μ'_0 exists if the quantities under square root are greater than or equal to zero. To choose the admissible value of μ'_0 , it should be remembered that $0 \leq \mu'_0 \leq 1$. All other values of μ'_0 are inadmissible. Substituting the value of μ'_0 from (14) in (13), we have

$$M(T_{p_1})_{\text{opt}} = \frac{(\alpha'_1)^2 + \alpha'_1 \alpha'_2 \mu'_0}{n \left[\alpha'_1 + \alpha'_2 (\mu'_0)^2 \right]} \quad (15)$$

6. Comparison of mean square error and efficiency

Now we compare the optimum MSE of the proposed estimator T_{p_1} with (i) the optimum MSE of the estimator proposed by Singh (2005), (ii) \bar{y}_n , i.e. mean per unit estimator, and (iii) \bar{y}_2 (Cochran, 1977) when no auxiliary information is used at any occasion. In each case, we obtain conditions under which the estimator T_{p_1} is better than the three estimators mentioned above.

(i) Comparison with Singh's (2005) estimator.

First, we consider the estimator $T = (\phi)T_1 + (1-\phi)T_2$ which is due to Singh (2005). The variance of this estimator, to the first order of approximation and for large population, is given as

$$M(T) = \frac{1}{n\mu(1-\mu)} \left[(\phi)^2 \alpha_1 (1-\mu) + (1-\phi)^2 (\alpha_1 \mu + \alpha_2 \mu^2) \right]$$

where

$$\alpha_1 = A + C, \alpha_2 = B - C, A = \bar{Y}^2 C_y^2, B = \bar{Y}^2 (C_x^2 - 2\rho_{yx} C_y C_x), C = \bar{Y}^2 (C_z^2 - 2\rho_{yz} C_y C_z)$$

The variance of T is minimized for

$$\phi_{opt} = \frac{\mu(\alpha_1 + \alpha_2\mu)}{(\alpha_1 + \alpha_2\mu^2)}$$

Thus, the resulting minimum variance of T is

$$M(T)_{opt} = \frac{\alpha_1^2 + \alpha_1\alpha_2\mu}{n[\alpha_1 + \alpha_2\mu^2]}$$

Under the assumption of $C_x = C_y = C_z = C$, minimizing $M(T)_{opt}$ with respect to μ , the optimum value of μ (fraction of a sample to be taken afresh at the second occasion) is given by

$$\left[\frac{-\alpha_1 \pm \sqrt{\alpha_1^2 + \alpha_1\alpha_2}}{\alpha_2} \right] = \frac{-(1 - \rho_{yz}) \pm \sqrt{(1 - \rho_{yz})(1 - \rho_{yx})}}{(\rho_{yz} - \rho_{yx})} = \mu_0 \text{ (say)}$$

Therefore,

$$M(T)_{opt} = \frac{\alpha_1^2 + \alpha_1\alpha_2\mu_0}{n[\alpha_1 + \alpha_2\mu_0^2]}$$

The proposed combined estimator T_{p1} is better than the combined estimator proposed by Singh (2005) if

$$\begin{aligned} M(T)_{opt} - M(T_{p1})_{opt} &> 0 \\ \Rightarrow \frac{\alpha_1^2 + \alpha_1\alpha_2\mu_0}{n[\alpha_1 + \alpha_2\mu_0^2]} - \frac{(\alpha_1')^2 + \alpha_1'\alpha_2'\mu_0'}{n[\alpha_1' + \alpha_2'(\mu_0')^2]} &> 0 \\ \Rightarrow (1 - \rho)(1 - \mu_0') &> 0 \end{aligned}$$

(See Appendix E(i) for derivation of the above result)

Since $0 \leq \mu_0' \leq 1$, $(1 - \mu_0')$ is always positive and $(1 - \rho)$ is always positive, which indicates that the above condition is always valid. That is, T_{p1} is a better estimator than the estimator T proposed by Singh (2005).

(ii) Comparison with mean per unit estimator.

Next, we consider the mean per unit estimator \bar{y}_n as follows:

$\bar{y}_n = \frac{\sum_{i=1}^n y_i}{n}$, under the assumption of $C_y = C$, the variance of mean per unit estimator is given by

$$V(\bar{y}_n) = \frac{\bar{Y}^2 C_y^2}{n} = \frac{\bar{Y}^2 C^2}{n}$$

The proposed combined estimator T_{p_1} is better than this estimator if

$$V(\bar{y}_n) - M(T_{p_1})_{\text{opt}} > 0$$

$$\Rightarrow \left[2 - (1 - \rho)(\mu_0')^2 \right] > 2(1 - \rho) \left[2 - (1 - \rho)\mu_0' \right] \quad (16)$$

(See Appendix E(ii) for derivation of the above result)

Hence, T_{p_1} is a better estimator than mean per unit estimator whenever equation (16) is satisfied. And it can be easily verified that the above condition is true for $\rho \geq 0.5$.

(iii) Comparison with Cochran's (1977) estimator.

Now we consider the Cochran's (1977) estimator \bar{y}'_2 as follows:

$\bar{y}'_2 = \phi_2 \bar{y}'_{2u} + (1 - \phi_2) \bar{y}'_{2m}$, i.e. combined estimator suggested by Cochran (1977) when no auxiliary information is used at any occasion.

Here, \bar{y}'_{2u} , \bar{y}'_{2m} are the unmatched and matched portions of the sample at the second occasion respectively, and

$$\phi_2 = \frac{W_{2u}}{W_{2u} + W_{2m}}$$

where W_{2u} and W_{2m} are the inverse variances, i.e.

$$\frac{1}{W_{2u}} = V(\bar{y}'_{2u}) = \frac{C_y^2 \bar{Y}^2}{u}, \quad \frac{1}{W_{2m}} = V(\bar{y}'_{2m}) = \frac{C_y^2 \bar{Y}^2 (1 - \rho_{yx}^2)}{m} + \rho_{yx}^2 \frac{C_y^2 \bar{Y}^2}{n}$$

By least square theory, the variance of \bar{y}_2' is

$$V(\bar{y}_2') = \frac{1}{W_{2u} + W_{2m}}$$

$$= \frac{C_y^2 \bar{Y}^2 (n - u\rho_{yx}^2)}{n^2 - u^2\rho_{yx}^2}$$

Minimizing $V(\bar{y}_2')$ with respect to u , the optimum value of u is given by

$$u_{opt} = \frac{n}{1 + \sqrt{1 - \rho_{yx}^2}} \text{ and } m_{opt} = \frac{n(\sqrt{1 - \rho_{yx}^2})}{1 + \sqrt{1 - \rho_{yx}^2}}$$

Under the assumption of $C_y = C$, the optimum variance of the above estimator is given as

$$V(\bar{y}_2')_{opt} = \left[1 + \sqrt{1 - \rho_{yx}^2}\right] \frac{\bar{Y}^2 C^2}{2n}$$

Then, the proposed combined estimator T_{p_1} is better than this estimator if

$$V(\bar{y}_2')_{opt} - M(T_{p_1})_{opt} > 0$$

$$\Rightarrow \left(1 + \sqrt{1 - \rho^2}\right) \left[2 - (1 - \rho)(\mu_0')^2\right] > 4(1 - \rho) \left[2 - (1 - \rho)\mu_0'\right] \quad (17)$$

(See Appendix E(iii) for derivation of the above result)

Hence, T_{p_1} is a better estimator than Cochran's (1977) estimator whenever equation (17) is satisfied. And it can be easily verified that the above condition is true for $\rho \geq 0.5$.

Further, conditions (16) and (17) are true for all values of $\rho \geq 0.5$. This is justified since ρ_{yx} expresses the relationship between current occasion and previous occasion study variable, ρ_{xz} for previous occasion study variable and auxiliary variable and ρ_{yz} for current occasion study variable and auxiliary variable.

As an example, the percent relative efficiency of the proposed estimator T_{p_1} with respect to (i) T , the estimator proposed by Singh (2005), (ii) \bar{y}_n , i.e. mean per

unit estimator, and (iii) \bar{y}'_2 (Cochran, 1977) when no auxiliary information is used on any occasion, have been computed for some assumed values of ρ_{yx} , ρ_{yz} and ρ_{zx} . Since \bar{y}_n and \bar{y}'_2 are unbiased estimators of \bar{Y} , their variances for large N are respectively given by

$$V(\bar{y}_n) = \frac{\bar{Y}^2 C_y^2}{n}$$

and

$$V(\bar{y}'_2)_{\text{opt}} = \left[1 + \sqrt{1 - \rho_{xy}^2} \right] \frac{\bar{Y}^2 C_y^2}{2n}$$

Here, E_1 , E_2 , E_3 are designated as the percent relative efficiencies of the proposed estimator T_{p_1} with respect to Singh's (2005) estimator T , \bar{y}_n and \bar{y}'_2 respectively.

Further, the expression of the optimum μ , i.e. μ'_0 and the percent relative efficiencies E_1 , E_2 and E_3 are in terms of population correlation coefficients. We assumed that all the correlations involved in the expressions of μ , E_1 , E_2 and E_3 are equal, i.e. $\rho_{xy} = \rho_{xz} = \rho_{yz} = \rho$. Accordingly, computed values of μ'_0 , E_1 , E_2 and E_3 for different choices of high positive correlations are shown in Table 1.

Table 1. Relative Efficiency (%) of T_{p_1} with respect to estimators T , \bar{y}_n and \bar{y}'_2 .

ρ	0.5	0.6	0.7	0.8	0.9
μ'_0	0.5359	0.52786	0.5203	0.513167	0.506411
E_1	107.18	112.194	117.3	122.5148	127.8537
E_2	107.18	131.966	173.435	256.5835	506.4113
E_3	100	118.769	148.646	205.2668	363.5754

From Table 1, it is clear that the proposed estimator is more efficient than T , \bar{y}_n and \bar{y}'_2 . Also, when the correlation coefficient $\rho \geq 0.5$ is increasing, the gain in precision of the proposed estimator T_{p_1} over T , \bar{y}_n and \bar{y}'_2 is also increasing. Further, it may be noticed that the maximum gain in efficiency occurs while comparing it with mean per unit estimator, which is very obvious.

7. Illustration using real life data

In order to illustrate the comparison of relative efficiency of the proposed estimator with respect to (i) T , the estimator proposed by Singh (2005), (ii) \bar{y}_n , i.e. mean per unit estimator, and (iii) \bar{y}'_2 (Cochran,1977) when no auxiliary information is used on any occasion, using real life approach, the data from the Census of India (2001) and (2011) was considered. We define the variables $x(y)$ as the total number of workers in villages in the state of Mizoram, India in 2001 (2011) and z is defined as an auxiliary variable which is the total number of literate people in villages in the state of Mizoram, India.

Using successive sampling as defined in section 2 for the above data set we take $n = 70$, $m = 35$ and $u = 35$. The following table shows the values of the different estimators as computed from the sample along with their corresponding mean square errors and efficiency of the proposed estimator T_{p_1} with respect to T , \bar{y}_n and \bar{y}'_2 .

Table 2. Relative Efficiency (%) of T_{p_1} with respect to estimators T , \bar{y}_n and \bar{y}'_2 using real life data.

Estimators	Estimates	MSE	Efficiency %
T_{p_1}	354	249.98	100
T	378	269.39	107.76
\bar{y}_n	300	1808.14	723.31
\bar{y}'_2	332	1240.26	496.14

The above table shows that the conclusions are the same as those of Table 1, that is the proposed estimator T_{p_1} is more efficient than T , \bar{y}_n and \bar{y}'_2 with maximum gain in efficiency occurring while comparing it with mean per unit estimator, which is very obvious.

8. Conclusion

In this study we have proposed a new chain ratio-to-regression estimator in successive sampling. The bias of the proposed estimator was computed up to the second order of approximation. The optimum replacement policy of the sampling fraction was obtained and considering bias up to the first order of approximation, the optimum mean square error of the proposed estimator was also obtained. The optimum mean square error of the proposed estimator was compared with that of

Singh's (2005) estimator and it was found that the proposed estimator is always better than Singh's (2005) estimator. Further, the proposed estimator was compared with mean per unit estimator and Cochran's (1977) estimator when no auxiliary information is used on any occasion. It was found that the proposed estimator is better than both of these estimators for $\rho \geq 0.5$, which is entirely justified. An example was considered by assuming different values of $\rho \geq 0.5$ to illustrate the above facts. At the end, a real life study was also done to demonstrate that the proposed estimator is more efficient than the other three existing estimators. Hence, the proposed estimator is recommended for further use.

REFERENCES

- CENSUS OF INDIA, (2001). www.cesusindiagov.in.
- CENSUS OF INDIA, (2011). www.cesusindiagov.in.
- COCHRAN, W. G., (1977). *Sampling Techniques*. Third edition. Wiley Eastern Ltd.
- DAS, A. K., (1982). Estimation of population ratio on two occasions. *Journal of the Indian Society of Agricultural Statistics*. 34: 1–9.
- GUPTA, P. C., (1979). Sampling on two successive occasions. *Journal of Statistical Research*. 13: 7–16.
- PATTERSON, H. D., (1950). Sampling on successive occasions with partial replacement of units. *Journal of the Royal Statistical Society*. 12(B): 241–255.
- RAO, J. N. K., GRAHAM J., E., (1964). Rotation designs for sampling on repeated occasions. *Journal of American Statistical Association*. 59: 492–509.
- SEN, A. R., (1971). Successive sampling with two auxiliary variables. *Sankhya*. 33(B): 371–378.
- SEN, A. R., (1972). Successive sampling with $p(p \geq 1)$ auxiliary variables, *The Annals of Mathematical Statistics*. 43: 2031–2034.
- SEN, A. R., (1973). Theory and application of sampling on repeated occasions with several auxiliary variables. *Biometrics*. 29: 381–385.
- SINGH, G. N., (2003). Estimation of population mean using auxiliary information on recent occasion in h occasions successive sampling. *Statistics in Transition*. 6(4): 523–532.
- SINGH, G. N., (2005). On the use of chain-type ratio estimator in successive sampling. *Statistics in Transition*. 7(1): 21–26.

- SINGH, G. N., PRIYANKA, K., (2006). On the use of chain-type ratio to difference estimator in successive sampling. *International Journal of Applied Mathematics and Statistics*. 5(S06): 41–49.
- SINGH, G. N., PRIYANKA, K., (2007). On the use of auxiliary information in search of good rotation patterns on successive occasions. *Bulletin of Statistics and Economics*. 1(A07): 42–60.
- SINGH, G. N., PRIYANKA, K., (2008). On the use of several auxiliary variates to improve the precision of estimates at current occasion. *Journal of the Indian Society of Agricultural Statistics*. 62(3): 253–265.
- SINGH, G. N., SINGH, V. K., (2001). On the use of auxiliary information in successive sampling. *Journal of the Indian Society of Agricultural Statistics*. 54(1): 1–12.
- SINGH, H. P., VISHWAKARMA, G. K., (2007). A general class of estimators in successive sampling. *Metron*. LXV(2): 201–227.
- SINGH, P., TALWAR, H. K., (1991). Estimation of population regression coefficient in successive sampling. *Biometrical Journal*. 33: 599–605.

APPENDICES

Appendix A

Bias of the proposed estimator

$$T_p = \frac{\bar{y}_m}{\bar{x}_m} \left[\bar{x}_m + b_{xz} (\bar{z}_n - \bar{z}_m) \right] \frac{\bar{z}}{\bar{z}_n}$$

$$\text{Bias}(T_p) = E(T_p) - \bar{Y}$$

Let

$$\bar{y}_m = \bar{Y} + \varepsilon_0', \quad \bar{z}_m = \bar{Z} + \varepsilon_1', \quad \bar{z}_n = \bar{Z} + \varepsilon_1, \quad \bar{x}_m = \bar{X} + \varepsilon_2', \quad s_{xz} = S_{xz} + \varepsilon_3, \quad s_z^2 = S_z^2 + \varepsilon_4,$$

$$\beta_{xz} = \frac{S_{xz}}{S_z^2}$$

such that $E(\varepsilon_i) = 0$ assuming the terms having order higher than 2 in ε 's are negligible, then

$$\begin{aligned} T_p &= \left(\frac{\bar{Y} + \varepsilon_0'}{\bar{X} + \varepsilon_2'} \right) \left[\bar{X} + \varepsilon_2' + \left(\frac{S_{xz} + \varepsilon_3}{S_z^2 + \varepsilon_4} \right) (\bar{Z} + \varepsilon_1 - \bar{Z} - \varepsilon_1') \right] \left(\frac{\bar{Z}}{\bar{Z} + \varepsilon_1} \right) \\ &= \frac{1}{\bar{X}} (\bar{Y} + \varepsilon_0') \left(1 - \frac{\varepsilon_2'}{\bar{X}} + \frac{(\varepsilon_2')^2}{\bar{X}^2} \right) \left[\bar{X} + \varepsilon_2' + \left(1 + \frac{\varepsilon_3}{S_{xz}} \right) \left(1 - \frac{\varepsilon_4}{S_z^2} + \frac{\varepsilon_4^2}{S_z^4} \right) \beta_{xz} (\varepsilon_1 - \varepsilon_1') \right] \left(1 - \frac{\varepsilon_1}{\bar{Z}} + \frac{\varepsilon_1^2}{\bar{Z}^2} \right) \\ &= \left\{ \frac{1}{\bar{X}} \left(\bar{Y} + \varepsilon_0' - \frac{\bar{Y}}{\bar{X}} \varepsilon_2' - \frac{1}{\bar{X}} \varepsilon_0' \varepsilon_2' + \frac{\bar{Y}}{\bar{X}^2} (\varepsilon_0')^2 - \bar{Y} \varepsilon_1 - \varepsilon_1 \varepsilon_0' + \frac{\bar{Y}}{\bar{X}} \varepsilon_1 \varepsilon_2' + \frac{\bar{Y}}{\bar{Z}^2} \varepsilon_1^2 \right) \right. \\ &\quad \left. \left[\bar{X} + \varepsilon_2' + \beta_{xz} \left(\varepsilon_1 - \frac{\varepsilon_1 \varepsilon_4}{S_z^2} + \frac{\varepsilon_1 \varepsilon_3}{S_{xz}} - \varepsilon_1' + \frac{\varepsilon_1' \varepsilon_4}{S_z^2} + \frac{\varepsilon_1' \varepsilon_3}{S_{xz}} \right) \right] \right\} \\ &= \left\{ \frac{1}{\bar{X}} \left[\bar{Y} \bar{X} + \beta_{xz} \bar{Y} \left(-\frac{\varepsilon_1 \varepsilon_4}{S_z^2} + \frac{\varepsilon_1 \varepsilon_3}{S_{xz}} + \frac{\varepsilon_1' \varepsilon_4}{S_z^2} + \frac{\varepsilon_1' \varepsilon_3}{S_{xz}} \right) + \beta_{xz} (\varepsilon_0' \varepsilon_1 - \varepsilon_0' \varepsilon_1') \right] \right. \\ &\quad \left. - \beta_{xz} \frac{\bar{Y}}{\bar{X}} (\varepsilon_2' \varepsilon_1 - \varepsilon_1' \varepsilon_2') - \beta_{xz} (\varepsilon_1^2 - \varepsilon_1 \varepsilon_1') - \bar{X} \varepsilon_0' \varepsilon_1 + \frac{\bar{X} \bar{Y}}{\bar{Z}^2} \varepsilon_1^2 \right] \right\} \end{aligned}$$

Therefore,

$$E(T_p) = \bar{Y} + \frac{1}{\bar{X}} \left\{ \begin{aligned} &\beta_{xz} \bar{Y} \left(-\frac{\text{Cov}(\bar{z}_n, s_z^2)}{S_z^2} + \frac{\text{Cov}(\bar{z}_n, s_{xz})}{S_{xz}} + \frac{\text{Cov}(\bar{z}_m, s_z^2)}{S_z^2} + \frac{\text{Cov}(\bar{z}_m, s_{xz})}{S_{xz}} \right) \\ &+ \beta_{xz} \left[\text{Cov}(\bar{y}_m, \bar{z}_n) - \text{Cov}(\bar{y}_m, \bar{z}_m) \right] - \frac{\bar{Y}}{\bar{X}} \text{Var}(\bar{x}_m) \\ &- \beta_{xz} \frac{\bar{Y}}{\bar{X}} \left[\text{Cov}(\bar{x}_m, \bar{z}_n) - \text{Cov}(\bar{x}_m, \bar{z}_m) \right] - \bar{X} \text{Cov}(\bar{y}_m, \bar{z}_n) + \frac{\bar{X}\bar{Y}}{\bar{Z}^2} \text{Var}(\bar{z}_n) \end{aligned} \right\} \tag{18}$$

Hence, $\text{Bias}(T_p) = E(T_p) - \bar{Y}$

$$= \frac{1}{\bar{X}} \left\{ \begin{aligned} &\beta_{xz} \bar{Y} \left(-\frac{\text{Cov}(\bar{z}_n, s_z^2)}{S_z^2} + \frac{\text{Cov}(\bar{z}_n, s_{xz})}{S_{xz}} + \frac{\text{Cov}(\bar{z}_m, s_z^2)}{S_z^2} + \frac{\text{Cov}(\bar{z}_m, s_{xz})}{S_{xz}} \right) \\ &+ \beta_{xz} \left[\text{Cov}(\bar{y}_m, \bar{z}_n) - \text{Cov}(\bar{y}_m, \bar{z}_m) \right] - \frac{\bar{Y}}{\bar{X}} \text{Var}(\bar{x}_m) \\ &- \beta_{xz} \frac{\bar{Y}}{\bar{X}} \left[\text{Cov}(\bar{x}_m, \bar{z}_n) - \text{Cov}(\bar{x}_m, \bar{z}_m) \right] - \bar{X} \text{Cov}(\bar{y}_m, \bar{z}_n) + \frac{\bar{X}\bar{Y}}{\bar{Z}^2} \text{Var}(\bar{z}_n) \end{aligned} \right\}$$

Appendix B

Mean Square Error of the proposed estimator

$$T_p = \frac{\bar{y}_m}{\bar{x}_m} \left[\bar{x}_m + b_{xz} (\bar{z}_n - \bar{z}_m) \right] \frac{\bar{z}}{\bar{z}_n}$$

From (18), we can see that up to the first order of approximation $E(T_p) = \bar{Y}$.

Therefore, mean square error up to the first order of approximation is given by

$$M(T_p) = E(T_p)^2 - \bar{Y}^2.$$

Now,

$$T_p = \left(\frac{\bar{Y} + \varepsilon_0'}{\bar{X} + \varepsilon_2'} \right) \left[\bar{X} + \varepsilon_2' + \left(\frac{S_{xz} + \varepsilon_3}{S_z^2 + \varepsilon_4} \right) (\bar{Z} + \varepsilon_1 - \bar{Z} - \varepsilon_1') \right] \left(\frac{\bar{Z}}{\bar{Z} + \varepsilon_1} \right)$$

$$= \frac{1}{\bar{X}}(\bar{Y} + \varepsilon_0') \left(1 - \frac{\varepsilon_2'}{\bar{X}}\right) \left[\bar{X} + \varepsilon_2' + \left(1 + \frac{\varepsilon_3}{S_{xz}}\right) \left(1 - \frac{\varepsilon_4}{S_z^2}\right) \beta_{xz} (\varepsilon_1 - \varepsilon_1') \right] \left(1 - \frac{\varepsilon_1}{\bar{Z}}\right)$$

Therefore,

$$(T_p)^2 = \frac{1}{\bar{X}^2} \left\{ \left(\bar{Y} + \varepsilon_0' - \frac{\bar{Y}}{\bar{X}} \varepsilon_2' \right) \left[\bar{X} + \varepsilon_2' + \left(1 + \frac{\varepsilon_3}{S_{xz}} - \frac{\varepsilon_4}{S_z^2}\right) \beta_{xz} (\varepsilon_1 - \varepsilon_1') \right] \left(1 - \frac{\varepsilon_1}{\bar{Z}}\right) \right\}^2$$

Now again,

$$E(T_p)^2 = \frac{1}{\bar{X}^2} E \left\{ \bar{X}\bar{Y} + \bar{X}\varepsilon_0' + \beta_{xz} \bar{Y}\varepsilon_1 - \beta_{xz} \bar{Y}\varepsilon_1' - \frac{\bar{X}\bar{Y}}{\bar{Z}} \varepsilon_1 \right\}^2$$

Assuming the terms up to the second order of approximation and neglecting terms higher than 2nd power of ε 's, we have

$$\begin{aligned} E(T_p)^2 &= \frac{1}{\bar{X}^2} E \left\{ \left(\bar{X}\bar{Y} \right)^2 + \bar{X}^2 (\varepsilon_0')^2 + \beta_{xz}^2 \bar{Y}^2 \varepsilon_1^2 + \beta_{xz}^2 \bar{Y}^2 (\varepsilon_1')^2 + \left(\frac{\bar{X}\bar{Y}}{\bar{Z}} \right)^2 \varepsilon_1^2 + 2\beta_{xz} \bar{X}\bar{Y}\varepsilon_0'\varepsilon_1 \right\} \\ &\quad \left\{ -2\beta_{xz} \bar{X}\bar{Y}\varepsilon_0'\varepsilon_1' - 2\frac{\bar{X}^2\bar{Y}}{\bar{Z}} \varepsilon_0'\varepsilon_1 - 2\beta_{xz}^2 \bar{Y}^2 \varepsilon_1'\varepsilon_1 - 2\beta_{xz} \frac{\bar{X}\bar{Y}}{\bar{Z}} \varepsilon_1^2 + 2\beta_{xz} \frac{\bar{X}\bar{Y}}{\bar{Z}} \varepsilon_1'\varepsilon_1' \right\} \\ &= \bar{Y}^2 \left\{ 1 + \frac{C_y^2}{m} + \frac{\rho_{xz}^2 C_x^2}{n} + \frac{\rho_{xz}^2 C_x^2}{m} + \frac{C_z^2}{n} + 2\frac{\rho_{xz}\rho_{yz} C_x C_y}{n} - 2\frac{\rho_{xz}\rho_{yz} C_x C_y}{m} \right\} \\ &\quad \left\{ -2\frac{\rho_{yz} C_y C_z}{n} - 2\frac{\rho_{xz}^2 C_x^2}{n} \right\} \\ &= \bar{Y}^2 \left\{ 1 + \frac{C_y^2}{m} + \left(\frac{1}{m} - \frac{1}{n} \right) (\rho_{xz}^2 C_x^2 - 2\rho_{xz}\rho_{yz} C_x C_y) + \frac{1}{n} (C_z^2 - 2\rho_{yz} C_y C_z) \right\} \end{aligned}$$

Therefore,

$$\begin{aligned} M(T_p) &= E(T_p)^2 - \bar{Y}^2 = \\ &= \bar{Y}^2 \left\{ \frac{C_y^2}{m} + \left(\frac{1}{m} - \frac{1}{n} \right) (\rho_{xz}^2 C_x^2 - 2\rho_{xz}\rho_{yz} C_x C_y) + \frac{1}{n} (C_z^2 - 2\rho_{yz} C_y C_z) \right\} \end{aligned}$$

Appendix C

Minimum Mean Square Error of T_{p1}

From equation (8), we have

$$M(T_{p1}) = \psi^2 M(T_1) + (1-\psi)^2 M(T_p) \quad (19)$$

Since $M(T_{p_1})$ is a function of an unknown constant ψ , therefore, it is minimized with respect to ψ , i.e. differentiating $M(T_{p_1})$ with respect to ψ and equating them to zero, we get

$$\Rightarrow \psi_{opt} = \frac{M(T_p)}{M(T_1) + M(T_p)}$$

Then, substituting ψ_{opt} from equation (19), the minimum MSE of T_{p_1} as

$$M(T_{p_1})_{opt} = \frac{M(T_1)M(T_p)}{[M(T_1) + M(T_p)]}$$

For simplification, let

$$A' = \bar{Y}^2 C_y^2, \quad B' = \bar{Y}^2 (\rho_{xz}^2 C_x^2 - 2\rho_{xz}\rho_{yz} C_x C_y), \quad C' = \bar{Y}^2 (C_z^2 - 2\rho_{yz} C_y C_z)$$

Therefore,

$$M(T_1) = \frac{A' + C'}{u} = \frac{A' + C'}{n\mu}$$

$$M(T_p) = \frac{A'}{m} + \left(\frac{1}{m} - \frac{1}{n}\right)B' + \frac{C'}{n} = \frac{A'}{n(1-\mu)} + \frac{B'\mu}{n(1-\mu)} + \frac{C'}{n} \quad \text{since } m = n - u = n - n\mu = n(1-\mu)$$

Now,

$$M(T_{p_1})_{opt} = \frac{\frac{A' + C'}{n} [A' + B'\mu + C'(1-\mu)]}{\frac{(A' + C')(1-\mu) + A'\mu + B'\mu^2 + C'\mu(1-\mu)}{(1-\mu)}}$$

$$= \frac{(A' + C')^2 + (A' + C')(B' - C')\mu}{n[(A' + C') + (B' - C')\mu^2]}$$

or
$$M(T_{p_1})_{opt} = \frac{(\alpha_1')^2 + \alpha_1' \alpha_2' \mu}{n[\alpha_1' + \alpha_2' \mu^2]} \tag{20}$$

where

$$\alpha_1' = A' + C', \quad \alpha_2' = B' - C'$$

Appendix D

Optimum Replacement Policy

Differentiating $M(T_{P_1})_{opt}$ from equation (20) with respect to μ and equating them to zero, we get

$$\frac{1}{n} \left[\frac{(\alpha'_1 + \alpha'_2 \mu^2)(\alpha'_1 \alpha'_2) - ((\alpha'_1)^2 + \alpha'_1 \alpha'_2 \mu)(2\alpha'_2 \mu)}{[\alpha'_1 + \alpha'_2 \mu^2]^2} \right] = 0$$

$$\Rightarrow (\alpha'_1 \alpha'_2) \mu^2 + 2(\alpha'_1 \alpha'_2) \mu - (\alpha'_1)^2 \alpha'_2 = 0$$

$$\text{Then, } \mu = \left[\frac{-\alpha'_1 \pm \sqrt{(\alpha'_1)^2 + (\alpha'_1 \alpha'_2)}}{\alpha'_2} \right]$$

Now, assuming $C_x = C_y = C_z = C$ (say)

$$\begin{aligned} \alpha'_1 &= A' + C' = \bar{Y}^2 C_y^2 + \bar{Y}^2 (C_z^2 - 2\rho_{yz} C_y C_z) \\ &= 2\bar{Y}^2 C^2 (1 - \rho_{yz}) \\ \alpha'_2 &= B' - C' = \bar{Y}^2 [\rho_{xz}^2 C_x^2 - 2\rho_{xz} \rho_{yz} C_x C_y] - \bar{Y}^2 (C_z^2 - 2\rho_{yz} C_y C_z) \\ &= \bar{Y}^2 C^2 [(1 - \rho_{xz})(2\rho_{yz} - 1 - \rho_{xz})] \\ \alpha'_1 \alpha'_2 &= (A' + C')(B' - C') \\ &= 2\bar{Y}^2 C^2 (1 - \rho_{yz}) \left\{ \bar{Y}^2 C^2 [(1 - \rho_{xz})(2\rho_{yz} - 1 - \rho_{xz})] \right\} \\ &= (\bar{Y}^2 C)^2 [2(1 - \rho_{yz})(1 - \rho_{xz})(2\rho_{yz} - 1 - \rho_{xz})] \end{aligned}$$

$$\text{Therefore, } \mu = \frac{-\alpha'_1 \pm \sqrt{(\alpha'_1)^2 + (\alpha'_1 \alpha'_2)}}{\alpha'_2}$$

$$= \frac{-2(1 - \rho_{yz}) \pm \sqrt{2(1 - \rho_{yz}) \{2(1 - \rho_{yz}) + (1 - \rho_{xz})(2\rho_{yz} - \rho_{xz} - 1)\}}}{(1 - \rho_{xz})(2\rho_{yz} - \rho_{xz} - 1)} = \mu'_0 \text{ (say)}$$

Appendix E

Comparison of mean square error

(i). $M(T)_{opt} - M(T_{p_1})_{opt} > 0$

$$\begin{aligned} &\Rightarrow \frac{\alpha_1^2 + \alpha_1 \alpha_2 \mu_0}{n[\alpha_1 + \alpha_2 \mu_0^2]} - \frac{(\alpha_1')^2 + \alpha_1' \alpha_2' \mu_0'}{n[\alpha_1' + \alpha_2' (\mu_0')^2]} > 0 \\ &\Rightarrow \frac{[(1-\rho_{yz}) + (\rho_{yz} - \rho_{xz}) \mu_0]}{[(1-\rho_{yz}) + (\rho_{yz} - \rho_{xz}) \mu_0^2]} - \frac{[2(1-\rho_{yz}) + (1-\rho_{xz})(2\rho_{yz} - 1 - \rho_{xz}) \mu_0']}{[2(1-\rho_{yz}) + (1-\rho_{xz})(2\rho_{yz} - 1 - \rho_{xz}) (\mu_0')^2]} > 0 \end{aligned} \tag{21}$$

Let us assume that $\rho_{xz} = \rho_{yz} = \rho$ (say), equation (21) will become

$$\begin{aligned} &\frac{[(1-\rho) + (0) \mu_0]}{[(1-\rho) + (0) \mu_0^2]} - \frac{[2(1-\rho) + (1-\rho)(\rho-1) \mu_0']}{[2(1-\rho) + (1-\rho)(\rho-1) (\mu_0')^2]} > 0 \\ &\Rightarrow 1 - \frac{[2-(1-\rho) \mu_0']}{[2-(1-\rho) (\mu_0')^2]} > 0 \end{aligned}$$

Therefore,

$$M(T)_{opt} - M(T_{p_1})_{opt} > 0 \Rightarrow (1-\rho)(1-\mu_0') > 0$$

(ii). $V(\bar{y}_n) - M(T_{p_1})_{opt} > 0$

$$\Rightarrow \left\{ 1 - \frac{2(1-\rho_{yz})[2(1-\rho_{yz}) + (1-\rho_{xz})(2\rho_{yz} - 1 - \rho_{xz}) \mu_0']}{[2(1-\rho_{yz}) + (1-\rho_{xz})(2\rho_{yz} - 1 - \rho_{xz}) (\mu_0')^2]} \right\} > 0 \tag{22}$$

Assuming that $\rho_{xz} = \rho_{yz} = \rho$ (say), equation (22) will become

$$\begin{aligned} &1 - \frac{2(1-\rho)[2-(1-\rho) \mu_0']}{[2-(1-\rho) (\mu_0')^2]} > 0 \\ &\Rightarrow [2-(1-\rho) (\mu_0')^2] - 2(1-\rho)[2-(1-\rho) \mu_0'] > 0 \end{aligned}$$

Therefore,

$$\begin{aligned} & V(\bar{y}_n) - M(T_{p_1})_{opt} > 0 \\ \Rightarrow & \left[2 - (1-\rho)(\mu'_0)^2 \right] > 2(1-\rho) \left[2 - (1-\rho)\mu'_0 \right] \end{aligned}$$

(iii). $V(\bar{y}'_2)_{opt} - M(T_{p_1})_{opt} > 0$

$$\Rightarrow \left\{ \left[1 + \sqrt{1 - \rho_{xy}^2} \right] \frac{\bar{Y}^2 C^2}{2n} - \frac{2(\bar{Y}^2 C^2)^2 (1 - \rho_{yz}) \left[2(1 - \rho_{yz}) + (1 - \rho_{xz})(2\rho_{yz} - 1 - \rho_{xz})\mu'_0 \right]}{n(\bar{Y}^2 C^2) \left[2(1 - \rho_{yz}) + (1 - \rho_{xz})(2\rho_{yz} - 1 - \rho_{xz})(\mu'_0)^2 \right]} \right\} > 0 \quad (23)$$

Assuming $\rho_{xz} = \rho_{yz} = \rho$ (say), equation (23) will become

$$\Rightarrow \left(1 + \sqrt{1 - \rho^2} \right) \left[2 - (1 - \rho)(\mu'_0)^2 \right] - 4(1 - \rho) \left[2 - (1 - \rho)\mu'_0 \right] > 0$$

Therefore,

$$\begin{aligned} & V(\bar{y}'_2)_{opt} - M(T_{p_1})_{opt} > 0 \\ \Rightarrow & \left(1 + \sqrt{1 - \rho^2} \right) \left[2 - (1 - \rho)(\mu'_0)^2 \right] > 4(1 - \rho) \left[2 - (1 - \rho)\mu'_0 \right] \end{aligned}$$