Statistics and Applications {**ISSN 2452-7395 (online)**}
Volume 17, No. 2, 2019 (New Series), pp 11-21

Ratio Type HT Estimators for Negative Adaptive Cluster Double Sampling

Raosaheb V. Latpate

Department of Statistics, Savitribai Phule Pune University, Pune-411007, India

Received: July 02, 2019; Revised: October 05, 2019; Accepted: November 26, 2019

Abstract

When the population is rare and clustered, the traditional sampling design gives the poor estimate of population total/ mean. In this situation, the negative adaptive cluster double sampling (NACDS) design is useful to gather the information. It is assumed that auxiliary information is negatively correlated with study of interest variable and auxiliary information is abundant. Hence, information of interest variable is rare and clustered. Traditional, ratio estimator performs poor. Hence, exponential ratio type Horvitz Thompson (HT) estimator and log ratio type HT estimator is proposed. The performance of these two estimators are compared by using sample survey.

Key words: NACDS; HT Estimator; Ratio Type Estimators; NACS.

1. Introduction

While conducting a sample survey, a number of difficult sampling problems are encountered. One of them is the problem in estimating the population mean/total when it is rare or geographically uneven. If the population of interest is hidden or elusive then it becomes difficult to identify it for sampling. The researcher may find the conventional sampling design such as simple random sampling (SRS), stratified sampling etc. as inadequate for producing data from the sampling units while studying such type of population. If conventional sampling designs are applied to population that are rare and clustered then usually very few units possessing the characteristic of interest are selected in the sample. Even a very large conventional sample would be inadequate in such cases. Due to these reasons researchers have thought about the unconventional sampling designs.

According to Thompson and Seber (1996), designs that can redirect sampling effort during a survey in response to observed values are known as adaptive sampling designs. These designs use

Corresponding Author: Raosaheb V. Latpate
Email id: rvl@unipune.ac.in

information gathered during the survey to select the succeeding sampling units. This distinguishes adaptive sampling from conventional sampling designs.

Neyman (1938) introduced two-phase sampling (double sampling). Horvitz and Thompson (1952) provided a general method of dealing with sampling without replacement from a finite universe when variable probabilities of selection are used for the elements remaining prior to each draw. They proposed an unbiased linear estimator of the population total of the variable of interest. An unbiased estimator of the sampling variance of this estimator is also obtained. The above estimator of population total is applicable to one-stage design. The authors presented an extension of this method for two-stage design. But this has a serious disadvantage of the possibility of having negative variance. Murthy (1957) has shown that corresponding to any ordered estimator there exists a more efficient unordered estimator.

Cassel et al. (1977) reviewed adaptive designs under the term informative designs. Thompson and Ramsey (1983) analyzed the situations of adaptive sampling designs. Seber(1986) described the potential importance of adaptive designs for the estimation of animal abundance. Thompson (1987) described the examples analyzed by him and Ramsey (1981). In this paper, they proved a theorem for a non-adaptive design in which the entire sample is selected ahead of time, will be optimum if and only if there is some possible selection of second phase units which is best for every possible outcome of the first phase observations. Särndal and Swensson (1987) have discussed estimation in the case of two-phase sampling and in the case of non-response. In this paper, general results were given for two- phase sampling with emphasis on regression estimation and on the problem of variance estimation. The concept of inclusion probability proportional to size sampling plans excluding adjacent units separated by at most a distance of $m \geq 1$ units is introduced by Mandal (2008).

Thompson (1990, 1991) presented designs in which, whenever the observed value of a selected unit satisfies a condition of interest, additional units are added to the sample from the neighbourhood of that unit. In these designs, the selection procedure depends on observed values of the variable of interest. Latpate et al. (2018 a) evaluated the sample size for adaptive cluster sampling. Medina and Thompson (2004) presented a multi-phase variant of ACS. They combined the ideas of double sampling and ACS. They called this new design as adaptive cluster double sampling. In this design the authors assumed the availability of an inexpensive and easy to measure auxiliary variable.

Latpate and Kshirsagar (2019) proposed negative adaptive cluster sampling (NACS) design. In this design, the variable of interest is negatively correlated with auxiliary variable. The adaptive procedure is used by using auxiliary variable. The condition of adaptation is on the auxiliary variable. The population is rare and clustered for interest variable. Because of negative correlation between interest and auxiliary variable, there is abundance of auxiliary information. It is easy to measure and less costly as compared to interest variable. The networks are formed for the rare occurrence of the auxiliary variable. It means the selected network has abundant information of interest variable. There is substantial expected sample size and cost reduction for the interest variable. Also, the auxiliary information is used at estimation and design stage. They have proposed the modified ratio and regression estimator. Latpate and Kshirsagar (2020) presented the two-stage negative adaptive cluster sampling design. This design is a combination of two-stage cluster sampling and NACS. Also, the modified ratio and regression estimator is proposed. In this design, auxiliary information is used at design and estimation stage.

Latpate and Kshirsagar (2018 b) presented the negative adaptive cluster double sampling (NACDS) design. It is a combination of NACS and double sampling. In this design,they assume that the auxiliary information is easily available and less expensive. The nature of underlined population is rare and patchy. To exploit the auxiliary information at design and estimation stage, auxiliary information must be abundantly available, easy to measure and less costly. The procedure of NACDS is as follows. Let $U = \{u_1, u_2, ..., u_N\}$ be a finite population of N units. Let Y and X be the interest and auxiliary variable respectively. They are known to be highly negatively correlated. Let X_i and $Y_i, i = 1, 2, ..., N$ be the values of X and Y respectively associated with the unit u_i. It is assumed that the information on auxiliary variable can be obtained from all the units selected in the sample. The goal is to estimate the population total of Y, given by $\tau_Y = \sum_{i=1}^{N} Y_i$.

An initial sample of size n units is drawn from the population by using SRSWOR. We denote this initial sample drawn as S_0. From S_0, obtain an adaptive cluster sample S_1 by using the following procedure: Denote the condition of interest with respect to X values by C_X. According to the negative correlation the condition is reversed for adaptation. Now following the procedure given by Thompson (1990), we add the neighbors of the units in S_0 that satisfy the condition C_X. The units to the right, left, above and below a unit are called as the neighbors of that unit. If any of these neighbors satisfy C_X then their neighbors are also added to the sample. This is continued till the neighbors not satisfying C_X are obtained. The units added to the sample S_0 adaptively which satisfy the condition C_X constitute a network. The units added to the sample S_0 adaptively which do not satisfy the condition C_X are called as the edge units. The set of units in a network along with its edge units is called as a cluster. The set of units included in all such clusters is called as an adaptive cluster sample. We denote it by S_1. Thus, indirectly we are assuming that the condition C_X for the additional sampling and a set of neighboring units for each $u_i \in U$ have been defined. Let K denote the number of distinct clusters formed by S_0. Mark the corresponding K clusters in the Y population and drop down the edge units to get K networks. This completes the first phase of the design. From each of these selected networks draw a sample by using SRSWOR. The sizes of these samples may be different. Suppose m_i denotes the number of units selected from the i^{th} selected network. Collection of all these units selected be denoted by S_2. This completes the second phase of sampling design. Now, note the values of X and Y for all the units included in S_2. This data is used to estimate the population parameter. In this design, the X value associated with every unit in the adaptive cluster sample S_1 has to be measured. Hence, the procedure does not control the number of observations on the auxiliary variable, but only the number of observations on the survey variable.

Bahl and Tuteja (1991) proposed the ratio and product type exponential estimators. Using this approach, I proposed the ratio and exponential ratio type HT estimator and log ratio type HT estimators. These estimators are useful to handle the problem of rare/clustered population. Särandal et al.(1992) proposed ratio type estimator. But, this estimator is less precise. These two estimators are compared by using monte carlo simulation method. The sample survey is presented for the comparison purpose. The proposed estimators are presented in section 2. Section 3, the sample survey is conducted by using NACDS. The results and discussion are added in section 4. The concluding remarks are incorporated in section 5.

2. The Proposed Estimators

The first phase sample S_I of size n_0 from population U is drawn by using simple random sampling without replacement and units are added by using Thompson [1990] procedure. The first order inclusion probabilities are as follows:

π_i : probability that unit i is included in $S_I = \sum_{i \in S_I} P(S_I)$
π_{ij} : probability that units i and j is included in $S_I = \sum_{i,j \in S_I} P(S_I)$
with $\pi_{ii} = \pi_i$ and $\pi_i > 0$ for all i and $\pi_{ij} > 0$ for all $i \neq j$.

Again, the second phase sample S_{II} of size n_1 from S_I is drawn by using simple random sampling without replacement. The second order conditional inclusion probabilities are as follows:

$\pi_{i|S_I}$: probability that i^{th} unit is included in S_{II} given $S_I = \sum_{i \in S_{II}} P(S_{II}|S_I)$
$\pi_{ij|S_I}$: probability that i^{th} and j^{th} units is included in S_{II} given $S_I = \sum_{i,j \in S_{II}} P(S_{II}|S_I)$
with $\pi_{ii|S_I} = \pi_{i|S_I}$ and for any S_I, $\pi_{i|S_I} > 0$ for all i and $\pi_{ij|S_I} > 0$ for all $i \neq j$.

The π^* estimator can be expressed as follows (Särandal et al.[1992]).

$$\pi_i^* = \begin{cases} \pi_i & ;if \quad i \in S_I \\ \pi_i \pi_{i|S_I} & ;if \quad i \in S_I \quad and \quad i \in S_{II} \end{cases}$$

and

$$\pi_{ij}^* = \begin{cases} \pi_{ij}\pi_{ij|S_I} & ;if \quad i,j \in S_I \quad and \quad i,j \in S_{II} \\ \pi_{ij}\pi_{i|S_I} & ;if \quad i,j \in S_I \quad and \quad i \in S_{II} \\ \pi_{ij}\pi_{j|S_I} & ;if \quad i,j \in S_I \quad and \quad j \in S_{II} \\ \pi_{ij} & ;if \quad i,j \in S_I \end{cases}$$

The HT estimator for the interest variable Y is,

$(\hat{\tau}_{IIy})_{HT} = \sum_{i \in S_{II}} y_i^*/\pi_i^*$
where, y_i^* the sum of the selected units at the second stage of the networks which includes the i^{th} unit. The HT estimator for the auxiliary variable X at second phase and first phase respectively are,

$(\hat{\tau}_{IIx})_{HT} = \sum_{i \in S_{II}} x_i^*/\pi_i^*$.
where, x_i^* the sum of the selected units at the second stage of the networks which includes the i^{th} unit.

$(\hat{\tau}_{Ix})_{HT} = \sum_{i \in S_I} x_i^*/\pi_i^*$ where, x_i^* the sum of the selected first stage of the networks which includes the i^{th} unit. We assume the large sample approximation to obtain the MSE.
$e_y = \frac{(\hat{\tau}_{IIy})_{HT} - \tau_y}{\tau_y}$
$e_x = \frac{(\hat{\tau}_{IIx})_{HT} - \tau_x}{\tau_x}$
$e_{x'} = \frac{(\hat{\tau}_{Ix})_{HT} - \tau_x}{\tau_x}$
We get
$E(e_y) = E_I(E_{II}(e_y|S_I)) = 0$, $E(e_x) = E_I(E_{II}(e_x|S_I)) = 0$, $E(e_{x'}) = E_I(E_{II}(e_{x'}|S_I)) = 0$;

$$V(e_y) = V_I(E_{II}(\tfrac{(\hat{\tau}_{IIy})_{HT}-\tau_y}{\tau_y}|S_I)) + E_I(V_{II}(\tfrac{(\hat{\tau}_{IIy})_{HT}-\tau_y}{\tau_y}|S_I))$$
$$= V_I\big(\tfrac{(\hat{\tau}_{Iy})_{HT}-\tau_y}{\tau_y}\big) + \tfrac{1}{2\tau_y^2}E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I} - \pi_{ij|S_I})(\tfrac{y_i}{\pi_i^*} - \tfrac{y_j}{\pi_j^*})^2\big)$$
$$= \tfrac{1}{2\tau_y^2}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j - \pi_{ij})(\tfrac{y_i}{\pi_i} - \tfrac{y_j}{\pi_j})^2 + E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I} - \pi_{ij|S_I})(\tfrac{y_i}{\pi_i^*} - \tfrac{y_j}{\pi_j^*})^2\big)\big]$$
$$V(e_x) = V_I(E_{II}(\tfrac{(\hat{\tau}_{IIx})_{HT}-\tau_x}{\tau_x}|S_I)) + E_I(V_{II}(\tfrac{(\hat{\tau}_{IIx})_{HT}-\tau_x}{\tau_x}|S_I))$$
$$= V_I\big(\tfrac{(\hat{\tau}_{Ix})_{HT}-\tau_x}{\tau_x}\big) + \tfrac{1}{2\tau_x^2}E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I} - \pi_{ij|S_I})(\tfrac{x_i}{\pi_i^*} - \tfrac{x_j}{\pi_j^*})^2\big)$$
$$= \tfrac{1}{2\tau_x^2}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j - \pi_{ij})(\tfrac{x_i}{\pi_i} - \tfrac{x_j}{\pi_j})^2 + E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I} - \pi_{ij|S_I})(\tfrac{x_i}{\pi_i^*} - \tfrac{x_j}{\pi_j^*})^2\big)\big]$$
$$V(e_{x'}) = V_I(E_{II}(\tfrac{(\hat{\tau}_{Ix})_{HT}-\tau_x}{\tau_x}|S_I)) + E_I(V_{II}(\tfrac{(\hat{\tau}_{Ix})_{HT}-\tau_x}{\tau_x}|S_I))$$
$$= V_I\big(\tfrac{(\hat{\tau}_{Ix})_{HT}-\tau_x}{\tau_x}\big)$$
$$= \tfrac{1}{2\tau_x^2}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j - \pi_{ij})(\tfrac{x_i}{\pi_i} - \tfrac{x_j}{\pi_j})^2\big]$$
$$Cov(e_x, e_y) = Cov_I(E_{II}(\tfrac{(\hat{\tau}_{IIy})_{HT}-\tau_y}{\tau_y}|S_I), E_{II}(\tfrac{(\hat{\tau}_{IIx})_{HT}-\tau_x}{\tau_x}|S_I)) +$$
$$E_I(Cov_{II}(\tfrac{(\hat{\tau}_{IIy})_{HT}-\tau_y}{\tau_y}, \tfrac{(\hat{\tau}_{IIx})_{HT}-\tau_x}{\tau_x}|S_I))$$
$$= Cov_I\big(\tfrac{(\hat{\tau}_{Iy})_{HT}-\tau_y}{\tau_y}, \tfrac{(\hat{\tau}_{Ix})_{HT}-\tau_x}{\tau_x}\big) + \tfrac{1}{2\tau_x\tau_y}E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I} - \pi_{ij|S_I})(\tfrac{x_i}{\pi_i^*} - \tfrac{y_j}{\pi_j^*})^2\big)$$
$$= \tfrac{1}{2\tau_x\tau_y}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j - \pi_{ij})(\tfrac{x_i}{\pi_i} - \tfrac{y_j}{\pi_j})^2 + E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I} - \pi_{ij|S_I})(\tfrac{x_i}{\pi_i^*} - \tfrac{y_j}{\pi_j^*})^2\big)\big]$$
$$Cov(e_{x'}, e_x) = Cov_I(E_{II}(\tfrac{(\hat{\tau}_{Ix})_{HT}-\tau_x}{\tau_x}|S_I), E_{II}(\tfrac{(\hat{\tau}_{IIx})_{HT}-\tau_x}{\tau_x}|S_I)) +$$
$$E_I(Cov_{II}(\tfrac{(\hat{\tau}_{Ix})_{HT}-\tau_x}{\tau_x}, \tfrac{(\hat{\tau}_{IIx})_{HT}-\tau_x}{\tau_x}|S_I))$$
$$= Cov_I\big(\tfrac{(\hat{\tau}_{Ix})_{HT}-\tau_x}{\tau_x}, \tfrac{(\hat{\tau}_{Ix})_{HT}-\tau_x}{\tau_x}\big)$$
$$= \tfrac{1}{2\tau_x^2}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j - \pi_{ij})(\tfrac{x_i}{\pi_i} - \tfrac{x_j}{\pi_j})^2$$
$$Cov(e_{x'}, e_y) = Cov_I(E_{II}(\tfrac{(\hat{\tau}_{Ix})_{HT}-\tau_x}{\tau_x}|S_I), E_{II}(\tfrac{(\hat{\tau}_{IIy})_{HT}-\tau_y}{\tau_y}|S_I)) +$$
$$E_I(Cov_{II}(\tfrac{(\hat{\tau}_{Ix})_{HT}-\tau_x}{\tau_x}, \tfrac{(\hat{\tau}_{IIy})_{HT}-\tau_y}{\tau_y}|S_I))$$
$$= Cov_I\big(\tfrac{(\hat{\tau}_{Ix})_{HT}-\tau_x}{\tau_x}, \tfrac{(\hat{\tau}_{Iy})_{HT}-\tau_y}{\tau_y}\big)$$
$$= \tfrac{1}{2\tau_x\tau_y}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j - \pi_{ij})(\tfrac{x_i}{\pi_i} - \tfrac{y_j}{\pi_j})^2$$

(i). Exponential Ratio Type HT Estimator: The exponential ratio type HT estimator for NACDS is as follows.

$$\hat{\tau}_{RADE} = (\hat{\tau}_{IIy})_{HT}\ exp\big[\tfrac{(\hat{\tau}_{Ix})_{HT}-(\hat{\tau}_{IIx})_{HT}}{(\hat{\tau}_{Ix})_{HT}+(\hat{\tau}_{IIx})_{HT}}\big]$$

Using the large sample approximation and neglecting the higher order terms we get. $\hat{\tau}_{RADE} = \tau_y\ \big[1 + e_y + \frac{1}{2}(e_{x'} - e_x) + \frac{1}{2}(e_ye_{x'} - e_ye_x) - \frac{1}{8}e_{x'}^2 + \frac{3}{8}e_x^2 - \frac{1}{4}e_{x'}e_x\big]$

$\hat{\tau}_{RADE}$ is a biased estimator.The bias of $\hat{\tau}_{RADE}$ is as follows.

$$Bias(\hat{\tau}_{RADE}) = \tau_y\big(\tfrac{3}{16\tau_x^2}\big[E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I} - \pi_{ij|S_I})(\tfrac{x_i}{\pi_i^*} - \tfrac{x_j}{\pi_j^*})^2\big)\big]$$
$$-\tfrac{1}{4\tau_x\tau_y}\big[E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I} - \pi_{ij|S_I})(\tfrac{x_i}{\pi_i^*} - \tfrac{y_j}{\pi_j^*})^2\big)\big]\big)$$

The mean square error of $\hat{\tau}_{RADE}$ can be expressed as follows.

$$MSE(\hat{\tau}_{RADE}) = \tau_y^2\ Var\big[e_y + \tfrac{1}{2}(e_{x'} - e_x)\big]$$
$$= \tau_y^2\ E\big[e_y^2 + \tfrac{1}{4}e_{x'}^2 - \tfrac{1}{2}e_{x'}e_x + \tfrac{1}{4}e_x^2 + e_ye_{x'} - e_ye_x\big]$$
$$MSE(\hat{\tau}_{RADE}) = \tau_y^2\ \big[\tfrac{1}{2\tau_y^2}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j - \pi_{ij})(\tfrac{y_i}{\pi_i} - \tfrac{y_j}{\pi_j})^2 + E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I} - \pi_{ij|S_I})(\tfrac{y_i}{\pi_i^*} - \tfrac{y_j}{\pi_j^*})^2\big)\big] + \tfrac{1}{8\tau_x^2}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j - \pi_{ij})(\tfrac{x_i}{\pi_i} - \tfrac{x_j}{\pi_j})^2\big] - \tfrac{1}{4\tau_x^2}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j - \pi_{ij})(\tfrac{x_i}{\pi_i} - \tfrac{x_j}{\pi_j})^2\big] +$$

$\frac{1}{8\tau_x^2}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j-\pi_{ij})(\frac{x_i}{\pi_i}-\frac{x_j}{\pi_j})^2+E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I}-\pi_{ij|S_I})(\frac{x_i}{\pi_i^*}-\frac{x_j}{\pi_j^*})^2\big)\big]+$
$\frac{1}{2\tau_x\tau_y}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j-\pi_{ij})(\frac{x_i}{\pi_i}-\frac{y_j}{\pi_j})^2\big]-\frac{1}{2\tau_x\tau_y}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j-\pi_{ij})(\frac{x_i}{\pi_i}-\frac{y_j}{\pi_j})^2+$
$E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I}-\pi_{ij|S_I})(\frac{x_i}{\pi_i^*}-\frac{y_j}{\pi_j^*})^2\big)\big]\big]$

After Simplification, the MSE of $\hat{\tau}_{RADE}$ is,

$$MSE(\hat{\tau}_{RADE})=\tau_y^2\,\big[\frac{1}{2\tau_y^2}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j-\pi_{ij})(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j})^2+E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I}-\pi_{ij|S_I})(\frac{y_i}{\pi_i^*}-\frac{y_j}{\pi_j^*})^2\big)\big]+\frac{1}{8\tau_x^2}\big[E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I}-\pi_{ij|S_I})(\frac{x_i}{\pi_i^*}-\frac{x_j}{\pi_j^*})^2\big)\big]-\frac{1}{2\tau_x\tau_y}\big[E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I}-\pi_{ij|S_I})(\frac{x_i}{\pi_i^*}-\frac{y_j}{\pi_j^*})^2\big)\big]\big] \quad (1)$$

(ii). Log Ratio Type HT Estimator: The log ratio type HT estimator for NACDS is as follows.

$$\hat{\tau}_{RADL}=(\hat{\tau}_{IIy})_{HT}\big(1+log\big[\frac{(\hat{\tau}_{IIx})_{HT}}{(\hat{\tau}_{Ix})_{HT}}\big]\big)$$

Using the large sample approximation and neglecting the higher order terms we get. $\hat{\tau}_{RADL}=\tau_y(1+e_y-e_{x'}+e_x+e_{x'}^2-e_xe_{x'}-e_{x'}e_y+e_xe_y)$

$\hat{\tau}_{RADL}$ is a biased estimator. The bias of $\hat{\tau}_{RADL}$ is as follows.

$$Bias(\hat{\tau}_{RADL})=\tau_y\big(\frac{1}{2\tau_x^2}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j-\pi_{ij})(\frac{x_i}{\pi_i}-\frac{x_j}{\pi_j})^2\big]+\frac{1}{2\tau_x\tau_y}\big[E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I}-\pi_{ij|S_I})(\frac{x_i}{\pi_i^*}-\frac{y_j}{\pi_j^*})^2\big)\big]\big)$$

The mean square error of $\hat{\tau}_{RADL}$ is as follows.

$$MSE(\hat{\tau}_{RADL})=V\big(\tau_y(1+e_y+e_{x'}-e_x)\big)$$
$$=\tau_y^2\big[V(e_y)+V(e_{x'})+V(e_x)-2Cov(e_x,e_{x'})-2Cov(e_y,e_{x'})+2Cov(e_y,e_x)\big]$$
$$MSE(\hat{\tau}_{RADL})=\tau_y^2\big[\frac{1}{2\tau_y^2}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j-\pi_{ij})(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j})^2+E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I}-\pi_{ij|S_I})(\frac{y_i}{\pi_i^*}-\frac{y_j}{\pi_j^*})^2\big)\big]+\frac{1}{2\tau_x^2}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j-\pi_{ij})(\frac{x_i}{\pi_i}-\frac{x_j}{\pi_j})^2\big]+\frac{1}{2\tau_x^2}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j-\pi_{ij})(\frac{x_i}{\pi_i}-\frac{x_j}{\pi_j})^2+E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I}-\pi_{ij|S_I})(\frac{x_i}{\pi_i^*}-\frac{x_j}{\pi_j^*})^2\big)\big]-\frac{1}{\tau_x^2}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j-\pi_{ij})(\frac{x_i}{\pi_i}-\frac{x_j}{\pi_j})^2-\frac{1}{\tau_x\tau_y}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j-\pi_{ij})(\frac{x_i}{\pi_i}-\frac{y_j}{\pi_j})^2+\frac{1}{\tau_x\tau_y}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j-\pi_{ij})(\frac{x_i}{\pi_i}-\frac{y_j}{\pi_j})^2+E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I}-\pi_{ij|S_I})(\frac{x_i}{\pi_i^*}-\frac{y_j}{\pi_j^*})^2\big)\big]\big]$$

After Simplification, the MSE of $\hat{\tau}_{RADL}$ is,

$$MSE(\hat{\tau}_{RADL})=\tau_y^2\big[\frac{1}{2\tau_y^2}\big[\sum\sum_{i\neq j\in U}(\pi_i\pi_j-\pi_{ij})(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j})^2+E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I}-\pi_{ij|S_I})(\frac{y_i}{\pi_i^*}-\frac{y_j}{\pi_j^*})^2\big)\big]+\frac{1}{2\tau_x^2}\big[E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I}-\pi_{ij|S_I})(\frac{x_i}{\pi_i^*}-\frac{x_j}{\pi_j^*})^2\big)\big]+\frac{1}{\tau_x\tau_y}\big[E_I\big(\sum\sum_{i\neq j\in S_I}(\pi_{i|S_I}\pi_{j|S_I}-\pi_{ij|S_I})(\frac{x_i}{\pi_i^*}-\frac{y_j}{\pi_j^*})^2\big)\big]\big] \quad (2)$$

3. Sample Survey

A sample survey was conducted by using NACDS. The area of 400 acres in the Tamhini Ghat, Maharashtra, India was divided into 400 plots each of size 1 acre. A random sample of 12 plots was drawn from this area by using SRSWOR. The percentage of silica content of the soil (X) was measured on these selected plots. Silica is abundant in the soil from Tamhini Ghat to Mumbai. But, there are intermediate patches of laterite where the occurrence of evergreen plants is more. We considered the condition C_X = $\{X \leq 20\}$ = $\{Percentage of silica \leq 20\}$. Further the plots in the sample satisfying C_X were located. Then the clusters were formed around these plots by using the procedure given by Thompson (1990). Each plot with $\{X > 20\}$ and selected in the initial sample formed a cluster of size one. Here the clusters were formed by using auxiliary information and the domain knowledge of silica content and evergreen plants. The two variables, percentage of silica content X and number of evergreen plants Y are negatively correlated. After forming such clusters in the X population, the edge units of clusters of size more than one were dropped to get networks. The networks are formed by using percent of silica content. The corresponding networks of number of evergreen plants are located. Figures 1 and 2 illustrate this methodology.

These plots formed the first phase sample S_1. Let K denote the number of distinct networks represented in this sample. A random sample of m_i (say), $(i = 1, 2, \cdots, K)$ units was drawn from the i^{th} network among these K networks by using SRSWOR. The collection of all so selected units formed the second phase sample S_2. In our study there were 12 networks formed in S_1. We took $m_1 = m_2 = m_3 = 2, m_4 = 3, m_5 = 4, m_6 = 2, m_7 = 2, m_8 = 2, m_9 = m_{10} = 4, m_{11} = 3$ and $m_{12} = 0$.

This set of units formed S_2. Values of the variables X and Y corresponding to the plots included in the second phase sample were recorded together to form a bivariate data. Using this data, the total number of evergreen plants in that area was estimated by using the proposed estimators.

4. Results and Discussion

For the computational efficiency in estimation, r number of repetitions were performed; where r varied as 5000, 10000 and 20000. We considered the initial sample sizes as 5, 10, 15, 20 and 25 for each repetition.

The estimated population total over r repetitions is given by:
$\hat{\tau}_Y = \frac{\sum_{i=1}^{r} \hat{\tau}_{Y_i}}{r}$
where $\hat{\tau}_{Y_i}$ denotes the estimated value of an estimator of the population total of the variable Y for the i^{th} repetition.

The estimated mean square error of the estimator of population total of the variable Y is given by:
$\widehat{MSE}(\hat{\tau}_Y) = \frac{\sum_{i=1}^{r} (\hat{\tau}_{Y_i} - \tau_Y)^2}{r}$

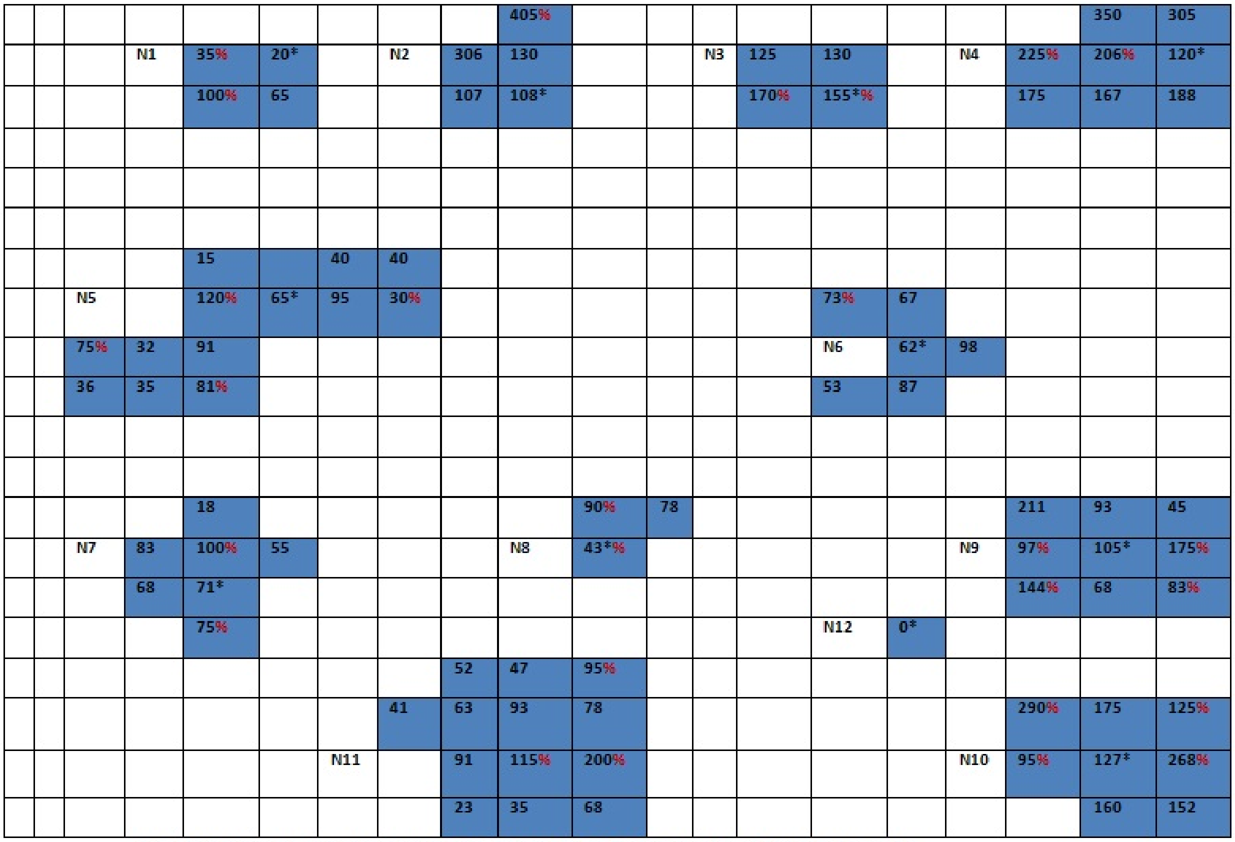
24	25	86	60	52	35	65	50	60	1	22	23	83	48	30	56	43	52	1	4
40	30	30	75	18	19*	55	30	4	14	38	27	27	14	14	49	23	6	6	10*
45	48	56	23	15	17	53	30	13	12*	43	45	53	11	12*	47	23	7	8	7
47	47	23	25	80	60	45	45	35	70	45	44	20	76	55	39	38	27	61	34
48	50	25	35	57	68	40	23	80	40	46	47	22	53	63	34	26	72	31	37
49	43	36	65	58	58	90	45	90	30	47	40	33	54	53	84	38	82	21	30
45	35	56	85	19	30	18	18	40	50	43	32	53	25	25	42	41	32	41	22
48	53	65	55	13	16*	15	18	30	60	46	50	62	29	17	18	51	22	51	40
70	30	17	18	15	48	44	44	35	50	68	27	24	29	43	19*	12	28	41	27
30	30	18	17	15	43	36	50	80	36	28	27	25	24	22	14	43	72	27	27
29	31	93	68	61	45	66	52	63	25	27	29	93	42	32	59	47	57	27	42
45	36	37	83	27	29	56	32	37	48	43	36	37	88	26	52	27	41	43	58
50	54	63	31	20	27	54	32	76	77	16	18	63	85	24	50	27	7	14	20
52	53	30	15	18	20	46	47	38	75	18*	21	20	70	57	42	42	12	11*	9
53	57	32	19	18*	70	41	25	83	45	51	20	32	47	65	37	20	10	18	15
54	50	43	73	16	78	91	47	93	24	26	42	43	59	55*	87	42	87	48	38
50	42	63	93	67	68	29	28	20	19	12	24	63	33	27	35	35	37	34	30
53	60	72	63	28	40	26	20	19*	13	11	23	70	35	29	65	85	5	8	10
75	37	24	26	22	26	45	46	12	10	9	28	24	45	45	22	26	14	9*	5
35	37	25	25	24	58	37	52	18	16	14	48	25	52	24	77	47	77	9	11

* in a square indicates selection in initial sample.

Figure 1: Silica (SiO_2) percentage on the different plots in the region.

Table 1: Estimated Values of Population Total of interest variable and its SE.

No. of Repetitions	Initial Sample Size	Exp Ratio Estimator		Ratio Estimator		Log Ratio Estimator	
r	n	$\hat{\tau}_{RADE}$	$\hat{SE}(\hat{\tau}_{RADE})$	$(\hat{\tau}_y)_{Ratio}$	$\hat{SE}(\hat{\tau}_y)_{Ratio}$	$\hat{\tau}_{RADL}$	$\hat{SE}(\hat{\tau}_{RADL})$
5000	5	9073.83	9594.65	9555.85	10461.68	8875.46	9212.46
	10	9094.24	6558.31	9206.55	6778.47	9044.51	6443.12
	15	9099.11	5200.63	9230.12	5373.81	9048.99	5085.60
	20	9075.71	4401.83	9106.03	4595.43	9031.09	4330.44
	25	9060.59	3878.48	9144.12	3992.87	8994.69	3800.33
10000	5	9429.05	9738.40	9458.85	10383.6	9053.38	9375.20
	10	9018.30	6452.53	9223.40	6738.29	8900.90	6384.41
	15	9055.01	5186.37	9114.49	5328.22	8936.08	5076.77
	20	9096.61	4416.34	9173.93	4577.09	9012.60	4317.32
	25	9082.23	3830.18	9110.27	3964.64	9014.61	3758.87
20000	5	9189.86	9676.48	9421.81	10328.04	8812.90	9277.79
	10	9071.97	6507.64	9224.23	6724.55	8962.86	6347.45
	15	9047.53	5197.31	9097.89	5313.16	8945.107	5112.01
	20	9108.18	4406.09	9047.11	4529.84	9025.19	4316.37
	25	9078.90	3868.39	9047.08	3947.24	9006.03	3781.37



N1 to N12 denote the network numbers. % in a plot indicates the selection at phase two.

Figure 2: Number of evergreen plants observed on the plots in the population.

The estimated values are presented in Table 1. It showed that, as the initial sample size increases, the standard error decreases. The proposed exponential ratio type HT estimator performs better as compared to ratio estimator proposed by Särandal et al.(1992). The log ratio type HT estimator is more efficient as compared to ratio estimator proposed by Särandal et al.(1992) and exponential ratio type HT estimator. Even though, exponential ratio type HT estimator and log ratio type HT estimator are biased with minimum MSE. Log ratio type HT estimator is negatively biased and exponential ratio type HT estimator is positively biased estimator.

Remark: If there is positive correlation between X and Y. Then, exponential ratio type HT estimator is more efficient as compared to ratio estimator proposed by Särandal et al.(1992) and log ratio type HT estimator.

Theoretically, it clearly shows that equation 1 and 2 of MSE. The covariance term is added in equation 1 and subtracted in equation 2.

5. Conclusions

The auxiliary variable and interest variable are negatively correlated. The auxiliary information is used at design and estimation stage. Using auxiliary information in NACDS, the network of interest information is identified and random samples are selected from these selected networks. The inclusion probabilities are evaluated. The proposed log ratio type HT estimator is efficient as compared to traditional ratio estimator and proposed exponential ratio type HT estimator. When, we employ this methodology for positively correlated variables. The exponential ratio type HT estimator is efficient as compared to log ratio type HT estimator and traditional ratio estimator. This is the important features of these estimators. These estimators can be useful for estimation of population total for the sample surveys in ecology, environmental science, health science and forestry.

References

Bahl, S. and Tuteja, R. K. (1991). Ratio and product type exponential estimators. *Journal of Information and Optimization Sciences*, **12**, 159-163.

Cassel, C. M., Särndal, C. E. and Wretman, J. H. (1977). *Foundation of Inference in Survey Sampling*. Wiley, New York.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of American Statistical Association*, **47**, 663-685.

Latpate, R. Kshirsagar, J. and Gore, S. (2018 a). Sample size consideration in adaptive cluster sampling. *Bulletin of the Marathwada Mathematical Society*, **19(1)**, 32-41.

Latpate, R. and Kshirsagar, J. (2018 b). Negative adaptive double cluster sampling. *Journal of Statistics and Applications*, **16(2)**, 117-134.

Latpate, R. and Kshirsagar, J. (2019). Negative adaptive cluster sampling. *Model Assisted Statistics and Applications*, **14**, 65–81.

Latpate, R. and Kshirsagar, J. (2020). Two stage negative adaptive cluster sampling. *Communications in Mathematics and Statistics*,**8**, 1-21.

Mandal, B. N., Parsad, R. and Gupta, V. K. (2008). IPPS sampling plans excluding adjacent units. *Communications-in-Statistics: Theory and Methods*, **37(16)**, 2532-2550.

Medina, M. H. F. and Thompson, S. K. (2004). Adaptive cluster double sampling. *Biometrika*, **91**, 877-891.

Murthy, M. N. (1957). Ordered and unordered estimators in sampling without replacement. *Sankhya*, **18**, 379-390.

Neyman, J. (1938). Contribution to the theory of sampling human populations. *Journal of American Statistical Association*, **33**, 101-116.

Ramsey, F. L. and Scott, J. M. (1981). Estimating number of terrestrial birds. *Studies in Avian Biology*, **6**, 341-345.

Särandal, C. E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York, Inc.

Seber, G. A. F. (1986). A review of estimating animal abundance. *Biometrics*, **42(2)**, 267-292.

Thompson, S. K. and Ramsey, F. L. (1983). *Adaptive Sampling of Animal Populations*. Technical report 82, Corvallis, OR: Department of Statistics, Oregon State University.

Thompson, S. K. (1987). Sample size for estimating multinomial proportions. *The American Statistician*, **41(1)**, 42-46.

Thompson S. K. (1990). Adaptive cluster sampling. *Journal of American Statistical Association*, **85(412)**, 1050-1058.

Thompson, S. K.(1991). Adaptive cluster sampling: Designs with primary and secondary units. *Biometrics*, **47(3)**, 1103-1115.

Thompson, S. K. and Seber, G. A. F. (1996). *Adaptive Sampling*. John Wiley and Sons, Inc., New York.