ISSN 1684-8403
Journal of Statistics
Volume 23, 2016. pp. 141-157

Comparison of Estimators in Case of Low Correlation in Adaptive Cluster Sampling

Muhammad Shahzad Chaudhry[1] *and Muhammad Hanif* [2]

Abstract

In this paper, two Regression-Cum-Exponential Estimators have been proposed to estimate the population mean using population mean of single auxiliary variable in Adaptive Cluster Sampling. The expressions for the Mean Square Error and Bias of the proposed Estimators have been derived. A simulation study has been carried out to demonstrate and compare the efficiencies and precisions of the Estimators. The proposed Estimators have been compared with Ratio, Regression, Exponential Ratio Estimators in usual sampling, the Hansen-Hurwitz and the Ratio Estimators in Adaptive Cluster Sampling when there is low positive correlation between study variable and auxiliary variable.

Keywords

Transformed population, Expected final sample size, Within network variances, Estimated relative bias, Estimated percentage relative efficiency

1. Introduction

In Survey Sampling, to assess the thickness of entities that are clustered has been a main problem. Examples of such clustered population comprise: plants and animals of rare and dying out species, fisheries, drug users, HIV and AIDS patients, etc. The Conventional Random Sampling Designs may be useless and often fail to give samples with meaningful information for such population.

[1] Government College University, Lahore, Pakistan
Email: almoeed@hotmail.com
[2] National College of Business Administration and Economics, Lahore, Pakistan
Email: drmianhanif@gmail.com

The Adaptive Cluster Sampling (ACS) method is suitable for the rare and clustered population. In ACS an initial sample is selected by a Conventional Sampling Design then the surrounding area of each unit selected in the first sample is measured. A condition *C* is defined to take in a unit in the sample. All adjacent units is added and investigated if the predefined condition *C* (usually $C_y > 0$, where *Y* is the study variable) is fulfilled and this procedure keep on. This process stops when a new unit does not meet the condition. All the units studied (including the initial sample) composed the final sample. The set consisting of those units that met the condition is known as a network. The units that do not satisfy the condition are known as edge units. A cluster is a combination of network units with connected edge units.

Auxiliary information's are used to increase the precision of the Estimators of the population mean of *Y*. Survey statisticians frequently exercise auxiliary information to estimate population parameters; mean and variance. To reduce the error in estimates surveyors are always probing for efficient estimation methods. The development goes on in various forms of Estimators. The available Estimators, in the case of ACS are relatively simple and most of these are based upon information of single auxiliary variable.

Thompson (1990) first proposed the idea of the ACS scheme and introduced modified Hansen-Hurwitz (1943) and Horvitz-Thompson (1952) type Estimators and Rao-Blackwell versions of Estimators. As an indispensable issue related to sampling technique, the sampling efficiency was examined with examples. This article launched a variety of subsequent research efforts on ACS. However, how those adaptive design factors including the predefined condition (or magnitude of the critical value), the definition of the neighborhood affects the efficiency of ACS in comparison with the non-adaptive design (simple random sampling), and possible challenges arising from case to case are not concretely touched.

Dryver (2003) found that ACS performs well in a uni-variate setting. The efficiency of ACS depends on the relationship of the variables with one another in a multivariate setting. The simulation on real data of blue-winged and red-winged results shows that Horvitz-Thompson Type Estimator was the most efficient Estimator using the condition of one type of duck to estimate that type of duck. The Simple Random Sample can be more efficient than an adaptive one when the prediction was on one type of duck and the condition was on another type of duck. The ACS is more efficient than SRS when the condition was a function of two types of ducks. For highly correlated variables the ACS performs well for the parameters of interest.

Chao (2004) proposed the Ratio Estimator in ACS and showed that it produces better estimation results than the original Estimator of Adaptive Cluster Sampling, and a Ratio Estimator under a comparable Conventional sampling design. This article is meant to be an initial investigation of the utilization of the auxiliary information in ACS.

Dryver and Chao (2007) discussed the Classical Ratio Estimator in ACS and proposed two new Ratio Estimators under ACS, "one of which is unbiased for ACS designs. The result shows that the proposed Estimators can be considered as a robust alternative of the conventional ratio Estimator, especially when the correlation between the variable of interest and the auxiliary variable is not high enough for conventional ratio Estimator to have satisfactory performance".

1.1 Adaptive Cluster Sampling Process: Consider a finite population of *N* units is labelled 1,2,3,…,N and denoted as u = $\{u_1, u_2, \ldots, u_N\}$. Consider a small initial sample s_o of size *n* with $n < N$ which is selected by simple random sample without replacement (SRSWOR). The first sample is chosen by traditional sampling process in an ACS procedure and then the predefined neighboring units for all the units of the first sample is considered for a particular condition C, say $y > 0$. If any of the units in the initial sample satisfy condition C ($y_i \leq c$ or $y_i \geq c$ where c is a constant), there neighboring units are added to the sample and observed. In general, if the characteristic of interest is found at a particular area then we continue to locate around that area for more information. Further, if any neighboring unit satisfies the condition then its neighborhoods are also sampled and the process goes on. This iterative process stops when the new unit does not satisfy condition *C*. The neighborhood can be defined by social and institutional relationships between units and can be decided in two ways. The first-order neighborhood consists of the sampling unit itself and four adjacent units denoted as east, west, north, and south. The second-order neighborhood (Figure 1) consists of first-order neighboring units and the units including northeast, northwest, southeast, and southwest units. There are in total eight neighborhood quadrates including the first-order neighborhood and the second order neighborhood. All the units including the initial sample composed the final sample.

A network is a set consisting of units that satisfy the specified condition (usually y =1). A unit that does not satisfy the predefined condition in the first sample is the network of size one. The edge units are those which do not satisfy the specified criteria. A cluster is a mixture of network units with associated edge units. Clusters may have overlapping edge units. The networks do not have common elements such that the union of the networks becomes the population. Thus, it is possible to partition the population of all units into a set of exclusive and complete networks. The networks related to clusters can be denoted by A_1, A_2, A_3 ,..., A_n or they can be shown with darker lines around the quadrates. These may be shaded as well. The edge units can be denoted with open circles (○). The entire region is partitioned into N rectangular or square units of equal size that can be set in a lattice system. The rectangular or square units are called quadrates. The units ($u_1,u_2,\ldots,u_N$) form a disjoint and exhaustive partition of the entire area so that units labels (1,2,…,N) categories the position of N quadrates. In ACS the population is measured in terms of quadrates only. The unit i which is related to the study variable y_i has a population vector of y-values, y = ($y_1, y_2, y_3, \ldots, y_N$). Now consider an example to understand the ACS procedure using first-order neighborhoods and the condition C that $y_i \geq 1$.

Figure 2 shows that the region is partitioned into N =50 quadrates of equal size and that the population is divided into three clusters of size greater than one. Figure 3 shows that the three networks are denoted with A_1, A_2, and A_3 and the edge units are denoted with open circles (○). According to the network symmetry assumption any unit in A_i , a network will lead to the selection of all units in that network A_i .

There are 35 networks of size one as any unit that does not meet the condition is a network of size one. There are three clusters. Each cluster can be decomposed into a network (that satisfies C) and individual networks of size 1 (that do not satisfy C) i.e. edge units. Thus, network A_1 has 6 units with 8 edge units. Network A_2 has 4 units with 9 edge units. Network A_3 has 5 units with 7 edge units. Thus, the cluster containing A_1 has 6 + 8 =14 units. The cluster containing A_2 has 4 + 9 =13 units and cluster containing A_3 has 5 + 7 =12 units. Clusters are not necessarily disjoint and may have common edge units. The clusters containing A_2 and A_3 have one overlapping edge unit.

Figure 4 shows that the cluster containing A_1 and the edge units are shaded, the overlapping edge unit of A_2 and A_3 is also shaded.

2. Some Estimators in Simple Random Sampling

Let N be the total number of units in the population. A random sample of size n is selected by using Simple Random Sampling WithOut Replacement. The study variable and auxiliary variable are denoted by y and x with their population means $\bar{Y}$ and $\bar{X}$, population standard deviation S_y and S_x, coefficient of variation C_y and C_x respectively. Also ρ_{xy} represent population correlation coefficient between X and Y, and $\theta = \frac{1}{n} - \frac{1}{N}$.

Cochran (1940) and Cochran (1942) proposed the Classical Ratio and Regression Estimators

$$t_1 = \bar{y}\left[\frac{\bar{X}}{\bar{x}}\right] \tag{2.1}$$

$$t_2 = \bar{y} + \beta_{yx}\left(\bar{X} - \bar{x}\right) \tag{2.2}$$

The Mean Square Error (MSE) of the Estimators of eq. (2.1) and eq. (2.2) are

$$MSE(t_1) = \theta\bar{Y}^2\left[C_y^2 + C_x^2 - 2\rho_{xy}C_xC_y\right] \tag{2.3}$$

$$MSE\left(t_2\right) = \theta\bar{Y}^2C_y^2\left(1-\rho_{xy}^2\right) \tag{2.4}$$

respectively.

Bahl and Tuteja (1991) proposed the Exponential Ratio and Exponential Product Estimators to estimate the population mean.

$$t_3 = \bar{y}\exp\left[\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right] \tag{2.5}$$

The Mean Square Error and Bias of the Exponential Ratio Estimator t_3 are

$$\text{MSE}(t_3) \approx \theta\bar{Y}^2\left[C_y^2 + \frac{C_x^2}{4} - \rho_{xy}C_xC_y\right] \tag{2.6}$$

$$Bias(t_3) \approx \theta\bar{Y}\left[\frac{3}{8}C_x^2 - \frac{\rho_{xy}C_xC_y}{2}\right] \tag{2.7}$$

3. Some Estimators in Adaptive Cluster Sampling

Let an initial sample of n units is selected with a Simple Random Sample WithOut Replacement (SRSWOR). The neighboring units are added to the sample if the y-value of a sampled unit meets a certain condition *C*, for example $y_i \geq c$ where *c* is a constant. If A_i is the network including unit *i*, then unit *i* will be included in the final sample either, any unit of network A_i is selected as part of the initial Simple Random Sample or any unit of a network for which unit *i* is an edge unit is selected. Suppose a finite population of size *N* is labeled as *1,2,3,…,N* . Let the auxiliary variable x_i be correlated with the variable of interest y_i, such that $y = \{y_1, y_2, \ldots . y_N\}$ and $x = \{x_1, x_2, \ldots . x_N\}$.

Let w_{yi} and w_{xi} denotes the average y-value and average x-value in the network which includes unit *i* such that, $w_{yi} = \frac{1}{m_i} \sum_{j \in A_i} y_j$ and $w_{xi} = \frac{1}{m_i} \sum_{j \in A_i} x_j$ respectively, where m_i is the size of the cluster A_i. ACS can be considered as Simple Random Sampling WithOut Replacement when the averages of networks are considered (Dryver and Chao, 2007 and Thompson, 2002). So, consider the notations $\bar{w}_y$ and $\bar{w}_x$ are the sample means of the study and auxiliary variables in the transformed population respectively, such that, $\bar{w}_y = \frac{1}{n} \sum_{i=1}^{n} w_{yi}$ and $\bar{w}_x = \frac{1}{n} \sum_{i=1}^{n} w_{xi}$. Consider C_{wy} and C_{wx} represents population coefficient of variations of the study and auxiliary variables respectively and ρ_{wxwy} represent population correlation coefficient between w_x and w_y in the ACS i.e. when average of the network assumed. Let us define,

$$\bar{e}_{wy} = \frac{\bar{w}_y - \bar{Y}}{\bar{Y}} \text{ and } \bar{e}_{wx} = \frac{\bar{w}_x - \bar{X}}{\bar{X}}. \tag{3.1}$$

where,

$\bar{e}_{wy}$ and $\bar{e}_{wx}$ are the Relative Sampling Errors of the study and auxiliary variables respectively, such that,

$$E\left(\bar{e}_{wy}\right) = E\left(\bar{e}_{wx}\right) = 0 \text{ and } E\left(\bar{e}_{wx}\bar{e}_{wy}\right) = \theta\rho_{wxwy}C_{wx}C_{wy} \tag{3.2}$$

$$E\left(\bar{e}_{wy}^2\right) = \theta C_{wy}^2 \quad \text{and} \quad E\left(\bar{e}_{wx}^2\right) = \theta C_{wx}^2 \tag{3.3}$$

Thompson (1990) developed an Unbiased Estimator for population mean $\bar{Y}$ in ACS based on a modification of the Hansen-Hurwitz Estimator which can be used when sampling is with replacement or without replacement. Units that do not satisfy *C* are ignored if they are not in the initial sample. In terms of the *n* networks (which may not be unique), intersected by the initial sample.

$$t_4 = \frac{1}{n}\sum_{i=1}^{n} w_{yi} = \bar{w}_y, \quad (3.4)$$

where,

$w_{yi} = \frac{1}{m_i}\sum_{j \in A_i} y_j$ is the mean of the m_i observations in A_i. The variance of t_4 is,

$$Var(t_4) = \frac{\theta}{N-1}\sum_{i=1}^{N}\left(w_{yi} - \bar{Y}\right)^2 \quad (3.5)$$

Dryver and Chao (2007) proposed a Modified Ratio Estimator for the population mean keeping in view Adaptive Cluster Sampling.

$$t_5 = \left[\frac{\sum_{i \in s_0} w_{yi}}{\sum_{i \in s_0} w_{xi}}\right]\bar{X} = \hat{R}\bar{X} \quad (3.6)$$

The Mean Square Error of t_5 is

$$MSE(t_5) = \frac{\theta}{N-1}\sum_{i=1}^{N}\left(w_{yi} - Rw_{xi}\right)^2 \quad (3.7)$$

where,

R is the population ratio between w_{xi} and w_{yi} in the transformed population.

Chutiman (2013) proposed a Modified Regression Estimator for the population mean of the study variable in Adaptive Cluster Sampling.

$$t_6 = \bar{w}_y + \beta_w(\bar{X} - \bar{w}_x) \quad (3.8)$$

where,

$$\beta_w = \frac{S_{wxwy}}{S_{wx}^2} = \frac{\rho_{wxwy}S_{wx}S_{wy}}{S_{wx}^2} = \frac{\bar{Y}}{\bar{X}}\frac{\rho_{wxwy}C_{wy}}{C_{wx}} \quad (3.9)$$

The approximate Mean Square Error of t_6 is

$$MSE(t_6) = \theta S_{wy}^2(1-\rho_{wxwy}^2) = \theta\bar{Y}C_{wy}^2(1-\rho_{wxwy}^2) \quad (3.10)$$

Shahzad and Hanif (2015) proposed a Generalized Exponential Estimator for the population mean in ACS using two auxiliary variables.

$$t_{GE} = \bar{w}_y \exp\left[\alpha\frac{\bar{X} - \bar{w}_x}{\bar{X} + (a-1)\bar{w}_x} + \lambda\frac{\bar{Z} - \bar{w}_z}{\bar{Z} + (b-1)\bar{w}_z}\right] \quad (3.11)$$

For $\alpha = 1,\ \lambda = 0,\ and\ a = 2$, the Estimator t_{GE} may be obtained as

$$t_7 = \bar{w}_y \exp\left[\frac{\bar{X} - \bar{w}_x}{\bar{X} + \bar{w}_x}\right] \tag{3.12}$$

The Bias and Mean Square Error of the Estimator t_7 are

$$\text{Bias}(t_7) \approx \theta\bar{Y}\left(\frac{3C_{wx}^2}{8} - \frac{\rho_{wxwy}C_{wx}C_{wy}}{2}\right) \tag{3.13}$$

$$\text{MSE}(t_7) = E\left(t_7 - \bar{Y}\right)^2 \approx \theta\bar{Y}^2\left(C_{wy}^2 + \frac{C_{wx}^2}{4} - \rho_{wxwy}C_{wx}C_{wy}\right) \tag{3.14}$$

4. Proposed Estimators in Adaptive Cluster Sampling

Following the Bahl and Tuteja (1991) and Shahzad and Hanif (2015) the Proposed Modified Regression-Cum-Exponential Ratio Estimators in ACS with one auxiliary variable are

$$t_8 = \bar{w}_y + \beta\exp\left(\frac{\bar{X} - \bar{w}_x}{\bar{X} + \bar{w}_x}\right) \tag{4.1}$$

$$t_9 = t_4 + \beta t_7 = \bar{w}_y + \beta\bar{w}_y\exp\left(\frac{\bar{X} - \bar{w}_x}{\bar{X} + \bar{w}_x}\right) \tag{4.2}$$

4.1 Bias and Mean Square Error of Estimator t_8: The Estimator eq. (4.1) may be written as

$$t_8 = \bar{Y}\left(1 + \bar{e}_{wy}\right) + \beta\exp\left[\frac{\bar{X} - \bar{X}(1 + \bar{e}_{wx})}{\bar{X} + \bar{X}(1 + \bar{e}_{wx})}\right] \tag{4.1.1}$$

$$t_8 - \bar{Y} = \bar{Y}\bar{e}_{wy} + \beta\exp\left[\frac{-\bar{e}_{wx}}{2}\left(1 + \frac{\bar{e}_{wx}}{2}\right)^{-1}\right] \tag{4.1.2}$$

$$t_8 - \bar{Y} = \bar{Y}\bar{e}_{wy} + \beta\exp\left[\frac{-\bar{e}_{wx}}{2} + \frac{\bar{e}_{wx}^2}{4}\right] \tag{4.1.3}$$

$$t_8 - \bar{Y} = \bar{Y}\bar{e}_{wy} + \beta\left[1 - \frac{\bar{e}_{wx}}{2} + \frac{\bar{e}_{wx}^2}{4} + \frac{\bar{e}_{wx}^2}{8}\right] \tag{4.1.4}$$

$$t_8 - \bar{Y} = \bar{Y}\bar{e}_{wy} + \beta\left[1 - \frac{\bar{e}_{wx}}{2} + \frac{3\bar{e}_{wx}^2}{8}\right] \tag{4.1.5}$$

Applying expectations on both sides of eq. (4.1.5), and using notations eq. (3.2), we get,

$$\text{Bias}(t_8) = E\left(t_8 - \bar{Y}\right) = \beta\left[1 + \frac{3\theta C_{wx}^2}{8}\right] \quad (4.1.6)$$

In order to derive Mean Square Error of eq. (4.1), we have eq. (4.1.3) by ignoring the term degree 2 or greater as

$$t_8 - \bar{Y} = \bar{Y}\bar{e}_{wy} + \beta\left[1 - \frac{\bar{e}_{wx}}{2}\right] \quad (4.1.7)$$

Taking square and expectations on the both sides of eq. (4.1.7), and using notations eq. (3.2 and 3.3),

$$\text{MSE}(t_8) = E(t_8 - \bar{Y})^2 \approx \bar{Y}^2\theta C_{wy}^2 + \beta^2 + \beta^2\frac{\theta C_{wx}^2}{4} - \bar{Y}\beta\theta\rho_{wxwy}C_{wx}C_{wy}. \quad (4.1.8)$$

Differentiate w.r.t to β and equate to zero, we get,

$$\beta = \frac{\bar{Y}\theta\rho_{wxwy}C_{wx}C_{wy}}{2\left(1 + \frac{\theta C_{wx}^2}{4}\right)}. \quad (4.1.9)$$

Substitute eq. (4.1.9) into eq. (4.1.8), we get minimum Mean Square Error,

$$MSE(t_8)_{\min} \approx \bar{Y}^2\theta C_{wy}^2\left[1 - \frac{\theta\rho_{wxwy}^2 C_{wx}^2}{2\left(1 + \frac{\theta C_{wx}^2}{4}\right)}\right] \quad (4.1.10)$$

4.2 Bias and Mean Square Error of Estimator t_9***:*** The Estimator eq. (4.2) may be written as

$$t_9 = \bar{Y}\left(1 + \bar{e}_{wy}\right) + \beta\bar{Y}\left(1 + \bar{e}_{wy}\right)\exp\left[\frac{\bar{X} - \bar{X}(1 + \bar{e}_{wx})}{\bar{X} + \bar{X}(1 + \bar{e}_{wx})}\right] \quad (4.2.1)$$

$$t_9 = \bar{Y} + \bar{Y}\bar{e}_{wy} + \beta\bar{Y}(1 + \bar{e}_{wy})\exp\left[\frac{-\bar{e}_{wx}}{2} + \frac{\bar{e}_{wx}^2}{4}\right] \quad (4.2.2)$$

$$t_9 - \bar{Y} = \bar{Y}\bar{e}_{wy} + \beta\bar{Y}(1 + \bar{e}_{wy})\left[1 - \frac{\bar{e}_{wx}}{2} + \frac{\bar{e}_{wx}^2}{4} + \frac{\bar{e}_{wx}^2}{8}\right] \quad (4.2.3)$$

$$t_9 - \bar{Y} = \bar{Y}\bar{e}_{wy} + \beta\bar{Y}\left[1 - \frac{\bar{e}_{wx}}{2} + \frac{\bar{e}_{wx}^2}{4} + \frac{\bar{e}_{wx}^2}{8} + \bar{e}_{wy} - \frac{\bar{e}_{wx}\bar{e}_{wy}}{2}\right] \quad (4.2.4)$$

Applying expectations on both sides of eq. (4.2.4), we get,

Bias (t_9) = $E\left(t_9-\bar{Y}\right)=\beta\bar{Y}\left[1+\frac{3\theta C_{wx}^2}{8}-\frac{\theta\rho_{wxwy}C_{wx}C_{wy}}{2}\right]$ (4.2.3)

In order to derive mean square error of eq. (4.2), we have eq. (4.2.2) by ignoring the term degree 2 or greater as

$$t_9-\bar{Y}=\bar{Y}\bar{e}_{wy}+\beta\bar{Y}\left(1+\bar{e}_{wy}\right)\exp\left(-\frac{\bar{e}_{wx}}{2}\right) \quad (4.2.4)$$

$$t_9-\bar{Y}=\bar{Y}\bar{e}_{wy}+\beta\bar{Y}+\beta\bar{Y}\bar{e}_{wy}-\frac{\beta\bar{Y}\bar{e}_{wx}}{2} \quad (4.2.5)$$

Taking square and expectations on the both sides of eq. (4.2.5)

$$MSE(t_9)=(t_9-\bar{Y})^2\approx\bar{Y}^2\theta C_{wy}^2+\beta^2\bar{Y}^2+\beta^2\bar{Y}^2\theta C_{wy}^2+\frac{\beta^2\bar{Y}^2\theta C_{wx}^2}{4}$$
$$+2\beta\theta\bar{Y}^2C_{wy}^2-\beta\bar{Y}^2\theta\rho_{wxwy}C_{wx}C_{wy}-\beta^2\bar{Y}^2\theta\rho_{wxwy}C_{wx}C_{wy}. \quad (4.2.6)$$

Differentiate w.r.t to β and equate to zero, we get,

$$\beta=\frac{-\theta C_{wy}^2+\frac{\theta\rho_{wxwy}C_{wx}C_{wy}}{2}}{\left(1+\theta C_{wy}^2+\frac{\theta C_{wx}^2}{4}-\theta\rho_{wxwy}C_{wx}C_{wy}\right)}. \quad (4.2.7)$$

Substitute eq. (4.2.7) into eq. (4.2.6), we get minimum Mean Square Error,

$$MSE(t_9)_{\min}\approx\bar{Y}^2\theta C_{wy}^2\left[1-\frac{\theta C_{wy}^2-\frac{\theta\rho_{wxwy}C_{wx}C_{wy}}{2}}{\theta C_{wy}^2\left(1+\theta C_{wy}^2+\frac{\theta C_{wx}^2}{4}-\theta\rho_{wxwy}C_{wx}C_{wy}\right)}\right]. \quad (4.2.8)$$

5. Results and Discussion

To compare the efficiency of proposed Estimators with the other Estimators, a population is used and performed simulations for the thorough study. The condition C for added units in the sample is $y > 0$. The y-values are obtained and averaged for keeping the sample network according to the condition and for each sample network *x*-values are obtained and averaged. For the simulation study ten thousands iteration was run for each Estimator to get accuracy estimates with the simple random sampling without replacement and the initial sample sizes of 5, 10, 15, 20 and 25.

The expected final sample size is sum of the probabilities of inclusion of all quadrates and it varies from sample to sample in ACS. Let, *E(v)* denotes the expected final sample size in ACS. For the comparison, the sample mean from a SRSWOR based on *E(v)* has variance using the usual formula.

$$Var(\bar{y}) = \frac{\sigma^2 (N - E(v))}{NE(v)} \quad (5.1)$$

The estimated Mean Square Error of the estimated mean is

$$\hat{MSE}(t_*) = \frac{1}{r}\sum_{i=1}^{r}(t_* - \bar{Y})^2 \quad (5.2)$$

where,

t_* is the value for the relevant Estimator for sample *i*, and *r* is the number of iterations. The Estimated Relative Bias is defined as

$$\hat{RBias}(t_*) = \frac{\frac{1}{r}\sum_{i=1}^{r}(t_*) - \bar{Y}}{\bar{Y}} \quad (5.3)$$

The percentage relative efficiency is,

$$PRE = \frac{Var(\bar{y})}{\hat{MSE}(t_*)} \times 100 \quad (5.4)$$

5.1 Population: This real population contains ring-necked ducks (RND) and blue-winged teal (BWT) data collected by Smith et al., (1995) are counts of two species of waterfowl in 50 100-km^2 quadrates in central Florida. The ring-necked ducks (Table 1) is taken as the study variable the blue-winged teal (Table 2) is taken as auxiliary variable. There found a very low correlation 0.0978 between the both types of fowl and this correlation increases to 0.2265 in the transformed population (Table 3 and 4). Thus, there is a low correlation between the sampling unit level and high correlation at the network (region) level. Dryver and Chao (2007) showed that usual Estimators in SRSWOR perform better than ACS Estimators for strong correlation at unit level but performs worse when having the strong correlation at network level.

The overall variance of the study variable is 3972532 and for auxiliary variable the variance is 3716168 while in the transformed population these variances reduce to 747163.9 and 3716156, respectively. The within network variance of the study variable for the network (4000, 20, 13500, 200, 234, 75, 1335, 4, 97) is 19755495 and for the corresponding values of the auxiliary variable (0, 0, 0, 3, 5, 24, 14, 2) the within network variance is 69.25. The overall variances are found to low as compare to the within network variances in the study variable population. ACS is preferable than the Conventional sampling if the within network variances are large enough as compare to overall variance (Dryver and Chao 2007).

6. Conclusion

The Estimated Relative Bias (Table 5) of the Regression Estimators in ACS goes quickly to zero when sample size increases as compare to the other Estimators in usual sampling methods. The Percentage Relative Efficiency (Table 6) of the ACS Estimators is much higher than the SRS Estimators. The Regression Estimator in ACS has maximum Percentage Relative Efficiency for the initial sample size and starts increasing for comparable sample sizes, while proposed Exponential Regression Estimator t_8 also has higher PRE than the Ratio Estimator and Exponential Ratio Estimator in ACS. Thus, the Regression and Exponential Regression Estimators in ACS perform much better than the other Conventional Estimators and Ratio and Exponential Ratio Estimators in ACS even though there is a very low positive correlation among the study and the auxiliary variables, under the given conditions.

Dryver and Chao (2007) treated 0/0 as zero for the Ratio Estimator. The usual Ratio Estimator and Ratio Estimator in ACS did not perform and return no value (*) and infinity (**) for the initial sample sizes 5, 10, 15 and 20, respectively. In this simulation study 0/0 is not assumed as 0. The use of Regression and Exponential Regression Estimators is better in ACS than assuming an unlikely assumption for the Ratio Estimator in Adaptive Cluster Sampling. Thus, Regression and Exponential Regression Type Estimators are much suitable and vigorous for patchy, rare and clustered population. The results of relative efficiencies have shown the poor performance of usual Estimators under SRSWOR and the Exponential Ratio Estimators because of the weak correlation between study variable and auxiliary variables at the unit level. For the future research, the case of low negative correlation between the study and auxiliary variables may also be explored in ACS. Moreover, the logarithmic, trigonometric, inverse trigonometric and hyperbolic Estimators may also be studied and explored in Conventional and Adaptive Sampling. Some Estimators are given in Table 8 for the future research.

Acknowledgments

The authors are grateful to Yves G. Berger**,** University of Southampton, UK for the programming advice to simulation studies and valuable suggestions regarding the improvement of this paper.

Southeast	East	Northeast
South	Sampling Unit	North
Southwest	West	Northwest

Figure1: Second-order neighbourhoods (Diagonal cells)

0	0	0	0	0	0	0	0	0	0
0	4	3	0	0	3	0	0	2	0
1	3	4	0	0	6	0	3	5	2
0	2	0	0	1	2	0	0	3	0
0	0	0	0	0	0	0	0	0	0

Figure 2: Population with three clusters

	○	○			○			○	
○	A_1	A_1	○	○	A_2	○	○	A_3	○
A_1	A_1	A_1	○	○	A_2	○	A_3	A_3	A_3
○	A_1	○	○	A_2	A_2	○	○	A_3	○
	○			○	○			○	

Figure 3: Networks with edge units

0	0	0	0	0	0	0	0	0	0
0	4	3	0	0	3	0	0	2	0
1	3	4	0	0	6	0	3	5	2
0	2	0	0	1	2	0	0	3	0
0	0	0	0	0	0	0	0	0	0

Figure 4: Cluster containing $A_{1,}$ Common edge unit of A_2 and A_3

Table1: Ring-necked ducks data (Smith et al., 1995) as study variable(y) for population

0	200	200	75	0	0	0	0	675	0
4000	13500	234	1335	4	0	0	35	0	55
0	0	0	0	97	0	4	0	1815	0
0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	1283

Table2: Blue-winged teal data (Smith et al., 1995) as auxiliary variable(x) for population

0	0	3	5	0	0	0	0	0	0
0	0	0	24	14	0	0	10	103	0
0	0	0	0	2	3	2	0	13639	1
0	0	0	0	0	0	0	0	14	122
0	0	0	0	0	0	2	0	0	177

Table3: Average of the network (wy) of study variable(y) for population

0	2162.78	2162.78	2162.78	0	0	0	0	675	0
2162.78	2162.78	2162.78	2162.78	2162.78	0	0	35	0	55
0	0	0	0	2162.78	0	4	0	1815	0
0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	1283

Table4: Average of the network (wx) of auxiliary variable for population

0	5.33	5.33	5.33	0	0	0	0	0	0
5.33	5.33	5.33	5.33	5.33	0	0	10	103	0
0	0	0	0	5.33	3	2	0	13639	1
0	0	0	0	0	0	0	0	14	122
0	0	0	0	0	0	2	0	0	177

Table 5: Estimated Relative Bias of the population for different sample sizes

n	E(v)	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
5	16.14	*	-0.016	1.245	0.003	*	-0.003	1.233	0.015	0.230
10	24.31	*	0.012	0.999	0.006	*	0.006	1.040	0.003	0.211
15	29.33	**	-0.011	0.856	-0.001	**	-0.001	0.885	0.001	0.195
20	33.17	**	-0.012	0.710	-0.003	29.805	0.004	0.723	0.001	0.174
25	36.54	20.667	-0.009	0.594	-0.002	20.077	-0.003	0.586	0.005	0.158

Table 6: Percentage Relative Efficiencies for the population for the Estimators based on E(v)

	Simple Random Sampling				Adaptive Cluster Sampling					
E(v)	$\bar{y}$	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
16.14	100	*	101.12	14.98	123.12	*	133.92	14.47	124.44	78.60
24.31	100	*	100.66	16.01	139.26	*	147.81	12.90	140.85	90.20
29.33	100	0	100.93	15.92	161.11	0.00	172.29	12.12	163.94	100.39
33.17	100	0	101.08	16.09	179.03	0.01	192.51	11.55	182.46	108.95
36.54	100	0	101.02	15.83	194.83	0.01	205.33	10.82	199.06	117.60

Table 7: Descriptive measures of the population

$\bar{X} = 282.42$	$\sigma_x^2 = 3716168$	$C_x = 682.5779$	$\rho_{xy} = 0.0979$
$\bar{Y} = 466.66$	$\sigma_y^2 = 3972532$	$C_y = 427.1035$	$\rho_{wxy} = 0.2265$
$\bar{w}_x = 282.42$	$\sigma_{wx}^2 = 3716156$	$C_{wx} = 682.577$	$N = 50$
$\bar{w}_y = 466.64$	$\sigma_{wy}^2 = 747163.9$	$C_{wy} = 185.2362$	$C > 0$

Table 8: Some Proposed Estimators for future research

SRS	ACS	SRS	ACS	SRS	ACS
$\bar{y}\ln\left[\frac{\bar{X}}{\bar{x}}\right]$	$\bar{w}_y\ln\left[\frac{\bar{X}}{\bar{w}_x}\right]$	$\bar{y}\sin\left[\frac{\bar{X}}{\bar{x}}\right]$	$\bar{w}_y\sin\left[\frac{\bar{X}}{\bar{w}_x}\right]$	$\bar{y}\sin^{-1}\left[\frac{\bar{X}}{\bar{x}}\right]$	$\bar{w}_y\sin\left[\frac{\bar{X}}{\bar{w}_x}\right]$
$\bar{y}\ln\left[\frac{\bar{X}}{\bar{x}}-1\right]$	$\bar{w}_y\ln\left[\frac{\bar{X}}{\bar{w}_x}-1\right]$	$\bar{y}\sin\left[\frac{\bar{X}}{\bar{x}}-1\right]$	$\bar{w}_y\sin\left[\frac{\bar{X}}{\bar{w}_x}-1\right]$	$\bar{y}\sin^{-1}\left[\frac{\bar{X}}{\bar{x}}-1\right]$	$\bar{w}_y\sin\left[\frac{\bar{X}}{\bar{w}_x}-1\right]$
$\bar{y}\ln\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]$	$\bar{w}_y\ln\left[\frac{\bar{X}-\bar{w}_x}{\bar{X}+\bar{w}_x}\right]$	$\bar{y}\sin\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]$	$\bar{w}_y\sin\left[\frac{\bar{X}-\bar{w}_x}{\bar{X}+\bar{w}_x}\right]$	$\bar{y}\sin^{-1}\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]$	$\bar{w}_y\sin^{-1}\left[\frac{\bar{X}-\bar{w}_x}{\bar{X}+\bar{w}_x}\right]$
$\bar{y}\sinh\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]$	$\bar{w}_y\sinh\left[\frac{\bar{X}-\bar{w}_x}{\bar{X}+\bar{w}_x}\right]$	$\bar{y}\sinh^{-1}\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]$	$\bar{w}_y\sinh^{-1}\left[\frac{\bar{X}-\bar{w}_x}{\bar{X}+\bar{w}_x}\right]$	$\bar{y}\cosh\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]$	$\bar{w}_y\cosh^{-1}\left[\frac{\bar{X}-\bar{w}_x}{\bar{X}+\bar{w}_x}\right]$

In General SRS	In General ACS
$\bar{y}\ln\left[\alpha\frac{\bar{X}-\bar{x}}{\bar{X}+(a-1)\bar{x}}+\lambda\frac{\bar{Z}-\bar{z}}{\bar{Z}+(b-1)\bar{z}}\right]$	$\bar{w}_y\ln\left[\alpha\frac{\bar{X}-\bar{w}_x}{\bar{X}+(a-1)\bar{w}_x}+\lambda\frac{\bar{Z}-\bar{w}_z}{\bar{Z}+(b-1)\bar{w}_z}\right]$
$\bar{y}TRG\left[\alpha\frac{\bar{X}-\bar{x}}{\bar{X}+(a-1)\bar{x}}+\lambda\frac{\bar{Z}-\bar{z}}{\bar{Z}+(b-1)\bar{z}}\right]$	$\bar{w}_yTRG\left[\alpha\frac{\bar{X}-\bar{w}_x}{\bar{X}+(a-1)\bar{w}_x}+\lambda\frac{\bar{Z}-\bar{w}_z}{\bar{Z}+(b-1)\bar{w}_z}\right]$
$\bar{y}INVTRG\left[\alpha\frac{\bar{X}-\bar{x}}{\bar{X}+(a-1)\bar{x}}+\lambda\frac{\bar{Z}-\bar{z}}{\bar{Z}+(b-1)\bar{z}}\right]$	$\bar{w}_yINVTRG\left[\alpha\frac{\bar{X}-\bar{w}_x}{\bar{X}+(a-1)\bar{w}_x}+\lambda\frac{\bar{Z}-\bar{w}_z}{\bar{Z}+(b-1)\bar{w}_z}\right]$
$\bar{y}TRGH\left[\alpha\frac{\bar{X}-\bar{x}}{\bar{X}+(a-1)\bar{x}}+\lambda\frac{\bar{Z}-\bar{z}}{\bar{Z}+(b-1)\bar{z}}\right]$	$\bar{w}_yTRGH\left[\alpha\frac{\bar{X}-\bar{w}_x}{\bar{X}+(a-1)\bar{w}_x}+\lambda\frac{\bar{Z}-\bar{w}_z}{\bar{Z}+(b-1)\bar{w}_z}\right]$
$\bar{y}INVHTRG\left[\alpha\frac{\bar{X}-\bar{x}}{\bar{X}+(a-1)\bar{x}}+\lambda\frac{\bar{Z}-\bar{z}}{\bar{Z}+(b-1)\bar{z}}\right]$	$\bar{w}_yINVHTRG\left[\alpha\frac{\bar{X}-\bar{w}_x}{\bar{X}+(a-1)\bar{w}_x}+\lambda\frac{\bar{Z}-\bar{w}_z}{\bar{Z}+(b-1)\bar{w}_z}\right]$

where,
TRG = Trigonometric functions
INVTRG = Inverse Trigonometric functions
TRG H = Hyperbolic Trigonometric functions
INVHTRG = Inverse Hyperbolic Trigonometric functions

References

1. Bahl, S. and Tuteja, R. K. (1991). Ratio and Product Type Exponential Estimator. *Information and Optimization Sciences*, **12**, 159-163.
2. Chao, C. T. (2004). Ratio estimation on Adaptive Cluster Sampling. *Journal of Chinese Statistical Association*, **42(3)**, 307–327.
3. Chaudhry, M. S. and Hanif, M. (2015). Generalized Exponential-Cum-Exponential Estimator in Adaptive Cluster Sampling. *Pakistan Journal of Statistics and Operation Research*, **11(4)**, 553-574.
4. Chutiman, N. (2013). Adaptive Cluster Sampling using auxiliary variable. *Journal of Mathematics and Statistics*, **9(3)**, 249-255.
5. Chutiman, N. and Kumphon, B. (2008). Ratio Estimator using two auxiliary variables for Adaptive Cluster Sampling. *Thailand Statistician*, **6(2)**, 241-256.
6. Cochran, W. G. (1977). *Sampling Techniques.* John Wiley, New York.
7. Cochran, W. G. (1940). The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *Journal of Agricultural Science*, **30**, 262-275.
8. Dryver, A. L. (2003). Performance of Adaptive Cluster Sampling Estimators in a multivariate setting. *Environmental and Ecological Statistics*, **10**, 107–113.
9. Dryver, A. L. and Chao, C. T. (2007). Ratio Estimators in Adaptive Cluster Sampling. *Environmetrics*, **18**, 607–620.
10. Hansen, M. M. and Hurwitz, W. N. (1943). On the theory of sampling from finite population. *The Annals of Mathematical Statistics*, **14**, 333–362.
11. Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal* of the *American* Statistical *Association*, **47**, 663–685.
12. Smith, D. R., Conroy, M. J. and Brakhage, D. H. (1995). Efficiency of Adaptive Cluster Sampling for estimating density of wintering waterfowl. *Biometrics*, **51**, 777-788.
13. Thompson, S. K. (1990). Adaptive Cluster Sampling. *Journal of American Statistical Association*, **85**, 1050–1059.
14. Thompson, S. K. and Seber, G. A. F. (1996). *Adaptive Sampling*. Wiley, New York.