ISSN 1684-8403
Journal of Statistics
Volume 22, 2015. pp. 257-273

Regression-cum-Exponential Ratio Estimators in Adaptive Cluster Sampling

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

Abstract

In this paper, Regression-cum-Exponential Ratio Estimators have been proposed for estimating the population mean using two auxiliary variables 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 Conventional Ratio, Regression, Exponential Ratio Estimators both under Simple Random Sampling With-Out Replacement (SRSWOR) and Adaptive Cluster Sampling.

Keywords

Transformed population, Simulated population, Expected final sample size, Comparative percentage relative efficiency, Unlikely assumption

1. Introduction

The Adaptive Cluster Sampling (ACS) is suitable and efficient for the rare and clustered population. In Adaptive Cluster Sampling ,the initial sample is selected by a Conventional sampling design such as Simple Random Sampling then the neighborhood of each unit selected is considered if the value of the study variable from the sampled unit meet a pre-defined condition C usually y > 0. The neighboring unit is added and examined if the condition is satisfied and the process continues until the new unit meets the condition.

[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 final sample comprises all the units studied and the initial sample. A network consists of those units that meet the predefined condition. The units that do not meet the specified condition are known as edge units. A cluster is a combination of network and edge units. The neighborhood can be defined by social and institutional relationships between units. The first-order neighborhood consists of the sampling unit itself and four adjacent units denoted as east (above), west (below), north (right), and south (left). The second-order neighborhood consists of first-order neighboring units and the units including northeast, northwest, southeast, and southwest units i.e. diagonal quadrats.

Consider the following example to understand the Adaptive Cluster Sampling process using first-order neighborhoods and the condition C that $y_i > 0$. Figure 1 show that the study region is divided into 50 quadrats and that the population is divided into three clusters according to the condition. There are 35 networks of size one as any unit that does not meet the condition is a network of size one. Each cluster can be divided into a network (that satisfies C) and individual networks of size 1(that do not satisfy C) i.e. edge units. Thus, first network has 6 units with 8 edge units. Network 2 has 5 units with 10 edge units. Network 3 has 4 units with 7 edge units.

0	0	0	0	0	8	0	0	0	0
0	4	3	0	0	3	0	0	2	0
1	3	4	0	0	6	0	0	5	2
0	2	0	0	1	2	0	0	3	0
0	0	0	0	0	0	0	0	0	0

Figure 1: The three networks are shaded according to condition

Thompson (1990) first introduced the idea of the Adaptive Cluster Sampling to estimate the rare and clustered population and proposed four unbiased Estimators in Adaptive Cluster Sampling. Smith et al. (1995) studied the efficiency of Adaptive Cluster for estimating density of wintering water fowl and found that the efficiency is highest as compare to Simple Random Sampling design when the within network variance is close to population variance. Dryver (2003) found that ACS performs well in a Univariate setting. The simulation on real data of blue-winged and red-winged results shows that Horvitz-Thompson Type Estimator (1952) was the most efficient Estimator using the condition of one type of duck to estimate that type of duck. For highly correlated variables the ACS performs well for the parameters of interest. Chao (2004) proposed the Ratio Estimator in

Adaptive Cluster Sampling and showed that it produces better estimation results than the original Estimator of Adaptive Cluster Sampling. Dryver and Chao (2007) suggested the Classical Ratio Estimator in Adaptive Cluster Sampling (ACS) and proposed two new Ratio Estimators under ACS. Chutiman and Kumphon (2008) proposed a Ratio Estimator in Adaptive Cluster Sampling using two auxiliary variables. Chutiman (2013) proposed Ratio Estimators using population coefficient of variation and coefficient of Kurtosis, Regression and difference Estimators by using single auxiliary variable.

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 With-Out 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.

Cochran (1940) and Cochran (1942) proposed the Classical Ratio and Regression Estimators are given by:

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

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

The Mean Square Error (MSE) of the Estimators of equation (2.1) and (2.2) are as follows:

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

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

respectively.

where $\theta = \frac{1}{n} - \frac{1}{N}$

Bahl and Tuteja (1991) proposed the Exponential Ratio Estimator to estimate the population mean is given by:

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

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

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

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

3. Some Estimators in Adaptive Cluster Sampling

Suppose, a finite population of size *N* is labeled as *1,2,3,…,N* and an initial sample of n units is selected with a Simple Random Sample With-Out Replacement. Let, w_{yi} , w_{xi} and w_{zi} denotes the average y-value , average x-value and average z-value in the network which includes unit *i* such that, $w_{yi} = \frac{1}{m_i} \sum_{j \in A_i} y_j$, $w_{xi} = \frac{1}{m_i} \sum_{j \in A_i} x_j$ and $w_{zi} = \frac{1}{m_i} \sum_{j \in A_i} z_j$, respectively.

Adaptive Cluster Sampling can be considered as Simple Random Sample With-Out Replacement when the averages of networks are considered (Dryver and Chao, 2007 and Thompson, 2002). Consider the notations $\bar{w}_y$, $\bar{w}_x$ and $\bar{w}_z$ are the sample means of the study and auxiliary variables in the transformed population respectively, such that, $\bar{w}_y = \frac{1}{n} \sum_{i=n}^{n} w_{yi}$, $\bar{w}_x = \frac{1}{n} \sum_{i=1}^{n} w_{xi}$ and $\bar{w}_z = \frac{1}{n} \sum_{i=1}^{n} w_{zi}$

Consider C_{wy} , C_{wx} and C_{wz} represents population coefficient of variations of the study and auxiliary variables respectively. Let ρ_{wxwy} and ρ_{wzwy} represent population correlation coefficients between w_x and w_y , and, w_z and w_y respectively. Let us define,

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

where

$\bar{e}_{wy}$, $\bar{e}_{wx}$ and $\bar{e}_{wz}$ are the sampling errors of the study and auxiliary variables respectively, such that

$$E\left(\bar{e}_{wy}\right) = E\left(\bar{e}_{wx}\right) = E\left(\bar{e}_{wz}\right) = 0 \tag{3.2}$$

$$E\left(\bar{e}_{wx}\bar{e}_{wy}\right) = \theta \rho_{wxwy} C_{wx} C_{wy} \ \text{and} \ E\left(\bar{e}_{wz}\bar{e}_{wy}\right) = \theta \rho_{wzwy} C_{wz} C_{wy} \tag{3.3}$$

$$E\left(\bar{e}_{wy}^2\right) = \theta C_{wy}^2 \ , \ \ E\left(\bar{e}_{wx}^2\right) = \theta C_{wx}^2 \ \text{and} \ E\left(\bar{e}_{wz}^2\right) = \theta C_{wz}^2 \tag{3.4}$$

Thompson (1990) developed an unbiased Estimator for population mean μ_y in ACS based on a modification of the Hansen-Hurwitz (1943) Estimator which can be used when sampling is with replacement or without replacement. In terms of the *n* networks (which may not be unique) intersected by the initial sample

$$t_4 = \frac{1}{n}\sum_{i=1}^{n} w_{yi} = \bar{w}_y$$

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

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

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

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

The Mean Square Error of t_5 is,

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

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 as follows:

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

where

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

The approximate Mean Squared Error of t_6 is given by:

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

Noor and Hanif (2014) proposed a Modified Exponential Ratio Estimator for the population mean in ACS, following the Bahl and Tuteja (1991) as given by:

$$t_7 = \bar{w}_y \exp\left[\frac{\bar{X} - \bar{w}_x}{\bar{X} + \bar{w}_x}\right] \tag{3.11}$$

The Bias and Mean Square Error of the Estimator are as follows:

$$\text{Bias}(t_7) \approx \theta\bar{Y}\left(\frac{3C_{wx}^2}{8} - \frac{\rho_{wxwy}C_{wx}C_{wy}}{2}\right) \tag{3.12}$$

$$\text{MSE}(t_7) = \approx \theta\bar{Y}^2\left(C_{wy}^2 + \frac{C_{wx}^2}{4} - \rho_{wxwy}C_{wx}C_{wy}\right) \tag{3.13}$$

4. Proposed Estimators in Adaptive Cluster Sampling

Following the Bahl and Tuteja (1991), Chutiman (2013) and Noor and Hanif (2014) the proposed Modified Regression-cum-Exponential Ratio Estimators in Adaptive Cluster Sampling with two auxiliary variables are given by:

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

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

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

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

or

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

or

$$t_8 = \left[\bar{Y} + \bar{Y}\bar{e}_{wy} - \beta\bar{Z}\bar{e}_{wz}\right]\exp\left[\frac{-\bar{e}_{wx}}{2\left(1 + \frac{\bar{e}_{wx}}{2}\right)}\right] \tag{4.1.3}$$

or

$$t_8 = \left[\bar{Y} + \bar{Y}\bar{e}_{wy} - \beta\bar{Z}\bar{e}_{wz}\right]\exp\left[\frac{-\bar{e}_{wx}}{2}\left(1+\frac{\bar{e}_{wx}}{2}\right)^{-1}\right] \tag{4.1.4}$$

Expanding the series up to the first order, rewriting equation (4.1.4) as:

$$t_8 = \left[\bar{Y} + \bar{Y}\bar{e}_{wy} - \beta\bar{Z}\bar{e}_{wz}\right]\exp\left[\frac{-\bar{e}_{wx}}{2}\left(1-\frac{\bar{e}_{wx}}{2}\right)\right] \tag{4.1.5}$$

or

$$t_8 = \left[\bar{Y} + \bar{Y}\bar{e}_{wy} - \beta\bar{Z}\bar{e}_{wz}\right]\exp\left[\frac{-\bar{e}_{wx}}{2}+\frac{\bar{e}_{wx}^2}{4}\right] \tag{4.1.6}$$

Expanding the exponential term up-to the second degree, we get equation (4.1.6) as:

$$t_8 = \left[\bar{Y} + \bar{Y}\bar{e}_{wy} - \beta\bar{Z}\bar{e}_{wz}\right]\left[1-\frac{\bar{e}_{wx}}{2}+\frac{\bar{e}_{wx}^2}{4}+\frac{1}{2}\left(\frac{\bar{e}_{wx}}{2}+\frac{\bar{e}_{wx}^2}{4}\right)^2\right] \tag{4.1.7}$$

or

$$t_8 = \left[\bar{Y} + \bar{Y}\bar{e}_{wy} - \beta\bar{Z}\bar{e}_{wz}\right]\left[1-\frac{\bar{e}_{wx}}{2}+\frac{\bar{e}_{wx}^2}{4}+\frac{\bar{e}_{wx}^2}{8}\right] \tag{4.1.8}$$

or

$$t_8 = \bar{Y}\left[1-\frac{\bar{e}_{wx}}{2}+\frac{3\bar{e}_{wx}^2}{8}\right]+\bar{Y}\bar{e}_{wy}\left[1-\frac{\bar{e}_{wx}}{2}+\frac{3\bar{e}_{wx}^2}{8}\right]-\beta\bar{Z}\bar{e}_{wz}\left[1-\frac{\bar{e}_{wx}}{2}+\frac{3\bar{e}_{wx}^2}{8}\right] \tag{4.1.9}$$

Simplifying, ignoring the terms with degree three or greater

$$t_8 - \bar{Y} = \left[\bar{Y}\bar{e}_{wy} - \frac{\bar{Y}\bar{e}_{wx}}{2} + \frac{3\bar{Y}\bar{e}_{wx}^2}{8} - \frac{\bar{Y}\bar{e}_{wy}\bar{e}_{wx}}{2} - \beta\bar{Z}\bar{e}_{wz} + \frac{\beta\bar{Z}\bar{e}_{wz}\bar{e}_{wx}}{2}\right] \tag{4.1.10}$$

Applying expectations on both sides of equation (4.1.10), and using notations of equation (3.3-3.5), we get:

$$\mathrm{E}(t_8 - \bar{Y}) = \left[\frac{3\theta\bar{Y}C_{wx}^2}{8} - \frac{\bar{Y}\theta\rho_{wxwy}C_{wx}C_{yx}}{2} + \frac{\beta\theta\bar{Z}\rho_{wxwz}C_{wx}C_{wz}}{2}\right] \tag{4.1.11}$$

In order to derive Mean Square Error of equation (4.1), we have equation (4.1.6) by ignoring the term degree 2 or greater as:

$$t_8 = \left[\bar{Y} + \bar{Y}\bar{e}_{wy} - \beta\bar{Z}\bar{e}_{wz}\right]\exp\left[\frac{-\bar{e}_{wx}}{2}\right] \tag{4.1.12}$$

or

$$t_8 = \left[\bar{Y} + \bar{Y}\bar{e}_{wy} - \beta\bar{Z}\bar{e}_{wz}\right]\left[1 - \frac{\bar{e}_{wx}}{2}\right] \tag{4.1.13}$$

or

$$t_8 = \left[\bar{Y} + \bar{Y}\bar{e}_{wy} - \beta\bar{Z}\bar{e}_{wz} - \frac{\bar{Y}\bar{e}_{wx}}{2} - \frac{\bar{Y}\bar{e}_{wy}\bar{e}_{wx}}{2} + \frac{\beta\bar{Z}\bar{e}_{wz}\bar{e}_{wx}}{2}\right] \tag{4.1.14}$$

or

$$(t_8 - \bar{Y}) = \left[\bar{Y} + \bar{Y}\bar{e}_{wy} - \beta\bar{Z}\bar{e}_{wz} - \frac{\bar{Y}\bar{e}_{wx}}{2} - \frac{\bar{Y}\bar{e}_{wy}\bar{e}_{wx}}{2} + \frac{\beta\bar{Z}\bar{e}_{wz}\bar{e}_{wx}}{2}\right] \tag{4.1.15}$$

Taking square and expectations on the both sides of equation (4.1.15), and using notations of equation (3.3-3.5)

$$\text{MSE}(t_8) \approx \left[\bar{Y}^2\theta C_{wy}^2 + \frac{\bar{Y}^2\theta C_{wx}^2}{4} - \beta^2\bar{Z}^2\theta C_{wz}^2 - \bar{Y}^2\theta\rho_{wxwy}C_{wx}C_{wy}\right.$$
$$\left. -2\beta\bar{Y}\bar{Z}\theta\rho_{wzwy}C_{wz}C_{wy} + \beta\bar{Y}\bar{Z}\theta\rho_{wxwz}C_{wx}C_{wz}\right] \tag{4.1.16}$$

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

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

or

$$t_9 = \left[\bar{Y}\left(1 + \bar{e}_{wy}\right) + \beta\left\{-\bar{Z}\bar{e}_{wz}\right\}\right]\exp\left[-\bar{e}_{wx}\right] \tag{4.2.2}$$

Expanding the Exponential term up-to the second degree, we get equation (4.2.2) as:

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

or

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

Simplifying, ignoring the terms with degree three or greater

$$t_9 - \bar{Y} = \left[\bar{Y}\bar{e}_{wy} - \bar{Y}\bar{e}_{wx} + \bar{Y}(\frac{\bar{e}_{wx}^2}{2}) - \bar{Y}\bar{e}_{wy}\bar{e}_{wx} - \beta\bar{Z}\bar{e}_{wz} + \beta\bar{Z}\bar{e}_{wz}\bar{e}_{wx} \right] \quad (4.2.5)$$

Applying expectations on both sides of equation (4.2.5), and using notations of equation (3.3-3.5), we get:

$$\mathrm{E}(t_9 - \bar{Y}) = \left[\theta\bar{Y}\frac{C_{wx}^2}{2} - \theta\bar{Y}\rho_{wxwy}C_{wx}C_{yx} + \beta\theta\bar{Z}\rho_{wxwz}C_{wx}C_{wz} \right] \quad (4.2.6)$$

In order to derive Mean Square Error of equation (4.2), we have equation (4.2.3) by ignoring the term degree 2 or greater as:

$$t_9 = \left[\bar{Y} + \bar{Y}\bar{e}_{wy} - \beta\bar{Z}\bar{e}_{wz} \right]\left[1 - \bar{e}_{wx}\right] \quad (4.2.7)$$

or

$$t_9 = \left[\bar{Y} + \bar{Y}\bar{e}_{wy} - \beta\bar{Z}\bar{e}_{wz} - \bar{Y}\bar{e}_{wx} - \bar{Y}\bar{e}_{wy}\bar{e}_{wx} + \beta\bar{Z}\bar{e}_{wz}\bar{e}_{wx} \right] \quad (4.2.8)$$

or

$$t_9 - \bar{Y} = \left(\bar{Y}\bar{e}_{wy} - \beta\bar{Z}\bar{e}_{wz} - \bar{Y}\bar{e}_{wx} - \bar{Y}\bar{e}_{wy}\bar{e}_{wx} + \beta\bar{Z}\bar{e}_{wz}\bar{e}_{wx} \right) \quad (4.2.9)$$

Taking square and expectations on the both sides of equation (4.2.9), and using notations of equation (3.3-3.5), we get:

$$\mathrm{MSE}(t_9) \approx \left[\begin{array}{l} \bar{Y}^2\theta C_{wy}^2 + \bar{Y}^2\theta C_{wx}^2 + \beta^2\bar{Z}^2\theta C_{wz}^2 - 2\beta\overline{YZ}\theta\rho_{wzwy}C_{wz}C_{wy} \\ -2\bar{Y}^2\theta\rho_{wxwy}C_{wx}C_{wy} + 2\beta\overline{YZ}\theta\rho_{wxwz}C_{wx}C_{wz} \end{array} \right] \quad (4.2.10)$$

5. Conclusion

To compare the efficiency of proposed Estimators with the other Estimators, a simulated population is used and performed simulations for the 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 10,20,30,40 and 50.

In the ACS the expected final sample size varies from sample to sample. Let, E(v) denotes the expected final sample size in ACS, is sum of the probabilities of inclusion of all quadrats. In the ACS the expected final sample size varies from sample to sample. 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 Squared Error of the estimated mean is given by:

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

Where t_* is the value for the relevant Estimator for sample i, and r is the number of iterations.

The Percentage Relative Efficiency is given by:

$$PRE = \frac{Var(\bar{y})}{\hat{MSE}(t_*)} \times 100 \quad (5.3)$$

5.1. Population: In this population, the two rarely clustered populations from the Thompson (1992) have been considered as the auxiliary variables in Table 1 and Table 2 as x and z, respectively. Dryver and Chao (2007) generated the values for the variable of interest using the following two models.

$$y_i = 4x_i + \varepsilon_i \quad \text{where} \quad \varepsilon_i \sim N(0, x_i) \quad (5.1.1)$$

$$y_i = 4w_{xi} + \varepsilon_i \quad \text{where} \quad \varepsilon_i \sim N(0, w_{xi}) \quad (5.1.2)$$

The variability of the study variable is proportional to the auxiliary variable itself in model (5.1.1) whereas it is proportional to the within-network mean level of the auxiliary variable in model (5.1.2). Consequently, the within network variances of the study variable in the two networks consisting of more than one units are much larger in the population generated by model (5.1.1).

To evaluate the performance of proposed Estimators, we simulate the values for variable of interest using the model (5.1.3), variable x and z have been used as auxiliary variables. Let y_i, w_{xi} and w_{zi} denote the i^{th} value for the variable of interest y, averages of networks for the auxiliary variables w_x and w_z respectively.

In the given model, the averages of the networks (transformed population; Thompson, 2002) for variable x and z have been used as auxiliary variables showing the strong correlation at the network (region) level.

$$y_i = 4w_{xi} + 4w_{zi} + \varepsilon_i \quad \text{where} \quad \varepsilon_i \sim N(0, w_{xi} + w_{zi}) \quad (5.1.3)$$

In simulated population from model (5.1.3), the variance of study variable is proportional to the sum of the network means level of auxiliary variables. The simulated population using model (5.1.3) is given in Table 3. Chao et al. (2010) and Dryver and Chao (2007) have proved that Conventional Estimators in SRSWOR perform better than Adaptive Cluster Sampling Estimators for strong correlation at unit level. In this population, we evaluate the Estimators when having the correlation at network level not on unit level.

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 and z-values are obtained and averaged. The correlation coefficient between variable of interest and auxiliary variables x and z is 0.5138866 and 0.3654023, respectively. The correlation coefficient in the transformed population between variable of interest and auxiliary variables x and z is 0.7752588 and 0.576928, respectively. Thus, correlation is high at network (region) level than the sampling unit level. Dryver and Chao (2007) showed that usual Estimators in SRSWOR perform better than Adaptive Cluster Sampling Estimators for strong correlation at unit level but performs worse when having the strong correlation at network level. The Adaptive Estimators will be more efficient if there is a sufficiently high correlation in the transformed population.

Thompson (2002) investigated that Adaptive Cluster Sampling is preferable than the comparable Conventional sampling methods if the within network variance is sufficiently large as compared to overall variance of the study variable and presented the condition when the Modified Hansen-Hurwitz Estimator for Adaptive Cluster Sampling have lower variance than variance of the mean per unit for a simple random sampling without replacement of size $E(v)$ if and only if,

$$\left(\frac{1}{n}-\frac{1}{E(v)}\right)S_y^2 \le \frac{N-n}{Nn(N-1)}\sum_{k=1}^{K}\sum_{i\in A_i}(y_i-w_i)^2, \tag{5.1.4}$$

or

$$S_y^2 \le \frac{E(v)[N-n]}{N[E(v)-n]}S_{wy_D}^2. \tag{5.1.5}$$

where

within network variance is defined by:

$$S_{wy_D}^2 = \frac{1}{(N-1)}\sum_{k=1}^{K}\sum_{i\in A_i}(y_i-w_i)^2. \tag{5.1.6}$$

Using equation (5.1.6) the within network variance of the study variable is found to be 1.281415. Let, for the initial sample size n = 10, expected sample size will be $E(v)$ = 34.73with the condition of interest y > 0, using the condition (5.1.5), we have:

$$169.0594 \le \frac{34.73[400-10]}{400[34.73-10]}(1.281415) \cdot \tag{5.1.7}$$

$$169.0594 > 1.754514 \tag{5.1.8}$$

Thus, the overall variance is found to be very high as compare to the within network variance for the study variable. Adaptive Cluster Sampling is preferable than the Conventional sampling if the within network variance is large enough as compare to overall variance (Thompson 2002). So, Adaptive Estimators are expected to perform worse than the Estimators in Simple Random Sampling.

The within network variances of the study variable for the network (38, 31, 37, 30, 32, 30, 29, 30, 29, 30, 31, 32, 33) is 8.064103, for the network (38, 33, 35, 31, 29, 31, 33, 38, 36, 33, 35, 36) is 8, for network (22, 24, 20, 26, 19, 25) is 7.866667, for the network (37, 42, 43, 40, 38, 40, 36, 31, 42, 31, 41) is 18.45455, for the network (45, 52, 44, 49) is 13.66667, and for the network (47, 50, 46, 45, 49, 53, 49, 52, 53) is 8.75. The within network variances do not accounts a large portion of overall variance. Thus, Adaptive Estimators are expected to perform worse than the comparable usual Estimators. The Conventional Estimators are more efficient than the Adaptive Estimators if within-network variances do not account for a large portion of the overall variance (Dryver and Chao 2007).

Simulation experiment has been conducted for all the Adaptive Estimators and Simple Random Sampling Estimators. The estimated relative Bias (Table 4) of all the Estimators decreases by increasing the sample size. The Bias decreases by increasing the sample size as recommended that Bias decreases for large sample sizes (Lohr, 1999). The Percentage Relative Efficiency (Table 5) of all the ACS Estimators remains much lower than the SRS Estimators except the proposed Estimators t_8 and t_9. The proposed Regression-cum-Exponential Ratio Estimator t_8 has the maximum Percentage Relative Efficiency. The proposed regression-cum-exponential ratio Estimator t_9 also has greater percentage relative efficiency than all other Estimators. The regression Estimator t_6 in ACS did not perform well in term of percentage relative efficiency. Thus, the proposed Regression-cum-Exponential Ratio Type Estimators are more robust for rare Clustered population even when the efficiency conditions are not fulfilled. Dryver and Chao (2007) assumed 0/0 as zero for the Ratio Estimator. The usual Ratio Estimator and Ratio Estimator in ACS did not perform and return no value (*) for the all the sample sizes. In this simulation study, 0/0 is not assumed 0. The use of Regression-cum-Exponential Ratio Estimators is better in Adaptive Cluster Sampling than assuming an unlikely assumption for the Ratio Estimator in Adaptive Cluster Sampling. We did the simulation study of logarithmic estimators in ACS and found good. We recommend the detail study of ACS design with logarithmic estimators as future research.

Acknowledgements

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

Table 1: Auxiliary variable X (Thompson, 1992) for population

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	25	2	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	38	3	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	6	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	11	26	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	22	19	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	3	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	17	26	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	10	26	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	5	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 2: Auxiliary variable Z (Thompson, 1992) for population

0	0	0	0	5	13	3	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	2	11	2	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	3	1	0	0	0	0	0	0	0	0	0

0	0	0	0	0	0	0	0	5	39	10	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	5	13	4	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	2	22	3	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	10	8	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	7	22	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 3: Simulated values of y under model (5.1.3) using *Wx* and *Wy*

0	0	0	0	22	20	19	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	24	26	25	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	47	45	49	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	49	52	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	46	53	53	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	37	29	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
38	30	30	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	32	29	33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	30	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	38	42	0	0	0	0	0	0	0	0	0
0	0	0	0	29	36	0	0	42	40	31	0	0	0	0	0	0	0	0	0
0	0	0	33	31	33	0	0	43	36	41	0	0	0	0	0	0	0	0	0
0	0	38	35	33	35	0	37	40	31	0	0	0	0	0	0	0	0	0	0

0	0	0	31	38	36	0	0	0	0	0	0	45	44	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	52	49	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 4: Estimated Relative Bias for the Estimators

n	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
10	*	0.01	0.36	-0.01	*	0	0.16	-0.03	-0.26
20	*	0.01	0.26	0.00	*	0	0.11	-0.05	-0.17
30	*	0.00	0.19	0.00	*	0	0.06	-0.05	-0.14
40	*	0.00	0.14	0.00	*	0	0.04	-0.06	-0.12
50	*	0.00	0.12	0.00	*	0	0.04	-0.06	-0.11

Table 5: Comparative Percentage Relative Efficiencies for the Estimators

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
34.73	100	*	36.49	11.04	27.64	*	68.26	6.18	317.72	112.84
63.03	100	*	38.41	12.97	27.40	*	70.39	9.01	381.54	120.52
86.34	100	*	39.12	15.73	29.54	*	71.31	12.91	405.83	123.45
105.82	100	*	42.41	18.20	30.21	*	77.68	17.12	390.41	131.52
122.34	100	*	43.55	19.83	33.91	*	82.33	19.69	370.52	134.65

Table 6: Descriptive measures of the populations

$\bar{X} = 0.8075$	$\sigma_x^2 = 15.55934$	$C_x = 4.8849$	$\rho_{xz} = -0.0315$
$\bar{Z} = 0.4050$	$\sigma_z^2 = 6.9834$	$C_z = 6.5250$	$\rho_{xy} = 0.5139$
$\bar{Y} = 5.0275$	$\sigma_y^2 = 169.0594$	$C_y = 2.5862$	$\rho_{yz} = 0.3654$
$\bar{w}_x = 0.8098$	$\sigma_{wx}^2 = 7.3624$	$C_{wx} = 3.3509$	$\rho_{wxwz} = -0.0685$
$\bar{w}_z = 0.4751$	$\sigma_{wz}^2 = 4.3092$	$C_{wz} = 4.3696$	$\rho_{wxwy} = 0.7753$
$\bar{w}_y = 5.0273$	$\sigma_{wy}^2 = 167.7641$	$C_{wy} = 2.5764$	$\rho_{wywz} = 0.5770$

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. Chutiman, N. (2013). Adaptive Cluster Sampling using auxiliary variable. *Journal of Mathematics and Statistics*, **9(3)**, 249-255.
4. Chutiman, N. and Kumphon, B. (2008). Ratio Estimator using two auxiliary variables for Adaptive Cluster Sampling. *Thailand Statistician*, **6(2)**, 241-256.
5. 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 Sciences*, **30**, 262-275.
6. Cochran, W. G. (1942). Sampling theory when the sampling units are of unequal sizes. *Journal of the American Statistical Association*, **37**, 199-212.
7. Cochran, W.G. (1977). *Sampling Techniques.* John Wiley, New York.
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. Lohr, S. L. (1999). *Sampling Design and Analysis Pacific Grove*. Duxbury Press, CA.
11. Noor-ul-Amin and Hanif, M. (2013). Exponential Estimators in Survey Sampling. Unpublished Thesis. Submitted to National College of Business Administration and Economics, Lahore.
12. Hansen, M. M. and Hurwitz, W. N. (1943). On the theory of sampling from finite population. *The Annals of Mathematical Statistics*, **14**, 333–362.
13. 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.
14. 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.
15. Thompson, S. K. (1990). Adaptive Cluster Sampling. *Journal of American Statistical Association*, **85**, 1050–1059.
16. Thompson, S. K. (1992). *Sampling*. John Wiley and Sons, New York.
17. Thompson, S. K. (2002). *Sampling*. John Wiley and Sons, New York.