Proceedings of the Pakistan Academy of Sciences:
A. Physical and Computational Sciences 53 (4): 447–457 (2016)

ISSN: 2518-4245 (print), 2518-4253 (online)

Pakistan Academy of Sciences

Research Article

Product and Exponential Product Estimators in Adaptive Cluster Sampling under Different Population Situations

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

[1]Department of Statistics, Government College University, Lahore, Pakistan
[2]National College of Business Administration & Economics, Lahore, Pakistan

Abstract: In this paper, the product and exponential product estimators have been proposed for estimating the population mean using population mean of an auxiliary variable, when there is negative correlation between the variables, under adaptive cluster sampling (ACS) design. The expressions for mean squared error (MSE) and bias of the proposed estimators have been derived. Two simulated populations are used and simulation studies have been conceded out to reveal and match the efficiencies of the estimators. The proposed estimators have been matched with conventional estimators and estimators in ACS. The simulation results showed that the proposed product and exponential product estimators are more efficient as compare to conventional as well as Hansen-Hurwitz and ratio estimators in ACS.

Keywords: Auxiliary information, simulated population, transformed population, bivariate normal distribution, negative correlation, expected final sample size, comparable variance, estimated relative bias

1. INTRODUCTION

Sampling selects a part of population of interest to gain information about the whole. Data are often produced by sampling a population of individuals or objects. The inferences made will rely on the statistics gained from the data so these inferences can only be as good as the data. The sampling design refers to the method used to select the sample from the population. Deprived sampling design can produce deceptive conclusions. In conventional sampling designs the sample size is determined prior to the survey. In these designs the sampling units are independent on observations gathered throughout the survey.

A main problem that arises in survey sampling is estimation of the density of rare and clustered population, such as plants, birds and animals of scarce and dying out species, fisheries, patchy minerals exploration, toxic waste concentrations, drug addicted, and AIDS patients. The traditional random sampling designs may be inefficient and frequently not succeed to offer samples with useful information for rare population, because it is possible that the majority of the sampled elements give no information [1].

The Adaptive Cluster Sampling (ACS) design is appropriate for the scarce and bunched population. In ACS, a first sample is selected with a usual sampling design then the vicinity of every element selected in first sample is considered. If the observed values of the study variable satisfy a specific condition C, say y_i > C then the additional units in the neighbourhood of the *i*th units is sampled. Each neighbouring element is included and investigated if the predefined condition C is fulfilled and the procedure keep on until a new unit does not satisfy the condition. It is a type of network sampling, which provides improved estimates as compared to conventional designs in case of clustered rare population. All the units studied (including the initial sample) compose the final sample. The set consisting on those elements that met the condition are called a network. The elements that are fails to fulfil the condition are called edge units. Clusters are grouping of networks and edge units.

Thompson [2] first proposed the idea of the ACS scheme and introduced modified Hansen-Hurwitz [3]

Received, October 2015; Accepted, November 2016
*Corresponding author: Muhammad Shahzad Chaudhry; Email: almoeed@hotmail.com

and Horvitz-Thompson [4] form estimators. Dryver [5] established that ACS performed extraordinary in univariate situation but in the multivariate setting, the efficiency of ACS depended on the relationship of the variables. The simulation results for actual data of blue-winged and red-winged showed that Horvitz-Thompson [4] form of estimator was the mainly efficient estimator in certain conditions. Dryver and Chao [6] proposed the conventional ratio estimator in ACS and also proposed two more ratio estimators.

Adaptive Cluster Sampling Process

Consider a fixed population of N elements labelled *1,2,3,...,N* are denoted as u = $\{u_1,u_2,\ldots,u_N\}$. Consider a small initial sample of size n which is selected from N by simple random sample without replacement (srswor). The first sample is chosen by traditional sampling process in an ACS procedure and then the predefined neighbouring units for all the units of the first sample are considered for a particular condition C. If any of the elements in the first sample satisfy condition C there neighbouring elements are included 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 neighbouring unit satisfies the condition then its neighbourhoods are also sampled and the process goes on. This iterative process stops when the new unit does not satisfy condition C. The vicinity can be decided in two ways. The sampling element and four neighbouring elements are known as the first-order neighbourhood denoted by east, west, north, and south. The first-order neighbouring elements and the elements northeast, northwest, southeast, and southwest are known as second-order neighbourhood. There are in total eight neighbourhood quadrats including the first-order neighbourhood and the second order neighbourhood. All the units including the initial sample composed the final sample.

A network consists of elements that satisfy the specified condition (usually y =1). The networks of size one are those elements which fail to meet the predefined condition C in the first sample. 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 elements in a form of exclusive and entire 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 quadrats. 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 quadrats. The units $(u_1,u_2,\ldots,u_N)$ form a disjoint and comprehensive partition of the entire area so that units labels (1,2,…,N) categorise the position of N quadrats. In ACS the population is measured in terms of quadrats only.

2. MATERIALS AND METHODS

2.1 Some Estimators in Simple Random Sampling

Let N be the entire number of elements in the population. A random sample of size n is selected by using srswor. The study variable and auxiliary variable are represented by y and x with the population means $\overline{Y}$ and $\overline{X}$, population standard deviation S_y and S_x and coefficient of variation C_y and C_x respectively. Also let ρ_{xy} denote population correlation coefficient between X and Y. The sample means of the study and auxiliary variables are denoted by $\overline{y}$ and $\overline{x}$ respectively. Cochran [7] and Robson [8] proposed the classical ratio and classical product estimators, respectively, for estimating the population mean stated as follows:

$$t_1 = \overline{y}\left[\frac{\overline{X}}{\overline{x}}\right], \tag{1}$$

and $$t_2 = \overline{y}\left[\frac{\overline{x}}{\overline{X}}\right]. \tag{2}$$

The mean squared error (MSE) of the estimators (1) and (2) are given by:

$$\text{MSE}\,(t_1) \approx \theta \bar{Y}^2 \left[C_y^2 + C_x^2 - 2\rho_{xy} C_x C_y \right], \tag{3}$$

and
$$\text{MSE}\,(t_2) \approx \theta \bar{Y}^2 \left[C_y^2 + C_x^2 + 2\rho_{xy} C_x C_y \right], \tag{4}$$

respectively. Where $\theta = \frac{1}{n} - \frac{1}{N}$. The ratio and product estimators are design biased.

Bahl and Tuteja [9] proposed the exponential ratio and exponential product estimators to estimate the population mean are given by:

$$t_3 = \bar{y} \exp\left[\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}} \right], \tag{5}$$

$$t_4 = \bar{y} \exp\left[\frac{\bar{x} - \bar{X}}{\bar{X} + \bar{x}} \right]. \tag{6}$$

The MSE and bias of the exponential ratio estimator t_3 are as follows:

$$\text{MSE}(t_3) \approx \theta \bar{Y}^2 \left[C_y^2 + \frac{C_x^2}{4} - \rho_{xy} C_x C_y \right], \tag{7}$$

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

The MSE and bias of the exponential product estimator t_4 are given by:

$$\text{MSE}(t_4) \approx \theta \bar{Y}^2 \left[C_y^2 + \frac{C_x^2}{4} + \rho_{xy} C_x C_y \right], \tag{9}$$

$$Bias(t_4) \approx \theta \bar{Y} \left[-\frac{1}{8} C_x^2 + \frac{\rho_{xy} C_x C_y}{2} \right]. \tag{10}$$

2.2 Some Estimators in Adaptive Cluster Sampling

Let a preliminary sample of n elements is selected with a srswor from a finite population of size N categorized like $1,2,3,\ldots,N$. The average y-value and average x-value in the network which includes unit i are $w_{yi} = \frac{1}{m_i} \sum_{j \in A_i} y_j$ and $w_{xi} = \frac{1}{m_i} \sum_{j \in A_i} x_j$ respectively. ACS can be considered as srswor when the averages of networks are considered [1, 6]. The averages of networks are considered as transformed population. Transformed population is obtained with the replacement of the original values of the networks with the averages of the networks. In the case of transformed population is used, each sample of size one selected with srswor will be representative of the whole network if it intersect to any unit of a network. In the transformed population, the sample means of the study and auxiliary variables are $\bar{w}_y = \frac{1}{n} \sum_{i=1}^{n} w_{yi}$ and $\bar{w}_x = \frac{1}{n} \sum_{i=1}^{n} w_{xi}$ respectively. Consider C_{wy} and C_{wx} represents population coefficient of variations of the study variable and auxiliary variable in the transformed population respectively and ρ_{wxwy} represent population correlation coefficient between w_x and w_y in the ACS. Let us define,

$$\bar{e}_{wy} = \frac{\bar{w}_y - \bar{Y}}{\bar{Y}} \text{ and } \bar{e}_{wx} = \frac{\bar{w}_x - \bar{X}}{\bar{X}}. \tag{11}$$

Where $\bar{e}_{wy}$ and $\bar{e}_{wx}$ are relative sampling errors of the study variable and auxiliary variable respectively, such that:

$$E\left(\bar{e}_{wy}\right) = E\left(\bar{e}_{wx}\right) = 0 \text{ and } E\left(\bar{e}_{wx}\bar{e}_{wy}\right) = \theta\rho_{wxwy}C_{wx}C_{wy} \tag{12}$$

$$E\left(\bar{e}_{wy}^2\right) = \theta C_{wy}^2 \quad \text{and} \quad E\left(\bar{e}_{wx}^2\right) = \theta C_{wx}^2 \tag{13}$$

Thompson [2] proposed an unbiased modified Hansen-Hurwitz [3] estimator for population mean in ACS and can be used as sampling done with replacement or without replacement. Elements that do not meet the condition *C* are ignored if these elements are not selected in the preliminary sample. In the form of *n* networks (which possibly will not be exclusive) overlapped by the preliminary sample (because transformed population is used, ACS becomes srswor and each unit selected will represent a whole network) is stated as follows:

$$t_5 = \frac{1}{n}\sum_{i=1}^{n} w_{yi} = \bar{w}_y \,. \tag{14}$$

Where $w_{yi} = \frac{1}{m_i}\sum_{j\in A_i} y_j$ is the average of the number of elements m_i in the network A_i.

The variance of t_5 is given by:

$$Var(t_5) = \frac{\theta}{N-1}\sum_{i=1}^{N}\left(w_{yi} - \bar{Y}\right)^2 . \tag{15}$$

Dryver and Chao [6] proposed a modified ratio estimator to estimate the population mean in ACS is given by:

$$t_6 = \left[\frac{\sum_{i\in s_0} w_{yi}}{\sum_{i\in s_0} w_{xi}}\right]\bar{X} = \hat{R}\bar{X} \,. \tag{16}$$

Where $\hat{R}$ is the sample ratio between w_{yi} and w_{xi}.

The MSE of t_6 is given by:

$$MSE(t_6) = \frac{\theta}{N-1}\sum_{i=1}^{N}\left(w_{yi} - Rw_{xi}\right)^2 . \tag{17}$$

Where *R* is the population ratio between w_{yi} and w_{xi}.

2.3 Proposed Estimators in Adaptive Cluster Sampling

The proposed modified classical product estimator in ACS with one auxiliary variable is stated as follows:

$$t_7 = \bar{w}_y \frac{\bar{w}_x}{\bar{X}} . \tag{18}$$

Following the Bahl and Tuteja [9], the proposed exponential product estimator in ACS with one auxiliary variable is stated as follows:

$$t_8 = \bar{w}_y \exp\left[\frac{\bar{w}_x - \bar{X}}{\bar{w}_x + \bar{X}}\right]. \tag{19}$$

2.3.1 Bias and Mean Square Error of Proposed Product Estimator t_7

Using (11) the estimator (18) may be written as follows:

$$t_7 = \bar{Y}\left(1+\bar{e}_{wy}\right)\frac{\bar{X}(1+\bar{e}_{wx})}{\bar{X}}, \tag{20}$$

so, $$t_7 = \bar{Y}\left[1+\bar{e}_{wy}+\bar{e}_{wx}+\bar{e}_{wx}\bar{e}_{wy}\right]. \tag{21}$$

Applying expectations on both sides of (21), and using the notations (12) we get as follows:

$$\text{Bias}\, t_7 = E\left(t_7-\bar{Y}\right) = \bar{Y}\theta\rho_{wxwy}C_{wx}C_{wy}. \tag{22}$$

In order to derive MSE of (18), we have (23) by ignoring the term degree 2 or greater as follows:

$$t_7 = \bar{Y}\left[1+\bar{e}_{wy}+\bar{e}_{wx}\right], \tag{23}$$

$$t_7-\bar{Y} = \left[\bar{e}_{wy}+\bar{e}_{wx}\right]. \tag{24}$$

Taking square and expectations on the both sides of (24), the obtained as follows:

$$\text{MSE}\left(t_7\right) = E\left(t_7-\bar{Y}\right)^2 \approx \bar{Y}^2\left(\theta C_{wy}^2+\theta C_{wx}^2+2\theta\rho_{wxwy}C_{wx}C_{wy}\right). \tag{25}$$

2.3.2 Bias and Mean Square Error of the Proposed Exponential Product Estimator t_8

Using (11) the estimator (19) may be written as follows:

$$t_8 = \bar{Y}\left(1+\bar{e}_{wy}\right)\exp\left[\frac{\bar{X}(1+\bar{e}_{wx})-\bar{X}}{\bar{X}(1+\bar{e}_{wx})+\bar{X}}\right], \tag{26}$$

$$t_8 = \bar{Y}\left(1+\bar{e}_{wy}\right)\exp\left[\frac{\bar{e}_{wx}}{2}\left(1+\frac{\bar{e}_{wx}}{2}\right)^{-1}\right], \tag{27}$$

or $$t_8 = \bar{Y}\left(1+\bar{e}_{wy}\right)\exp\left[\frac{\bar{e}_{wx}}{2}\left(1-\frac{\bar{e}_{wx}}{2}+\frac{\bar{e}_{wx}^2}{4}\right)\right], \tag{28}$$

or $$t_8 = \bar{Y}\left(1+\bar{e}_{wy}\right)\exp\left[\frac{\bar{e}_{wx}}{2}-\frac{\bar{e}_{wx}^2}{4}\right]. \tag{29}$$

Expanding the exponential term up-to the second degree, we get (29) as follows:

$$t_8 \approx \bar{Y}\left(1+\bar{e}_{wy}\right)\left[1+\frac{\bar{e}_{wx}}{2}-\frac{\bar{e}_{wx}^2}{4}+\frac{\bar{e}_{wx}^2}{8}\right]. \tag{30}$$

Simplifying, ignoring the terms with degree three or greater we get as follows:

$$t_8-\bar{Y} \approx \bar{Y}\left(\bar{e}_{wy}+\frac{\bar{e}_{wx}}{2}-\frac{\bar{e}_{wx}^2}{4}+\frac{\bar{e}_{wx}^2}{8}+\frac{\bar{e}_{wy}\bar{e}_{wx}}{2}\right). \tag{31}$$

Applying expectations on both sides of (31) as follows:

$$\text{Bias}\left(t_8\right) = E\left(t_8-\bar{Y}\right) \approx \bar{Y}\theta\left(-\frac{C_{wx}^2}{4}+\frac{C_{wx}^2}{8}+\frac{\rho_{wxwy}C_{wx}C_{wy}}{2}\right). \tag{32}$$

In order to derive MSE of (19), we have (32) as follows:

$$t_8 \approx \bar{Y}\left(1+\bar{e}_{wy}\right)\exp\left[\frac{\bar{e}_{wx}}{2}(1-\frac{\bar{e}_{wx}}{2})\right]. \tag{33}$$

Ignoring the terms with power two or greater, we get (33) as follows:

$$t_8 \approx \bar{Y}\left(1+\bar{e}_{wy}\right)\exp\left[\frac{\bar{e}_{wx}}{2}\right]. \tag{34}$$

Opening the exponential term, ignoring terms with power two or more we get (34) as follows:

$$t_8 \approx \bar{Y}\left(1+\bar{e}_{wy}\right)\left[1+\frac{\bar{e}_{wx}}{2}\right], \tag{35}$$

or

$$t_8 - \bar{Y} \approx \bar{Y}\left(\bar{e}_{wy}+\frac{\bar{e}_{wx}}{2}+\frac{\bar{e}_{wx}\bar{e}_{wy}}{2}\right). \tag{36}$$

Taking square and expectations on the both sides of (36), and using notations (12 &13) we get as follows:

$$\mathrm{MSE}\left(t_8\right)=E\left(t_8-\bar{Y}\right)^2 \approx \theta\bar{Y}^2\left(C_{wy}^2+\frac{C_{wx}^2}{4}+\rho_{wxwy}C_{wx}C_{wy}\right). \tag{37}$$

3. RESULTS AND DISCUSSION

3.1 Simulation Study

To evaluate and match the efficiency of suggested estimators with the already existing estimators, two different types of simulated populations are used and executed simulations for the comprehensive study. The condition C for included elements in the sample is defined. To get the transformed population, the y-values are acquired and averaged for keeping the sample network with respect to the condition and for every sample network parallel *x*-values are obtained and averaged, then y-values and x-values are replaced with their networks averages, accordingly. For the simulation study ten thousands iteration was executed to get MSE and bias for each estimator with the srswor and the initial sample sizes of 5,10,15,20 and 25 for populations 1 and 2. In ACS, the ultimate sample size is generally larger than the preliminary sample size. Let E(*v*) denote the expected final sample size in ACS, this is the sum of the probabilities of inclusion of all quadrats. The expected final sample size fluctuates from one sample to another sample in ACS. For the comparison, the sample mean from a srswor based on E(*v*) has variance using the formula stated as follows:

$$Var(\bar{y})=\frac{\sigma^2(N-E(v))}{NE(v)} \tag{38}$$

The estimated relative bias is defined as:

$$\hat{RBias}\left(t_*\right)=\frac{\frac{1}{r}\sum_{i=1}^{r}(t_*)-\bar{Y}}{\bar{Y}} \tag{39}$$

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

The estimated MSE of the estimated mean is given by:

$$\hat{MSE}\left(t_*\right)=\frac{1}{r}\sum_{i=1}^{r}(t_*-\bar{Y})^2 \tag{40}$$

The percentage relative efficiency is given by:

$$PRE = \frac{Var(\bar{y})}{MSE(\hat{t}_*)} \times 100 \tag{41}$$

3.2 Population 1: Study Variable is Clustered and Auxiliary Variable is Binary

Dryver and Chao [6] used blue-winged teal data (Fig. 1) collected by Smith et al. [10] as an auxiliary variable for proposed ratio estimators under ACS and compared their efficiency with conventional ratio estimator in srswor. Let y_i and x_i denote the i[th] value for the variable of interest *y*, auxiliary variables *x* (say blue-winged teal), Dryver and Chao [6] generated the values for the variable of interest using the following two models:

$$y_i = 4x_i + \varepsilon_i \quad \text{where} \quad \varepsilon_i \sim N(0, x_i) \tag{42}$$

$$y_i = 4w_{xi} + \varepsilon_i \quad \text{where} \quad \varepsilon_i \sim N(0, w_{xi}) \tag{43}$$

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

Fig. 1. Blue-Winged Teal data [10] for population.

The variability of the variable of interest y is proportional to the auxiliary variable itself in model (42) while it is proportional to the within-network mean level of the auxiliary variable in model (43). Therefore, the within network variances of the variable of interest in the networks are greatly bigger in the population produced with model (42). In this simulation study the blue-winged teal (BWT) is taken as the study variable y. To generate a data set for the auxiliary variable *x* (Fig. 2) that contributes high negative correlation (-0.999) correspond to the BWT data which is the variable of interest (*y*), the following model is used:

$$x_i = (-1) y_i + \varepsilon_i \quad \text{where} \quad \varepsilon_i \sim N(0, y_i) \tag{44}$$

1	1	1	1	1	1	1	1	1	1
1	1	1	0	1	1	1	1	0	1
1	1	1	1	1	1	1	1	0	1
1	1	1	1	1	1	1	1	1	0
1	1	1	1	1	1	1	1	1	0

Fig. 2. Simulated *x* values, based on model (43) and using BWT data for population 1.

In model (44), the study variable y is treated as independent variable just to generate a data set for the auxiliary variable x. Then 50 is added to and divided by each value to avoid the maximum number of negative values and the remaining negative values are treated as zero. Doing so, the correlation decreases to -0.442. Now changing the role of variables again (i.e., BWT is treated as study variable and not as an auxiliary variable) the condition for the variable of interest BWT is set as C ≥10 to added unit in the sample. There remaining only two networks with this condition but the correlation found to be -0.908 between the average values of the networks (Fig. 3, 4). Thus, there is a low correlation at a unit level but high correlation at the network level. Dryver and Chao [6] demonstrated that classical estimators in srswor execute well than ACS estimators for high correlation at unit level while execute poorer when contain the high correlation at network level.

0	0	3	5	0	0	0	0	0	0
0	0	0	19	19	0	0	2344.2	2344.2	0
0	0	0	0	2	3	2	0	2344.2	1
0	0	0	0	0	0	0	0	2344.2	2344.2
0	0	0	0	0	0	2	0	0	2344.2

Fig. 3. Transformed population-1 with average values of the networks (Wy).

1	1	1	1	1	1	1	1	1	1
1	1	1	0.5	0.5	1	1	0.33	0.33	1
1	1	1	1	1	1	1	1	0.33	1
1	1	1	1	1	1	1	1	0.33	0.33
1	1	1	1	1	1	1	1	1	0.33

Fig. 4. Transformed Population-1 with average values of the networks (Wx).

The variability of the auxiliary variable is proportional to the variable of interest in the model (44). The whole variance of the variable of interest is 3716168 whereas in the transformed population the variance decreased to 591498.701. The within network variance of the variable of interest for the network (10, 103, 13639, 14, 122, 177) is 30621746.971. An enormous part of whole variance is accounts by within network variance. Therefore, estimators in ACS are likely to be more efficient than the equivalent estimators in srs. The estimators in srs are more efficient than the estimators in ACS if within-network variances do not report a great part of the overall variance [6].

The comparative percentage relative efficiency (Table 1) of the ACS estimators is much higher than their counterpart SRS estimators. The proposed modified product t_7 and exponential modified product estimator t_8 in ACS has maximum percentage relative efficiency for the initial sample size and percentage relative efficiency starts increasing rapidly for comparable sample sizes. Thus, the proposed product estimator in ACS perform much superior than the other usual estimators, the ratio, and the Hansen-Hurwitz estimators in ACS when there is a high negative correlation among the study and the auxiliary variables, under the given conditions.

Table 1. Comparative percentage relative efficiencies for population1 based on E(v).

$\bar{y}$	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
100	59	146	79	123	110	72	158	137
100	81	117	91	109	113	88	143	130
100	88	109	95	105	123	96	152	135
100	93	106	96	103	138	114	168	151
100	95	104	98	102	159	134	189	171

The estimated relative bias (Table 2) of the estimators decreases as sample sizes increases in the ACS as well as in the srs. For ACS, as like srswor, it is suggested a bigger sample size for a small bias [11].

Table 2. Estimated relative bias for population 1 for different sample sizes.

n	E(v)	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
5	19.22	0.23	-0.17	0.12	-0.11	0	0.12	-0.12	-0.07
10	29.09	0.09	-0.09	0.04	-0.03	0	0.07	-0.05	-0.04
15	34.36	0.02	-0.05	0.03	-0.03	0	0.05	-0.03	-0.02
20	37.54	0.04	-0.03	0.02	0.00	0	0.02	-0.03	-0.01
25	39.90	0.02	-0.02	0.01	0.00	0	0.02	-0.01	-0.01

3.3 Population 2: Study Variable is Rare Clustered and Auxiliary Variable is Abundant

The population 2 (Fig. 5 & Fig. 6) has been generated from a bivariate normal distribution with the mean vector μ and covariance matrix Σ, this is $\begin{pmatrix} Y \\ X \end{pmatrix} \sim N(\mu, \Sigma)$. In particular, we assumed $\mu = (0,10)$ and $\Sigma = \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix}$. The condition C to included elements in the sample is $y > 0$ for population 2. The correlation in the pair of random variables of the population was found to be -0.912. When negative values are assumed zeroes in simulated population the correlation reduces to -0.722. The correlation when the averages of the networks (Fig. 7, 8) assumed increases to -0.733.

0	0	0	0	0	0	0	0	2	0
0	0	0	0	1	0	2	0	1	1
0	1	2	0	0	1	1	0	1	1
0	1	0	0	0	1	1	0	1	0
2	0	1	0	0	0	1	2	0	0

Fig. 5. Simulated y values for population 2, based on bivariate normal distribution.

8	14	9	10	13	13	9	9	7	11
11	14	10	12	8	14	8	10	6	10
9	9	6	10	12	8	9	11	9	8
11	9	11	15	12	11	8	14	8	10
6	12	8	13	11	13	10	6	9	11

Fig. 6. Simulated x values for population 2, based on bivariate normal distribution.

0	0	0	0	0	0	0	0	1.16	0
0	0	0	0	1	0	1.29	0	1.16	1.16
0	1.33	1.33	0	0	1.29	1.29	0	1.16	1.16
0	1.33	0	0	0	1.29	1.29	0	1.16	0
2	0	1	0	0	0	1.29	1.29	0	0

Fig. 7. Transformed population-2 with average values of the networks (Wy).

8	14	9	10	13	13	9	9	8	11
11	14	10	12	8	14	8.57	10	8	8
9	8	8	10	12	8.57	8.57	11	8	8
11	8	11	15	12	8.57	8.57	14	8	10
6	12	8	13	11	13	8.57	8.57	9	11

Fig. 8. Transformed population-2 with average values of the networks (Wx).

The whole variance of the variable of interest is 0.459 while this variance reduced to 0.399 in the transformed population. The within network variances of the study variable for the network (1,1,3) is 0.333, for the network (1,1,2,1,1,1,2) is 0.238, and for the network (2,1,1,1,1,1) is 0.167. The network variances do not accounts a huge part of whole variance. The overall variance is found to be high as compare to the within network variances for the study variable. Thus, adaptive estimators are expected to perform worse than the comparable usual estimators. The usual estimators will be more efficient than the adaptive estimators if within-network variances do not account for a huge part of the overall variance [6].

The percentage relative efficiency (Table 3) of the ACS estimators is much lower than the SRS estimators, as expected. The usual product estimator has maximum percentage relative efficiency, while usual exponential product estimator has higher percentage relative efficiency than the other conventional and adaptive estimators in ACS. The bias of all the estimators decreases by increasing the sample size. The estimated relative bias is given in Table 4. The bias decreases by increasing the sample size as recommended that bias decreases for large sample sizes [11].

Table 3. Comparative percentage relative efficiencies for population 2 based on E(v).

$\bar{y}$	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
100	71.16	131.7	84.69	116.6	22.84	16.46	29.13	25.44
100	76.32	127.3	87.16	113.6	21.29	16.82	27.09	24.16
100	77.67	125.2	88.22	113.3	20.93	16.40	26.56	23.87
100	78.21	126.7	88.83	112.6	21.38	16.76	27.04	24.61
100	79.32	125.6	88.90	112.3	22.21	17.86	28.05	25.53

Table 4. Estimated relative bias for population 2 for different sample sizes.

N	E(ν)	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
5	18.25	0.05	-0.05	0.02	-0.01	0	0.04	-0.04	-0.02
10	28.79	0.02	-0.02	0.01	-0.01	0	0.02	-0.02	-0.01
15	35.05	0.01	-0.02	0.00	-0.01	0	0.02	-0.01	0.00
20	39.03	0.01	-0.01	0.00	-0.01	0	0.01	-0.01	-0.01
25	41.82	0.00	0.00	0.00	0.00	0	0.00	-0.01	0.00

4. CONCLUSIONS

The performance of proposed product estimator and exponential product is better than all the other estimators including conventional as well as Hansen-Hurwitz and ratio estimators in ACS sampling for population 1. The proposed estimators become more efficient as initial sample size increases for the population 1, under the given conditions. Thus, the proposed product estimator and proposed exponential product estimator should be employed for rare and clustered population when there is negative correlation between the study variable and the auxiliary variable. The product and exponential product estimators may be studied for the population variance for negatively correlated study and auxiliary variables in ACS. Moreover, some logarithmic form of estimators may also be derived as a future research in ACS.

5. ACKNOWLEDGEMENTS

The authors are indebted to Yves G. Berger**,** University of Southampton, UK for the programming guidance to produce populations, simulations and precious suggestions about the upgrading of this paper.

6. REFERENCES

1. Thompson, S. K. *Sampling*. Wiley, New York (1992).
2. Thompson, S.K. Adaptive cluster sampling. *Journal of American Statistical Association* 85: 1050–1059 (1990).
3. Hansen, M.M. & W.N. Hurwitz. On the theory of sampling from finite population. *Annals of Mathematical Statistics* 14: 333–362 (1943).
4. Horvitz, D.G. & D.J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of American Statistical Association* 47: 663–685 (1952).
5. Dryver, A.L. Performance of adaptive cluster sampling estimators in a multivariate setting. *Environmental and Ecological Statistics* 10: 107–113 (2003).
6. Dryver, A.L. & C.T. Chao. Ratio estimators in adaptive cluster sampling. *Environmetrics* 18: 607–620 (2007).
7. Cochran, W.G. 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 (1940).
8. Robson, D.S. Application of multivariate polykays to the theory of unbiased ratio type estimators. *Journal of American Statistical Association* 52: 511-522 (1957).
9. Bahl, S. & R.K. Tuteja. Ratio and product type exponential estimator. *Information and Optimization Sciences* 12: 159-163 (1991).
10. Smith, D.R., M.J. Conroy, & D.H. Brakhage. Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. *Biometrics* 51: 777-788 (1995).
11. Lohr, S.L. *Sampling Design and Analysis*. Pacific Grove, Duxbury Press, CA (1999).