Thailand Statistician
January 2020; 18(1): 65-76
http://statassoc.or.th
Contributed paper

# Joint Influence of Exponential Ratio and Exponential Product Estimators for the Estimation of Clustered Population Variance in Adaptive Cluster Sampling

**Muhammad Nouman Qureshi*[a], [c], Naureen Riaz [b], [c] and Muhammad Hanif [c]**
[a] School of Statistics, University of Minnesota, Minneapolis, USA.
[b] Department of Statistics, Lahore Garrison University, Lahore, Pakistan.
[c] Department of Statistics, National College of Business Administration & Economics, Lahore, Pakistan.
*Corresponding author; e-mail: queri006@umn.edu

Received: 27 July 2018
Revised: 24 March 2019
Accepted: 10 April 2019

**Abstract**

Adaptive cluster sampling (ACS) is considered to be the most efficient sampling design for the estimation of statistical parameters of rare and clustered populations. In this paper, we proposed an estimator that jointly incorporate the exponential ratio and exponential product type estimators using single auxiliary variable based on the averages of the networks in adaptive cluster sampling. The expressions of approximate bias and mean square error of the proposed estimator are derived. A numerical study is carried out using real and artificial populations to demonstrate and compare the efficiency of the proposed estimator over the traditional variance estimator under simple random sampling (SRS). The results of relative efficiencies show that the proposed estimator is more efficient than all the adaptive and non-adaptive estimators considered in this paper.

**Keywords**: Auxiliary Information, clustered population, Hansen-Hurwitz estimation, within network variances, variance estimation.

## 1. Introduction

Adaptive cluster sampling first proposed by Thompson (1990) is considered to be the most suitable and efficient sampling design for the estimation of rare and clustered population parameters. Such clustered populations includes plants and animals of rare and endangered species, flocking, fisheries, epidemiology of sporadic diseases, noise problems, pollution concentrations, criminal and hotspot investigations, drug users, AIDS and HIV patients. The use of auxiliary information is the most constructive to get the better estimation results for rare and clustered populations. In many environmental surveys, if the count of particular species in a locality is known as survey variable, the availability of food, habitat, rainfall and temperature of the same locality would be considered as auxiliary variable for the estimation of population parameters, like mean, total, variance etc.

Thompson (1990, 1991a, 1991b, 1992) suggested several unbiased estimators for the estimation of rare and clustered population mean and variance under ACS design. Smith et al. (1995) applied ACS for estimating the density of wintering waterfowl and concluded that the efficiency of ACS estimator is highest as compare to the simple random sampling (SRS) design when the within-network variance is close to the overall population variance. Chao (2004a) and Dryver and Chao (2007) proposed ratio estimators based on the modification of Hansen-Hurwitz (1943) and Horvitz-Thompson (1952) type estimators for the estimation of population mean under ACS.

In this paper, we propose an estimator that combines the exponential ratio and exponential product estimators using probability weighting approach for the estimation of rare and highly clustered population variance under the framework of ACS design. The methodology of ACS design with a brief example is described in Section 2. Some existing estimators under SRS and ACS designs with basic notations are addressed in Section 3. The expressions of approximate bias and minimum mean square error (MSE) of the proposed estimator are derived in Section 4 with some special cases in which the proposed estimator is reduced to exponential ratio and exponential product type estimators on different values of optimization constant. An improved version of the proposed estimator is also discussed in the same section. A numerical study is conducted in Section 5 on real and artificial clustered populations in order to check the performance of the proposed estimator under various initial and expected sample sizes. Conclusions and remarks on the paper are given in Section 6.

## 2. Methodology of Adaptive Cluster Sampling

Consider a study region $\Omega$ that can be spatially partitioned into a grid of $N$ equal sized rectangular units labelled from 1 to $N$. The response of the survey variable $(y_i)$ together with auxiliary variable $(x_i)$ is associated with each unit $i=1,2,...,N.$ The population vectors of $y$-values and $x$-values are define as $Y=(y_1,y_2,...,y_N)$ and $X=(x_1,x_2,...,x_N),$ respectively. Let the population is divided into $K$ exhaustive networks $(A's)$ and $A_i$ denotes the $i^{\text{th}}$ network with $m_i$ units. The dataset $d$ is selected by using initial sample size from the sampled units say $s,$ with the associated values of the survey variable $(y_s)$ along with the corresponding values of the auxiliary variable $(x_s)$ from $Y$ and $X,$ respectively. Hence the dataset $d$ is defined as $d=\{s,y_s,x_s\}.$

In ACS the initial sample of size $n$ is selected by a conventional sampling design such as simple random sampling (SRS), stratified sampling or systematic sampling. If the value of the survey variable from the sampled unit satisfies $C,$ usually $C=\{y;\, y>0\},$ the first-order neighboring units (up, down, to the left, to the right) will be added to the sample and examined. If the neighboring units of first neighbors satisfy the $C,$ then their neighboring units will be added to the sample and observed and the process remain continuous until no new units met the condition. The final sample comprises the initially selected units and all the studied units that satisfied the condition. A network consists of those units that satisfied $C.$ The units that do not meet condition are known as edge units. A cluster is a combination of network and edge units.

Figure 1 illustrates the idea of a cluster in which the unit having the star is an initially selected unit. The shaded units are adaptively added units that satisfying a pre-defined condition, $C.$ The units that do not meet $C$ (bold) are known as edge units.

| | | | | | |
|---|---|---|---|---|---|
| 0 | 0 | **0** | **0** | 0 | 0 |
| 0 | **0** | 1 | 8 | **0** | 0 |
| **0** | 5 | 9* | 7 | 2 | **0** |
| **0** | 2 | 12 | 5 | **0** | 0 |
| 0 | **0** | 4 | **0** | 0 | 0 |
| 0 | 0 | **0** | 0 | 0 | 0 |

**Figure 1** A cluster, a network with its edge units

## 3. Existing Estimators in Simple Random Sampling

Let a random sample of size $n$ is selected from a clustered population of $N$ units by using simple random sampling without replacement (SRSWOR). The study and auxiliary variables are denoted by $y$ and $x$ with their respective means and, whereas the standard deviations $S_y$ and $S_x$ and coefficient of variations, $C_y$ and $C_x$, respectively. Further, $\rho_{yx}$ represents the population correlation coefficient between $Y$ and $X$. The following notations are used for the expressions of bias and MSE of the existing estimators

$$\lambda_{pq} = \frac{\mu_{pq}}{\mu_{20}^{p/2}\ \mu_{02}^{q/2}},\ \mu_{pq} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\mu_y)^p(x_i-\mu_x)^q \text{ and } \gamma = \frac{1}{E(v)},$$

where $p$ and $q$ are non-negative integers and $E(v)$ is the expected sample size. The quantities, $\mu_{20}$ and $\mu_{02}$ be the second order moments and $\lambda_{pq}$ is known as moment ratio's.

The usual unbiased sample estimator for population variance with the expression of its variance is given as

$$\hat{S}_0^2 = s_y^2,$$

$$Var(\hat{S}_0^2) = \gamma\sigma_y^2(\lambda_{40}-1),$$

where $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{Y})^2$.

Isaki (1983) first proposed the ratio estimator with single auxiliary variable for the estimation of finite population variance in SRS. The estimator with the expression of its MSE is

$$\hat{S}_1^2 = \frac{s_y^2}{s_x^2}\sigma_x^2,$$

$$MSE(\hat{S}_1^2) = \gamma\sigma_y^2\left[(\lambda_{40}-1)+(\lambda_{04}-1)-2(\lambda_{22}-1)\right].$$

Singh et al. (2011) proposed an exponential ratio-type estimator having a single auxiliary variable for population variance. The estimator with the expression of its MSE is

$$\hat{S}_2^2 = s_y^2\exp\left(\frac{\sigma_x^2-s_x^2}{\sigma_x^2+s_x^2}\right),$$

$$MSE(\hat{S}_2^2) \approx \gamma\sigma_y^2\left[(\lambda_{40}-1)+0.25(\lambda_{04}-1)-(\lambda_{22}-1)\right].$$

Let the initial sample of $n$ units is selected with SRSWOR from a clustered population. Suppose $w_{yi}$ and $w_{xi}$ denoted the average values of $y$ and $x$ in the network which includes unit $i$ such that; $w_{yi} = m_i^{-1}\sum_{j\varepsilon A_i} y_j$ and $w_{xi} = m_i^{-1}\sum_{j\varepsilon A_i} x_j$, respectively. Adaptive cluster sampling can be considered as SRSWOR when the averages of networks are considered (Dryver and Chao 2007; Thompson 2012).

Consider the notations $\bar{w}_y$ and $\bar{w}_x$ are the sample means whereas $s_{wy}^2$ and $s_{wx}^2$ are the sample variances of the survey and auxiliary variable based on the transformed population, respectively such that

$$\bar{w}_y = \frac{1}{n}\sum_{i=1}^{s} w_{yi}, \qquad s_{wy}^2 = \frac{1}{n-1}\sum_{i=1}^{s}(w_{yi} - \bar{w}_y)^2,$$

and

$$\bar{w}_x = \frac{1}{n}\sum_{i=1}^{s} w_{xi}, \qquad s_{wx}^2 = \frac{1}{n-1}\sum_{i=1}^{s}(w_{xi} - \bar{w}_x)^2.$$

In ACS, we consider the following notations to obtain the expressions of bias and MSE of the proposed estimator. Let us define $\xi_{wy}$ and $\xi_{wx}$ are the error terms of the survey and auxiliary variables, respectively

$$\left.\begin{aligned} &\lambda_{\bar{w}rs} = \frac{\mu_{\bar{w}rs}}{\mu_{\bar{w}20}^{r/2}\ \mu_{\bar{w}02}^{s/2}},\ \mu_{\bar{w}rs} = \frac{1}{N-1}\sum_{i=1}^{N}(w_{yi}-\mu_y)^r(w_{xi}-\mu_x)^s \text{ and } \gamma_{\bar{w}} = \frac{1}{n} \\ &s_{wy}^2 = \sigma_{wy}^2(1+\xi_{wy}), s_{wx}^2 = \sigma_{wx}^2(1+\xi_{wx});\ \text{such that } E(\xi_{wy}) = E(\xi_{wx}) = 0 \\ &E(\xi_{wy}^2) = \gamma_{\bar{w}}(\lambda_{\bar{w}40}-1), E(\xi_{wx}^2) = \gamma_{\bar{w}}(\lambda_{\bar{w}04}-1), E(\xi_{wy}\xi_{wx}) = \gamma_{\bar{w}}(\lambda_{\bar{w}22}-1), \end{aligned}\right\} \quad (1)$$

where $r$ and $s$ be the non-negative integers. The quantities $\mu_{\bar{w}20}$ and $\mu_{\bar{w}02}$ be the second order moments and $\lambda_{\bar{w}pq}$ is known as moment ratio for the adaptive estimators.

Thompson (1992) defined an unbiased estimator for finite population variance in ACS based on the modification of Hansen-Hurwitz (HH) type estimator. The estimator with its variance is defined as

$$\hat{S}_3^2 = s_{wy}^2,$$

$$Var(\hat{S}_3^2) = \gamma_{\bar{w}}\sigma_{wy}^2(\lambda_{\bar{w}40} - 1).$$

Chaudhry et al. (2016) modified classical ratio and exponential ratio estimator with single auxiliary variable in ACS design as

$$\hat{S}_4^2 = \frac{s_{wy}^2}{s_{wx}^2}\sigma_{wx}^2,$$

$$\hat{S}_5^2 = s_{wy}^2 \exp\left(\frac{\sigma_{wx}^2 - s_{wx}^2}{\sigma_{wx}^2 + s_{wx}^2}\right).$$

The expressions of MSE's of $\hat{S}_4^2$ and $\hat{S}_5^2$ are given as follows:

$$MSE(\hat{S}_4^2) = \gamma_{\bar{w}}\sigma_{wy}^2\left[(\lambda_{\bar{w}40}-1) + (\lambda_{\bar{w}04}-1) - 2(\lambda_{\bar{w}22}-1)\right],$$

$$MSE(\hat{S}_5^2) = \gamma_{\bar{w}}\sigma_{wy}^2\left[(\lambda_{\bar{w}40}-1) + 0.25(\lambda_{\bar{w}04}-1) - (\lambda_{\bar{w}22}-1)\right].$$

## 4. Proposed Estimator

In this section, an estimator is developed that incorporate the exponential ratio and exponential product estimator simultaneously using probability weighting approach for the estimation of highly clumped population variance under ACS design. The estimator is defined as

$$\hat{S}_{Rp}^2 = s_{wy}^2\left[\alpha_w \exp\left(\frac{\sigma_{wx}^2 - s_{wx}^2}{\sigma_{wx}^2 + s_{wx}^2}\right) + (1-\alpha_w)\exp\left(\frac{s_{wx}^2 - \sigma_{wx}^2}{\sigma_{wx}^2 + s_{wx}^2}\right)\right],$$

where $\alpha_w$ is constant that need to be optimized and estimated for the expression of minimum value of the MSE of the proposed generalized estimator.

Using the notations (1), we may rewrite the proposed estimator as

$$\hat{S}_{Rp}^2 = \sigma_{wy}^2(1+\xi_{wy})\left[\alpha_w \exp\left(\frac{\sigma_{wx}^2 - \sigma_{wx}^2(1+\xi_{wx})}{\sigma_{wx}^2 + \sigma_{wx}^2(1+\xi_{wx})}\right) + (1-\alpha_w)\exp\left(\frac{\sigma_{wx}^2(1+\xi_{wx}) - \sigma_{wx}^2}{\sigma_{wx}^2 + \sigma_{wx}^2(1+\xi_{wx})}\right)\right]. \quad (2)$$

Applying Taylor series on (2) and expanding the exponential terms up to the second order approximation, we have

$$\hat{S}_{Rp}^2 \approx \sigma_{wy}^2(1+\xi_{wy})\left[\alpha_w\left(1 - \frac{\xi_{wx}}{2} + \frac{\xi_{wx}^2}{4} + \frac{\xi_{wx}^2}{8}\right) + (1-\alpha_w)\exp\left(1 + \frac{\xi_{wx}}{2} - \frac{\xi_{wx}^2}{4} + \frac{\xi_{wx}^2}{8}\right)\right]. \quad (3)$$

Simplifying and applying expectation on (3), we have

$$Bias(\hat{S}_{Rp}^2) = E(\hat{S}_{RR}^2 - \sigma_{wy}^2) \approx \frac{\gamma_{\overline{w}}}{8}\sigma_{wy}^2\left[(\lambda_{\overline{w}04} - 1) + 4(1-2\alpha_w)(\lambda_{\overline{w}22} - 1)\right]. \quad (4)$$

In order to obtain the expression of MSE of the proposed estimator, we may write the proposed estimator using first order approximation as

$$\hat{S}_{Rp}^2 - \sigma_{wy}^2 \approx \sigma_{wy}^2\left[\xi_{wy} + \left(\frac{1}{2} - \alpha_w\right)\xi_{wx}\right]. \quad (5)$$

The expression of MSE of the proposed estimator may be obtain by squaring and taking expectation on both sides of (5), we have

$$E(\hat{S}_{Rp}^2 - \sigma_{wy}^2)^2 \approx \gamma_{\overline{w}}\sigma_{wy}^4\left[(\lambda_{\overline{w}40} - 1) + \left(\frac{1}{2} - \alpha_w\right)^2(\lambda_{\overline{w}04} - 1) + 2\left(\frac{1}{2} - \alpha_w\right)(\lambda_{\overline{w}22} - 1)\right]. \quad (6)$$

In order to obtain the expression of minimum MSE of the proposed estimator, differentiating (6) with respect to '$\alpha_w$' we have

$$\hat{\alpha}_{w(opt)} = 0.5 + (\lambda_{\overline{w}22} - 1)(\lambda_{\overline{w}04} - 1)^{-1}.$$

The final expression of minimum MSE of the proposed estimator is

$$MSE_{min}(\hat{S}_{Rp}^2) \approx \gamma_{\overline{w}}\sigma_{wy}^4(\lambda_{\overline{w}40} - 1)(1-\rho^2),$$

where $\rho = \dfrac{(\lambda_{\overline{w}22} - 1)}{\sqrt{(\lambda_{\overline{w}40} - 1)}\sqrt{(\lambda_{\overline{w}04} - 1)}}$.

It is clearly seen that the expression of minimum MSE of the proposed estimator is same as the approximate variance of linear regression estimator for population variance. It is noticed that the proposed estimator is reduced to exponential ratio and exponential product estimator for the values '1' and '0' of the optimization constant $(\alpha w)$, respectively. Improved version of the proposed estimator

$$\hat{S}_{IRp}^2 = \eta_w s_{wy}^2\left[\alpha_w \exp\left(\frac{\sigma_{wx}^2 - s_{wx}^2}{\sigma_{wx}^2 + s_{wx}^2}\right) + (1-\alpha_w)\exp\left(\frac{s_{wx}^2 - \sigma_{wx}^2}{\sigma_{wx}^2 + s_{wx}^2}\right)\right],$$

where $\eta_w$ (i.e., $0 < \eta_w < 1$) is a suitable chosen scalars whose value is to be estimated so that MSE of $\hat{S}_{IRp}^2$ is minimized. The expressions of bias and MSE are given as, respectively

$$Bias(\hat{t}^{*}_{wGE}) \approx \sigma^2_{wy}\left[(\eta_w - 1) + \frac{\gamma_{\overline{w}}\eta_w}{8}\left[(\lambda_{\overline{w}04} - 1) + 4(1-2\alpha_w)(\lambda_{\overline{w}22} - 1)\right]\right].$$

or

$$Bias(\hat{S}^2_{wIRp}) \approx (\hat{\eta}_w - 1)\sigma^2_{wy} + \hat{\eta}_w Bias(\hat{S}^2_{wRp}) \tag{7}$$

and

$$MSE(\hat{S}^2_{wIRp}) \approx \sigma^4_{wy}\left[(\eta_w - 1)^2 + \gamma_{\overline{w}}\eta^2_w(\lambda_{\overline{w}40} - 1)(1-\rho^2)\right],$$

or

$$MSE(\hat{S}^2_{wIRp}) \approx (\eta_w - 1)^2\sigma^4_{wy} + \eta^2_w MSE(\hat{S}^2_{wRp}). \tag{8}$$

In order to obtain the minimum MSE of $\hat{S}^2_{wRp}$, differentiating (8) with respect to $\eta_w$, we have

$$\hat{\eta}_w = \frac{\sigma^4_{wy}}{\sigma^4_{wy} + MSE(\hat{S}^2_{wRp})}.$$

The expression of minimum MSE of the improved version of proposed estimator is

$$MSE_{min}(\hat{S}^2_{wIRp}) \approx (\hat{\eta}_w - 1)^2\sigma^4_{wy} + \hat{\eta}^2_w MSE(\hat{S}^2_{wRp}). \tag{9}$$

## 5. Numerical Study

Two types of numerical studies are conducted in order to deal with the situation in which the ACS estimators work more efficiently than the conventional SRS estimators. In ACS, the expected sample size usually depends on the initial sample size and varies according to the adaptation of the neighboring units. The expected sample size denoted by '$E(v)$' is the sum of the probabilities of inclusion of all the quadrates (Thompson 2012) define as

$$E(v) = \sum_{i=1}^{N} \pi_i.$$

For the fair comparison, the usual formula of sample variance for SRSWOR based on the expected sample size is

$$Var(\sigma^{2*}_y) = \gamma\sigma^4_y(\lambda_{40} - 1).$$

The general expressions of absolute relative bias (ARB) and MSE of the proposed estimator are given as respectively

$$ARB(\hat{S}^2_i) = \frac{\left|\frac{1}{r}\sum_{i=1}^{r}\hat{S}^2_i - \sigma^2_{wy}\right|}{\sigma^2_{wy}},$$

and

$$MSE(\hat{S}^2_i) = \frac{1}{r}\sum_{i=1}^{r}(\hat{S}^2_i - \sigma^2_{wy})^2,$$

where $\hat{S}^2_*$ is the value of the relevant estimators presented in this paper and $r$ is the total number of iterations. Due to the high variability in the data, one hundred thousand iterations $(r = 100{,}000)$ have been made for all the estimators to get the high accuracy on various initial and their corresponding expected sample sizes.

The evaluation of the proposed estimator over the traditional variance estimator under SRS is based on the relative efficiency (R.E.). The R.E. of the estimators considered in this paper, compared to the usual variance estimator is denoted by

$$R.E.(\hat{S}_i^2) = \frac{Var(\sigma_y^{2*})}{MSE(\hat{S}_i^2)},$$

where $i = 0,1,2,...,5,$ Rp and IRp.

### 5.1. Application using real population

A real population of blue-winged teal (Smith et al. 1995) is used to examine the performance of the proposed estimator in comparison of the other competing estimators. The total study area of five thousand km$^2$ has been divided into 50-100-km$^2$ equal quadrates in central Florida as shown in Table 1. The total number count of blue-winged teal is considered as survey variable $(y)$ whereas the auxiliary variable $(x)$ is generated using the linear model (Chutiman and Chiangpradit 2014) summarized in Table 2.

$$x_i = 4y_i - \varepsilon_i,$$

where $\varepsilon_i \sim N(0, y_i)$.

**Table 1** Blue-winged teal data, *y* (Smith et al. 1995)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 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 | 13,639 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 177 |
| 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 122 |

**Table 2** Simulated *x*-values generated from model

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 13 | 19 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 93 | 59 | 0 | 0 | 37 | 419 | 0 |
| 0 | 0 | 0 | 0 | 9 | 10 | 8 | 0 | 45,621 | 6 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 59 | 493 |
| 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 691 |

The transformed populations for both survey and auxiliary variables are summarized in Table 3 and Table 4, respectively.

**Table 3** Average values of the networks of blue-winged teal data

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 7.57 | 7.57 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 7.57 | 7.57 | 0 | 0 | 2009.4 | 2,009.4 | 0 |
| 0 | 0 | 0 | 0 | 7.57 | 7.57 | 7.57 | 0 | 2,009.4 | 2,009.4 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2,009.4 | 2,009.4 |
| 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 2,009.4 |

**Table 4** Average values of simulated auxiliary variable, $x$

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 204.14 | 204.14 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 204.14 | 204.14 | 0 | 0 | 6760.86 | 6760.86 | 0 |
| 0 | 0 | 0 | 0 | 204.14 | 204.14 | 204.14 | 0 | 6760.86 | 6760.86 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6760.86 | 6760.86 |
| 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 6760.86 |

Thompson (2012) suggested that the conventional estimators perform worse than the ACS estimator if the within-network variances are high enough as compare to the overall variance of the survey variable. The standard HH type variance estimator for ACS will be more efficient than the usual unbiased estimator for SRS if

$$S_y^2 < \eta S_{wy}^2, \tag{10}$$

where the within-network variance is defined as

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

The correlation coefficients between survey and auxiliary variables in actual and transformed population are 0.9999 and 0.9996, respectively. The variance of the survey variable is 3,716,168 whereas the variances within the network are 70.29 and 26,302,470, respectively, satisfying (10) for different comparable expected sample size given in Table 11 (Appendix B). The within-network variances accounting a large portion of the overall variance of survey variable indicating that the adaptive estimators perform better than the usual conventional estimators. In this simulation, the ACS estimators were calculated on different values of the initial sample sizes $n$ whereas, for the estimators under SRS, the size of the sample was set according to the expected sample size. The initial sample sizes of SRSWOR along with their corresponding expected sample sizes are summarized in Table 5 and Table 6 together with the amount of ARB and RE's of all the adaptive and non-adaptive estimators.

**Table 5** Absolute relative bias for all conventional and ACS estimators

| Sample Sizes | | Non-adaptive Estimators | | | Adaptive Estimators | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | $E(v)$ | $\hat{S}_0^2$ | $\hat{S}_1^2$ | $\hat{S}_2^2$ | $\hat{S}_3^2$ | $\hat{S}_4^2$ | $\hat{S}_5^2$ | $\hat{S}_{Rp}^2$ | $\hat{S}_{IRp}^2$ |
| 5 | 18.36 | 6.7586 | * | 4.5768 | 0.2954 | * | 0.7765 | 0.6798 | 0.4928 |
| 10 | 28.77 | 5.7364 | * | 3.4756 | 0.1184 | * | 0.5467 | 0.4454 | 0.1743 |
| 15 | 34.12 | 4.8763 | * | 1.9989 | 0.0475 | * | 0.2211 | 0.1232 | 0.0776 |
| 20 | 37.44 | 2.9867 | 7.7564 | 0.7634 | 0.0112 | 6.8873 | 0.0574 | 0.0765 | 0.0231 |
| 25 | 39.91 | 1.4675 | 7.2473 | 0.4487 | 0.0072 | 4.5768 | 0.0101 | 0.0454 | 0.0076 |

* : (0/0) = undefined

**Table 6** Relative efficiencies for all conventional and ACS estimators

| Sample Sizes | | Non-adaptive Estimators | | | Adaptive Estimators | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | $E(v)$ | $\hat{S}_0^2$ | $\hat{S}_1^2$ | $\hat{S}_2^2$ | $\hat{S}_3^2$ | $\hat{S}_4^2$ | $\hat{S}_5^2$ | $\hat{S}_{Rp}^2$ | $\hat{S}_{IRp}^2$ |
| 5 | 18.36 | 6.5465 | * | 11.6768 | 13.3452 | * | 102.3515 | 112.8987 | 135.4784 |
| 10 | 28.77 | 6.9365 | * | 12.1232 | 13.7768 | * | 112.1943 | 126.3324 | 151.5989 |
| 15 | 34.12 | 7.5665 | * | 12.9809 | 14.6785 | * | 138.8573 | 147.3445 | 176.8134 |
| 20 | 37.44 | 7.9684 | 5.5823 | 14.6522 | 15.8674 | 155.7869 | 151.2341 | 178.8734 | 214.6481 |
| 25 | 39.91 | 8.0024 | 5.7379 | 16.7684 | 18.5467 | 187.6512 | 154.4595 | 199.2354 | 239.0825 |

* : (0/0) = undefined

From the results of simulation study on all the adaptive and non-adaptive estimators, it can be seen that the amount of ARB's converges to zero by increasing the sample size except the ratio estimators as shown in Table 5. The R.E.'s presented in Table 6 indicating that all the adaptive estimators are more efficient than the non-adaptive estimators. The conventional ratio estimator performs better than the unbiased sample variance estimator and exponential ratio estimator in SRS for large sample sizes. Nevertheless, it is not appropriate for the estimation of rare and clustered population. Further, the proposed estimator has maximum R.E. among all the estimators considered in this paper. The adaptive ratio estimator performs better than the typical ratio estimator under SRS on the large sample.

### 5.2. Application using artificial population

A clustered population is taken from Thompson (2012), in which the area of the population have been partitioned over a study region $20\times 20 = 400$ square units as the auxiliary variable $(x)$ given in Table 9 (Appendix A). The pre-defined condition is considered as $y \geq 1$ with the same ACS process in which the immediate first-order neighboring units added to the sample. Dryver and Chao (2007) simulated the values of the survey variable using the following two models.

$$y_i = 4x_i + \varepsilon_i, \quad \text{where} \quad \varepsilon_i \sim N(0, x_i) \tag{11}$$

$$y_i = 4w_{xi} + \varepsilon_i, \quad \text{where} \quad \varepsilon_i \sim N(0, w_{xi}). \tag{12}$$

In model (11), the values of the survey variable were generated with original data whereas in model (12), the values were generated using averages of the networks of the auxiliary variable. In order to evaluate the performance of the proposed estimator, we simulated $y$-values in such a way that there exists an exponential relation between survey and auxiliary variables.

$$y_i = e^{0.4x_i + \varepsilon_i}, \quad \text{where} \quad \varepsilon_i \sim N(0, x_i). \tag{13}$$

In the given model (13), let $y_i$ be the $i^{\text{th}}$ values for the survey variable, which generated in such a way that the variability of the survey variable is exponentially proportional to the auxiliary variable given in Table 10. As the survey variable is not linearly related to the auxiliary variable, the coefficients of correlation at unit and network mean level are 0.4832 and 0.2504, respectively. The overall variance of the survey variable is 39,858,467,668 whereas the variances within the networks were found to be 89,418,237, 175,481,872 and $1{,}988{,}729\times 10^5$, respectively, satisfying (10) for different comparable expected sample size given in Appendix B. The within network variances accounting a large portion of the overall variance of survey variable indicating that the adaptive estimators performs better than the usual conventional estimators. The initial sample sizes of SRSWOR with their corresponding expected sample sizes are summarized in first two columns of Table 7 and Table 8 together with the amount of ARB and R.E.'s of all the adaptive and non-adaptive estimators.

**Table 7** Absolute relative bias for all conventional and ACS estimators

| Sample Sizes | | Non-adaptive Estimators | | | Adaptive Estimators | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | $E(v)$ | $\hat{S}_0^2$ | $\hat{S}_1^2$ | $\hat{S}_2^2$ | $\hat{S}_3^2$ | $\hat{S}_4^2$ | $\hat{S}_5^2$ | $\hat{S}_{Rp}^2$ | $\hat{S}_{IRp}^2$ |
| 10 | 26.709 | 2.7681 | * | 3.7419 | 0.0913 | * | 0.5718 | 0.4763 | 0.3312 |
| 20 | 48.779 | 1.4512 | * | 2.6131 | 0.0411 | * | 0.2520 | 0.2713 | 0.1211 |
| 30 | 67.315 | 0.3451 | * | 1.8221 | 0.0041 | * | 0.0911 | 0.0741 | 0.0053 |
| 40 | 83.175 | 0.0176 | 6.1571 | 0.2312 | 0.0000 | 4.4459 | 0.0022 | 0.0019 | 0.0001 |
| 50 | 97.019 | 0.0010 | 4.0421 | 0.0506 | 0.0000 | 2.4563 | 0.0000 | 0.0000 | 0.0000 |

* : (0/0) = undefined

**Table 8** Absolute relative bias for all conventional and ACS estimators

| Sample Sizes | | Non-adaptive Estimators | | | Adaptive Estimators | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | $E(v)$ | $\hat{S}_0^2$ | $\hat{S}_1^2$ | $\hat{S}_2^2$ | $\hat{S}_3^2$ | $\hat{S}_4^2$ | $\hat{S}_5^2$ | $\hat{S}_{Rp}^2$ | $\hat{S}_{IRp}^2$ |
| 10 | 26.709 | 9.0437 | * | 13.5695 | 14.1819 | * | 111.3515 | 155.3412 | 205.0504 |
| 20 | 48.779 | 8.0924 | * | 16.4958 | 16.9827 | * | 126.1943 | 178.7385 | 235.9348 |
| 30 | 67.315 | 9.3872 | * | 16.8617 | 18.9855 | * | 148.8573 | 190.1170 | 250.9544 |
| 40 | 83.175 | 8.1195 | 11.523 | 14.5446 | 22.3281 | 71.7362 | 163.2341 | 233.8745 | 308.7143 |
| 50 | 97.019 | 9.6644 | 12.379 | 18.3706 | 24.4533 | 78.7163 | 188.4595 | 283.1345 | 373.7375 |

* : (0/0) = undefined

From the results given in Table 7, it can be seen that the amount of ARB's converges to zero as the sample size increases. Lohr (1999) recommended that the amount of bias converges to zero by increasing the sample size in ACS. The R.E.'s presented in Table 8 indicates that all the adaptive estimators are more efficient than the non-adaptive estimators in terms of R.E.'s. The usual sample variance estimator under ACS design performs slightly better than the ratio and exponential ratio estimator under SRS design. Further, the proposed estimator and its improved version have maximum *RE* among all the adaptive and non-adaptive estimators considered in this paper.

## 6. Conclusions

Unlike existing estimators, our proposed estimator performs considerably better for the estimation of clustered population variance. The results of the simulation study indicate that the estimators under ACS design perform better than the estimators under conventional sampling design on various initial and their corresponding expected sample sizes. Dryver and Chao (2007) assumed that 0/0 as zero for ratio estimators. In this simulation, 0/0 is not considered as zero, therefore the conventional and adaptive ratio estimators did not perform and return no value (*) for small sample sizes. The within-network variances of the survey variable in both populations accounting a large portion of the overall variance indicating that the conventional estimators perform worse than the adaptive estimators. Further, the amount of ARB tends to zero and the proposed estimator performs better as the sample size increases. Thus, it is suggested that the proposed estimator is useful for variance estimation of clustered populations like rare and clustered species, fisheries, flocking, HIV and AIDS patients.

## References

Chao CT. Ratio estimation on adaptive cluster sampling. J Chinese Stat Assoc. 2004a; 42(3): 307-327.

Chaudhry MS, Tahir S, Hanif M. Ratio estimators for population variance in adaptive cluster sampling. Pakistan J Sci. 2016; 68(3): 308-314.

Chutiman N, Chiangpradit M. Ratio estimator in adaptive cluster sampling without replacement of networks. J Prob Stat. 2014; Article ID 726398, 6 pages http://dx.doi.org/10.1155/2014/726398.

Dryver AL, Chao CT. Ratio estimators in adaptive cluster sampling. Environmetrics. 2007; 18: 607-620.

Hansen MM, Hurwitz WN. On the theory of sampling from finite population. Ann Math Stat. 1943; 14: 333-362.

Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. J Am Stat Assoc. 1952; 47: 663-685.

Isaki C. Variance estimation using auxiliary information. J Am Stat Assoc. 1983; 78: 117-123.

Lohr SL. Sampling: design and analysis. Pacific Grove: Duxbury Press; 1999.

Singh R, Chauhan P, Sawan N, Smarandache F. Improved exponential estimator for population variance using two auxiliary variables. Ital J Pure Appl Math. 2011; 28: 101-108.

Smith DR, Conroy MJ, Brakhage DH. Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. Biometrics. 1995; 51: 777-788.

Thompson SK. Adaptive cluster sampling. J Am Stat Assoc. 1990; 85: 1050-1059.

Thompson SK. Adaptive cluster sampling; designs with primary and secondary units. Biometrika. 1991a; 47(3): 1103-1115.

Thompson SK. Stratified adaptive cluster sampling. Biometrika. 1991b; 78(2): 389-397.

Thompson SK. Sampling. 1st ed. John Wiley & Sons: New York; 1992.

Thompson SK. Sampling. 3rd ed. John Wiley & Sons: New York; 2012.

**Appendix A**

**Table 9** Auxiliary variable *x* (Thompson 2012) 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 10** Simulated $y$-values using model (13)

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **275** | **22,017** | **7** | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **6,632** | **392,788** | **8** | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **5** | **11** | **3** | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | **7** | **6** | **10** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **13** | **84** | **32,862** | **7** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **13** | **6,635** | **2,001** | **17** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | **9** | **15** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | **7** | **2** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | **903** | **32,853** | **41** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | **11** | **54** | **32,855** | **18** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | **5** | **3** | **7** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

## Appendix B

**Table 11** Thompson efficiency condition for real and artificial populations

| Real Population | | | | Artificial Population | | | |
|---|---|---|---|---|---|---|---|
| $n$ | $E(v)$ | $S_y^2$ | $\eta S_{wy}^2$ | $n$ | $E(v)$ | $S_y^2$ | $\eta S_{wy}^2$ |
| 5 | 18.93 | 3,716,168 | 846,002,409 | 10 | 26.769 | 39,858,550,873 | $239{,}532\times10^{7}$ |
| 10 | 28.77 | 3,716,168 | 1,730,250,857 | 20 | 48.779 | 39,858,550,873 | $7{,}306{,}704\times10^{6}$ |
| 15 | 34.12 | 3,716,168 | 2,169,959,542 | 30 | 67.315 | 39,858,550,873 | $1{,}249{,}543\times10^{7}$ |
| 20 | 37.44 | 3,716,168 | 2,092,823,318 | 40 | 83.175 | 39,858,550,873 | $168{,}845\times10^{8}$ |
| 25 | 39.91 | 3,716,168 | 1,811,654,534 | 50 | 97.019 | 39,858,550,873 | $2{,}120{,}333\times10^{8}$ |