

Generalized Exponential Estimator for the Estimation of Clustered Population Variance in Adaptive Cluster Sampling

Muhammad Nouman Qureshi^{1,3,*}, Ayesha Iftikhar^{2,3}, Muhammad Hanif³

¹ School of Statistics, University of Minnesota, Minneapolis, 55455, USA

² Lahore Business School, University of Lahore, Lahore, 54000, Pakistan

³ Department of Statistics, National College of Business Administration and Economics, Lahore, 54000, Pakistan

ARTICLE INFO

Article History

Received 02 Dec 2018

Accepted 28 Oct 2019

Keywords

Adaptive cluster sampling

Auxiliary information

Exponential estimators

Hansen-Hurwitz estimation

Simulated population and Poisson
clustered process

Mathematics Subject Classification:
62D05

ABSTRACT

In this paper, we proposed a generalized exponential estimator with two auxiliary variables for the estimation of highly clumped population variance under adaptive cluster sampling design. The expressions of approximate bias and minimum mean square error are derived. A family of exponential ratio and exponential product estimator is obtained by using different values of generalized and optimized constants. A numerical study is carried out on real and artificial populations to examine the performance of the proposed estimator over the competing estimators. Related results show that the proposed generalized exponential estimator is able to provide considerably better results over the competing estimators for the estimation of rare and highly clustered population variance.

© 2019 The Authors. Published by Atlantis Press SARL.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. INTRODUCTION

Adaptive cluster sampling (ACS), first proposed by Thompson [1] is considered to be the most capable design for providing summary information on occurrence, structure and abundance of rare and clustered populations. The main feature of ACS is that it can select more meaningful samples and provide more efficient estimates for the field investigators to estimate the rare and clustered population parameters (total, mean, variance, proportion) as compare to the other conventional sampling designs. Auxiliary information is commonly used to improve the efficiency of the estimators. In many environmental surveys, the presence of rare species in some communities is always associated with some other community densities which are more readily observable than the presence of a rare species. So the presence of such community abundance can be used as an auxiliary variable to inform sampling of rare species.

Various authors used ACS at design stage and estimation stage in the past and applied it in many real life situations. Thompson [1–3] introduced different techniques of ACS design in which the initial sample is selected by some classical sampling designs (such as stratified sampling, strip and systematic sampling) and proposed the standard ACS estimators based on the modification of Horvitz-Thompson (HT) [4] and Hansen-Hurwitz (HH) [5] type estimators. Chao [6,7] is considered to be the first, who utilized auxiliary information to propose the ratio and regression estimators based on the modification of HT estimator and discussed the confidence interval as well. Dryver and Chao [8] suggested a ratio estimator based on the modification of HH type estimators in ACS and concluded that the adaptive estimators perform considerably better than traditional ratio estimator if the variability of the survey variable is proportional to the within-network means level of the auxiliary variable.

In this article, we proposed a generalized exponential estimator having two auxiliary variables for the estimation of rare and highly clustered population variance under ACS design. The methodology of ACS design along with some existing estimators in SRS and ACS are presented in Section 2. The expressions of approximate bias and minimum mean square error (MSE) of the proposed estimator are derived in Section 3. Some special cases are also discussed in the same section in which the proposed estimator reduces to the several exponential ratio and exponential product-type estimators for different values of generalize and optimize constants. A numerical study is conducted in Section 4 based on the real data application taken from Smith *et al.* [9] together with the artificial clustered populations generated by Poisson cluster process (Diggle [10]) at different level of rarity and aggregation. Final summary and remarks on the paper are given in Section 5.

*Corresponding author. Email: queri006@umn.edu

2. METHODOLOGY OF ACS

The ACS design typically starts with random sample of size n , which can be taken by simple random sampling (Thompson [1]), stratified sampling (Thompson [3]; Brown [11]), strip or systematic sampling (Thompson [2]; Acharya *et al.* [12]) method and extended it straightforwardly to the ACS.

Prior to the ACS procedure, a threshold value is chosen, say C , and if any of the initially selected sample meet or exceed this threshold, i.e., $y_i \geq C$, the spatially surrounding (denoted as north $\uparrow$, south $\downarrow$, east $\rightarrow$ and west $\leftarrow$) units will be added to the sample and investigated. If any of these neighboring units meet this threshold, their neighboring units will be selected and so on. The final sample is the collection of clusters that consist of the units detected in the initial sample, including the units that meet the threshold. A collection of initially sampled units along with the neighborhood units that met the threshold is called a network. The units that found below the threshold are called edge units. Together, a network with its edge units makes up a cluster. Any initially sampled unit that does not meet the condition is considered to be a network of size one.

Figure 1 illustrates the idea of ACS with condition $\{y; y \geq 1\}$. The total study region is divided into $5 \times 10 = 50$ quadrants in which the population is divided into three clusters that satisfying C . The units having star (*) are initially selected units and all the shaded units are adapted units that satisfied C . The bold units are known as edge units that does not meet the threshold. There are 37 networks of size one including edge units. Each cluster is based on a network (shaded) and its associated edge units (bold). Thus the first network based 5 units and 8 edge units. Second network has also 5 units but 7 edge units. Network 3 has 3 units and 5 edge units.

As in the usual finite population sampling situation, the population is considered to consist of N units that labeled from 1 to N , denoted as $u = \{1, 2, 3, \dots, N\}$. Associated with each unit i , the values of survey variable together with the auxiliary variables are denoted as y_i , x_i and z_i respectively. Let the population is divided into K ($k = 1, 2, \dots, K$) exhaustive networks (A_k) and A_i denotes the i^{th} network having m_i units. The data set d is selected by using initial sample size from the sampled units say s , where $s = \{i_1, i_2, \dots, i_n\}$. Hence the data set is $d = \{s, y_s, x_s, z_s\}$. The population mean and variance of the study variable and auxiliary variables may be computed as, respectively

$$\mu_y = \frac{1}{N} \sum_{i=1}^N y_i, \mu_x = \frac{1}{N} \sum_{i=1}^N x_i \text{ and } \mu_z = \frac{1}{N} \sum_{i=1}^N z_i,$$

and

$$\sigma_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \mu_y)^2, \sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2 \text{ and } \sigma_z^2 = \frac{1}{N-1} \sum_{i=1}^N (z_i - \mu_z)^2.$$

In ACS the initial sample of size n units is selected with simple random sampling without replacement (SRSWOR). Let w_{yj} , w_{xj} and w_{zj} defined as new variables of the survey and auxiliary variables in the i^{th} network having m_i units, which may be estimated as, respectively.

$$w_{y_j} = \frac{1}{m_i} \sum_{j \in A_i} y_j, w_{x_j} = \frac{1}{m_i} \sum_{j \in A_i} x_j, \text{ and } w_{z_j} = \frac{1}{m_i} \sum_{j \in A_i} z_j,$$

ACS can be considered as SRSWOR if the average values of the networks are used instead of original values (Thompson [13]; Dryver and Chao [8]). Consider the notations of $\bar{w}_y$, $\bar{w}_x$ and $\bar{w}_z$ are the sample means whereas s_{wy}^2 , s_{wx}^2 and s_{wz}^2 are the sample variances of the survey and auxiliary variables based on the transformed populations respectively, such that

$$\bar{w}_y = \frac{1}{n} \sum_{i=1}^s w_{yi}, \bar{w}_x = \frac{1}{n} \sum_{i=1}^s w_{xi} \text{ and } \bar{w}_z = \frac{1}{n} \sum_{i=1}^s w_{zi}.$$

and

$$s_{wy}^2 = \frac{1}{n-1} \sum_{i=1}^s (w_{yi} - \bar{w}_y)^2, s_{wx}^2 = \frac{1}{n-1} \sum_{i=1}^s (w_{xi} - \bar{w}_x)^2 \text{ and } s_{wz}^2 = \frac{1}{n-1} \sum_{i=1}^s (w_{zi} - \bar{w}_z)^2.$$

0	0	0	0	0	0	0	0	0	0
0	23*	2	0	0	0	0*	0	0	7
1	62	0	0	27	0	0	0	12*	1
0	4	0	0	3105	9	0	0	0	0
0	0	0	2	10*	0	0	0	0	1*

Figure 1 | The example of 3 clusters according to the condition.

In ACS, we consider the following notations to obtain the expressions of bias and MSE of the proposed estimators. Let us define ξ_{wy} , ξ_{wx} and ξ_{wz} are the error terms of the survey variable and auxiliary variables

$$\left. \begin{aligned} \lambda_{wpr} &= \frac{\mu_{wpr}^{-pqr}}{\mu_{w200}^{p/2} \mu_{w020}^{q/2} \mu_{w002}^{r/2}}, \mu_{wpr}^{-pqr} = \frac{1}{N-1} \sum_{i=1}^N (w_{yi} - \mu_y)^p (w_{xi} - \mu_x)^q (w_{zi} - \mu_z)^r \\ s_{wi}^2 &= \sigma_{wi}^2 (1 + \xi_{wi}) \text{ and } E(\xi_{wi}) = 0, E(\xi_{wy}^2) = \gamma(\lambda_{w400} - 1), E(\xi_{wx}^2) = \gamma(\lambda_{w040} - 1), E(\xi_{wz}^2) = \gamma(\lambda_{w004} - 1) \\ E(\xi_{wy}\xi_{wx}) &= \gamma(\lambda_{w220} - 1), E(\xi_{wy}\xi_{wz}) = \gamma(\lambda_{w202} - 1), E(\xi_{wx}\xi_{wz}) = \gamma(\lambda_{w022} - 1) \text{ and } \gamma = \frac{1}{n} \text{ where } \begin{matrix} i = y, x, z \\ j = y, x, z \end{matrix} \begin{matrix} i \neq j \end{matrix} \end{aligned} \right\}. \quad (1)$$

The traditional sample variance estimator of the population variance in SRS is defined as $\hat{S}_{w0}^2 = s_y^2$. The estimators $\hat{S}_{w1}^2$ and $\hat{S}_{w2}^2$ are ratio and exponential ratio estimators with two auxiliary variables in SRS design. The standard unbiased variance estimator without having auxiliary variable in ACS is define as $\hat{S}_{w3}^2 = s_{wy}^2$ whereas $\hat{S}_{w4}^2$ and $\hat{S}_{w5}^2$ are ratio and exponential ratio estimators with single auxiliary variable.

3. GENERALIZED EXPONENTIAL ESTIMATOR

In this section, motivated by Khan [14], we proposed a generalized exponential type estimator for the estimation of rare and highly clustered population variance in ACS. The exponential estimator with two auxiliary variables is

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

which can be generalize by introducing the positive real constant “a and b” ($a \geq 0, b \geq 0$) and α and β which can assume values $-1, 0$ and 1 . The generalized estimator is

$$\hat{S}_{wGE}^2 = s_{wy}^2 \exp \left[\alpha \left\{ 1 - \frac{as_{wx}^2}{\sigma_{wx}^2 + (a-1)s_{wx}^2} \right\} \right] \exp \left[\beta \left\{ 1 - \frac{bs_{wz}^2}{\sigma_{wz}^2 + (b-1)s_{wz}^2} \right\} \right]. \quad (2)$$

In order to obtain the expression of approximate bias, we may re-write (2) using the notations given in (1) as

$$\hat{S}_{wGE}^2 \approx \sigma_{wy}^2 (1 + \xi_{wy}) \exp \left[\alpha \frac{\sigma_{wx}^2 - \sigma_{wx}^2 (1 + \xi_{wx})}{\sigma_{wx}^2 + (a-1)\sigma_{wx}^2 (1 + \xi_{wx})} \right] \exp \left[\beta \frac{\sigma_{wz}^2 - \sigma_{wz}^2 (1 + \xi_{wz})}{\sigma_{wz}^2 + (b-1)\sigma_{wz}^2 (1 + \xi_{wz})} \right].$$

Simplifying and applying Taylor series and taking expectation on both sides, we have

$$E(\hat{S}_{wGE}^2 - \sigma_{wy}^2) \approx \gamma \sigma_{wy}^2 (\lambda_{w040} - 1) \left[\frac{\alpha}{a} \left(\frac{a-1}{a} + \frac{\alpha}{2a} \right) + \frac{\beta}{b} \left(\frac{b-1}{b} + \frac{\beta}{2b} \right) f_{004}^x - \frac{\alpha}{a} f_{220}^x - \frac{\beta}{b} f_{202}^x \right],$$

where $f_{004}^x = (\lambda_{w004} - 1) / (\lambda_{w040} - 1)$, $f_{220}^x = (\lambda_{w220} - 1) / (\lambda_{w040} - 1)$ and $f_{202}^x = (\lambda_{w202} - 1) / (\lambda_{w040} - 1)$.

The final expression of approximate bias of the proposed generalized exponential estimator is

$$\text{bias}(\hat{S}_{wGE}^2) \approx \frac{\gamma}{2} \sigma_{wy}^2 (\lambda_{w040} - 1) [a^{-1} \phi_1 \{2(a-1) + \alpha\} - 2af_{220}^x] + b^{-1} \phi_2 \{2(b-1) + \beta\} f_{004}^x - 2bf_{202}^x],$$

where $\phi_1 = \alpha/a$ and $\phi_2 = \beta/b$.

In order to obtain the expression of MSE, we may re-write the (2) using first-order approximation as

$$\hat{S}_{wGE}^2 \approx \sigma_{wy}^2 (1 + \xi_{wy}) (1 - \phi_1 \xi_{wx} - \phi_2 \xi_{wz}).$$

Simplifying, squaring and taking expectation on both sides, we have

$$E(\hat{S}_{wGE}^2 - \sigma_{wy}^2)^2 \approx \sigma_{wy}^4 E(\xi_{wy} - \phi_1 \xi_{wx} - \phi_2 \xi_{wz})^2.$$

The expression of MSE of the proposed generalized exponential type estimator is

$$MSE_{\min}(\hat{S}_{wGE}^2) \approx \gamma \sigma_{wy}^4 (\lambda_{w400} - 1) [1 + \phi_1^2 f_{040}^y + \phi_2^2 f_{004}^y - 2\phi_1 f_{220}^y - 2\phi_2 f_{202}^y + 2\phi_1 \phi_2 f_{022}^y], \quad (3)$$

where $f_{040}^y = \frac{(\lambda_{w040} - 1)}{(\lambda_{w400} - 1)}$, $f_{004}^y = \frac{(\lambda_{w004} - 1)}{(\lambda_{w400} - 1)}$, $f_{220}^y = \frac{(\lambda_{w220} - 1)}{(\lambda_{w400} - 1)}$, $f_{202}^y = \frac{(\lambda_{w202} - 1)}{(\lambda_{w400} - 1)}$ and $f_{022}^y = \frac{(\lambda_{w022} - 1)}{(\lambda_{w400} - 1)}$.

In order to obtain the expression of minimum MSE, differentiating (3) with respect to ϕ_1 and ϕ_2 , we have

$$\hat{\phi}_{1(opt)} = \frac{(\lambda_{w220}^- - 1)(\lambda_{w004}^- - 1) - (\lambda_{w202}^- - 1)(\lambda_{w022}^- - 1)}{(\lambda_{w040}^- - 1)(\lambda_{w004}^- - 1)(1 - \rho_{wxz}^2)} \text{ and } \hat{\phi}_{2(opt)} = \frac{(\lambda_{w202}^- - 1)(\lambda_{w040}^- - 1) - (\lambda_{w220}^- - 1)(\lambda_{w022}^- - 1)}{(\lambda_{w040}^- - 1)(\lambda_{w004}^- - 1)(1 - \rho_{wxz}^2)},$$

$$\text{where } \rho_{wxz} = \frac{(\lambda_{w022}^- - 1)}{\sqrt{(\lambda_{w040}^- - 1)}\sqrt{(\lambda_{w004}^- - 1)}}.$$

Final expression of the minimum MSE of the proposed generalized exponential type estimator is

$$MSE_{\min}(\hat{S}_{wGE}^2) \approx \gamma \sigma_{wy}^4 (\lambda_{w400}^- - 1) \left[1 + \hat{\phi}_{1(opt)}^2 f_{040}^y + \hat{\phi}_{2(opt)}^2 f_{004}^y - 2\hat{\phi}_{1(opt)} f_{220}^y - 2\hat{\phi}_{2(opt)} f_{202}^y + 2\hat{\phi}_{1(opt)}\hat{\phi}_{2(opt)} f_{022}^y \right].$$

It is noticed that we may get various forms of exponential ratio and exponential product-type estimators for different values of a , b , α and β as new family of the proposed estimator summarized in Appendix A. Here we consider only two special cases of the proposed estimator.

i. $a = 2$, $b = 2$, $\alpha = 1$ and $\beta = 1$.

$$\hat{S}_{wrr}^2 = s_{wy}^2 \exp \left[1 - \frac{2s_{wx}^2}{\sigma_{wx}^2 + s_{wx}^2} \right] \exp \left[1 - \frac{2s_{wz}^2}{\sigma_{wz}^2 + s_{wz}^2} \right].$$

ii. $a = 1$, $b = 1$, $\alpha = 1$ and $\beta = 1$.

$$\hat{S}_{wRR}^2 = s_{wy}^2 \exp \left[1 - \frac{s_{wx}^2}{\sigma_{wx}^2} \right] \exp \left[1 - \frac{s_{wz}^2}{\sigma_{wz}^2} \right].$$

4. PERFORMANCE EVALUATION

In this section, we set up a numerical study to examine the performance of the proposed estimator over the competing estimators on different initial and their corresponding expected sample sizes. The threshold value is chosen as 1 with the same ACS process in which the immediate first-order neighboring units added to the sample. In ACS, the expected sample size is usually depends on the initial sample size and varies according to the adaptation of neighboring units of the initial sample size. The expected sample size denoted by “ $E(v)$ ” is the sum of the probabilities of inclusion of all the quadrates (Thompson [13]) define as

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

$$\text{Note that: } \pi_i = 1 - \binom{N - m_i}{n} / \binom{N}{n}.$$

The general expression for the computation of absolute relative bias (ARB) and MSE of all the estimators are given as

$$ARB(\hat{S}_{wi}^2) = \frac{|\sum_{i=1}^R \hat{S}_{wi}^2 - \sigma_{wy}^2|}{\sigma_{wy}^2},$$

and

$$MSE(\hat{S}_{wi}^2) = \frac{1}{R} \sum_{i=1}^R (\hat{S}_{wi}^2 - \sigma_{wy}^2)^2,$$

where $\hat{S}_{wi}^2$ is the value of the relevant estimators presented in this paper and ‘ R ’ is the total number of iterations.

The evaluation of the proposed estimators over the sample variance estimator under SRS is based on the relative efficiency (RE), denoted as

$$RE(\hat{S}_{wi}^2) = \frac{Var(\hat{S}_{w0}^2)}{MSE(\hat{S}_{wi}^2)},$$

where $i = 1, 2, 3, 4, 5$, rr , RR , and GE .

4.1. Simulation Study

Two rarely clustered artificial populations were generated as a realization of the Poisson cluster process over a study region partitioned into $20 \times 20 = 400$ squared units at different level of rarity and aggregation. Under this process the number of clusters (parents) were randomly distributed within the whole study area and each individual unit (offspring) were randomly placed around each cluster centers. For the simulation, the average number of parents (λ_1) is fixed as 5 and the average number of offspring (λ_2) is to be chosen 500.

In this simulation study, we generated the values of the survey variable using *ecodist*-package in R in such a way that the variability of the survey variable is exponentially proportional to the auxiliary variables (x and z). Figure 2 illustrates the behavior of the initial clustered populations of both auxiliary variables x and z . One hundred thousand Monte Carlo simulated ACS and SRSWOR samples were generated to compute the values of the estimators considered in this paper. The population size is 400 and the initial sample varies from 5 to 100. The amount of ARBs and REs of all the estimators considered in this paper are summarized in Tables 1 and 2 on different initial and their corresponding expected sample sizes.

The results summarized in Tables 1 and 2 indicates that the ACS estimators perform considerably better than the conventional sample variance, ratio and exponential ratio estimators in SRS design. The generalized exponential estimator performs considerably better than all the adaptive and non-adaptive estimators in term of RE. The classical ratio and exponential ratio estimators under SRS did not perform well even for the large sample due to the highly clustered behavior of populations. The amount of ARBs for all the estimators converges to zero by increasing the sample size.

4.2. Real Data Application

Two real data of waterfowl species has been taken from Smith *et al.* [9] to examine the performance of the proposed estimator over the sample variance estimator under conventional sampling design. The total area under study covered a region of 200 25-km² quadrates in

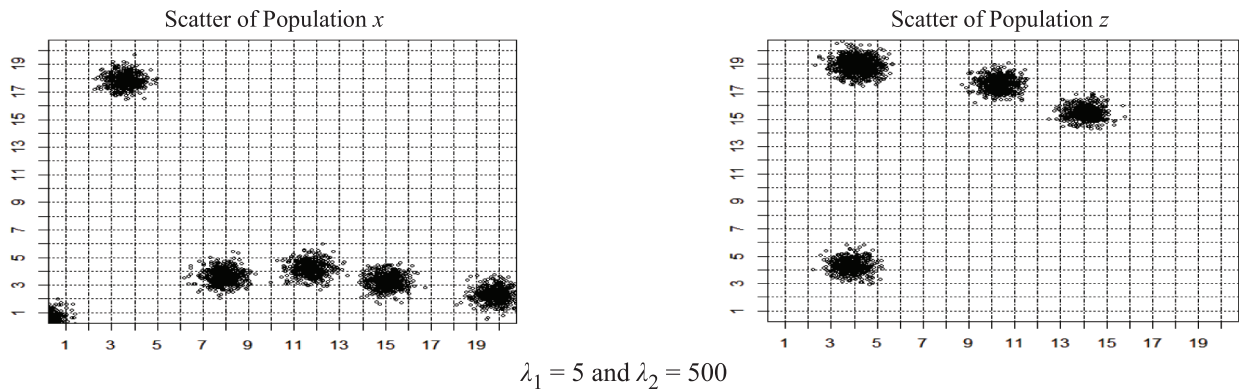


Figure 2 | Clustered populations for x and z generated by Poisson cluster process.

Table 1 | Amount of ARB's of SRS and ACS estimators for $\lambda_1 = 5$ and $\lambda_2 = 500$.

n	$E(v)$	$\hat{S}_{w1}^2$	$\hat{S}_{w2}^2$	$\hat{S}_{w3}^2$	$\hat{S}_{w4}^2$	$\hat{S}_{w5}^2$	$\hat{S}_{wrr}^2$	$\hat{S}_{wRR}^2$	$\hat{S}_{wGE}^2$
5	26.76	1.01	0.61	0.42	0.01	0.29	1.24	1.08	0.84
10	57.70	0.76	0.40	0.33	0.00	0.19	0.36	0.29	0.21
20	92.86	0.51	0.25	0.21	0.00	0.12	0.16	0.12	0.10
50	144.43	0.32	0.22	0.11	0.00	0.09	0.08	0.06	0.04
75	166.98	0.21	0.27	0.07	0.00	0.06	0.05	0.03	0.02
100	187.17	0.14	0.30	0.06	0.00	0.04	0.03	0.02	0.01

ARB, absolute relative bias; SRS, simple random sampling; ACS, adaptive cluster sampling.

Table 2 | Amount of RE's of SRS and ACS estimators for $\lambda_1 = 5$ and $\lambda_2 = 50$.

n	$E(v)$	$\hat{S}_{w1}^2$	$\hat{S}_{w2}^2$	$\hat{S}_{w3}^2$	$\hat{S}_{w4}^2$	$\hat{S}_{w5}^2$	$\hat{S}_{wrr}^2$	$\hat{S}_{wRR}^2$	$\hat{S}_{wGE}^2$
5	26.76	0.01	0.58	0.72	0.87	0.97	1.87	1.98	2.21
10	57.70	0.01	0.78	0.89	1.20	1.23	2.26	2.41	2.86
20	92.86	0.02	0.94	1.00	1.91	2.11	3.56	3.88	4.56
50	144.43	0.04	0.96	1.39	2.56	2.72	5.11	5.91	6.43
75	166.98	0.07	0.91	1.62	2.76	2.98	7.01	7.63	8.98
100	187.17	0.11	0.92	1.41	3.26	3.51	8.11	8.97	10.55

RE, relative efficiency; SRS, simple random sampling; ACS, adaptive cluster sampling.

Central Florida. The counts of Blue-winged teal (x) and Ring-necked teal (z) as shown in Figures 3 and 4 are considered to be the auxiliary variables.

In this simulation, we generate the values of the survey variable using the linear combination of the auxiliary variables, as

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

The variability of the survey variable is directly proportional to the sum of the linear combination of the auxiliary variables.

It is found that there are only 2 networks that satisfying C as shown in Figure 5. The overall variance of the survey variable is 81231457 and the two networks consisting of size 7 contains the within-network variances 1042.286 and 574239554.1 respectively. The within-network variances accounting a large portion of the overall variance indicating that the ACS estimators will perform better than the estimators under conventional sampling design. The transformed populations of survey and auxiliary variables are summarized in Figures 6–8.

Due to the high variability in the data, one hundred thousand iterations have been looped for all the estimators on various initial and their corresponding expected sample sizes. The amount of ARB and REs of all the adaptive and non-adaptive estimators are given in Tables 3 and 4.

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

Figure 3 | Blue-winged teal population (x).

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	35	75	0
0	0	0	0	0	0	0	0	2255	13
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	24

Figure 4 | Green-winged teal population (z).

0	0	12	20	0	0	0	0	0	0
0	0	0	93	54	0	0	173	713	0
0	0	0	0	8	11	9	0	63770	4
0	0	0	0	0	0	0	0	52	486
0	0	0	0	0	0	7	0	0	820

Figure 5 | Simulated y -values using model (3).

0	0	7.5	7.5	0	0	0	0	0	0
0	0	0	7.5	7.5	0	0	2009.4	2009.4	0
0	0	0	0	7.5	7.5	7.5	0	2009.4	2009.4
0	0	0	0	0	0	0	0	2009.4	2009.4
0	0	0	0	0	0	2	0	0	2009.4

Figure 6 | Transformed population of blue-winged teal (w_x).

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	343.14	343.14	0
0	0	0	0	0	0	0	0	343.14	343.14
0	0	0	0	0	0	0	0	343.14	343.14
0	0	0	0	0	0	0	0	0	343.14

Figure 7 | Transformed population of green-winged teal (w_z).

0	0	29.57	29.57	0	0	0	0	0	0
0	0	0	29.57	29.57	0	0	9429.7	9429.7	0
0	0	0	0	29.57	29.57	29.57	0	9429.7	9429.7
0	0	0	0	0	0	0	0	9429.7	9429.7
0	0	0	0	0	0	7	0	0	9429.7

Figure 8 | Transformed population of survey variable (w_y).

Table 3 | Amount of absolute relative bias conventional and ACS estimators.

n	$E(v)$	$\hat{S}_{w1}^2$	$\hat{S}_{w2}^2$	$\hat{S}_{w3}^2$	$\hat{S}_{w4}^2$	$\hat{S}_{w5}^2$	$\hat{S}_{wrr}^2$	$\hat{S}_{wRR}^2$	$\hat{S}_{wGE}^2$
2	9.80	28.61	17.34	0.08	2.82	2.41	2.23	2.10	2.04
5	18.90	22.44	12.67	0.06	2.26	1.82	1.63	1.23	1.23
10	28.76	19.65	8.43	0.07	1.87	1.22	1.17	1.08	1.00
15	34.13	13.40	5.84	0.06	1.42	1.17	1.10	0.87	0.79
20	37.44	8.80	4.69	0.02	1.21	1.09	1.03	0.70	0.64
25	39.91	6.49	3.20	0.03	1.01	0.83	0.77	0.49	0.41

ACS, adaptive cluster sampling.

Table 4 | Amount of relative efficiency conventional and ACS estimators.

n	$E(v)$	$\hat{S}_{w1}^2$	$\hat{S}_{w2}^2$	$\hat{S}_{w3}^2$	$\hat{S}_{w4}^2$	$\hat{S}_{w5}^2$	$\hat{S}_{wrr}^2$	$\hat{S}_{wRR}^2$	$\hat{S}_{wGE}^2$
2	9.80	0.00	1.42	90.76	107.10	153.23	287.54	359.54	383.18
5	18.90	0.03	1.73	103.75	116.27	173.61	316.89	611.20	636.52
10	28.76	0.04	2.03	105.21	130.07	186.77	667.43	926.54	987.38
15	34.13	0.03	3.33	106.63	173.45	237.15	915.93	1533.76	1611.65
20	37.44	0.04	5.57	116.04	194.43	248.12	1209.65	1746.02	1804.44
25	39.91	0.03	5.63	125.23	210.43	264.64	1506.06	1896.32	2151.65

ACS, adaptive cluster sampling.

The simulation results of $ARBs$ and REs summarized in Tables 3 and 4 support the findings obtained from the artificial populations. The REs of the estimators, $\hat{S}_{wrr}^2$, $\hat{S}_{wRR}^2$ and $\hat{S}_{wGE}^2$ is increasing and amount of $ARBs$ of all the adaptive and non-adaptive estimators decline to zero by increasing the sample sizes as shown in Tables 3 and 4. The typical ratio estimator under conventional SRS design did not perform well due to the clustered tendency in population units. The exponential estimators, $\hat{S}_{wrr}^2$ and $\hat{S}_{wRR}^2$ performs better than the sample variance, ratio and exponential ratio estimators having single auxiliary variable under ACS design. The proposed estimator $\hat{S}_{wGE}^2$ works significantly among all the adaptive and non-adaptive estimators.

5. CONCLUSION

As the numerical study illustrated in Section 4, we infer that the proposed generalized exponential estimator is able to provide better results for the estimation of rare and clustered population variance at different level of rarity and aggregation. The performance of the proposed estimator is better than all the adaptive and non-adaptive estimators considered in this paper. In simulation study performed on artificial populations, it is observed that the estimators, $\hat{S}_{wrr}^2$, $\hat{S}_{wRR}^2$ and $\hat{S}_{wGE}^2$ performs well on various sample sizes. The conventional estimators did not perform well due to the uneven and clustered behavior in population units. The amount of ARB of all the estimators converges to zero as the sample size increases. In real data application, the results support the findings obtained from the simulation study on artificial populations. The proposed estimator found to be highly efficient as compare to the other estimators on different initial and their corresponding expected sample sizes. The within-network variances of survey variable accounting a large portion of the overall variance indicated that the adaptive estimators are expected to perform better and more efficient than the comparable non-adaptive estimators.

CONFLICT OF INTEREST

The author declares that there is no conflict of interests regarding the publication of this paper.

Funding Statement

The authors received no financial support for the research, authorship, and/or publication of this article.

ACKNOWLEDGMENTS

The first author is thankful to the Higher Education Commission of Pakistan for awarding International Research Support Fellowship. The authors are grateful to the Editor in Chief and anonymous referees for their careful reading and constructive suggestions.

REFERENCES

1. S.K. Thompson, *J. Am. Stat. Assoc.* 85 (1990), 1050–1059.
2. S. Thompson, *Biometrika*. 47 (1991), 1103–1115.
3. S. Thompson, *Biometrika*. 78 (1991), 389–397.
4. D.G. Horvitz, D.J. Thompson, *J. Am. Stat. Assoc.* 47 (1952), 663–685.
5. M.H. Hansen, W.N. Hurwitz, *Ann. Math. Stat.* 14 (1943), 333–362.
6. C.T. Chao, *J. Chin. Stat. Assoc.* 42 (2004), 307–327.
7. C.T. Chao, Regression Estimation in Adaptive Cluster Sampling; Estimator, Confidence Interval and Sample Selection. National Science Council Report, 2004. NSC 92-2118-M-006-002.
8. A.L. Dryver, C.T. Chao, *Environmetrics*. 18 (2007), 607–620.
9. D.R. Smith, M.J. Conroy, D.H. Brakhage, *Biometrics*. 51 (1995), 777–788.
10. P.J. Diggle, Statistical Analysis of Spatial Point Patterns, Academic Press, London, 2013, <https://www.crcpress.com/Statistical-Analysis-of-Spatial-and-Spatio-Temporal-Point-Patterns/Diggle/p/book/9781466560239>.
11. J.A. Brown, *Aust. N. Z. J. Stat.* 41 (1999), 395–404.
12. G. Acharya, *Ecol. Econ.* 35 (2000), 63–74.
13. S. Thompson, *Sampling*, third ed., Wiley, New York, 2012.
14. M. Khan, *J. Probab. Stat.* 2016 (2016), 6.

APPENDIX A

Family of the proposed generalized exponential estimator

<i>Estimators</i>	α	β	a	b	<i>Estimators</i>	α	β	a	b
$t_{G1} = s_{wy}^2$	0	0	a	b	$t_{G16} = s_{wy}^2 \exp \left[\frac{\sigma_{wx}^2 - s_{wx}^2}{\sigma_{wx}^2} + \frac{\sigma_{wz}^2 - s_{wz}^2}{\sigma_{wz}^2 + s_{wz}^2} \right]$	1	1	1	2
$t_{G2} = s_{wy}^2 \exp \left[\frac{\sigma_{wx}^2 - s_{wx}^2}{\sigma_{wx}^2 + (a-1)s_{wx}^2} \right]$	1	0	a	b	$t_{G17} = s_{wy}^2 \exp \left[\frac{\sigma_{wx}^2 - s_{wx}^2}{\sigma_{wx}^2 + s_{wx}^2} + \frac{\sigma_{wz}^2 - s_{wz}^2}{\sigma_{wz}^2} \right]$	1	1	2	1
$t_{G3} = s_{wy}^2 \exp \left[\frac{\sigma_{wx}^2 - s_{wx}^2}{\sigma_{wx}^2} \right]$	1	0	1	b	$t_{wrr} = t_{G18} = s_{wy}^2 \exp \left[\frac{\sigma_{wx}^2 - s_{wx}^2}{\sigma_{wx}^2 + s_{wx}^2} + \frac{\sigma_{wz}^2 - s_{wz}^2}{\sigma_{wz}^2 + s_{wz}^2} \right]$	1	1	2	2
$t_{G4} = s_{wy}^2 \exp \left[\frac{\sigma_{wx}^2 - s_{wx}^2}{\sigma_{wx}^2 + s_{wx}^2} \right]$	1	0	2	b	$t_{G19} = s_{wy}^2 \exp \left[\frac{\sigma_{wx}^2 - s_{wx}^2}{\sigma_{wx}^2} + \frac{s_{wz}^2 - \sigma_{wz}^2}{\sigma_{wz}^2} \right]$	1	-1	1	1
$t_{G5} = s_{wy}^2 \exp \left[\frac{s_{wx}^2 - \sigma_{wx}^2}{\sigma_{wx}^2 + (a-1)s_{wx}^2} \right]$	-1	0	a	b	$t_{G20} = s_{wy}^2 \exp \left[\frac{\sigma_{wx}^2 - s_{wx}^2}{\sigma_{wx}^2} + \frac{s_{wz}^2 - \sigma_{wz}^2}{\sigma_{wz}^2 + s_{wz}^2} \right]$	1	-1	1	2
$t_{G6} = s_{wy}^2 \exp \left[\frac{s_{wx}^2 - \sigma_{wx}^2}{\sigma_{wx}^2} \right]$	-1	0	1	b	$t_{G21} = s_{wy}^2 \exp \left[\frac{\sigma_{wx}^2 - s_{wx}^2}{\sigma_{wx}^2 + s_{wx}^2} + \frac{s_{wz}^2 - \sigma_{wz}^2}{\sigma_{wz}^2} \right]$	1	-1	2	1
$t_{G7} = s_{wy}^2 \exp \left[\frac{s_{wx}^2 - \sigma_{wx}^2}{\sigma_{wx}^2 + s_{wx}^2} \right]$	-1	0	2	b	$t_{G22} = s_{wy}^2 \exp \left[\frac{\sigma_{wx}^2 - s_{wx}^2}{\sigma_{wx}^2 + s_{wx}^2} + \frac{\sigma_{wz}^2 - s_{wz}^2}{\sigma_{wz}^2 + s_{wz}^2} \right]$	1	-1	2	2
$t_{G8} = s_{wy}^2 \exp \left[\frac{\sigma_{wz}^2 - s_{wz}^2}{\sigma_{wz}^2 + (b-1)s_{wz}^2} \right]$	0	1	a	b	$t_{G23} = s_{wy}^2 \exp \left[\frac{s_{wx}^2 - \sigma_{wx}^2}{\sigma_{wx}^2} + \frac{\sigma_{wz}^2 - s_{wz}^2}{\sigma_{wz}^2} \right]$	-1	1	1	1
$t_{G9} = s_{wy}^2 \exp \left[\frac{\sigma_{wz}^2 - s_{wz}^2}{\sigma_{wz}^2} \right]$	0	1	a	1	$t_{G24} = s_{wy}^2 \exp \left[\frac{s_{wx}^2 - \sigma_{wx}^2}{\sigma_{wx}^2} + \frac{\sigma_{wz}^2 - s_{wz}^2}{\sigma_{wz}^2 + s_{wz}^2} \right]$	-1	1	1	2
$t_{G10} = s_{wy}^2 \exp \left[\frac{\sigma_{wz}^2 - s_{wz}^2}{\sigma_{wz}^2 + s_{wz}^2} \right]$	0	1	a	2	$t_{G25} = s_{wy}^2 \exp \left[\frac{s_{wx}^2 - \sigma_{wx}^2}{\sigma_{wx}^2 + s_{wx}^2} + \frac{\sigma_{wz}^2 - s_{wz}^2}{\sigma_{wz}^2} \right]$	-1	1	2	1
$t_{G11} = s_{wy}^2 \exp \left[\frac{s_{wz}^2 - \sigma_{wz}^2}{\sigma_{wz}^2 + (b-1)s_{wz}^2} \right]$	0	1	a	b	$t_{G26} = s_{wy}^2 \exp \left[\frac{s_{wx}^2 - \sigma_{wx}^2}{\sigma_{wx}^2 + s_{wx}^2} + \frac{\sigma_{wz}^2 - s_{wz}^2}{\sigma_{wz}^2 + s_{wz}^2} \right]$	-1	1	2	2
$t_{G12} = s_{wy}^2 \exp \left[\frac{s_{wz}^2 - \sigma_{wz}^2}{\sigma_{wz}^2} \right]$	0	1	a	1	$t_{G27} = s_{wy}^2 \exp \left[\frac{s_{wx}^2 - \sigma_{wx}^2}{\sigma_{wx}^2} + \frac{s_{wz}^2 - \sigma_{wz}^2}{\sigma_{wz}^2} \right]$	-1	-1	1	1
$t_{G13} = s_{wy}^2 \exp \left[\frac{s_{wz}^2 - \sigma_{wz}^2}{\sigma_{wz}^2 + s_{wz}^2} \right]$	0	1	a	2	$t_{G28} = s_{wy}^2 \exp \left[\frac{s_{wx}^2 - \sigma_{wx}^2}{\sigma_{wx}^2} + \frac{s_{wz}^2 - \sigma_{wz}^2}{\sigma_{wz}^2 + s_{wz}^2} \right]$	-1	-1	1	2
$t_{G14} = s_{wy}^2 \exp \left[\frac{\sigma_{wx}^2 - s_{wx}^2}{\sigma_{wx}^2 + (a-1)s_{wx}^2} + \frac{\sigma_{wz}^2 - s_{wz}^2}{\sigma_{wz}^2 + (b-1)s_{wz}^2} \right]$	1	1	a	b	$t_{G29} = s_{wy}^2 \exp \left[\frac{s_{wx}^2 - \sigma_{wx}^2}{\sigma_{wx}^2 + s_{wx}^2} + \frac{s_{wz}^2 - \sigma_{wz}^2}{\sigma_{wz}^2} \right]$	-1	-1	2	1
$t_{wRR} = t_{G15} = s_{wy}^2 \exp \left[\frac{\sigma_{wx}^2 - s_{wx}^2}{\sigma_{wx}^2} + \frac{\sigma_{wz}^2 - s_{wz}^2}{\sigma_{wz}^2} \right]$	1	1	1	1	$t_{G30} = s_{wy}^2 \exp \left[\frac{s_{wx}^2 - \sigma_{wx}^2}{\sigma_{wx}^2 + s_{wx}^2} + \frac{s_{wz}^2 - \sigma_{wz}^2}{\sigma_{wz}^2 + s_{wz}^2} \right]$	-1	-1	2	2