MODIFIED MULTIVARIATE RATIO ESTIMATOR FOR ADAPTIVE CLUSTER SAMPLING

***Wale-Orojo, Oluwaseun Ayobami[1], Ajayi, Abayomi Olumuyiwa[1], Olayiwola, Olaniyi Mathew[1] and Atanda, Omolola Dorcas[2]**

[1]Department of Statistics, Federal University of Agriculture Abeokuta, Ogun State, Nigeria.

[2]Mathematics Unit, St. Andrews College, Cambridge, England

Corresponding e-mail: wale-orojooa@funaab.edu.ng , +2348037394875.

Abstract

This study developed an improved measure of the Multivariate Ratio Estimator (MRE), which accommodates the Median and Kurtosis of auxiliary variable, in a network sampling. The adequacy of the estimator obtained for rare events via Adaptive Cluster (Networking) Sampling (ACS), was established via the estimation of its biasness and mean square error (MSE). Several estimators were compared to the obtained MRE via efficiency to ascertain the performance of MRE. Simulated and Secondary data from Nigeria Meteorological Agency was used to demonstrate the usefulness of the obtained estimator. The proposed modified Multivariate estimator is flexible as it accommodates several population parameters of the auxiliary information thereby increasing efficiency.

Keywords: Auxiliary Variable, Bias, Kurtosis, Mean Square Error, Median, Rare Event

1 Introduction

Thompson (1991) refined the procedure of sampling rare events from un-clustered population. The procedure which required initial selection of units by simple random sampling (SRS) method, involved the identification of variable of interest from each of the sampled unit before further identification of neighbouring units that contains the variable of interest. The collection of all the sampled units and its neighbouring units that contains the variable of interest are said to meet the criteria for inclusion in the study and are linked together and called Network, while units that did not meet the condition of inclusion are called Edges. The combination of network and its associated edge units are then identified as Clusters. Many researchers have worked on several estimators (ratio, regression, multivariate ratio, product estimators etc.) from prior auxiliary information in a survey sampling while preserving the skewness, correlation, coefficient of variation measures of the initial population (Olkin, 1958; Singh, 1967; Khoshnevisan et al. (2007); Hanif and Ahmad, 2011; Olayiwola et al. (2015); Subramani, 2018). A similar approach from generalized exponential estimator which retained the population variance was proposed by Qureshi et al. (2019) and Younis (2019). Akingbade and Okafor (2019), developed some linear combination for the class of ratio estimators that accommodates the identity of the initial population mean. Also, Selk et al. (2021) showed the uniform consistency, rates of convergence of the considered non-parametric estimator for both the multivariate deterministic design points and multivariate random covariates. Hence this study developed MRE from auxiliary variables while keeping the Median and Kurtosis identity of the population intact.

2 Materials and Method

The Hansen-Hurwitz estimator of the population according to Thompson 1990

$$\bar{y}_{ac} = \frac{1}{n}\sum_{i=1}^{n} (W_y)i \tag{1}$$

where $(w_y)i$ is the average of the variable of interest in the network that includes unit i of the initial

$$(W_y)i = \frac{1}{m_i}\sum_{j\epsilon\psi i} y_j \tag{2}$$

Variance of $\bar{y}_a c$ is:

$$v(\bar{y}_a c) = \frac{N-n}{N(N-1)}\sum_{i=1}^{N} ((W_y)i - \mu_y)^2 \tag{3}$$

2.1 Multivariate Ratio Estimator for Adaptive Cluster Sampling

Suppose the variable of interest y defined on the finite population $\Psi = 1,\ 2,\ \ldots,\ N$ consists of N units, then for each unit i, we have variable y_i and its associated auxiliary variables $X_1, X_2, \ldots, X_p$. Suppose our $p = 2$, then the population mean of y is $\mu_y = \frac{1}{N}\sum_{i=1}^{N} y_i$ and the population means of X_1 and X_2 are $\mu_{x_1} = \frac{1}{N}\sum_{i=1}^{N} x_1$ and $\mu_{x_2} = \frac{1}{N}\sum_{i=1}^{N} x_2$, then $\bar{y}$ is the sample mean of the variable of interest, while $\bar{x}_1$ and $\bar{x}_2$ are the sample mean of the auxiliary variables all under SRS. Let $\bar{y}_{acs}$ be the proposed adaptive cluster sampling estimator of the population mean then the network (φ) which includes unit *i* and size n_i denotes the number of units in the network. Thus, the proposed estimator $\bar{Y}$ using ACS designs for p-auxiliary variables was defined as follows;

$$T_{m(ac)} = w_1\bar{y}_{ac}\left[\frac{a_{1(W_x)}\bar{X}_1 + b_{1(W_x)}}{a_{1(W_x)}\bar{x}_{1(ac)} + b_{1(W_x)}}\right]^g + w_2\bar{y}_{ac}\left[\frac{a_{2(W_x)}\bar{X}_2 + b_{2(W_x)}}{a_{2(W_x)}\bar{x}_{2(ac)} + b_{2(W_x)}}\right]^g + \ldots$$
$$+ w_p\bar{y}_{ac}\left[\frac{a_{p(W_x)}\bar{X}_p + b_{p(W_x)}}{a_{p(W_x)}\bar{x}_{p(ac)} + b_{p(W_x)}}\right]^g \tag{4}$$

where $w_1 + w_2 + \ldots + w_p = 1$ are weights, $a_{1(W_x)}, a_{2(W_x)}, b_{1(W_x)}, b_{2(W_x)}, \ldots$ are some population parameters of the auxiliary variables. For simplification, a bivariate version $p = 2$ of the proposed multivariate estimator in adaptive cluster sampling is;

$$T_{m(ac)} = w_1\bar{y}_{ac}\left[\frac{a_{1(W_x)}\bar{X}_1 + b_{1(W_x)}}{a_{1(W_x)}\bar{x}_{1(ac)} + b_{1(W_x)}}\right]^g + w_2\bar{y}_{ac}\left[\frac{a_{2(W_x)}\bar{X}_2 + b_{2(W_x)}}{a_{2(W_x)}\bar{x}_{2(ac)} + b_{2(W_x)}}\right]^g \tag{5}$$

where $w_1 + w_2 = 1$
some members of the estimator $T_{m(ac)}$ are shown in Table 1

Table 1: Members of the family of Estimator

a_1	a_2	b_1	b_2	Proposed Multivariate Estimators
1	1	1	1	$T_{0m(ac)} = w_1\bar{y}_{ac}[\frac{\bar{X}_1}{\bar{x}_{1(ac)}}] + w_2\bar{y}_{ac}[\frac{\bar{X}_2}{\bar{x}_{2(ac)}}]$
1	1	C_{Wx}	C_{Wx}	$T_{1m(ac)} = w_1\bar{y}_{ac}[\frac{\bar{X}_1+C_{(Wx_1)}}{\bar{x}_{1(ac)}+C_{(Wx_1)}}] + w_2\bar{y}_{ac}[\frac{\bar{X}_2+C_{(Wx_2)}}{\bar{x}_{2(ac)}+C_{(Wx_2)}}]$
1	1	$B_{1(Wx)}$	$B_{1(Wx)}$	$T_{2m(ac)} = w_1\bar{y}_{ac}[\frac{\bar{X}_1+B_1(Wx_1)}{\bar{x}_{1(ac)}+B_1(Wx_1)}] + w_2\bar{y}_{ac}[\frac{\bar{X}_2+B_1(Wx_2)}{\bar{x}_{2(ac)}+B_1(Wx_2)}]$
1	1	$B_{2(Wx)}$	$B_{2(Wx)}$	$T_{3m(ac)} = w_1\bar{y}_{ac}[\frac{\bar{X}_1+B_{2(Wx_1)}}{\bar{x}_{1(ac)}+B_{2(Wx_1)}}] + w_2\bar{y}_{ac}[\frac{\bar{X}_2+B_{2(Wx_2)}}{\bar{x}_{2(ac)}+B_{2(Wx_2)}}]$
1	1	$md_{(Wx)}$	$md_{(Wx)}$	$T_{4m(ac)} = w_1\bar{y}_{ac}[\frac{\bar{X}_1+md_{(Wx_1)}}{\bar{x}_{1(ac)}+md_{(Wx_1)}}] + w_2\bar{y}_{ac}[\frac{\bar{X}_2+md_{(Wx_2)}}{\bar{x}_{2(ac)}+md_{(Wx_2)}}]$
$B_{1(Wx)}$	$B_{1(Wx)}$	$B_{2(Wx)}$	$B_{2(Wx)}$	$T_{5m(ac)} = w_1\bar{y}_{ac}[\frac{B_{1(Wx_1)}\bar{X}_1+B_{2(Wx_1)}}{B_{1(Wx_1)}\bar{x}_{1(ac)}+B_{2(Wx_1)}}] + w_2\bar{y}_{ac}[\frac{B_{1(Wx_1)}\bar{X}_2+B_{2(Wx_2)}}{B_{1(Wx_1)}\bar{x}_{2(ac)}+B_{2(Wx_2)}}]$
$B_{1(Wx)}$	$B_{1(Wx)}$	$C_{(Wx)}$	$C_{(Wx)}$	$T_{6m(ac)} = w_1\bar{y}_{ac}[\frac{B_{1(Wx_1)}\bar{X}_1+C_{(Wx_1)}}{B_{1(Wx_1)}\bar{x}_{1(ac)}+C_{(Wx_1)}}] + w_2\bar{y}_{ac}[\frac{B_{1(Wx_1)}\bar{X}_2+C_{(Wx_2)}}{B_{1(Wx_1)}\bar{x}_{2(ac)}+C_{(Wx_2)}}]$
$B_{1(Wx)}$	$B_{1(Wx)}$	$md_{(Wx)}$	$md_{(Wx)}$	$T_{7m(ac)} = w_1\bar{y}_{ac}[\frac{B_{1(Wx_1)}\bar{X}_1+md_{(Wx_1)}}{B_{1(Wx_1)}\bar{x}_{1(ac)}+md_{(Wx_1)}}] + w_2\bar{y}_{ac}[\frac{B_{1(Wx_2)}\bar{X}_2+md_{(Wx_2)}}{B_{1(Wx_2)}\bar{x}_{2(ac)}+md_{(Wx_2)}}]$
$B_{2(Wx)}$	$B_{2(Wx)}$	$md_{(Wx)}$	$md_{(Wx)}$	$T_{8m(ac)} = w_1\bar{y}_{ac}[\frac{B_{2(Wx_1)}\bar{X}_1+md_{(Wx_1)}}{B_{2(Wx_1)}\bar{x}_{1(ac)}+md_{(Wx_1)}}] + w_2\bar{y}_{ac}[\frac{B_{2(Wx_2)}\bar{X}_2+md_{(Wx_2)}}{B_{2(Wx_2)}\bar{x}_{2(ac)}+md_{(Wx_2)}}]$
$B_{2(Wx)}$	$B_{2(Wx)}$	$C_{(Wx)}$	$C_{(Wx)}$	$T_{9m(ac)} = w_1\bar{y}_{ac}[\frac{B_{2(Wx_1)}\bar{X}_1+C_{(Wx_1)}}{B_{2(Wx_1)}\bar{x}_{1(ac)}+C_{(Wx_1)}}] + w_2\bar{y}_{ac}[\frac{B_{2(Wx_2)}\bar{X}_2+C_{(Wx_2)}}{B_{2(Wx_2)}\bar{x}_{2(ac)}+C_{(Wx_2)}}]$
$C_{(Wx)}$	$C_{(Wx)}$	$md_{(Wx)}$	$md_{(Wx)}$	$T_{10m(ac)} = w_1\bar{y}_{ac}[\frac{C_{(Wx_1)}\bar{X}_1+md_{(Wx_1)}}{C_{(Wx_1)}\bar{x}_{1(ac)}+md_{(Wx_1)}}] + w_2\bar{y}_{ac}[\frac{C_{(Wx_2)}\bar{X}_2+md_{(Wx_2)}}{C_{(Wx_2)}\bar{x}_{2(ac)}+md_{(Wx_2)}}]$

where notations used are:

N - Population size,

n - Sample size,

$f = \frac{N_2(N-n)}{Nn}$

Y- variable of interest, X_1 and X_2- Auxiliary variables,

$\bar{Y}$ - Population means,
$\bar{X}_1$ and $\bar{X}_2$ - Auxiliary population means,
$\bar{x}_1$ and $\bar{x}_2$ -Auxiliary sample means,
$\bar{y}$ - Sample means,
S_y, S_{x_1} and S_{x_2} - Population standard deviations,
S_{yx_1}- Population covariance between Y and X_1,
S_{yx_2} - Population covariance between Y and X_2,
$S_{x_1x_2}$ - Population covariance between $X_1 and X_2$,
ρ_{yx_1} - Correlation Coefficient between Y and X_1,
ρ_{yx_2}- Correlation Coefficient between Y and X_2,
$\rho_{x_1x_2}$ - Correlation Coefficient between X_1 and X_2,
C_{x_1}, C_{x_2}, C_y- Coefficient of variations,
$\beta_{2(i)} = \frac{N(N+1)\sum_{i=1}^{N}(X_i-\bar{X})^4}{(N-1)(N-2)(N-3)S^4} - \frac{3(N-1)^2}{(N-2)(N-3)}$ - Coefficient of Kurtosis of the Auxiliary Variables,
$\beta_{2(1)}, \beta_{2(2)}$- Coefficient of Kurtosis of the Auxiliary Variables,
$Md_{(1)}$, and $Md_{(2)}$ - Median of the Auxiliary variables.

In deriving the bias and mean square error of the proposed estimator, let $e_0 = \frac{\bar{y}_{ac}}{\bar{Y}} - 1$,
$e_1 = \frac{\bar{x}_{1(ac)}}{\bar{X}_1} - 1,\ e_2 = \frac{\bar{x}_{2(ac)}}{\bar{X}_2} - 1$

such that
$E(e_0) = E(e_1) = E(e_2) = 0$

then
$E(e_0^2) = fC_{Wy}^2,\ E(e_1^2) = fC_{Wx_1}^2,\ E(e_2^2) = fC_{Wx_2}^2$

$E(e_0e_1) = f\rho_{Wyx_1}C_{Wy}C_{Wx_1},\ E(e_0e_2) = f\rho_{Wyx_2}C_{Wy}C_{Wx_2},\ E(e_1e_2) = f\rho_{Wx_1x_2}C_{Wx_1}C_{Wx_2}$

$\rho_{Wyx_1} = \frac{S_{Wyx_1}}{S_{Wy}S_{Wx_1}},\ \rho_{Wyx_2} = \frac{S_{Wyx_2}}{S_{Wy}S_{Wx_2}},\ \rho_{Wx_1x_2} = \frac{S_{Wx_1x_2}}{S_{Wx_1}S_{Wx_2}},$
$C_y = \frac{S_{Wy}}{\bar{Y}},\ C_{Wx_1} = \frac{S_{Wx_1}}{\bar{X}_1}, C_{Wx_2} = \frac{S_{Wx_2}}{\bar{X}_2}$

$S_{Wx_1}^2 = \frac{1}{N-1}\sum_{i=1}^{N}((W_{x_1})_i - \bar{X}_1)^2$

$S_{Wx_2}^2 = \frac{1}{N-1}\sum((W_{x_2})_i - \bar{X}_2)^2$

$S_{Wy}^2 = \frac{1}{N-1}\sum((W_y)_i - \bar{Y})^2$

$S_{Wx_1x_2}^2 = \frac{1}{N-1}\sum((W_{x_1})_i - \bar{X}_1)((W_{x_2})_i - \bar{X}_2)$
$S_{Wyx_1} = \frac{1}{N-1}\sum((W_{x_1})_i - \bar{X}_1)((W_y)_i - Y)$

$S_{Wyx_2} = \frac{1}{N-1}\sum((W_{x_2})_i - \bar{X}_2)((W_y)_i - \bar{Y})$

$$T_{m(ac)} = \bar{Y}(1 + e_0 - \lambda_2 e_2 + \lambda_2^2 e_2^2 - \lambda_2 e_0 e_2)$$

$$B(T_{m(ac)}) = \bar{Y}[\lambda_2^2 E(e_2^2) - \lambda_2 E(e_0 e_2)] + W\bar{Y}[\lambda_1^2 E(e_1^2) - \lambda_2^2 E(e_2^2) - \lambda_1 E(e_0 e_1) + \lambda_2 E(e_0 e_2)]$$

$$B(T_{m(ac)}) = f\bar{Y}[\lambda_2^2 C_{Wx_2}^2 - \lambda_2 e_{Wyx_2} C_{Wy} C_{Wx_2} + w(\lambda_1^2 C_{Wx_1}^2 - \lambda_2^2 C_{Wx_2}^2 + \lambda_2 \rho_{Wyx_2} C_{Wy} C_{Wx_2} - \lambda_1 \rho_{Wyx_1} C_{Wy} C_{Wx_1})]$$

$$MSE(T_{m(ac)}) = E(T_{acs} - \bar{Y})^2 \approx \bar{Y}^2 E[(e_0 - \lambda_2 e_2) + w(\lambda_2 e_2 - \lambda_1 e_1)]^2$$

$$= \bar{Y}^2 E[(e_0^2 + \lambda_2^2 e_2^2 - 2\lambda_2 e_0 e_2) + w^2(\lambda_2^2 e_2^2 + \lambda_1^2 e_1^2 - 2\lambda_1 \lambda_2 e_1 e_2) + 2w(\lambda_2 e_0 e_2 - \lambda_2^2 e_2^2 - \lambda_1 e_0 e_1 + \lambda_1 \lambda_2 e_1 e_2)]$$

$$= f\bar{Y}^2(C_{Wy}^2 + \lambda_2^2 C_{Wx_2}^2 - 2\lambda_2 \rho_{Wyx_2} C_{Wy} C_{Wx_2}) + w^2(\lambda_2^2 C_{Wx_2}^2 + \lambda_1^2 C_{Wx_1}^2 - 2\lambda_1 \lambda_2 \rho_{Wx_1x_2} C_{Wx_1} C_{Wx_2}) + 2w(\lambda_2 \rho_{Wyx_2} C_{Wy} C_{Wx_2} - \lambda_2^2 C_{Wx_2}^2 - \lambda_1 \rho_{Wyx_1} C_{Wy} C_{Wx_1} + \lambda_1 \lambda_2 \rho_{Wx_1x_2} C_{Wx_1} C_{Wx_2})$$

on differentiating the MSE with respect to w and equating to zero we obtained

$$w = -\frac{\lambda_2 \rho_{Wyx_2} C_{Wy} C_{Wx_2} - \lambda_1 \rho_{Wyx_1} C_{Wy} C_{Wx_1} - \lambda_2^2 C_{Wx_2}^2 + \lambda_1 \lambda_2 C_{Wx_1} C_{Wx_2} \rho_{Wx_1x_2}}{\lambda_2^2 C_{Wx_2}^2 - 2\lambda_1 \lambda_2 \rho_{Wx_1x_2} C_{Wx_1} C_{Wx_2} + \lambda_1^2 C_{Wx_1}^2} \tag{6}$$

inserting equation (6) we obtained the minimum mean square error

$$min.MSE(T_{m(ac)}) = \left(\frac{1-f}{n}\right)\bar{Y}^2[C_{Wy}^2 + \lambda_2^2 C_{Wx_2}^2 - 2\lambda_2 \rho_{Wyx_2} C_{Wy} C_{Wx_2}) - \frac{(\lambda_2 \rho_{Wyx_2} C_{Wy} C_{Wx_2} - \lambda_1 \rho_{Wyx_1} C_{Wy} C_{Wx_1} - \lambda_2^2 C_{Wx_2}^2 + \lambda_1 \lambda_2 C_{Wx_1} C_{Wx_2} \rho_{Wx_1x_2})^2}{\lambda_2^2 C_{Wx_2}^2 - 2\lambda_1 \lambda_2 \rho_{Wx_1x_2} C_{Wx_1} C_{Wx_2} + \lambda_1^2 C_{Wx_1}^2}] \tag{7}$$

inserting λ_1 and λ_2 shown in table 2 into equation (7) gives the minimum mean square error for each estimator in table 1.

Table 2: Parameters for minimum Mean square error given for the proposed estimators

a_1	a_2	b_1	b_2	λ_1	λ_2
1	1	1	1	1	1
1	1	C_{Wx}	C_{Wx}	$\frac{\bar{X}_1}{\bar{x}_1+C_{Wx_1}}$	$\frac{\bar{X}_2}{\bar{x}_2+C_{Wx_2}}$
1	1	$B_{1(Wx)}$	$B_{1(Wx)}$	$\frac{\bar{X}_1}{\bar{x}_1+B_{1(Wx_1)}}$	$\frac{\bar{X}_2}{\bar{x}_2+B_{1(Wx_2)}}$
1	1	$B_{2(Wx)}$	$B_{2(Wx)}$	$\frac{\bar{X}_1}{\bar{x}_1+B_{2(Wx_1)}}$	$\frac{\bar{X}_2}{\bar{x}_2+B_{2(Wx_2)}}$
1	1	$md_{(Wx)}$	$md_{(Wx)}$	$\frac{\bar{X}_1}{\bar{x}_1+md_{(Wx_1)}}$	$\frac{\bar{X}_2}{\bar{x}_2+md_{(Wx_2)}}$
$B_{1(Wx)}$	$B_{1(Wx)}$	$B_{2(Wx)}$	$B_{2(Wx)}$	$\frac{B_{1(Wx_1)}\bar{X}_1}{B_{1(Wx_1)}\bar{x}_1+C_{(Wx_1)}}$	$\frac{B_{1(Wx_2)}\bar{X}_2}{B_{1(Wx_2)}\bar{x}_2+C_{(Wx_2)}}$
$B_{1(Wx)}$	$B_{1(Wx)}$	$C_{(Wx)}$	$C_{(Wx)}$	$\frac{B_{1(Wx_1)}\bar{X}_1}{B_{1(Wx_1)}\bar{x}_1+C_{(Wx_1)}}$	$\frac{B_{1(Wx_2)}\bar{X}_2}{B_{1(Wx_2)}\bar{x}_2+C_{(Wx_2)}}$
$B_{1(Wx)}$	$B_{1(Wx)}$	$md_{(Wx)}$	$md_{(Wx)}$	$\frac{B_{1(Wx_1)}\bar{X}_1}{B_{1(Wx_1)}\bar{x}_1+md_{(Wx_1)}}$	$\frac{B_{1(Wx_2)}\bar{X}_2}{B_{1(Wx_2)}\bar{x}_2+md_{(Wx_2)}}$
$B_{2(Wx)}$	$B_{2(Wx)}$	$md_{(Wx)}$	$md_{(Wx)}$	$\frac{B_{2(Wx_1)}\bar{X}_1}{B_{2(Wx_1)}\bar{x}_1+md_{(Wx_1)}}$	$\frac{B_{2(Wx_2)}\bar{X}_2}{B_{2(Wx_2)}\bar{x}_2+md_{(Wx_2)}}$
$B_{2(Wx)}$	$B_{2(Wx)}$	$C_{(Wx)}$	$C_{(Wx)}$	$\frac{B_{2(Wx_1)}\bar{X}_1}{B_{2(Wx_1)}\bar{x}_1+C_{(Wx_1)}}$	$\frac{B_{2(Wx_2)}\bar{X}_2}{B_{2(Wx_2)}\bar{x}_2+C_{(Wx_2)}}$
$C_{(Wx)}$	$C_{(Wx)}$	$md_{(Wx)}$	$md_{(Wx)}$	$\frac{C_{(Wx_1)}\bar{X}_1}{C_{(Wx_1)}\bar{x}_1+md_{(Wx_1)}}$	$\frac{C_{(Wx_2)}\bar{X}_2}{C_{(Wx_2)}\bar{x}_2+md_{(Wx_2)}}$

2.2 Efficiency Comparison

To compare the efficiency of the suggested ratio estimator $MSE(T_{acs})$ with the existing estimator $MSE(T_{M(ac)}$ (Eq. 1), it can be seen that the MSEs contained three terms, the first without the weight (w), the second with the weight (w) and the third term with the square of the weight (w), The first term of the MSEs, since $0 < \lambda < 1$, then it is computed that $MSE(T_{MR}) - MSE(T_{m(ac)}) = C_{Wx_2}^2(1-\lambda_2^2) + 2\rho_{Wyx_2}C_{Wy}C_{Wx_2}(\lambda_2 - 1) > 0$, with other terms greater than zero. It can be seen that the ratio estimators are consistent estimators, since the proposed estimator with multiple auxiliary variables are linear combinations of consistent estimators it follows that the proposed estimator is also consistent.

3 Data Analysis and Presentation

3.1 **Simulated Data**

For the simulation studies we considered the dataset in figure 1 of Thompson (1991). This contains information on the y-values associated with units in a 20 by 20 unit grid. The aim was to use adaptive cluster sampling to estimate the number of points-objects in the study region of 400 units. The datasets which contains x. X coordinate, y. Y coordinate, Network ID, a unique number identifying the network to which the unit belongs, m. The number of units in the given network and y_value and the y-value associated with the unit (variable of interest, Y) were accessed on ACSampling package in r console. The values of $X_1 = 2Y + e$ and $X_2 = 0.1Y + e$ where $e \sim N(0, Y)$ were generated as shown in figures 2, 3, and 4. The final population now contains variables, Y, X_1 and X_2 .The population parameters for the variables were then computed.

The samples were selected using adaptive cluster sampling with sample sizes n= 5, 10, 50, 150 and 400. The proposed estimators were computed for each sample size. 10,000 iterations were use to obtain an accuracy estimate of the population mean. The estimated mean square error of the mean estimate is

$$\hat{R}MSE(\hat{T}_{acs}) = \frac{1}{10000}\sum_{i=1}^{10000} (\hat{T}_{acs} - \bar{Y})^2 \tag{8}$$

$$\hat{R}Bias(\hat{T}_{acs}) = \frac{1}{10000}\sum_{i=1}^{10000} (\hat{T}_{acs} - \bar{Y}) \tag{9}$$

The results were as shown in table 3 and 4 respectively

3.2 Real Life Data

Yearly measurements of rainfall, maximum temperature and relative humidity of 32 states across Nigeria from 1981–2018 was collected as secondary data of Nigeria Meteorological Agency from National Bureau of Statistics. An initial sample of size $n^1 = 100$ were selected using simple random sample without replacement from the rainfall data. High rainfall volume was taken as relatively rare in the population, and of the initial samples of 100, only 31 met the condition of greater or equal to the overall average rainfall volume. Using adaptive cluster sampling design to obtain adequate samples, the neighbourhood of the initially selected unit that met the criteria of greater or equal to the average rainfall volume was added. This was done until no unit in the neighbourhood of the selected unit met the condition again. Thus, starting with an initial sample of 100 out of which 31 satisfied the criteria of selection, the adaptive cluster sampling design led to final sample size of 256. Identifying the adaptive samples of interest location, it was realized that the rainfall data adaptively selected fell into the following states (Ogun, Delta, Edo, Niger, Cross-River, Enugu, Oyo, Rivers and Akwa-Ibom) which constituted the final samples. The corresponding maximum temperature and relative humidity of the rainfall data selected were used as auxiliary variables. The analysis was done using the R 4.0.5 package. The rainfall data from 1981 – 2018 of all the states that fell into our study is presented in table 7 (Appendix) and figure 1 also showed the rainfall pattern in each state over the same period of 1981 – 2018, table 1 shows the descriptive analysis. Also, for the existing estimators, 256 samples were selected using Adaptive Sampling techniques and then used it to compute existing biasness and MSE for multivariate ratios.

Table 3: Descriptive Statistics of Rainfall, Maximum Temperature and Relative Humidity

	Mean	SE	Med.	Std Dev.	Sam. Var.	Kurt.	Skewness	Range	Min.	Max.	Sum	Count	Conf . Lvl(95%)
Rainfall	274.59	8.69	215.98	139.01	19322.59	2.41	1.76	676.49	149.18	825.68	70294.17	256.00	17.11
Temp.	32.41	0.12	32.00	1.88	3.52	41.77	4.42	24.10	27.40	51.50	8296.47	256.00	0.23
R. H.	62.46	0.53	64.08	8.47	71.82	0.59	-1.04	39.90	34.10	74.00	15988.76	256.00	1.04

4 Results

The final samples of 256 adaptively selected as shown in appendix has its descriptive analysis of the variable of interest (Rainfall) and the auxiliary variables (maximum temperature and relative humidity) for the period of 1981 – 2018. Tables 5 showed the relative MSE of the estimators.

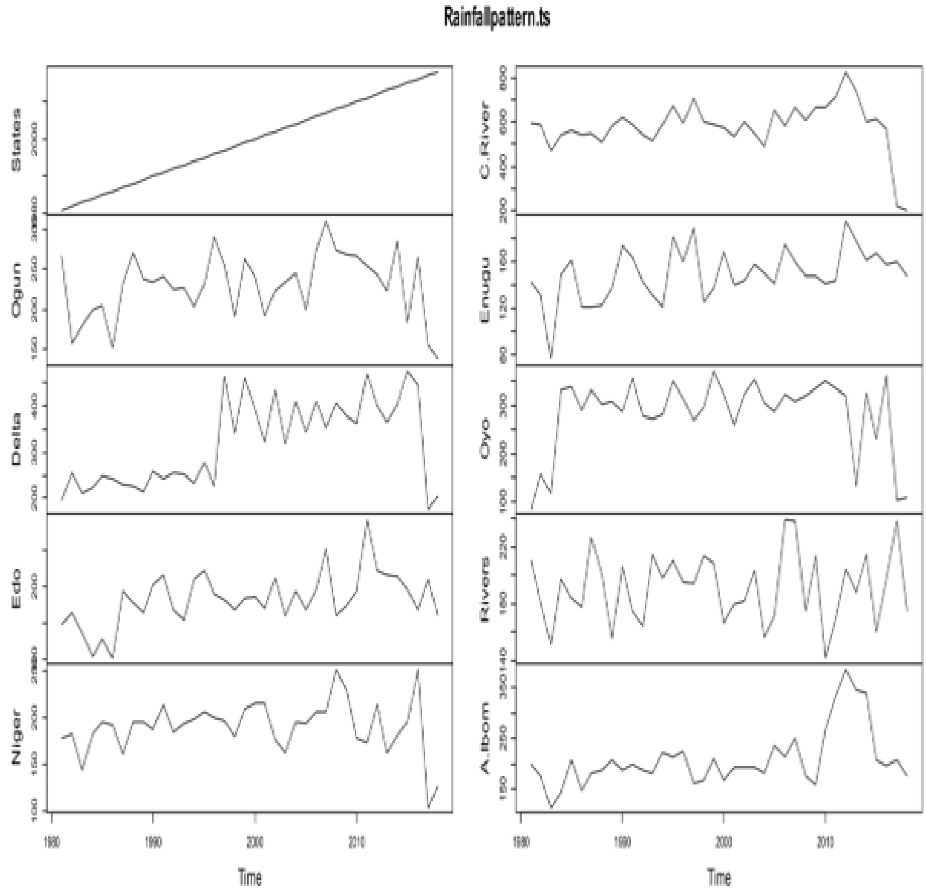

Figure 1: Rainfall pattern of the Adaptive locations.

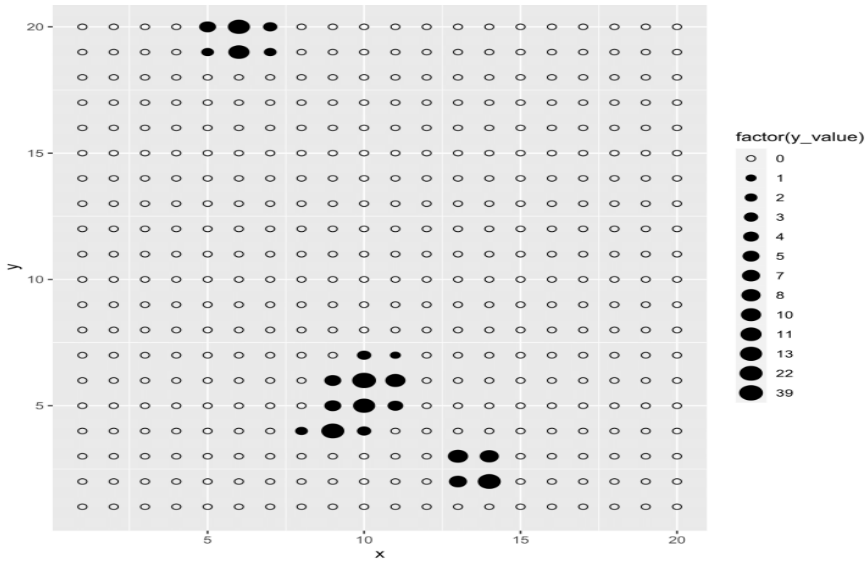

Figure 2: Simulated Y from Thompson (1991).

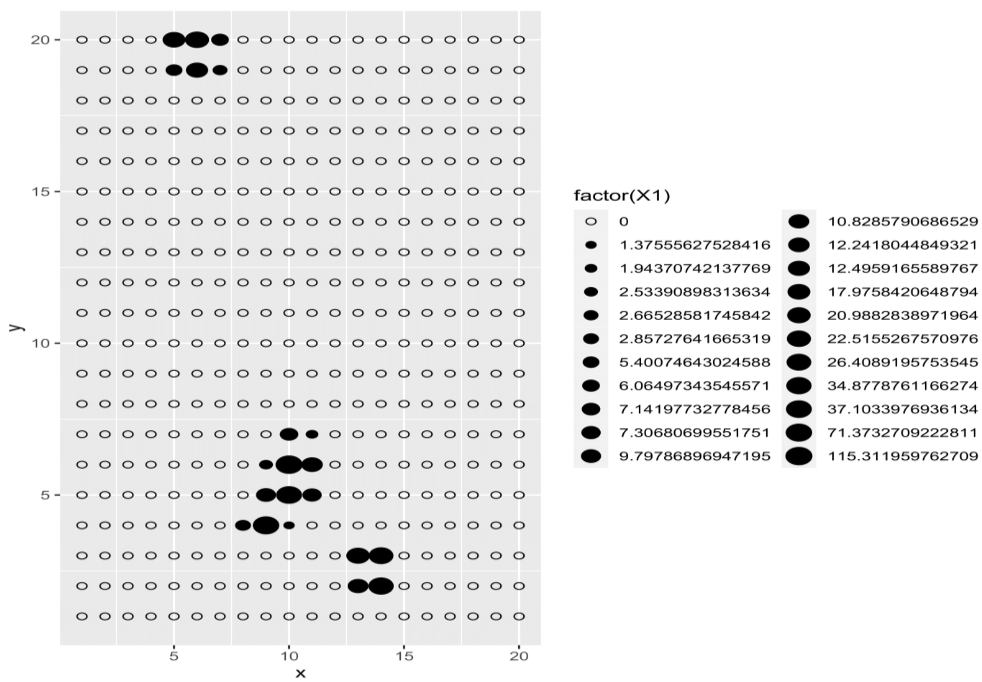

Figure 3: Simulated X_1.

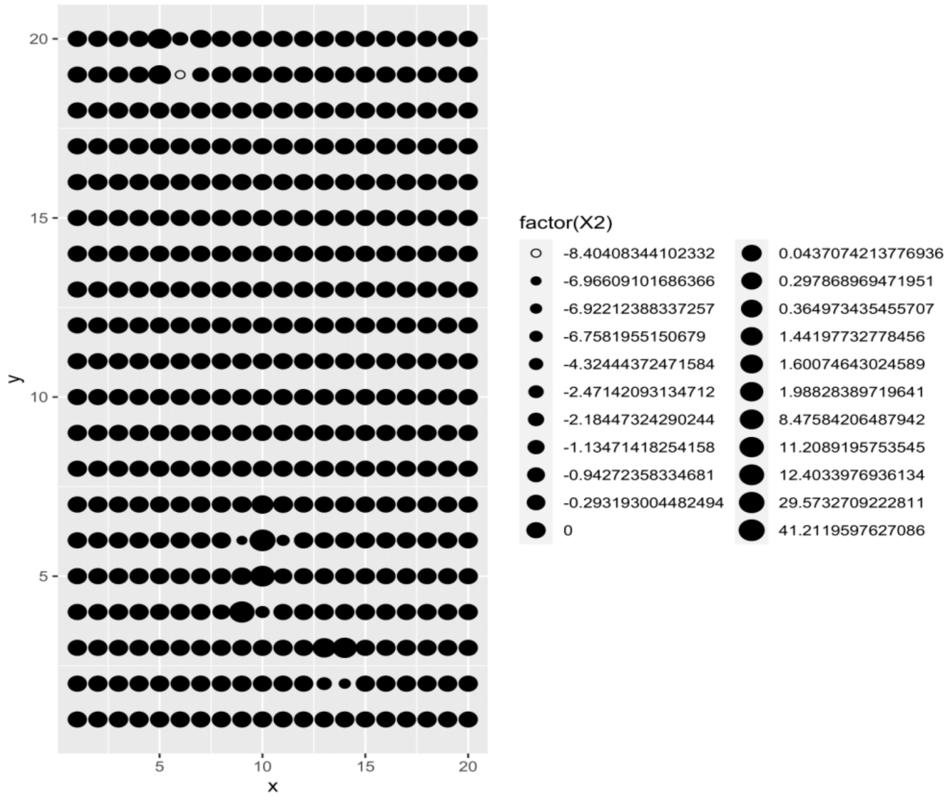

Figure 4: Simulated X_2.

***Table 4: Relative MSE (RMSE)* for Simulated data**

Estimator	5	10	50	150	300
0	901851	767004.49	32,884,178	34.19	0.03
1	350.84	2931.03	549192.11	79292.74	54.42
2	663.02	3865.03	381030.46	495.88	57.919
3	9.4	25.72	677	82.67	0.5256
4	11.494	0.161	0.146	0.011	0.00495
5	1070.86	1121327.59	19,189,627.05	108199.62	43.55
7	2.44	0.171	0.144	0.0096	0.00049
8	1.484	0.66	0.1712	0.0128	0.00049
9	29377	15839.14	2.70E+10	27695	0.333
10	418.23	12.78	0.311	0.0337	0.00116
Existing Non AC Sampling	27597577.635	229607.027	631.462	0.2968	0.0120

Table 5: Relative Bias for Simulated data

$T_{m(ac)}$	5	10	50	150	300
0	-56990.33	-231319.77	-1001161.5	-4222.52	624.19
1	29104.54	54446	-60193.36	82509.23	-1034.3
2	17732.98	25447.33	-52973.03	12041.14	733.72
3	16067.06	23971.87	30232.37	8742.86	2792.47
4	78.278	-599.265	-1126.465	-882.496	-219.234
5	32030.86	-115749.04	-587445.17	70382.14	298.53
7	-672.49	-751.46	-1272.88	-876.37	-220.76
8	-220.7	-246.81	-1003.6	-867.02	-219.32
9	-20774.65	-20774.65	-1003.6	18459.1	918.16
10	-894.59	2869.48	39.23	-655.04	-215.59
Existing Non AC Sampling	713830.084	66583.477	-2919.661	-1344.740	-159.346

Table 6: Real Life data

$T_{m(ac)}$	Estimate	M.S.E	PRE
0	281.9406	20.9913	2270.80314
1	282.0600	20.8138	2289.31574
2	282.1203	20.6214	2309.74231
3	281.9425	20.9185	2278.35791
4	278.3409	27.2424	1772.68596
5	273.7251	152.4652	398.874628
6	273.5340	150.9629	401.848865
7	284.9095	34272.3696	101.329584
8	306.0058	1313.2705	134.698092
9	304.6232	11323.7667	104.024101
10	274.6478	34.8549	1407.36224
Existing Non AC Sampling	278.556	45567.98	100

5 Conclusion

This study proposed a modified multivariate ratio estimator in adaptive cluster sampling for rare and clustered populations, this may provide better results than the conventional multivariate ratio estimator in simple random sampling (non-adaptive sampling). This proposed estimator is useful when the exact pattern of concentration of population is unknown. Some known parameters associated with the auxiliary variables have been used to produce various special cases of the proposed estimator (Table 2). From the results of simulation study, illustrated in Table 4 and 5, it was observed that the proposed class of ratio estimators provides better results in comparison of non-adaptive estimators, with Estimator 8, 7 and 4 being the least amongst suggested estimators for the simulation data while for the real-life data, all suggested estimator performed better with estimators 2, 1 and 3 performing best in terms of percentage relative efficiency. This also suggests that the choice of known parameters of the auxiliary variables influences the bias and mean square error of estimators. The amount of Relative bias converges to zero as the sample (initial and expected) sizes increase. In addition, one main reason behind the use of the known parameters of the auxiliary variable is to get the results on all sample sizes for the proposed class of estimators. Finally, it is inferred that the proposed class of estimators is the most efficient and useful estimators for rare and clustered populations. The amount of relative bias decreases and the proposed class of estimators performs better as the sample size increases. The adaptive estimators achieve higher efficiency than the non-adaptive estimators considered in this article for both real life population and an artificial population generated by the Thompson (1991).

References

[1] Akingbade, T. J., and Okafor, F.C. (2019). A Class of Ratio-Type Estimator Using Two Auxiliary Variables for Estimating the Population Mean with Some Known Population Parameters. Pakistan Journal of Statistics and Operation Research, 15, 329-340.

[2] Hanif, M., and Ahmad, M. (2011). A family of estimators for single and two-phase sampling using two auxiliary attributes. *Pakistan journal of statistics*, *27*(1).

[3] Khoshnevisan, M., Singh, R., Chauhan, P., Sawan, N., and Smarandache, F., (2007). A general family of estimators for estimating population mean using known value of some population parameter(s). Far East J. Theoretical Statistics, 22(2), 181-191. https://doi.org/10.48550/arXiv.math/0701243

[4] Olayiwola O.M., F.S. Apantaku, A. A. Adewara and A.F. Gboyega (2015) Modified Ratio Estimator Using Multiple Auxiliary Variables. The Pacific Journal of Science and Technology, Volume 16. Number 2. November 2015 (Fall), 108 -113

[5] Olkin, I. (1958). Multivariate Ratio Estimation for Finite Population. Biometrika, 45, 154-165. doi:10.1093/biomet/45.1-2.154.

[6] Qureshi, M. N., Iftikhar A., Hanif, M. (2019). Generalized Exponential Estimator for the Estimation of Clustered Population Variance in Adaptive Cluster Sampling. Journal of Statistical Theory and Applications, 18(4), pp. 416–424. DOI: https://doi.org/10.2991/jsta.d.191204.001.

[7] Selk L, Tillier C, Marigliano O. (2021). Multivariate boundary regression models. Scand J Statist.,1–27. https://doi.org/10.1111/sjos.12519

[8] Singh, M. P. (1967). Multivariate product method of estimation for finite populations. J. Indian Soc. Agri. Statist., 19, 1-10.

[9] Subramani J. (2018). Two parameter modified ratio estimators with two auxiliary variables for the estimation of finite population mean. Biometrics and Biostatistics International Journal 7(6): 559–568. DOI: 10.15406/bbij.2018.07.00259

[10] Thompson, S. K. (1991). Adaptive Cluster Sampling . Journal of the American Statistical Association, 85(412), 1050–1059. doi:10.1080/01621459.1990.10474975

[11] Younis, F. and Shabbir, J., (2019). Estimation of general parameter in adaptive cluster sampling using two auxiliary variables. Journal of the National Science Foundation of Sri Lanka, 47(1), pp.89–103. DOI: http://doi.org/10.4038/jnsfsr.v47i1.8933

Appendix I

Table 7: Rainfall data of all the states that fell into the sample location from 1981 – 2018

States	Ogun	Delta	Edo	Niger	C/River	Enugu	Oyo	Rivers	A/Ibom
1981	266.48	195.83	148.82	178.72	596.45	142.01	84.71	210.89	198.52
1982	156.61	254.74	164.08	183.13	587.09	130.55	156.98	181.14	175.44
1983	179.09	210.13	137.27	143.51	472.15	76.09	116.62	151.37	113.89
1984	199.60	224.79	104.12	182.96	540.19	148.28	331.41	196.88	144.48
1985	203.98	248.07	128.78	195.98	561.46	160.88	339.53	183.88	205.95
1986	152.09	242.98	102.35	192.02	540.24	120.88	290.95	177.80	146.65
1987	231.58	231.30	194.69	161.37	548.92	120.97	330.53	226.99	180.43
1988	269.83	225.73	179.09	195.60	509.20	121.75	300.29	201.74	186.57
1989	237.24	214.29	164.28	194.97	582.63	136.98	307.40	155.41	208.50
1990	234.81	258.23	203.06	187.23	617.42	173.48	287.66	206.89	187.85
1991	240.83	243.39	216.47	214.10	587.08	163.38	355.24	174.53	198.66
1992	224.19	255.03	168.04	184.13	544.18	142.08	278.61	164.25	187.99
1993	227.23	250.88	155.07	193.34	514.97	131.41	272.53	215.14	181.28
1994	202.98	233.92	210.96	199.02	586.50	121.24	281.70	197.85	222.36
1995	233.06	278.15	222.37	207.23	672.75	180.66	349.61	210.71	211.66
1996	291.50	225.48	190.53	200.27	595.24	159.95	315.33	194.98	223.48
1997	255.05	462.94	183.40	197.80	703.29	188.62	269.13	194.15	161.01
1998	190.93	339.50	168.23	179.33	600.71	124.68	294.26	214.09	167.53
1999	262.85	459.32	185.23	209.81	589.00	137.28	372.29	208.30	208.87
2000	238.73	390.73	186.65	215.63	575.88	168.90	322.96	166.19	166.24
2001	192.78	322.75	170.21	215.61	536.25	139.77	258.42	179.46	192.26
2002	223.30	436.71	213.27	175.96	603.37	143.39	321.73	182.14	192.98
2003	234.48	317.46	161.48	162.88	546.03	157.58	353.78	203.87	191.53
2004	245.23	410.59	194.03	194.65	486.27	149.18	303.38	156.46	180.17
2005	199.24	343.75	167.83	194.26	656.03	141.45	288.01	171.27	235.13
2006	274.17	411.49	196.54	205.71	579.94	174.69	322.07	239.05	213.23
2007	310.39	354.55	252.73	206.55	663.08	159.27	307.83	237.96	248.76
2008	274.24	406.49	160.53	252.41	607.47	147.38	319.63	175.08	175.52
2009	269.58	378.41	174.82	229.26	664.94	147.48	334.98	213.54	159.31
2010	266.51	361.16	194.95	178.18	665.40	140.58	351.20	141.78	267.15
2011	253.33	469.07	291.95	173.73	709.97	143.84	336.36	169.39	331.87
2012	242.59	401.54	222.97	214.75	825.68	194.34	319.72	204.42	385.01
2013	223.73	367.21	216.33	162.84	740.03	176.46	131.90	188.30	343.86
2014	284.40	404.49	215.28	180.99	601.55	160.80	325.33	214.63	339.81
2015	182.77	477.37	196.12	195.73	611.25	167.41	227.78	160.56	207.55
2016	264.94	445.28	168.86	251.72	567.26	156.88	361.68	196.79	194.48
2017	155.20	177.27	210.61	103.28	221.03	159.27	102.61	237.96	207.30
2018	137.12	203.25	160.53	126.20	202.49	147.38	106.54	175.08	175.52

Source: Nigeria Meteorological Agency